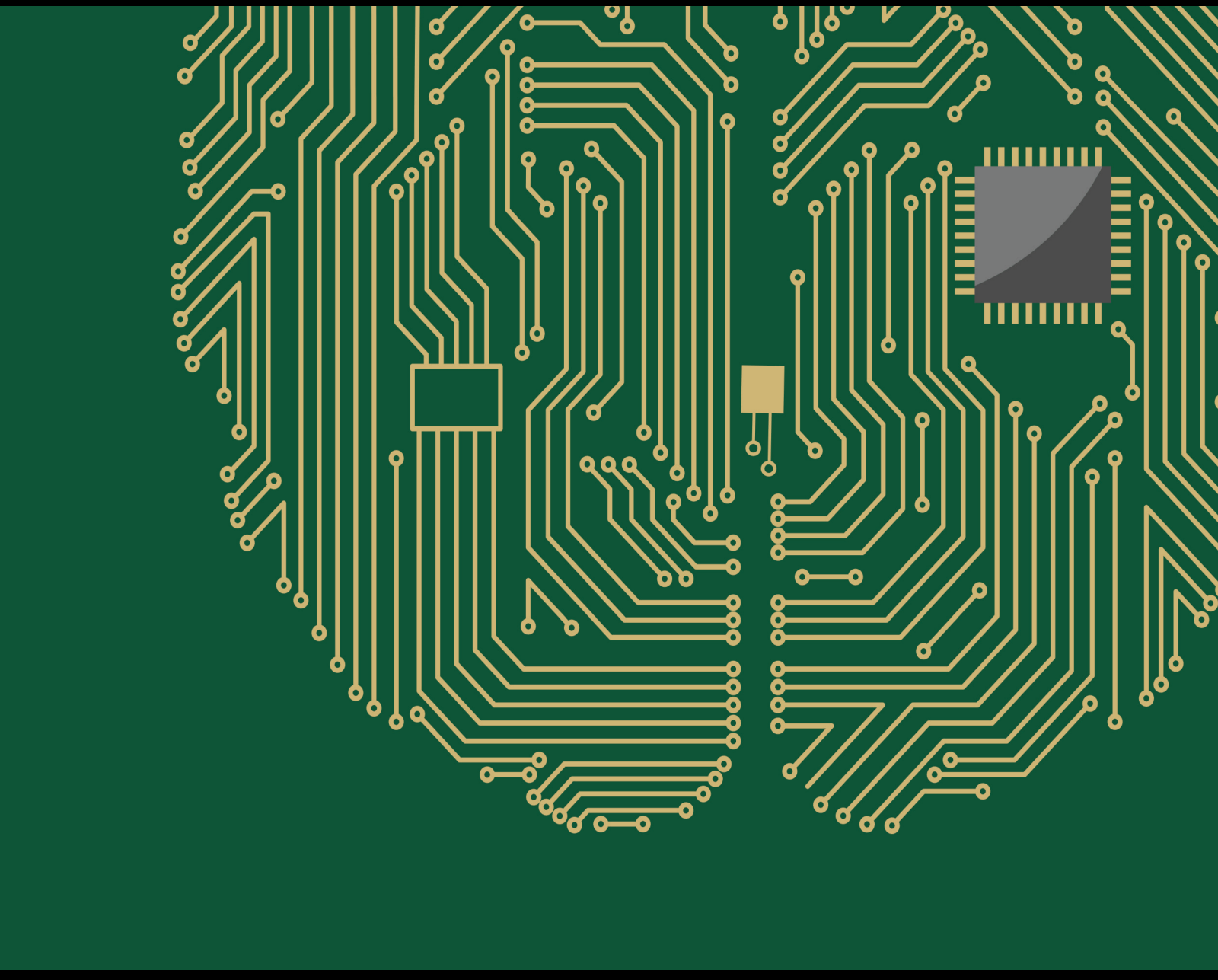
Computational Intelligence and Neuroscience

Deep Learning-based Domain Adaptation Methods and Applications

Lead Guest Editor: Wenming Cao
Guest Editors: Cheng Liu and Man-Fai Leung



Deep Learning-based Domain Adaptation Methods and Applications

Computational Intelligence and Neuroscience

Deep Learning-based Domain Adaptation Methods and Applications

Lead Guest Editor: Wenming Cao
Guest Editors: Cheng Liu and Man-Fai Leung

Contents

Contents

Contents

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9835236, 1 page
https://doi.org/10.1155/2023/9835236

Retraction

Retracted: Analysis of the Application Effect of Multidisciplinary Team Cooperation Model in Chronic Heart Failure under WeChat Platform

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Huang, Y. Su, and X. Mao, "Analysis of the Application Effect of Multidisciplinary Team Cooperation Model in Chronic Heart Failure under WeChat Platform," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4051955, 7 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9803857, 1 page
https://doi.org/10.1155/2023/9803857

Retraction

Retracted: Deep Learning of Subject Context in Ideological and Political Class Based on Recursive Neural Network

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Jiang and X. Gao, "Deep Learning of Subject Context in Ideological and Political Class Based on Recursive Neural Network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8437548, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9761741, 1 page
https://doi.org/10.1155/2023/9761741

Retraction

Retracted: Research on Kinetic Energy Recovery of Energy Vehicle ABS Solenoid Valve Based on the ELM Deep Learning Model

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Tu and L. Zhang, "Research on Kinetic Energy Recovery of Energy Vehicle ABS Solenoid Valve Based on the ELM Deep Learning Model," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6571085, 7 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9845814, 1 page
https://doi.org/10.1155/2023/9845814

Retraction

Retracted: Prediction and Analysis of Dynamic Changes of College Students' Ideological and Political Changes Based on Multiple Regression

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. Wang, "Prediction and Analysis of Dynamic Changes of College Students' Ideological and Political Changes Based on Multiple Regression," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5323699, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9825827, 1 page
https://doi.org/10.1155/2023/9825827

Retraction

Retracted: Research on Application of Ecological Sports Innovation in Efficient Development Based on DCN Deep Learning

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Hu, "Research on Application of Ecological Sports Innovation in Efficient Development Based on DCN Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9586509, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9784806, 1 page
https://doi.org/10.1155/2023/9784806

Retraction

Retracted: Strategies for Ideological and Political Education in Colleges and Universities Based on Deep Learning

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Sun, "Strategies for Ideological and Political Education in Colleges and Universities Based on Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5322677, 9 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 6116144, 9 pages
https://doi.org/10.1155/2023/6116144

Research Article

Volleyball Movement Standardization Recognition Model Based on Convolutional Neural Network

Baiyu Li and Ming Tian

Department of Physical Education and Research of Lanzhou University, Lanzhou, Gansu, China

Correspondence should be addressed to Ming Tian; tianming@lzu.edu.cn

Received 25 July 2022; Revised 18 September 2022; Accepted 5 October 2022; Published 25 January 2023

Academic Editor: Man Fai Leung

Artificial intelligence and deep learning have attracted much attention from researchers in industry and academia. The volleyball movement standardization and recognition model involve the application of artificial intelligence and deep learning. In order to solve the problem that human action in volleyball video is continuous and effective spatial and temporal features need to be extracted from the video stream, the Inception module is decoupled and heterogeneous, replacing the original 5×5 convolutional structures with two 3×3 convolutional structures, as well as replacing the 3×3 convolutional structures with 1×3 and a 3×1 convolutional structure with internal parameter optimization to ensure the accuracy of recognition. The model uses the input motion video RGB map as the spatial input and the optical flow map as the temporal input, and the two are weighted 1 : 1 for feature fusion. Experiments are conducted on the volleyball action video and homemade dataset in UCF101, and the experimental data show that the accuracy of the DNet volleyball action standardization recognition model proposed in this paper is 94.12%, which proves that the method improves the recognition ability of the model while speeding up the training speed. The research presented in this paper provides important theoretical guidance for artificial intelligence and deep learning.

1. Introduction

Artificial intelligence and deep learning have attracted much attention from researchers in industry and academia. The volleyball movement standardization and recognition model involve the application of artificial intelligence and deep learning. The research presented in this paper provides important theoretical guidance for artificial intelligence and deep learning. Based on the computer vision platform to achieve volleyball action standard recognition, which is an important trend in the application and development of artificial intelligence in sports with a very wide range of application prospects, the surveillance camera in real-time under the human action recognition technology support to identify the volleyball player's action standardization domain and record and make a warning can effectively improve the volleyball player's sports efficiency and volleyball sports level. In addition, through the analysis of the volleyball sports video, it can be clearly derived from the video whether player action standardization, volleyball sports tactical layout, and point-to-point analysis have an important role.

With the increasing maturity of convolutional neural network technology, human action recognition based on convolutional neural networks is widely concerned by domestic and foreign research scholars. Technical action detection in sports videos is an important application in the field of computer vision in sports, among which volleyball has distinct technical characteristics and has significant advantages for individual human action technology recognition and classification. Through the recognition and classification of both players' action techniques in sports videos, it can effectively improve the technical analysis ability and technical level of volleyball players. Therefore, the convolutional neural network volleyball action standard type recognition model has important research significance and application value.

For artificial intelligence and deep learning, in convolutional neural network-based volleyball action criterion recognition models, the extraction of effective action features as well as representation features from video sequences is an important part of action analysis, retrieval, and recognition, which directly affects the accuracy and robustness of its results. In the past decades, many video action recognition methods and action recognition datasets have emerged. With the successful application of convolutional neural networks in the image domain (image classification [1], target detection [2], and scene classification [3]), research on the use of convolutional neural networks for video action recognition has started to emerge, and successful applications from images to videos illustrate the superiority of convolutional neural network models. Therefore, convolutional neural network-based action recognition has become an active research area, and its algorithmic framework is shown in Figure 1.

2. Contribution

As an important application project of computer vision technology in the field of sports, volleyball player action standardness recognition has been a challenging research topic in the field of computer vision, and the volleyball standardness action recognition model not only requires the convolutional neural network to extract spatial information such as motion background and character appearance scene in the image but also needs to effectively extract motion information in the temporal dimension of the video. In summary, this paper combines the idea of a spatio-temporal dual channel and constructs a dual-stream DNet convolutional neural network volleyball action standardness recognition model based on Inception [4], which uses the idea of weighted fusion to connect the extracted spatio-temporal features so as to improve the training speed of the model and the accuracy of volleyball action standardness recognition. In addition to theoretical contributions, this paper also has some practical significance. As volleyball is a mainstream sport for the public, action recognition for volleyball players and volleyball fans can help improve people's ability to use volleyball and enhance the public's love for volleyball. This paper also has some general applicability and can be applied to other sports, in addition to volleyball, for action recognition.

3. Literature Review

For artificial intelligence and deep learning-aided volleyball movement standardization recognition models, human action recognition can be divided into behavior recognition based on traditional manual features, and behavior recognition based on depth features according to the way of feature extraction division. The specific division of behavior recognition is shown in Figure 2. Traditional feature extraction methods are generally designed manually to characterize action features through manual observation and design. Compared with traditional methods, deep learning methods do not use manual initiative to extract features, retain more valuable information in the video, and generally outperform traditional methods in terms of effectiveness. Deep learning methods applied to human behavior recognition should not only use the spatial information of the video but also the temporal information of the video, which is also the focus of the method research.

The essence of action recognition in volleyball belongs to the category of human action recognition classification. Convolutional neural networks were introduced to the video domain for action recognition. In 2014, Simonyan and Zisserman [5] proposed the first human action recognition method based on dual-stream neural networks, in which ImageNet was used to pretrain, RGB single frames and optical flow maps were used to train the dual-stream network, and finally fused for action recognition. The method makes full use of spatio-temporal feature information, improves the accuracy of action recognition, and lays a theoretical foundation for volleyball action recognition. In 2016, Wang et al. [6] proposed the time segment network (TSN), combining a sparse temporal sampling strategy with a video supervision method based on the classical dual-stream network. TSN is also composed of a spatial stream convolutional network and a temporal stream convolutional network. But unlike two-stream, which uses a single frame or a single stack of frames, TSN uses a series of short segments sparsely sampled from the entire video, each of which will give its own initial prediction of the behavior class, and the "consensus" of these segments to obtain the video-level prediction results. 2014 LSVRC challenge winner GoogleNet [7] network was proposed, with its structure of adding multiple Inception network models to the traditional deep convolutional neural network. The second deficiency is compensated by cross-practice, regularization, and data expansion techniques, and the network structure is named the temporal segmentation network (TSN). Due to the recent successful application of residual networks (ResNets) in deep network training [8], a novel spatial-temporal residual network model has been proposed that combines ResNets and a dual-flow model, which learns the spatio-temporal characteristics of the behavior hierarchically by connecting the residuals of the air-domain and time-domain flows.

4. Research Design

4.1. Convolutional Double-Stream Volleyball Action Standard Type Recognition Model Construction. For artificial intelligence and deep learning, the structure of a dual-stream DNet network can be divided into two feature extraction modes: spatial stream channel and temporal stream channel. The advantage of a DNet-based dual-stream convolutional neural network model over a traditional convolutional neural network is that it captures temporal information features in video data more clearly and efficiently. For single-frame RGB, spatial information generally expresses features mainly in the form of coordinate positions of things and fixed scenes, while for temporal information, more information about the continuous motion of the target can be

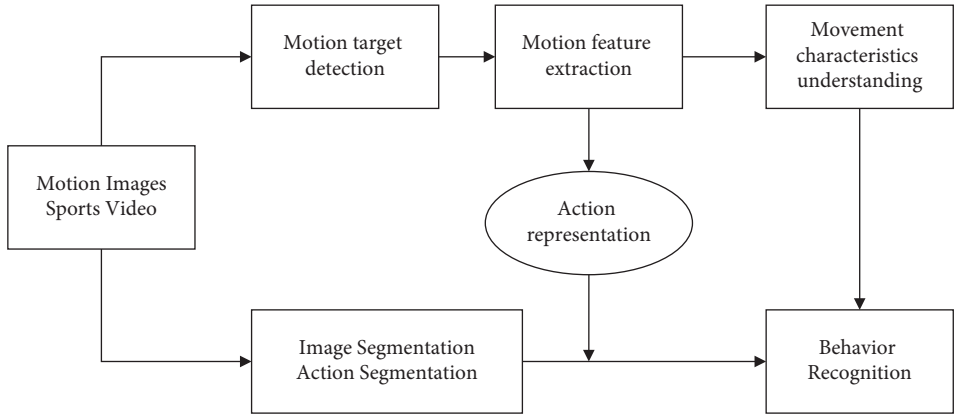


Figure 1: Flow chart of standard types of motion action recognition for artificial intelligence and deep learning.

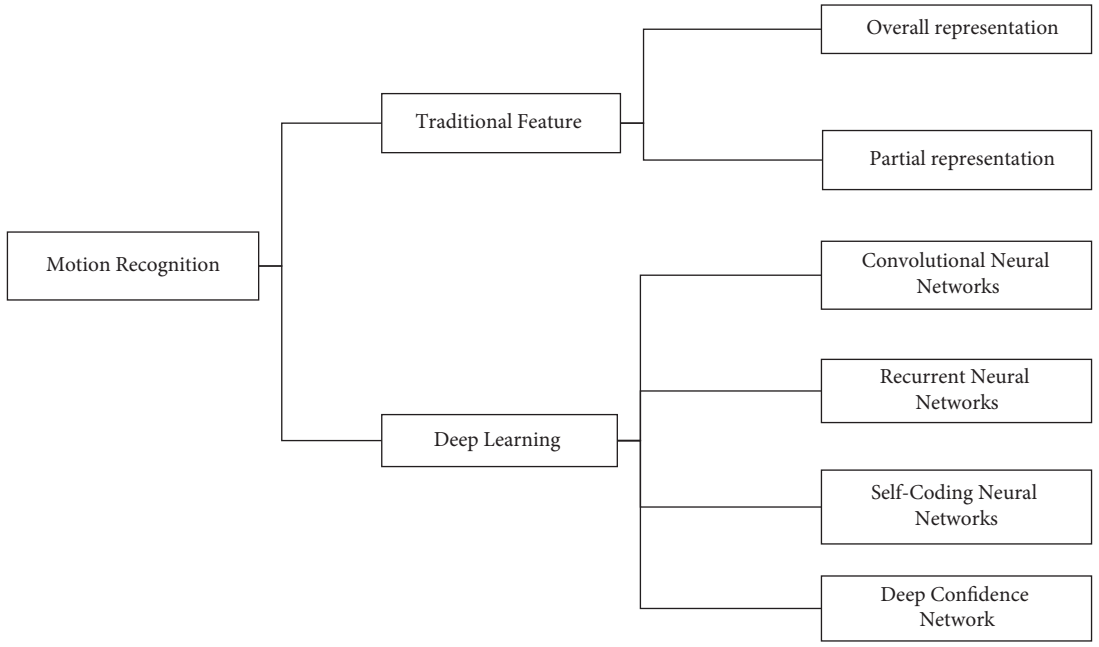


Figure 2: Classification of action recognition for artificial intelligence and deep learning.

conveyed in the form of the continuous motion of multiple optical flow frames. Both spatial and temporal stream channels are trained by applying the same network to the feature images, and subsequently, action recognition is achieved through spatio-temporal feature fusion.

The specific network structure of the dual-stream DNet network is shown in Figure 3, which uses a two-branch network architecture to capture the spatial and temporal information of the video, respectively. The air domain uses RGB images as input to extract appearance features, and the time domain uses optical flow information as input to extract temporal features, and then the two behavior recognition datasets are classified by weighted feature fusion of the two-channel features and by the multitask training method to remove overfitting and thus obtain better results.

4.2. GoogleNet Network Model. With artificial intelligence and deep learning, GoogleNet is a deep neural network model based on the Inception module introduced by Google, which won the ImageNet competition in 2014. GoogleNet differs from previous networks such as AlexNet4 and VGGl6 by increasing the depth of the network to obtain better training results, which deepens the core structure of the GoogleNet network since its Inception, as shown in Figure 4. Inception is to take multiple convolutional or pooling operations and put them together to assemble a network module, and design the neural network to assemble the whole network structure in a module. In networks that do not use this approach, we often layer only one operation, such as convolution or pooling [9], and the convolution kernel size for the convolution operation is also fixed size. However, in practical situations, different sizes of

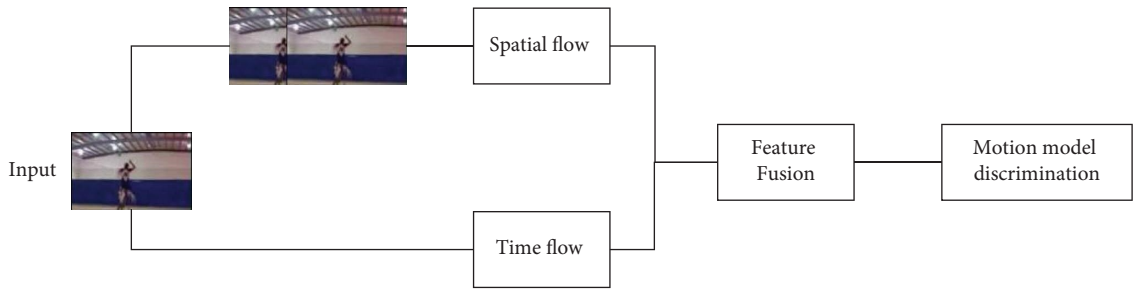


Figure 3: Structure of a dual-stream DNet convolutional neural network for artificial intelligence and deep learning.

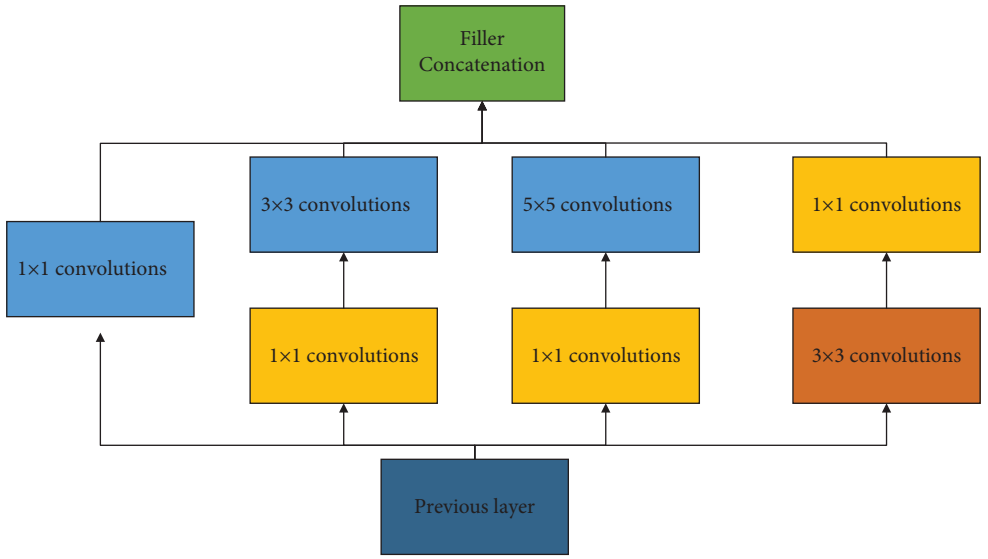


Figure 4: Inception module for artificial intelligence and deep learning.

convolutional kernels are needed at different scales so as to obtain the best performance. Moreover, for the same picture, different sizes of convolution kernels perform differently because they have different perceptual fields. Therefore, it is necessary to let the network choose the appropriate perceptual field by itself, and Inception can meet such a need. An Inception module provides multiple convolutional kernel operations in parallel, and the network chooses to use them by adjusting the parameters in the training process. The whole Inception structure is a series of multiple Inception modules.

The two main contributions of the Inception structure are the use of 1×1 convolution for lifting and dimensioning, as well as the simultaneous convolution and reaggregation at multiple dimensions. In Inception, a single convolutional kernel is changed to a model with three parallel convolutional kernels (1×1, 3×3, 5×5) and a single pooling layer. Each path is followed by a 1×1 convolution operation immediately before the convolution operation or after the pooling operation. More convolutions can be superimposed on the same size perceptual field to extract richer features, and the 1×1 convolution can also achieve the effect of dimensionality reduction, which reduces the computational complexity and improves the performance of the network. Therefore, in this paper, the GoogleNet network is selected as the base network for volleyball movement standardization and recognition.

4.3. Construction of DNet Network Model. With artificial intelligence and deep learning in this paper, we will use the improved GoogleNet, i.e., DNet, for volleyball action standardization and recognition research, and DNet optimizes the internal structure of the Inception module.

The DNet network structure improves the original three groups of parallel convolutional operations, named shallow EnInception-1 module, middle EnInception-2 module, and deep EnInception-3 module, respectively. Since the convolutional neural network with high accuracy, multiple parameters, and large computation has the problem of landing, the original convolutional structure is decoupled to improve the speed and accuracy of the deep convolutional neural network. In this paper, the original 5×5 convolutional structures of Inception-1 are decomposed into two 3×3 convolutional channels to reduce the model's computational parameters. In addition, the 3×3 convolutional structure of all Inceptions is modified into an asymmetric [10] convolutional pattern, i.e., two convolutional blocks of 1×3 and 3×1 are used instead of the original 3×3 convolutional structure to increase the stability of the model.

After improving the shallow Inception-1 module, the mid-layer Inception-2 module undergoes a larger structural improvement. Since the mid-layer module is more important for feature extraction than the shallow layer, the complete convolution operation is retained to make feature extraction more significant while accelerating the computational speed. First, the 5×5 sizes convolutional kernel in the traditional middle layer of the Inception-2 module structure is improved. A new convolutional layer is formed by connecting two 3×3 sized convolutional kernels in sequence, replacing the original large convolutional kernel structure with a new small convolutional network. The two "3×3" size convolutional kernels, which are closer to the output graph, are also replaced by "3×1" and "1×3" convolutional kernels in the middle Inception-2 module structure, as shown in the dashed part of Figure 5. The filter banks in the module are then extended to make the network structure wider rather than deeper, which reduces the dimensionality and results in a better performance of the neural network, which is more suitable for high-dimensional features such as the amplitude of the fine human movements in volleyball. Therefore, while the computation is faster, the accuracy of the model after adjustment is higher compared with the previous one, as shown in Figure 5. For the deeper Inception-3 structure, the original structure of the model is retained in this paper because the spatial concentration of the convolutional kernels decreases in the deeper layers of the network and the relatively larger convolutional kernels extract more abstract features and are, therefore, suitable for application in the deeper layers of the network.

In addition, an auxiliary classification layer batch is introduced to improve the convergence of the very deep network. The aim is to push useful gradients to lower layers to make them immediately useful and to improve convergence during training by resisting the vanishing gradient problem in very deep networks. The auxiliary classifier also plays the role of a regularization term. This is because the main classifier of the network performs better if the side branches are batch normalized or have a discarded layer. This also gives weak supporting evidence for the speculation that batch normalization acts as a regularization term [11].

4.4. Parameter Selection and Optimization

4.4.1. Auxiliary Classification Layer. The left panel of Figure 6 shows the traditional pooling method, which loses the information on the feature map, and the right panel shows the process of enlarging the feature map before pooling, which has the problem of computing too much brightness. In order to balance the gain and loss of both, one-half of the feature map is obtained by convolution first and one-half of the feature map is obtained by pooling, and finally stitching is performed, as shown in Figure 7. This Inception module is used for 35×35 down to 17×17 and 17×17 down to 8×8. The Inception module reduces the grid size while expanding the filter set. It is fast and avoids performance bottlenecks.

4.4.2. Introduce Dropout [12] Layer and Choose the Optimal Ratio. The two improved shallow Inception modules, the five middle Inception-4 modules with optimized structure, and the two high Inception-5 modules with larger convolutional kernels were sequentially entered after the maximum pooling operation, followed by a global average pooling, and the optimal dropout ratio of 0.5 was chosen by introducing the dropout layer at the fully connected layer.

4.4.3. Classification of Feature Images by AM-Softmax [13] Classifier Loses. Since this classifier has a larger class spacing and smaller intraclass distance for different action features, the AM-Softmax classifier with algorithm improvement based on Softmax [14] is chosen to be applied to volleyball action recognition, which makes the local action classification effect more obvious. The abovementioned improved DNet network for the GoogleNet network has more optimized performance compared with the original basic network.

4.5. Spatio-Temporal Feature Fusion Design. In this paper, the spatio-temporal fusion strategy chosen for action recognition of volleyball movement in the video is designed, and the fusion approach in the classification layer is selected for feature fusion. The fusion approach for the classification layer of the dual-stream network is to perform feature fusion in the later classification layer after the fully connected layer. The dual-stream network extracts the spatial and temporal features in the motion video separately, and the weights are shared between the two networks. After convolution, pooling, and fully connected layer operations in the DNet network, the features of the two streams are fused in the classification layer, as shown in Figure 8.

In this paper, the weighted fusion method is used for feature fusion at the classification layer. For the weighted fusion method, it can be expressed as summing spatial features and temporal features after defining different weight assignments. For feature fusion, the temporal flow network and the spatial flow network are fused using the weighted fusion method with a weighted fusion weight ratio of 5 : 5 for time: space, and the fusion timing is chosen for fusion at the classification layer.

5. Model Regression Results and Analysis

5.1. Introduction to the Dataset. The experimental data in this paper are the public dataset of UCF101 [15] and the homemade volleyball sports dataset, in which the video content of the homemade dataset includes a large number of volleyball standard action demonstrations with various angles in the video, various changes in lighting information, low pixels of the video, and a short video length. The UCF101 public is an upgraded version of the UCF50 [16] database, which contains 13,320 video samples and 101 action categories. The videos in each action category are divided into 25 groups, each containing 4–7 actions. There are about 120 videos of volleyball sports in the UCF101 dataset, and there are 134 homemade datasets for VolleyballSis. All the video

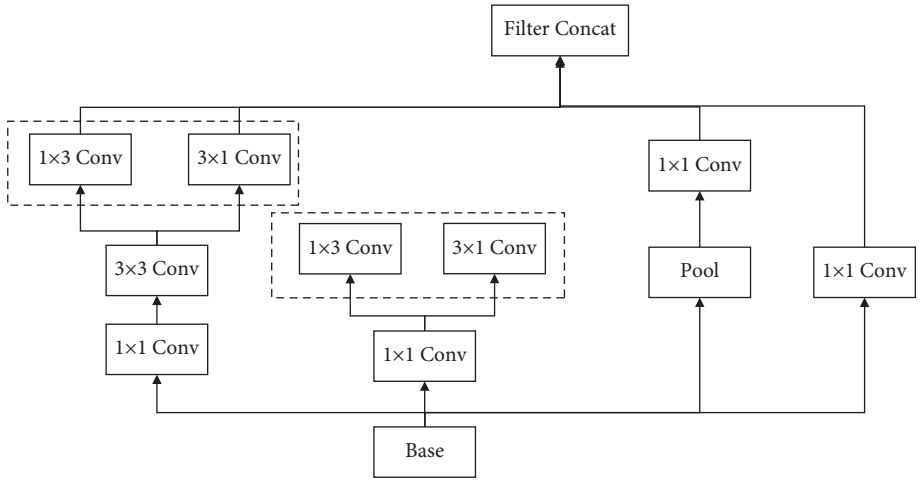


Figure 5: DNet Inception convolutional model for artificial intelligence and deep learning.

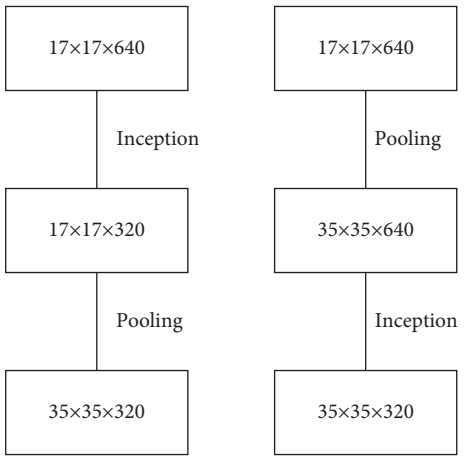


Figure 6: Traditional pooling method for artificial intelligence and deep learning.

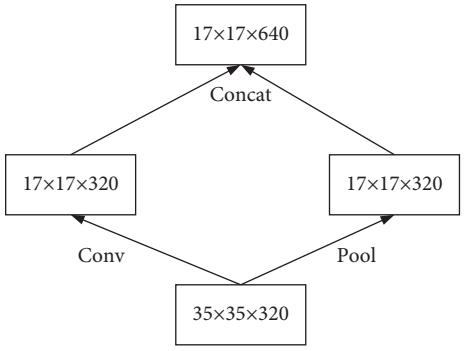


Figure 7: Auxiliary classification layer for artificial intelligence and deep learning.

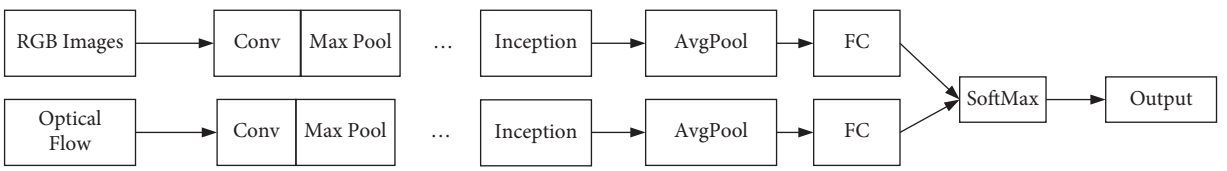


Figure 8: Spatio-temporal feature fusion design for artificial intelligence and deep learning.

contents include many single volleyball basic actions and volleyball sparring actions, with various angles in the videos, various changes of lighting information, low-pixel videos, and short video length. According to the analysis requirements of the technical characteristics of volleyball standard movements, the input data are divided into four categories: (1) serving; (2) matting; (3) passing; (4) dunking.

5.2. Introduction to the Experimental Environment. The experimental hardware environment is shown in Table 1.

The experimental environment in this paper is the PyTorch platform, in which the input dataset UCF101 and the homemade dataset are processed in RGB and optical flow formats, and for the GNet spatial branch, RGB format images with spatial step size step = 4 are used, and for the GNet temporal flow branch, $T = 8$ continuous optical flow images are overlaid. The size of each iteration in the model training process is 100, and the learning rate is 0.001. A total of 45 iterations are performed in this experiment.

5.3. Evaluation Indicators. Commonly used evaluation metrics are accuracy, detection rate, and recall, where accuracy represents the proportion of samples with correct predictions, as shown in the following equation:

$$Acc = \frac{TP + TN}{N}, \tag{1}$$

where TP and TN are the numbers of positive samples correctly identified as positive and negative samples correctly identified as negative, respectively, and N is the number of samples tested.

The detection rate is for a specific category, as shown in the following equation:

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{TP}{N}, \tag{2}$$

where N is the number of all detected target frames of a specific class, and the recall is calculated, which represents the ratio of correctly predicted target frames to all ground truths, as shown in the following equation:

$$\text{Recall} = \frac{TP}{TP + FN}. \tag{3}$$

To address the different requirements for the location deviation of target detection in different scenarios, an IOU threshold is usually given, above which the detection is considered successful. As well as considering the problem of category balancing, the performance of each category is usually found separately and then averaged for each category.

First, all the detection results are sorted, with higher scores being higher, and then the detection is judged to be successful in turn. First, find the intersection ratio of detection results (dets) and real targets (gt), i.e., IOU, and find the real target with the maximum IOU value for each detection result; if this maximum IOU exceeds a certain threshold, the det is matched against this gt and considered TP, noting that each gt can only be matched once. If the IOU of a det to all gt does not exceed the threshold or the gt with the maximum IOU exceeding the threshold has been matched, it is regarded as FP. After traversing all the dets from high to low, if there is still a gt that has not been matched, it is regarded as FN.

Table 1: Hardware environment configuration for artificial intelligence and deep learning.

Environment name	Specific configuration
Memory	64 GB
OS	Ubuntu20.04.1
GPU	GeForceRTX2080Ti11G
CPU	Intel®Core™i7-10700KF
Python/PyTorch	3.8/1.8.0
Cuda	10.2

After getting TP, FP, and FN, the precision and recall can be calculated. But different tasks have different requirements for precision and recall; some tasks require higher recall and wrong detection is acceptable, while other tasks require higher precision and missed detection is acceptable. To evaluate recall and precision as a whole, we select the first n of the ranked det ($n = 1, 2, 3, \ldots$) resulting in a recall-precision graph, and obtaining the AP value by calculating the area under this curve. The AP values for each category are averaged to obtain mAP.

5.4. Description of Experimental Results. The experimental results are shown in Table 2.

Figure 9 shows the line graph of the DNet recognition rate, where the horizontal coordinates represent each mainstream algorithm and the vertical coordinates represent the mAP values of each algorithm on the UCF101 dataset. From the experimental results, it can be analyzed that using the weighted fusion method by setting the weight ratio to time: space 5 : 5, decoupling and heterogeneity of the GoogleNet three-layer convolutional kernel, after using the auxiliary classification layer for fusion, the average accuracy of volleyball action standardization recognition of the feature fusion dual-stream DNet convolutional neural network reached 94.12% at this time, which was higher than other network models on the 80.45% on the homemade dataset VolleyballSis, which is higher than other network models. For the model structure of the algorithm, most of the algorithm data modalities used in this paper are RGB and optical stream OF. In the to-stream structure, DNet uses RGB and OF as input, respectively. On this basis, the three-layer branching Inception structure is split, and the original 5×5 convolutional structure is split into two 3×3 convolutional kernels to reduce the computational parameters and improve the computational speed. In addition, the heterogeneous convolutional network structure is introduced, and the 3×3 convolutional model is further split into two convolutional models of 1×3 and 3×1 to improve the accuracy of the algorithm model without increasing the computational effort, and good results are obtained.

Table 2: Comparison of experimental results of DNet for artificial intelligence and deep learning.

Network structure	UCF101 *map* (%)	VolleyballSis *map* (%)	Input of network
GoogleNet	82.94	73.62	RGB
DS VGG	90.66	74.43	RGB + OF
DS ResNet	94.52	70.65	RGB + OF
DS GoogleNet [17]	93.32	79.12	RGB + OF
DNet	94.12	80.45	RGB + OF

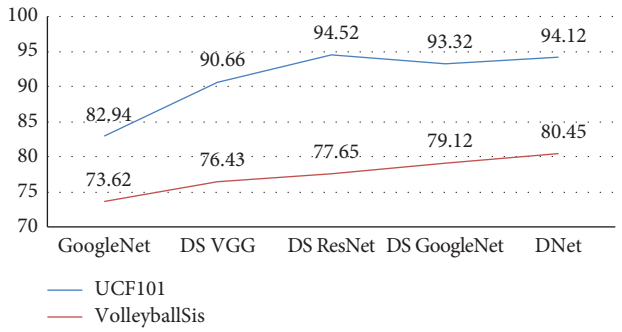


Figure 9: DNet accuracy comparison for artificial intelligence and deep learning.

6. Conclusion and Suggestion

Artificial intelligence and deep learning have attracted much attention from researchers in industry and academia. The volleyball movement standardization and recognition model involve the application of artificial intelligence and deep learning. The research presented in this paper provides important theoretical guidance for artificial intelligence and deep learning. The heterogeneous DNet convolutional neural network with spatio-temporal feature fusion for volleyball action standardization recognition proposed in this paper greatly reduces the training parameters of the model, thus achieving the goal of improving the model speed. The introduction of heterogeneous convolution also ensures the stability problem of the model with constant parameters. It effectively improves the recognition ability and generalization ability of the model. However, there are still some problems and difficulties in volleyball action, and standardization recognition.

(1) Strong dependence of the model on the motion viewpoint problem.

When the model is applied to the actual volleyball sports scenario, most of the captured videos are collected from random viewpoints. Therefore, in the actual application scenario of the model, the model is required to decouple the motion viewpoint and reduce the dependence problem of the model on the motion viewpoint. Nowadays, most of the data sets are focused on a single viewpoint or captured by a specific viewpoint, so the future improvement algorithm can cope with the change of complex viewpoints and handle multiple random viewpoints.

(2) Weakly supervised algorithm problem.

Supervised learning-based behavior recognition relies heavily on the annotation of video action tags, and the annotation of video images can consume a lot of resources. With the continuous development of deep learning, the cost of video image labeling becomes more and more expensive. How to use the algorithm itself to perform image labeling has become a hot topic of research nowadays. It is based on weakly supervised or unsupervised volleyball action standard-type recognition, which can be classified without manual labeling or with only a small amount of labeling.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Washington, DC;USA, June 2014.

[3] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1915–1929, 2013.

[4] P. Ballester and R. M. Araujo, "On the performance of GoogLeNet and AlexNet applied to sketches," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Brazil, February 2016.

[5] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[6] L. Wang, Y. Xiong, Z. Wang et al., "Temporal segment networks: towards good practices for deep action recognition," in *Proceedings of the European Conference on Computer Vision*, pp. 20–36, Springer, Cham, Heidelberg, Germany, 2016 October.

[7] Z. Zhong, L. Jin, and Z. Xie, "High performance offline handwritten Chinese character recognition using googlenet and directional feature maps," in *Proceedings of the 2015 13th International Conference on Document Analysis and*

Recognition (ICDAR), pp. 846–850, IEEE, Tunis, Tunisia, 2015, August.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[9] W. J. Liu, X. J. Liang, and H. C. Qu, "Learning performance of convolutional neural networks with different pooling models," *Journal of Image and Graphics*, vol. 21, no. 9, pp. 1178–1190, 2016.

[10] X. Ding, Y. Guo, G. Ding, and J. Han, "Acnet: strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1911–1920, Seoul, South Korea, October 2019.

[11] Y. Wen, J. Wang, B. Huang, and J. M. Zurada, "Convergence analysis of inverse iterative neural networks with L2 penalty," *Journal of Applied Computer Science Methods*, vol. 8, no. 2, pp. 85–98, 2016.

[12] P. Baldi and P. J. Sadowski, "Understanding dropout," *Advances in Neural Information Processing Systems*, vol. 26, 2013.

[13] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 926–930, 2018.

[14] S. Gold and A. Rangarajan, "Softmax to softassign: neural network algorithms for combinatorial optimization," *Journal of Artificial Neural Networks*, vol. 2, no. 4, pp. 381–399, 1996.

[15] K. Soomro, A. R. Zamir, and M. Shah, "A dataset of 101 human action classes from videos in the wild," *Center for Research in Computer Vision*, vol. 2, no. 11, 2012.

[16] K. K. Reddy and M. Shah, "Recognizing 50 human action categories of web videos," *Machine Vision and Applications*, vol. 24, no. 5, pp. 971–981, 2013.

[17] N. Yang, Z. Zhang, J. Yang, Z. Hong, and J. Shi, "A convolutional neural network of GoogLeNet applied in mineral prospectivity prediction based on multi-source geoinformation," *Natural Resources Research*, vol. 30, no. 6, pp. 3905–3923, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2576559, 9 pages
https://doi.org/10.1155/2022/2576559

Research Article

Research on the Guidance of Youth Labor Education Based on the "Combination of Education and Production Labor" Program Based on the Deep Learning Model

Linyu Xiao [1] **and Xiaoyi Liao** [2]

[1]*Hunan Agricultural University, Institute of Marxism, Changsha 410128, Hunan, China*
[2]*Hunan Agricultural University, College of Horticulture, Changsha 410128, Hunan, China*

Correspondence should be addressed to Xiaoyi Liao; liaoxiaoyi@hunau.edu.cn

Received 5 June 2022; Revised 27 June 2022; Accepted 30 June 2022; Published 11 October 2022

Academic Editor: Wenming Cao

At present, there is a lack of research on Marx's idea of "combining education and productive labor" and its guiding significance for youth labor education, and no effective teaching model has been formed. In response to this problem, this study proposes a semi-supervised deep learning model based on u-wordMixup (SD-uwM). When there is a shortage of labeled samples, semi-supervised learning uses a large number of unlabeled samples to solve the problem of labeling bottlenecks. However, since the unlabeled samples and labeled samples come from different fields, there may be quality problems in the unlabeled samples, which makes the generalization ability of the model worse., resulting in a decrease in classification accuracy. The model uses the u-wordMixup method to perform data augmentation on unlabeled samples. Under the constraints of supervised cross-entropy and unsupervised consistency loss, it can improve the quality of unlabeled samples and reduce overfitting. The comparative experimental results on the AGNews, THUCNews, and 20Newsgroups data sets show that the proposed method can improve the generalization ability of the model and also effectively improve the time performance. The study found that the SD-uwM model uses the u-wordMixup method to enhance the unlabeled samples and combines the idea of the Mean Teacher model, which can significantly improve the text classification performance. The SD-uwM model can improve the generalization ability and time performance of the model, respectively, 86.4 ± 1.3 and 90.5 ± 1.3. Therefore, the use of SD-uwM in Marx's program is of great practical significance for the guidance process of youth labor education.

1. Introduction

From the standpoint of the proletariat, Marx opposed the exploitation and oppression of workers by the bourgeoisie, combined with the actual development of the modern machinery industry to construct a theoretical system combining productive labor and education, and demonstrated its inevitability in the process of social development [1]. This theory also had a profound impact on Marx's educational thought and provided important guidance on how the proletariat should carry out educational activities. In our country, it is very necessary to excavate the connotation of Marx's thought of "combining education with productive labor," which has theoretical value and practical significance. This thought of Marx can guide our country to educate and cultivate laborers with professional quality and promote the improvement of productivity, thereby promoting social development [2]. Since the 18th National Congress of the Communist Party of China, General Secretary Xi Jinping has delivered many important speeches on the topic of labor based on Marx's discussion on labor. He highly recognized the important position of labor in socialist countries, and also clarified the importance of cultivating young people's correct labor values and loving physical labor in the new era. General Secretary Xi Jinping clearly stated at the 2018 National Education Conference: "We must educate and guide students to advocate and respect labor." [3] In today's society, we must attach importance to labor

education and give full play to the educating function of physical labor and use Marxist labor Ideology arm the minds of young people, guide them to love labor, and train young people into laborers who meet the needs of social development in practice, so that they can better serve my country's modernization drive. At the same time, the "Opinions on Comprehensively Strengthening Labor Education in Colleges and Elementary Schools in the New Era" issued by my country pointed out that labor education is an indispensable part and an important part of education in the education system of socialist countries. We must pay attention to the unique educational value of labor and strengthen modern labor education [4]. Through classroom teaching, labor practice, etc., effective measures are taken to change the attitude of young people towards labor and to curb the phenomena of extravagance, lack of progress, and greed for pleasure in the campus. In this context, it is more necessary to study Marx's thought of "combining education and productive labor" and its guiding significance for youth labor education.

The famous Russian educator Ushinsky believes that education is an important hand to guide students to respect and love labor, and it is also an indispensable part of cultivating students' labor habits [5, 6] Lenin attached great importance to the combination of education and productive labor. He inherited and developed Marx's idea and pointed out that the ideal realization of the future society cannot be separated from the combination of education and productive labor of the new generation. In a socialist country, labor courses should be an indispensable course in school education. Students can master labor skills by participating in voluntary labor and combine practice and theory to help themselves better master scientific knowledge and become socialist laborers [7]. Makarenko believes that labor has an important impact on a person's future development. Individual labor can improve one's living standard and enhance happiness; in collective labor, people help each other and can establish good interpersonal relationships [8]. In the book "Dedicate the whole heart to the child," Sukhomlinsky said that the beauty of human beings is the brightest in labor, so an important task of education is to make the child's surroundings full of the natural world and through labor a world created and built. Labor has the function of aesthetic education, and schools should make labor a need for students' spiritual life and become a powerful educational force on this basis. Lenin, Makarenko, and Sukhomlinsky all stood from the perspective of the proletariat and believed that the ultimate goal of labor education was to train laborers for a communist society. The discussion of this topic by British scholars can be traced back to the early utopian socialist Thomas More who required children to learn agricultural knowledge in school, go to work in the fields, and required every youth to learn at least one kind of handicraft. Owen also mentioned that people will form their own characters during labor, so labor education should be paid attention to in childhood [9]. Locke believes that labor can make people physically and mentally healthy and overcome many vices. Students can choose one of crafts, gardening, or agriculture to practice [10]. German educator Comenius said in "The Great Teaching Theory" that when realizing the ideal of education through labor, attention should be paid to adapting to the natural order, adapting to human nature and age characteristics. Pestalozzi, a Swiss educator, was the first educator to put the idea of combining education and labor into practice in the history of Western education. He believed that this method could coordinate and develop people's abilities in all aspects and advocated the integration of education and labor [11].

Through the consistency training framework [12], based on the wordMixup data enhancement method, this paper proposes a u-wordMixup (unlabeled sample word mixture) data enhancement method for unlabeled samples. Under the constraint of loss, the enhancement generates high-quality additional training samples and reduces overfitting. Based on the u-wordMixup method, a new semi-supervised deep learning model (SD-uwM) is proposed using the Mean Teacher model for consistent training. The u-wordMixup data augmentation method enhances unlabeled samples with the goal of reducing unsupervised consistency loss, constrains the quality of unlabeled training samples, and reduces model overfitting. The objective loss function combines supervised cross-entropy loss and unsupervised consistency loss and uses the MeanTeacher method for consistency training to improve the generalization ability of the model. Optimize the key words of the "Combination of Education and Productive Labor" program to achieve the optimal application of youth education.

2. Principles and Methods

2.1. wordMixup Data Enhancement. wordMixup is a data enhancement method for labeled samples. The idea is to interpolate the word embedding vectors of two samples to generate a new sample word embedding matrix as an enhanced sample [13]. Given a pair of labeled samples (x_i, y_i) and (x_j, y_j), perform word embedding to get (x_i, y_i) and (x_j, y_j), where $x_i \in \mathrm{RN} \times d$, $x_j \in \mathrm{RN} \times d$ is the word embedding matrix of text x_i and x_j, N is the number of words, d is the word vector dimension, and y_i and y_j are the corresponding class labels. Then perform interpolation according to formulas (1) and (2) to obtain a new sample (x_{ij}, y_{ij}), where x_{ij} is the word embedding matrix of the enhanced sample, and y_{ij} is its class label.

$$\tilde{\mathbf{x}}_{ij}^{k} = \lambda \mathbf{x}_i^k + (1-\lambda)\mathbf{x}_j^k, \quad k = 1 \ldots N, \tag{1}$$

$$\tilde{\mathbf{y}}_{ij} = \lambda \mathbf{y}_i + (1-\lambda)\mathbf{y}_j. \tag{2}$$

Among them, x_i^k and x_j^k represent the word vector of the kth word in the text x_i and x_j, respectively, $\lambda \in [0, 1]$ is the interpolation weight factor, and x_{ij}^k is the word vector of the kth word of the new sample generated by interpolation. Perform word vector interpolation on each word in (x_i, y_i) and (x_j, y_j) one-to-one to obtain the embedding matrix x_{ij} of the new sample, where y_{ij} is the class label of x_{ij}, and (x_{ij}, y_{ij}) is the enhanced additional training samples.

The wordMixup method achieves good results in supervised text classification. But unlabeled in semi-supervised learning has no labels, how to generate pseudo labels for its

interpolation? To this end, based on wordMixup, this paper proposes an improved data enhancement method u-word-Mixup for unlabeled samples, which will be introduced in the following chapters [14].

2.2. SD-uwM Semi-Supervised Deep Learning Model

2.2.1. SD-uwM Model. The framework of the semi-supervised deep learning model (SD-uwM) is shown in Figure 1 [15]. It uses the idea of the Mean Teacher model to construct a teacher model T and a student model S, using labeled samples and unlabeled samples, based on supervised cross-entropy loss and unsupervised consistency loss objective functions for semi-supervised deep learning [16].

As shown in Figure 1, $DL = \{(xli, yli)\}\mathrm{Nl\,i} = 1$ represents the labeled training sample set, $DU = \{(\mathrm{xui})\}\mathrm{Nui} = 1$ represents the unlabeled training sample set, Nl represents the labeled sample set number, and Nu represents the number of unlabeled samples. T is the teacher model, S is the student model, and T and S have the same network structure. LS stands for supervised loss, LUS stands for unsupervised loss, and L stands for consistency objective loss function [17].

SD-uwM trains and learns simultaneously on labeled samples and unlabeled samples based on a consistent target loss function L [18]. As shown in the left half of Figure 1, the student model S computes a supervised cross-entropy loss LS on DL. At the same time, as shown in the right half of Figure 1, the student model S uses the u-wordMixup method to enhance the unlabeled samples. According to the prediction of the enhanced samples by the student model S and the prediction of the unlabeled samples by the teacher model T, the calculation is unsupervised and consistent. The performance loss LUS, LS and LUS together constitute the consistency target loss function L of the SD-uwM model. After many iterations, the SD-uwM model is trained to output the parameters of the student model S as the parameters of the final classification model [19].

2.2.2. u-wordMixup Data Enhancement. Using the Mean Teacher model idea and consistency training [20], the target loss function L of SD-uwM takes into account both supervised cross-entropy loss and unsupervised consistency loss and is defined as shown in

$$L = L_S + \beta L_{US}, \tag{3}$$

where LS is the supervised cross-entropy loss on the labeled samples DL, LUS is the unsupervised consistency loss on the unlabeled samples DU, and β is the scale coefficient. LS is the supervised loss of the student model S on the labeled sample set DL, which is calculated as

$$L_S = E_{\mathbf{x}_i^l, y_i^l \in D_L}\left[-y_i^l \log\ p_{\theta_l}\left(\mathbf{x}_i^l\right)\right]. \tag{4}$$

Among them, y_i^l is the true label of the labeled sample x_i^l, θ' represents the parameter of the student model S, and $p'_\theta(x_i^l)$ represents the predicted pseudo-label of the sample x_i^l by the student model S, that is, y_i^l.

Based on wordMixup, a u-wordMixup data augmentation method for unlabeled samples is proposed as part of the SD-uwM model [21]. Different from the wordMixup method, the interpolation operation object of the u-word-Mixup method has no real class label. The u-wordMixup method is shown in Figure 2 [22].

As shown in Figure 2, x_i^u and x_j^u represent two unlabeled samples, where $x_i^u \in \mathrm{RN} \times d$, $x_j^u \in \mathrm{RN} \times d$, N is the number of words, and d is the word vector dimension. The feature interpolation of the word vector is performed for each word in x_i^u and x_j^u one-to-one, and the embedding matrix $\tilde{x}_{ij}^u$ of a new unlabeled sample is obtained as an additional training sample [23]. The teacher model T predicts the unlabeled samples x_i^u and x_j^u to generate pseudo-labels $\hat{y}_i^u$ and $\hat{y}_j^u$ and performs pseudo-label interpolation on $\hat{y}_i^u$ and $\hat{y}_j^u$ to obtain $\tilde{y}_{ij}^u$, which is used as the pseudo-label of the enhanced sample $\tilde{x}_{ij}^u$. Then the student model S predicts the enhanced sample $\tilde{x}_{ij}^u$ to get the predicted label $\hat{y}_{ij}^u$. Among them, the calculation of feature interpolation and pseudo-label interpolation are as follows:

$$\tilde{\mathbf{x}}_{ij}^{uk} = q_\lambda\left(\mathbf{x}_i^u, \mathbf{x}_j^u\right) = \lambda \mathbf{x}_i^{uk} + (1-\lambda)\mathbf{x}_j^{uk}, \quad k = 1 \ldots N, \tag{5}$$

$$\tilde{\mathbf{y}}_{ij}^u = q_\lambda\left(\hat{\mathbf{y}}_i^u, \hat{\mathbf{y}}_j^u\right) = \lambda \hat{\mathbf{y}}_i^u + (1-\lambda)\hat{\mathbf{y}}_j^u. \tag{6}$$

Among them, $q\lambda\ (x_i^u, x_j^u)$ represents the u-wordMixup data enhancement transformation, x_{ki}^u and x_{kj}^u represent the word vector of the kth word of the samples x_i^u and x_j^u, respectively, $\lambda \in [0, 1]$ is the interpolation weight factor, and $\tilde{x}_{kij}^u$ is the interpolation. The word vector of the kth word of the generated augmented sample. $\hat{y}_i^u$ is the predicted pseudo-label of x_i^u by the teacher model T, $\hat{y}_j^u$ is the predicted pseudo-label of x_i^u by the teacher model T, and $\tilde{y}_{ij}^u$ is the pseudo-label generated by interpolation, that is, the pseudo-label of $\tilde{x}_{ij}^u$.

Based on consistent training, the pseudo-label predicted by the student model S for the enhanced unlabeled sample $\tilde{x}_{ij}^u$ is $\hat{y}_{ij}^u$, which should be as consistent as possible with the pseudo-label $\tilde{y}_{ij}^u$ generated by interpolation, that is, $\tilde{y}^u{}_{ij} \approx \hat{y}^u{}_{ij}$, and ideally the two are equal. Therefore, the unsupervised consistency loss LUS is computed as

$$L_{US} = E_{\mathbf{x}_i^u, \mathbf{x}_j^u \in D_U} E_{\tilde{\mathbf{x}}_{ij}^u \sim q_\lambda\left(\mathbf{x}_i^u, \mathbf{x}_j^u\right), \lambda \in [0,1]} \left[-\left(\lambda p_\theta\left(\mathbf{x}_i^u\right) + (1-\lambda) p_\theta\left(\mathbf{x}_j^u\right)\right) \log\ p_{\theta'}\left(\tilde{\mathbf{x}}_{ij}^u\right)\right], \tag{7}$$

where θ denotes the parameters of the teacher model T, θ' denotes the parameters of the student model S, and θ is the moving average of θ'. $p_\theta(x_i^u)$ represents the predicted pseudo-label of the sample x_i^u by the teacher model T, $p_\theta(x_j^u)$ represents the predicted pseudo-label of the sample x_j^u by the teacher model T, and $p_{\theta'}(\tilde{x}_{ij}^u)$ represents the prediction of the newly generated sample $\tilde{x}_{ij}^u$ by the student model S Pseudo tags. $\lambda \in [0, 1]$ is the interpolation weight factor, and $q\lambda(x_i^u, x_j^u)$ represents the u-wordMixup data augmentation transformation [24].

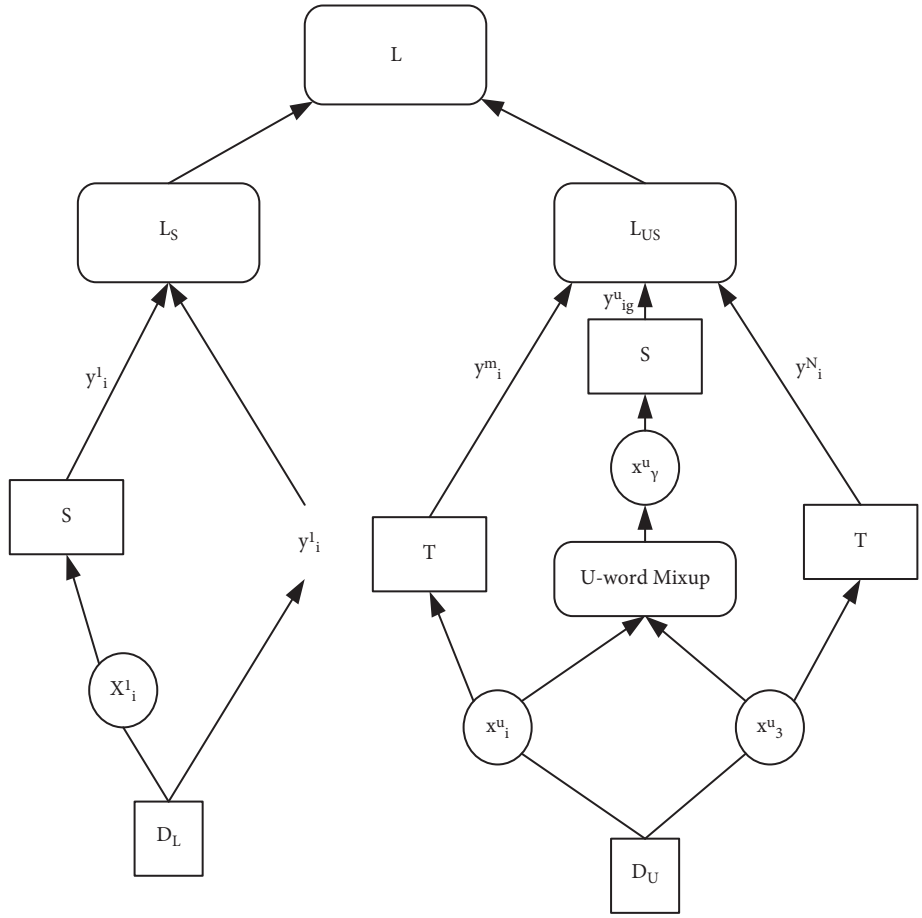


Figure 1: SD-uwM model.

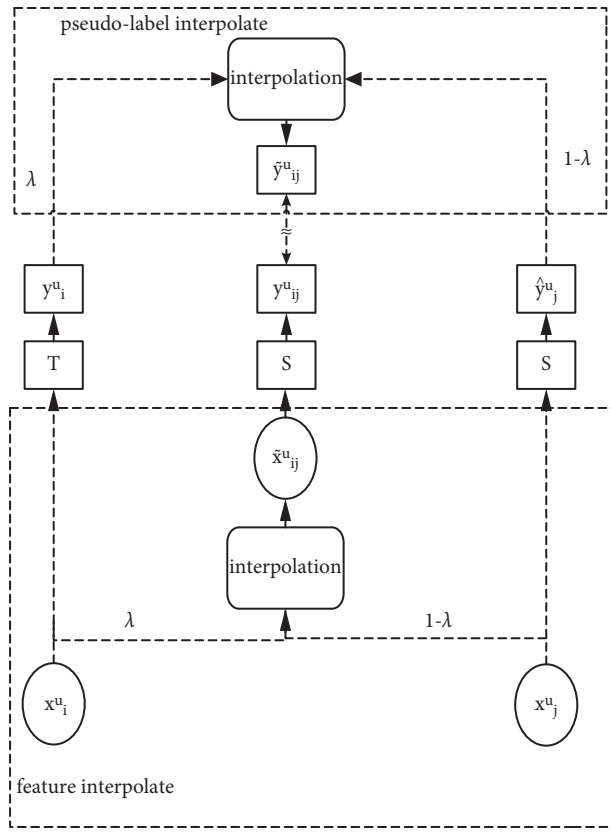


Figure 2: u-wordMixup method.

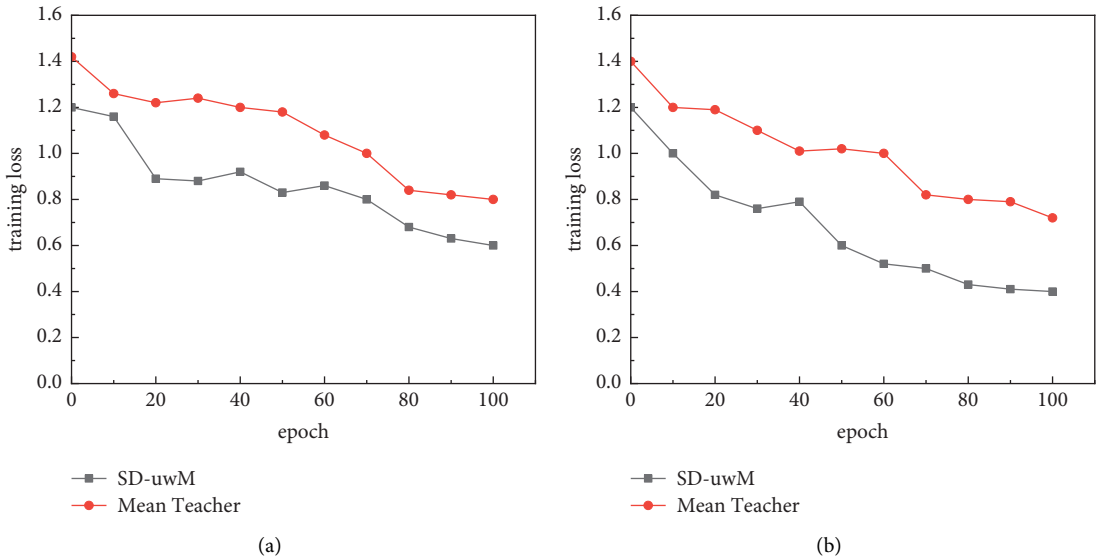


Figure 3: Comparison of training loss between SD-uwM and Mean Teacher when TextCNN is selected: (a) Nl = 200, Nu = 3000 (THUCNews data set); (b) Nl = 300, Nu = 5000 (AGNews data set).

The deep semi-supervised learning SD-uwM model uses the u-wordMixup method to enhance the unlabeled samples. The unsupervised consistency loss LUS aims to reduce the consistency loss and constrain the quality of the enhanced unlabeled samples. Combined with the Mean Teacher model, the teacher model T and the student model S are constructed, and the labeled samples and unlabeled samples are trained. The weighted summation of LS and LUS is used as the objective function L of the model SD-uwM [25].

2.2.3. SD-uwM Model Application. The algorithm description of SD-uwM is shown in Algorithm 1 [26]. The objective function L of the algorithm takes into account the supervised cross-entropy loss LS and the unsupervised consistency loss LUS, and the constraint enhancement generates unlabeled training samples [27]. In each iteration, according to the objective function L, the parameter θ' of the student model S is optimized. After many iterations, the parameter θ' of the optimal student model S is finally obtained.

3. Test Analysis

3.1. Data Set. This article selects three data sets: AGNews, 20Newsgroups and THUCNews. AGNews selects four categories of "world," "politics," "education" and "labor," and 20Newsgroups selects "alt.atheism," "soc.religion.Christian," "comp.graphics," and "sci.med" 4 categories, THUCNews selects 4 categories of "Education," "Labor," "Program," and "Technology." The scaling factor β is set to 1 in Algorithm 1.

3.2. Analysis of Experimental Results. The comparison method in the experiment is as follows:

(1) SD-uwM: the semi-supervised deep learning model based on u-wordMixup data augmentation proposed in this paper

(2) wM-SL: a supervised text classification method based on wordMixup data augmentation method in the literature

(3) SL: supervised text classification method without data augmentation

(4) Mean Teacher: semi-supervised method applied to image classification in the literature, modified for text classification tasks

3.2.1. Comparison of Target Loss between SD-uwM and Mean Teacher. In order to verify the effectiveness of the u-wordMixup method, a comparative experiment was conducted on SD-uwM and Mean Teacher models, and the training loss changes are shown in Figure 3.

As can be seen from Figure 3, compared to the Mean Teacher model, the SD-uwM model using the u-wordMixup data augmentation method has lower training loss. This is because the SD-uwM model objective loss function is more realistic, in which LUS is combined with the u-wordMixup method, aiming to reduce the unsupervised consistency loss, which can improve the quality of unlabeled samples, thereby improving the performance of the model.

3.2.2. Classification Comparison between SD-uwM Model and Other Methods

(1) Comparison of the Accuracy of SD-uwM Model with Other Methods. SD-uwM model and Mean Teacher, wM-SL, SL model on AGNews(Nl = 300, Nu = 5000),

Table 1: Classification comparison of four models on three classification results.

Model	Network structure	Accuracy		
		AGNews	THUCNews	20Newsgroups
SL	LSTM	75.4 ± 1.1	77.5 ± 1.3	71.5 ± 1.3
wM-SL		80.4 ± 1.3	83.2 ± 1.2	75.4 ± 1.3
Mean Teacher		82.1 ± 1.3	86.1 ± 1.3	77.5 ± 1.1
SD-uwM		90.4 ± 1.2	91.4 ± 1.3	85.4 ± 1.2
SL	TextCNN	76.4 ± 1.2	78.4 ± 1.4	71.2 ± 1.2
wM-SL		80.5 ± 1.2	84.5 ± 1.3	75.3 ± 1.2
Mean Teacher		83.6 ± 1.1	86.1 ± 1.5	78.1 ± 1.3
SD-uwM		91.2 ± 1.3	92.2 ± 1.3	86.2 ± 1.1

THUCNews(Nl = 300, Nu = 5000), and 20 Newsgroups (Nl = 200, Nu = 2000) The comparative experimental results are shown in Table 1.

As can be seen from Table 1, on the three data sets, regardless of whether the network structure is LSTM or TextCNN, the classification accuracy of the SD-uwM model is better than that of the SL, wM-SL, and Mean Teacher models, up to 90.4. The SL model is a supervised learning method, which requires a large number of labeled samples to achieve better performance. The wM-SL model only enhances the labeled samples, and the Mean Teacher model does not use the u-wordMixup method to enhance the samples. The SD-uwM model uses the u-wordMixup method to enhance the data of unlabeled samples and uses the unsupervised consistency loss constraint to improve the generalization ability of the model.

(2) Comparison of Classification Performance of SD-uwM Model with Increasing Iterations. The number of labeled samples and the number of unlabeled samples are fixed. As the number of iterations increases, the changes in the Macro-F1 value of the SD-uwM model and the SL, wM-SL, and Mean Teacher models are compared and analyzed. On AGNews (Nl = 300, Nu = 5000), THUCNews (Nl = 300, Nu = 5000), and 20Newsgroups (Nl = 200, Nu = 2000), the experimental results are shown in Figures 4 and 5.

It can be seen from Figures 4 and 5 that with the increase of the number of iterations, although the indicators of the SL, wM-SL, Mean Teacher, and SD-uwM models generally show an upward trend and converge to a certain upper limit, the SD-uwM model classification. The performance is significantly better than SL, wM-SL, and Mean Teacher models. As shown in Figure 5(a), using LSTM on AGNews, compared with Mean Teacher, wM-SL, and SL, the Macro-F1 of SD-uwM reaches 90.3%, which is increased by 8%, 9.9%, and 14.5%, respectively. This is because SD-uwM uses the u-wordMixup method to target unsupervised consistency loss for unlabeled sample enhancement, which can reduce overfitting and improve classification performance.

As can be seen from the above Figure 5, in order to verify the influence of unlabeled samples on the SD-uwM model, the number of labeled samples Nl = 300 is fixed on AGNews and THUCNews, the number of labeled samples Nl = 200 is fixed on 20Newsgroups, and the number of unlabeled samples is constantly increasing. Compare the classification results of SD-uwM model with SL, wM-SL, and Mean Teacher models. It can be seen that the indicators of SD-uwM and Mean Teacher show an upward trend with the increase of unlabeled samples, but the classification results of SD-uwM are significantly better than that of Mean Teacher, wM-SL, and SL. Using LSTM on THUCNews, compared with Mean Teacher, wM-SL, and SL, the Macro-F1 of SD-uwM has reached 91.4%, an increase of 5.3%, 8.2%, and 13.9%, respectively. It can be seen that since the SD-uwM model uses the u-wordMixup method to enhance the unlabeled samples and combines the idea of the Mean Teacher model, it can improve the text classification performance.

3.2.3. Time Performance Analysis of the SD-uwM Model. When selecting unlabeled samples, the usual semi-supervised learning algorithm needs to calculate the similarity matrix between the unlabeled samples and the labeled samples, which will increase the time complexity. The SD-uwM model in this paper is random sampling, and there is no need to calculate the similarity between the two samples. The temporal performance comparison between SD-uwM model and typical semi-supervised learning method Co-training is shown in Table 2.

As can be seen from Table 2, the classification accuracy of the SD-uwM model is 86.4% ± 1.3 and 90.5% ± 1.3 in the two different data sets, respectively, and the classification accuracy of the SD-uwM model is significantly higher than that of the co-training model. The classification accuracy was 83.3% ± 1.2 and 88.4% ± 1.2, respectively. At the same time, the time performance of SD-uwM model is significantly better than that of co-training, in which SD-uwM model time is basically maintained at 0.01–0.02 s, while co-training model has exceeded 30 s. The reason is that when selecting unlabeled training samples, the SD-uwM model is random sampling, and the

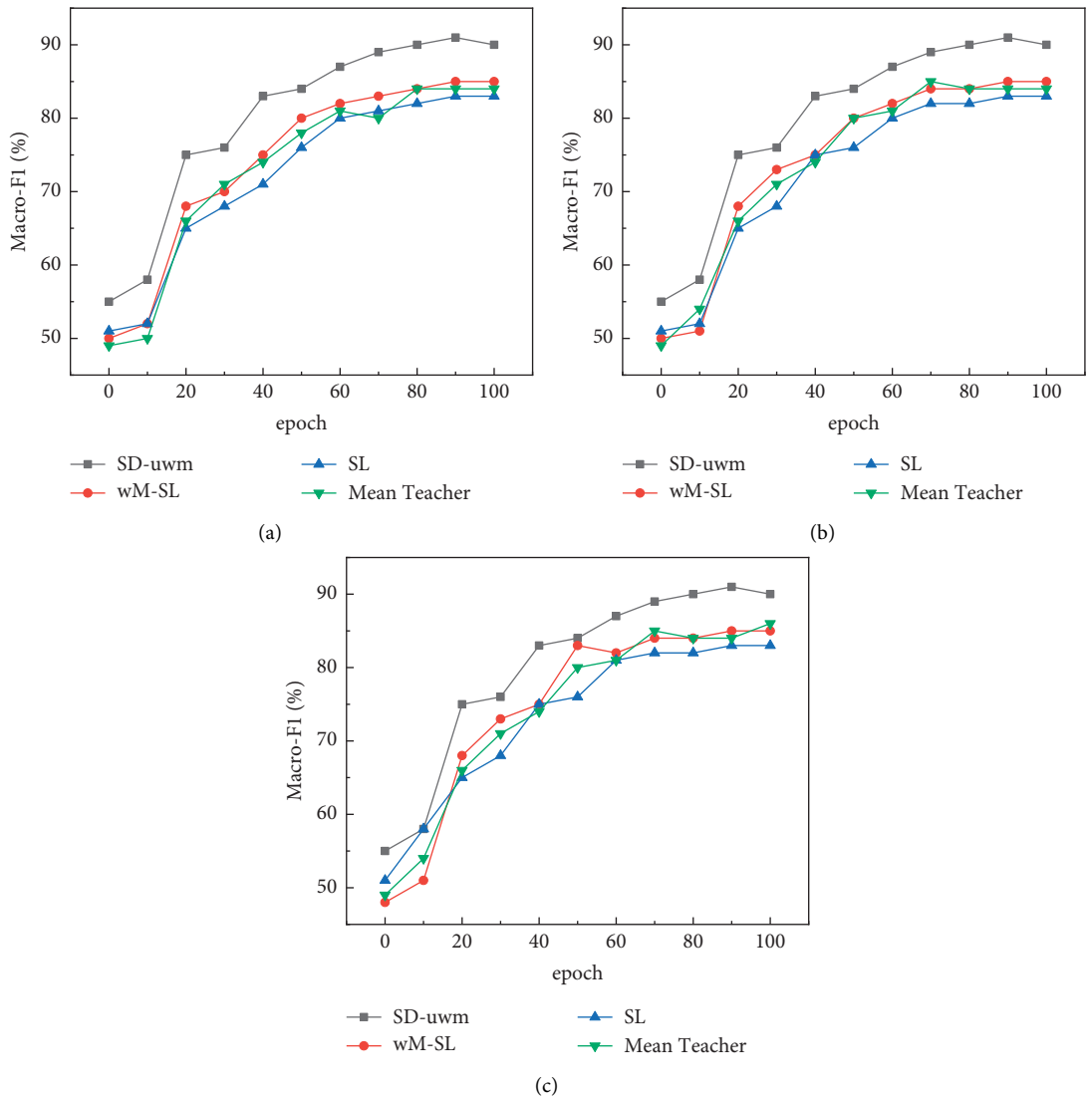


Figure 4: Comparison of macro-F1 value of each model with iteration times on three data sets using LSTM: (a) AGNews data set; (b) THUCNews data set; (c) 20Newsgroups data set.

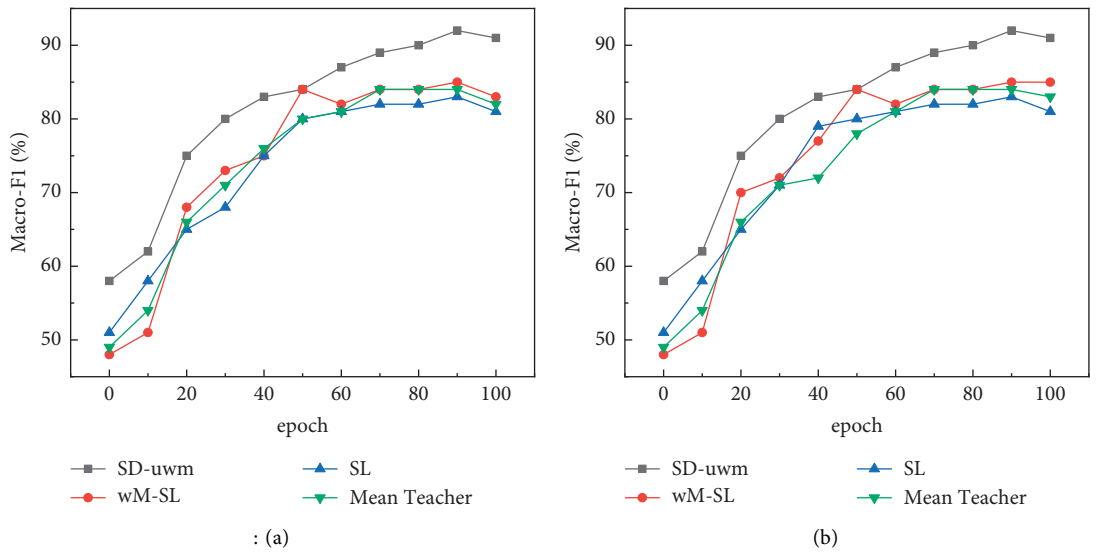
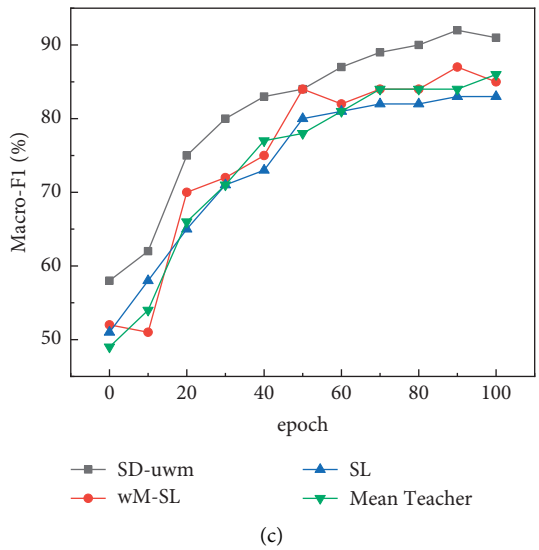


Figure 5: Continued.



(c)

Figure 5: Comparison of the macro-F1 value of each model with the number of iterations on three data sets using TextCNN: (a) AGNews data set; (b) THUCNews data set; (c) 20 Newsgroups data set.

Table 2: Time performance comparison of SN-uwM and co-training.

Dataset	Model	Accuracy (%)	Time (s)	Ratio of train time (SD-uwM/co-training)
20 Newsgroups	SD-uwM	86.4 ± 1.3	0.01	1/3000
Nl = 200, Nu = 2000	Co-training	83.3 ± 1.2	30	
THUCNews	SD-uwM	90.5 ± 1.3	0.02	1/2200
Nl = 300, Nu = 4000	Co-training	88.4 ± 1.2	44	

time complexity is O(1), while the co-training method needs to calculate the sample similarity matrix, and the time complexity is $O(Nl * Nu)$.

4. Conclusion

It is extremely necessary to study Marx's thought of "combining education with productive labor" and its guiding significance to youth labor education. This paper proposes a u-wordMixup method for the data augmentation of unlabeled samples and combines the consistent training framework and the Mean Teacher method to propose a semi-supervised deep learning model SD-uwM. The model uses the u-wordMixup method to enhance the data of unlabeled samples and takes into account the supervised cross-entropy loss and unsupervised consistency loss to construct a new objective function, so as to realize the teaching optimization of the combination of education and production labor. The findings of the study show that the experimental days are as follows:

(1) On the three data sets, regardless of whether the network structure is LSTM or TextCNN, the classification accuracy of the SD-uwM model is better than that of the SL, wM-SL, and Mean Teacher models

(2) SD-uwM model can improve the generalization ability and time performance of the model, which are 86.4 ± 1.3 and 90.5 ± 1.3, respectively

(3) Since the SD-uwM model uses the u-wordMixup method to enhance the unlabeled samples and combines the idea of the Mean Teacher model, it can improve the performance of text classification

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Research Project of Ideological and Political Work in Colleges and Universities in Hunan Province, under project number 19D08.

References

[1] G. S. Kostiuk, "Psychological Problems in the Combination of School Learning and Productive Work," *Russian Education & Society*, pp. 33–45, 2014.

[2] P. P. K. Chan, K. Chen, L. Xu, X. Hu, and D. S. Yeung, "Weakly supervised semantic segmentation with patch-based metric learning enhancement," in *Proceedings of the 30th International Conference on Artificial Neural Networks*, vol. 12893, pp. 471–482, Bratislava, Slovakia, September 2021.

[3] P. R. Agenor, "INFRASTRUCTURE, public education and growth with congestion costs," *Bulletin of Economic Research*, vol. 64, no. 4, pp. 449–469, 2012.

[4] S. Cunningham and T. D. Kendall, "Prostitution, hours, job amenities and education," *Review of Economics of the Household*, vol. 15, no. 4, pp. 1055–1080, 2017.

[5] A. İşçan, A. Ece, N. Vurgun, and F. Özcan, "Breast feeding habits of the labour-families in aydın," *The Journal of Tepecik Education and Research Hospital*, vol. 6, no. 1, pp. 97–103, 1996.

[6] L. Gori and M. Sodini, "Local and global bifurcations in an economic growth model with endogenous labour supply and multiplicative external habits," *Chaos*, vol. 24, no. 1, pp. 013122–013142, 2014.

[7] Z. Gao, "From student-centered pedagogy to student labor: Chinese education's transnational entanglements during the cold war," pp. 255–283, 2021.

[8] D. T. Yang, "Education in production: measuring labor quality and management," *American Journal of Agricultural Economics*, vol. 79, no. 3, pp. 764–772, 1997.

[9] A. Tarvainen and H. Valpola, "Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, December 2017.

[10] T. Miyato, S. I. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1979–1993, 2019.

[11] Y. Wang, "Implications of blended teaching based on theory of semantic wave for teaching English writing in high school," *Journal of Higher Education Research*, vol. 3, no. 2, pp. 166–168, 2022.

[12] Y. Tokozume, Y. Ushiku, and T. Harada, "Between-class learning for image classification," 2017, https://arxiv.org/abs/1711.10284.

[13] J. Miao, Z. Wang, X. Ning, N. Xiao, W. Cai, and R. Liu, "Practical and secure multifactor authentication protocol for autonomous vehicles in 5G. Software: practice and Experience," *Concurrency and Computation: Practice & Experience*, vol. 26, no. 4, pp. 987–1019, 2014.

[14] W. Chen, W. Xiang, A. Jz et al., "Uncertainty estimation for stereo matching based on evidential deep learning," *Pattern Recognition*, vol. 124, Article ID 108498, 2022.

[15] N. Papernot, M. Abadi, L. Erlingsson, I. Goodfellow, and K. Talwar, "Semi-supervised knowledge transfer for deep learning from private training data," 2016, https://arxiv.org/abs/1610.05755.

[16] Z. Yu, S. Li, L. N. U. Sun, L. Liu, and W. Haining, "Multi-distribution noise quantisation: an extreme compression scheme for transformer according to parameter distribution," *Connection Science*, .

[17] X. Ning, K. Gong, W. Li, and L. Zhang, "JWSAA: joint weak saliency and attention aware for person re-identification," *Neurocomputing*, vol. 453, pp. 801–811, 2021.

[18] C. Li, J. Zhu, and B. Zhang, "Max-margin Deep Generative Models for (Semi-)Supervised Learning," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 1, no. 99, 2016.

[19] J. Song, L. Gao, F. Zou, Y. Yan, and N. Sebe, "Deep and fast: deep learning hashing with semi-supervised graph construction," *Image and Vision Computing*, vol. 55, pp. 101–108, 2016.

[20] G. Yu, T. Xie, C. Xu et al., "Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images," *Cold Spring Harbor Laboratory*, vol. 12, 2020.

[21] W. A. Kamakura, M. Wedel, F. de Rosa, and J. A. Mazzon, "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction," *International Journal of Research in Marketing*, vol. 20, no. 1, pp. 45–65, 2003.

[22] S. Frühwirth-Schnatter, "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, vol. 15, no. 2, pp. 183–202, 1994.

[23] L. Ying, Z. Qian Nan, W. Fu Ping et al., L. Nam, Adaptive weights learning in CNN feature fusion for crime scene investigation image classification," *Connection Science*, vol. 33, no. 3, pp. 719–734, 2021.

[24] J. A. Royle and R. M. Dorazio, "Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models," *Journal of Ornithology*, vol. 152, no. S2, pp. 521–537, 2012.

[25] H. Guo, Y. Mao, and R. Zhang, "Augmenting data with mixup for sentence classification: an empirical study," 2019, https://arxiv.org/abs/1905.08941.

[26] N. G. Polson and S. L. Scott, "Rejoinder: data augmentation for support vector machines," *Bayesian Analysis*, vol. 6, no. 1, pp. 43–47, 2011.

[27] J. Gordon and J. M. Hernández-Lobato, "Combining deep generative and discriminative models for Bayesian semi-supervised learning," *Pattern Recognition*, vol. 100, Article ID 107156, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6707304, 12 pages
https://doi.org/10.1155/2022/6707304

Research Article

A Voice Cloning Method Based on the Improved HiFi-GAN Model

Zeyu Qiu,[1] **Jun Tang**,[1] **Yaxin Zhang**,[2] **Jiaxin Li**,[1] **and Xishan Bai**[3]

[1]*Information Engineering University, Zhengzhou 450001, China*
[2]*Handan Vocational College of Science and Technology, Handan 056000, China*
[3]*Yunnan Minzu University, Kunming 650504, Yunnan, China*

Correspondence should be addressed to Xishan Bai; 060899@ymu.edu.cn

Received 26 July 2022; Revised 6 September 2022; Accepted 28 September 2022; Published 11 October 2022

Academic Editor: Wenming Cao

With the aim of adapting a source Text to Speech (TTS) model to synthesize a personal voice by using a few speech samples from the target speaker, voice cloning provides a specific TTS service. Although the Tacotron 2-based multi-speaker TTS system can implement voice cloning by introducing a d-vector into the speaker encoder, the speaker characteristics described by the d-vector cannot allow for the voice information of the entire utterance. This affects the similarity of voice cloning. As a vocoder, WaveNet sacrifices speech generation speed. To balance the relationship between model parameters, inference speed, and voice quality, a voice cloning method based on improved HiFi-GAN has been proposed in this paper. (1) To improve the feature representation ability of the speaker encoder, the x-vector is used as the embedding vector that can characterize the target speaker. (2) To improve the performance of the HiFi-GAN vocoder, the input Mel spectrum is processed by a competitive multiscale convolution strategy. (3) The one-dimensional depth-wise separable convolution is used to replace all standard one-dimensional convolutions, significantly reducing the model parameters and increasing the inference speed. The improved HiFi-GAN model remarkably reduces the number of vocoder model parameters by about 68.58% and boosts the model's inference speed. The inference speed on the GPU and CPU has increased by 11.84% and 30.99%, respectively. Voice quality has also been marginally improved as MOS increased by 0.13 and PESQ increased by 0.11. The improved HiFi-GAN model exhibits outstanding performance and remarkable compatibility in the voice cloning task. Combined with the x-vector embedding, the proposed model achieves the highest score of all the models and test sets.

1. Introduction

Voice cloning [1] is a speech synthesis method that allows machines to synthesize the speech of a specific target speaker. It also provides a critical technical means for generating personalized speech. In personalized human-computer interaction scenarios, voice cloning technology possesses a wide range of applications in intelligent electronic terminal equipment like autonomous robots, Internet of Vehicles, Internet of Things, etc.

This technology can be divided into two categories based on the amount of target speaker corpus used. One method is a speech synthesis method that is based on a large amount of the target speaker corpus. The fundamental principle of this method is to train a speech synthesis system with a large amount of the target speaker's speech and synthesize the voice of the target speaker. The main disadvantage is that it needs to collect a large number of speech samples of a specific person, which is a tedious job in many cases. Hence, this method is rarely used. The second approach is based on a small number of samples. There are two ways to implement this method. One method involves using the speaker adaptation method [2]. The basic idea is to obtain a more matchable acoustic model by finetuning the parameters of the trained multispeaker generation model through an adaptive algorithm. Speaker adaptation entirely depends on adaptation parameters and leads to an increase in memory storage and serving costs. The second approach is to use the speaker encoding method [3]. The basic idea is to select an independent speaker encoder to extract the embedding vector of the target speaker and then splice the speaker embedding vector into the multispeaker speech generation

model for controlling the speech. Finally, the voice of the target speaker is synthesized by the vocoder. The advantage of this method is that the trained speaker encoding model does not require any finetuning, and the speaker embedding vector can be directly inferred from only a couple of speech samples of the target speaker, so the cloning speed is fast. The key part of this method is to design a good speaker encoder that has the capacity to extract the features that characterize the target speaker from a small number of speech fragments. The similarity of the cloned speech can be determined by the quality of extracted speaker features. Arik et al. made a detailed comparison between the speaker encoding and speaker adaption methods and both methods performed well in voice cloning tasks [1]. The speaker adaptation method requires thousands of finetuned steps to achieve a high-quality adaptive effect, making it more difficult for deployment in mobile devices without real-time synthesis. The cloning time and necessary memory for the speaker encoding method are really less, which is crucial for practical applications.

Although previous works in voice cloning have appropriately considered the limited speech samples in personalized voice, they have not completely addressed the key issues. They finetune the whole model [4] or the decoder part [5, 6], achieving good quality but leading to too many adaptation parameters. Reducing the number of adaptation parameters is crucial for the practical application of voice cloning tasks. Also, the memory storage can explode due to the increase in the number of users. Some works only finetune the speaker embedding or train the speaker encoder part [7, 8], which does not require any fine-tuning during voice cloning. Although these approaches lead to a lightweight and efficient adaptation, they provide poor cloning quality.

So far, Tacotron 2 [9], based on sequence-to-sequence architecture, has been a very popular model in the field of speech synthesis, possessing great development prospects and significant versatility. Tacotron 2 can be divided into two submodules: the acoustic feature prediction module and the vocoder module. At the time of training, the acoustic feature prediction module usually inputs the text sequence and outputs the acoustic features, while the vocoder module restores the predicted acoustic features to speech waveforms. However, Tacotron 2 cannot precisely control and synthesize diverse speech samples. To synthesize sounds closer to human beings, Tacotron 2 is usually extended and applied in several other tasks such as voice cloning, speech style control, speech prosody control, code-switching, etc. Wang et al. achieved prosody transfer and enhanced the emotional information of synthesized speech by adding a prosody encoder in the acoustic feature prediction module to model and learn the prosody features in both supervised as well as unsupervised ways [10, 11]. The speaker encoder can be simply regarded as a text-independent speaker recognition model. Kinnunen et al. completed speech conversion based on *i*-vector [12] in speaker recognition [13]. It is difficult to retain the nonlinear features in the original data when the *i*-vector uses a linear transformation to reduce the dimensions, which have been replaced by a *d*-vector with strong antinoise ability [14]. By adding the speaker encoder and extracting the *d*-vector with the target speaker features, the multispeaker TTS system [15] realizes the preliminary voice cloning. However, *d*-vector does not fully consider the context information of the entire utterance, which leads to the omission of speaker information. Snyder et al. first proposed a framework for extracting speaker embedding features based on the time-delay neural network (TDNN) [16] and successfully obtained the *x*-vector, which was applied to the speaker verification task and outperformed the traditional speaker vector. The application of the *x*-vector in the speaker encoder can effectively guide the prediction of acoustic features, which can significantly improve the similarity of the cloned speech. By combining WaveNet [17] based on the autoregressive model as a vocoder, Tacotron 2 can generate high-quality speech, but the sequential reasoning process of the autoregressive model makes it sluggish and inefficient to generate speech, which cannot meet the requirements of real-time applications. To address the limitations of autoregressive models, more and more researchers began focusing on nonautoregressive models based on generative adversarial networks (GAN). Parallel WaveGAN [18] and MelGAN [19] are the early attempts of GAN on vocoder. Although the model reasoning speed can be significantly improved, the speech they generated is not satisfactory in terms of quality. The appearance of HiFi-GAN [20] breaks the shackles not only by effectively modeling the long-term correlation of the speech waveform but, more importantly, by effectively modeling the periodic mode of the speech waveform. Besides, it achieves real-time and high-fidelity speech waveform generation. Moreover, as one of the most advanced vocoder networks, the HiFi-GAN model is used as the backend by many end-to-end speech synthesis systems to restore the predicted Mel spectrum to speech waveforms. However, there are still some shortcomings in HiFi-GAN, which fail to balance the speech quality with model parameters and inference speed. Therefore, the HiFi-GAN model must be improved for better application in the voice cloning task.

Although the multispeaker TTS model based on Tacotron 2 can expand the system architecture along with supporting the voice cloning function of multiple speakers, it is still slightly insufficient in terms of speaker feature extraction and synthesis speed. The speech information of the entire sentence is not taken into consideration by the d-vector, which affects the similarity of the cloned voice. The WaveNet vocoder can have a severe impact on the speed of speech generation. To better balance the relationship between speech quality, model parameters, and inference speed of the voice cloning system, the speaker features are extracted based on TDNN, and the output of each speech segment is aggregated after passing through the model through statistical pooling. This represents the feature vector of the target speaker and improves the quality of the generated speech. In this paper, a competitive multiscale convolution (CMSC) strategy and a depth-wise separable convolution (DSC) strategy are introduced to improve the HiFi-GAN model, which replaces the WaveNet vocoder, to significantly reduce the number of model parameters and further enhance the inference speed.

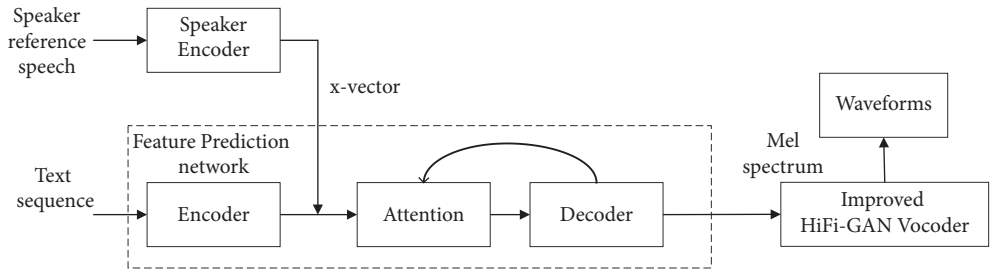


Figure 1: System architecture of voice cloning is based on improved HiFi-GAN.

2. Methods

The overall structure of the voice cloning system based on the improved HiFi-GAN model can be divided into three groups: speaker encoder network, feature prediction network, and vocoder network. As illustrated in Figure 1, a network based on speaker verification is adopted by the speaker encoder network. This encoder is trained by improving the performance of the speaker verification system. The feature prediction network follows the Tactron 2 architecture, which is implemented by an encoder-decoder architecture. First, the text sequence is converted into a semantic vector in this architecture by the encoder. After the semantic vector and the speaker embedding vector are spliced, it is processed by using the attention mechanism and forwarded to the decoder of the feature prediction network. Eventually, the decoder converts the sequence into the Mel spectrum. The vocoder network is implemented with an improved HiFi-GAN model, which quickly converts the Mel spectrum into speech waveforms while ensuring the quality of the generated speech.

2.1. Speaker Encoder Based on the X-Vector. The speaker encoder network is one of the core parts of the model, which determines the similarity of the cloned speech. In this paper, the speaker verification architecture is used to implement the speaker encoder based on the *x*-vector. In the speaker verification architecture, the input is the speaker's speech feature, and the output is the speaker's discrimination information . The network structure is shown in Figure 2 [16].

The first five layers of the network are frame-level layers. The output nodes of the other layers are 512, while the output nodes of the fifth layer are 1500. The actual input is a 20-dimensional MFCC speech feature. The current frame is spliced with the input of the first and last two frames and then sent to frame 1. The corresponding total context information is 5 frames, and the input is a 100-dimensional feature vector. The three frames of $\{t-2, t, t+2\}$ are spliced and sent to frame 2 as the output of the frame 1 layer. At this time, there are 9 frames including in-context information, and the splicing input is a 1536-dimensional feature vector. The operation of the frame 3 layer is similar, which comprises 15 frames' contextual information, and the splicing input is a 1536-dimensional feature vector. The output of the previous layer is directly used by frames 4 and 5 as input after batch normalization without splicing the context information. The statistical pooling layer receives the output of the

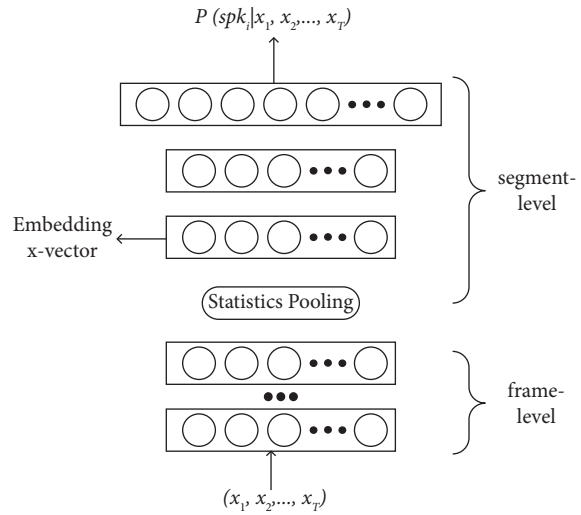


Figure 2: Network structure based on the *x*-vector extracted by the TDNN.

Table 1: Detailed parameters of TDNN.

Layer	Layer context	Total context	Input × output
Frame 1	$[t-2, t+2]$	5	100×512
Frame 2	$\{t-2, t, t+2\}$	9	1536×512
Frame 3	$\{t-3, t, t+3\}$	15	1536×512
Frame 4	$\{t\}$	15	512×512
Frame 5	$\{t\}$	15	512×1500
Sats pooling	$[0, T)$	T	$1500T \times 3000$
Segment 1	$\{0\}$	T	3000×512
Segment 2	$\{0\}$	T	512×512
SoftMax	$\{0\}$	T	$512 \times K$

frame 5 layer as input and computes the mean and standard deviation of all the frames $\{0, T\}$ of the input speech. These two statistics are spliced together to form a 3000-dimensional vector and sent to the next two segment-level processing layers. Considering different situations, the output dimensions of segments 6 and 7 can be set, and finally, the output is sent to the output layer to obtain the probability distributions of different speakers. The specific parameters of each layer of the network are mentioned in Table 1.

A multiclass cross-entropy objective function is used to train the network to distinguish different speakers. The probability distribution of the samples corresponding to each label is obtained after normalizing the output layer by the SoftMax function. Then, the cross-entropy loss function

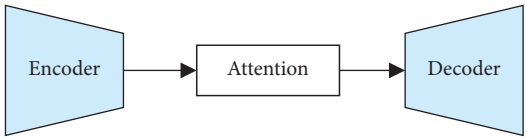


Figure 3: Encoder-decoder model structure.

is used to calculate the similarity of the results to the true sample probability distribution. The network parameters are continuously updated through backpropagation until convergence. As shown in (1), it is assumed that there are K speakers in the N training speech segments. $P(\text{spkr}_k \mid x_{1:T}^{(n)})$ denotes the probability that the input speech $x_1^{(n)}$, $x_2^{(n)}, \ldots, x_T^{(n)}$ of given T frames corresponds to the speaker k. If the corresponding speaker label of speech segment n is k, then d_{nk} is 1, else, it is 0.

$$E = -\sum_{n=1}^{N}\sum_{k=1}^{K} d_{nk} \ln\left(P\left(\text{spkr}_k \,\middle|\, x_{1:T}^{(n)}\right)\right). \tag{1}$$

2.2. Basic Structure of the Feature Prediction Network. The feature prediction network in this paper is based on the encoder-decoder model [21]. Its primary function is to direct the conversion of the input text into the Mel spectrum with the target speaker's characteristics after splicing with the x-vector vector output by the speaker encoder that describes the speaker's characteristics, so that the vocoder can restore waveforms. Figure 3 depicts its fundamental architectural principle.

The encoder first models the contextual information of the input text sequence with a 3-layer convolutional network, and the output of the final convolutional layer is fed into a bidirectional long-short-term memory network with 512 units to convert the input text sequence into a high-level feature sequence. The attention mechanism computes the weight of each element in the high-level feature sequence, assigns different weights to the encoder output, performs weighted summation, and then feeds it into the decoder. In this case, the attention network employs the location-sensitive attention mechanism, which extends the additional attention mechanism [22], alleviating potential error patterns caused by the decoder repeating or ignoring some subsequences. The decoder is a 5-layer convolution post-processing network with a 2-layer fully connected pre-processing network, a 2-layer unidirectional long short-term memory network, two linear mapping layers, and a 2-layer fully connected preprocessing network. The posterior probability of the output sequence and the output Mel spectrum are computed.

2.3. Improved HiFi-GAN Model. HiFi-GAN uses GAN as the basic generative model and includes a generator and two discriminators, which can efficiently convert the spectrum generated by the acoustic model into high-quality audio. HiFi-GAN is a vocoder commonly used in both academia and industry in recent years, but it still has some shortcomings. In order to reduce the model parameters of HiFi-GAN and improve the inference speed without sacrificing the speech quality, we use CMSC and DSC strategies to improve the HiFi-GAN model, and the details are described in the following submodules.

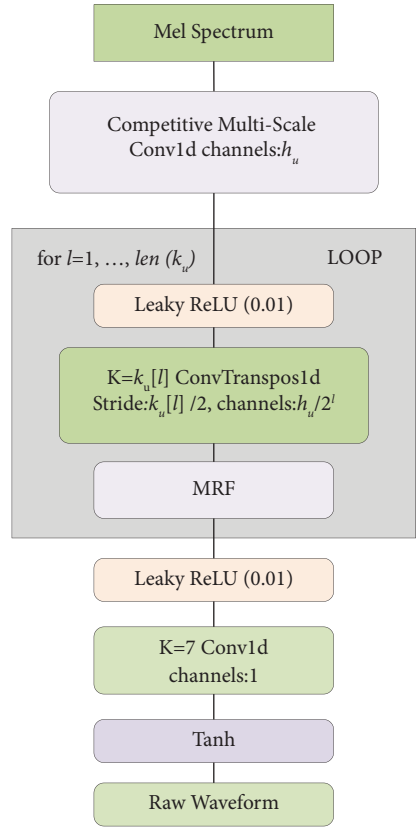


Figure 4: Structure diagram of the generator (K represents the size of the convolution kernel).

2.3.1. Generator. Figure 4 shows the structure of the generator, which adopts the Mel spectrum as input and continuously up-samples it by transposed convolution until the length of the output sequence is matched with the temporal resolution of the original waveform. Each transposed convolution is followed by a multireceptive field fusion (MRF) module. Figure 5 shows the specific structure of the MRF.

The sum of the outputs of multiple residual blocks (ResBlock) is accumulated by the MRF module. Each residual block is composed of a series of one-dimensional convolutions. These convolutions have different convolution kernels and dilation rates that form different sized receptive fields, effectively modeling the long-term correlations of speech waveforms.

Unlike the original generator network, a CMSC strategy is used to extract the features from the input Mel spectrum. The multisized convolution kernels are used to process the input Mel spectrum, and the sum of these processed results is returned. Compared with the original convolutional layer with a fixed convolution kernel size to extract features from the Mel spectrum, CMSC can better capture the local features between different frames and interframe correlations

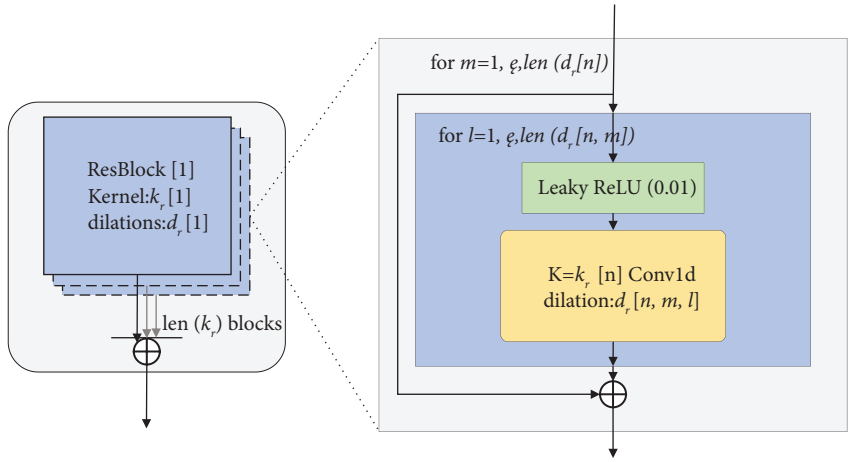


Figure 5: Structure diagram of MRF.

of the Mel spectrum and express the feature information extracted from the Mel spectrum in a better way while providing sufficient information for the subsequent network learning. It thus improves the learning ability of the model. Besides, the original generators are composed of standard 1D convolutional layers except for a few transposed convolutional layers for upsampling. Inspired by the DSCs in images [23], in this paper, these standard 1D convolutions are replaced with 1D DSCs, which is expected to further compress the model size and speed the model inference without compromising the quality of the generated speech; making it significant for applications with limited hardware. It must be noted that DSC using weight normalization is equivalent to the depth-wise convolution and the pointwise convolution, which adopt weight normalization [24].

2.3.2. Discriminator. For generative adversarial networks, the discriminator primarily plays an adversarial training role for the generator by guiding the generator to generate more realistic data. Here, the discriminator of the model basically adopts the original configuration of the HiFi-GAN model, which has two discriminators: a multiperiod discriminator and a multiscale discriminator.

2.3.3. Multiperiod Discriminator. Figure 6 highlights the structure of the multiperiod discriminator (MPD). The left represents the overall structure, and the right represents the network structure of the subdiscriminator. The feature map represents the feature output of each network layer and is used in the feature matching loss in the next section. It comprises multiple subdiscriminators with the same network structure, and each subdiscriminator can capture a part of the periodic signal of the input speech to detect various potential periodic patterns in the speech data.

To realize that the subdiscriminator captures the periodic pattern in the speech signal, the subdiscriminators do not directly process the speech waveform but pad and reshape the speech waveform. Figure 7 highlights the case when the period parameter p is 3. For ensuring that each subdiscriminator only accepts equally spaced sampling points of the input speech waveform, the interval is represented by p. Thus, the original one-dimensional speech of length T is processed into two-dimensional data with height T/p and width p. Therefore, the MPD needs to use a two-dimensional convolutional neural network to process these data. Other than the last network layer, the other layers use two-dimensional stride convolutions, which only stride in height, and each convolution layer uses weight normalization. In each convolutional layer of MPD, the size of the width axis of the convolution kernel is limited to 1. This leads to an independent processing of the periodic speech samples in the width axis direction. Thus, each subdiscriminator can capture the underlying periodic patterns that differ from each other in the speech by observing different parts of the speech waveforms.

2.3.4. Multiscale Discriminator. Figure 8 shows the structure of the multi-scale discriminator (MSD). The left represents the overall structure, and the right represents the network structure of the subdiscriminator. The MSD is a combination of three discriminators with the same network structure but working at different scales: processing the raw speech, ×2 average-pooled audio, and ×4 average-pooled audio. To allow the use of larger-sized kernels while keeping a smaller number of parameters, the subdiscriminator employs grouped convolutions. Apart from applying spectral normalization [25] in the first subdiscriminator for raw speech processing, which is used here to help stabilize training, the other two sub-discriminators apply weight normalization.

2.4. Loss Function. The loss function can be categorized into three parts: adversarial loss, feature matching loss, and Mel spectrum loss.

2.4.1. Adversarial Loss. For the adversarial training objectives of the generator and discriminator, the settings of LSGAN models are followed. The discriminator classifies the speech samples, where the real speech is classified as 1, and the speech generated by the generator is classified as 0. The

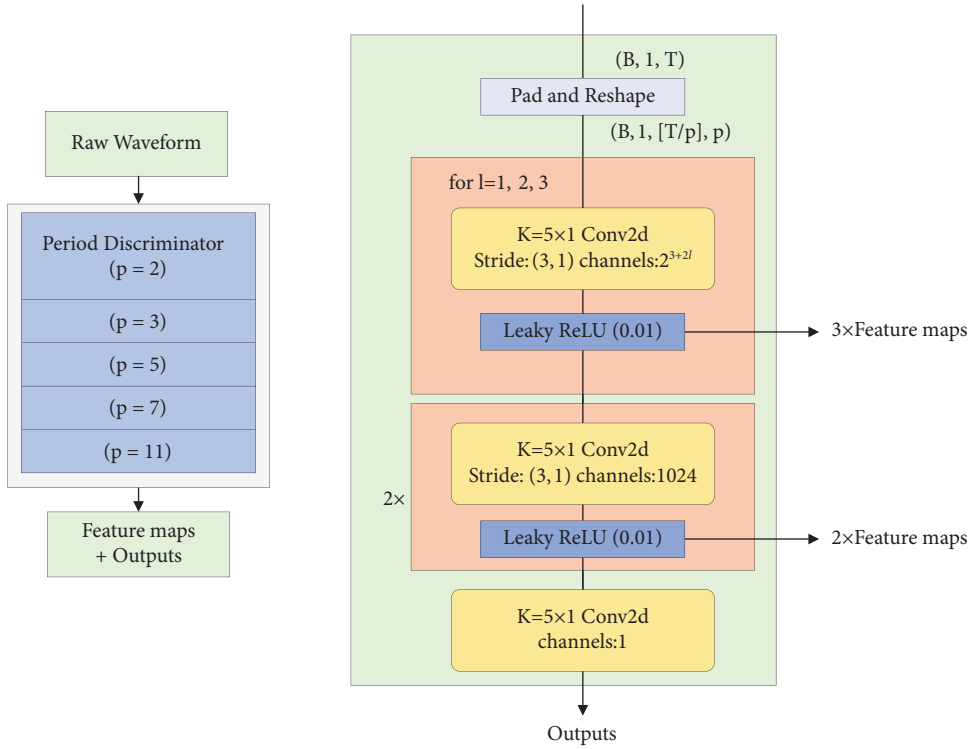


Figure 6: Schematic diagram of the multiperiod discriminator.

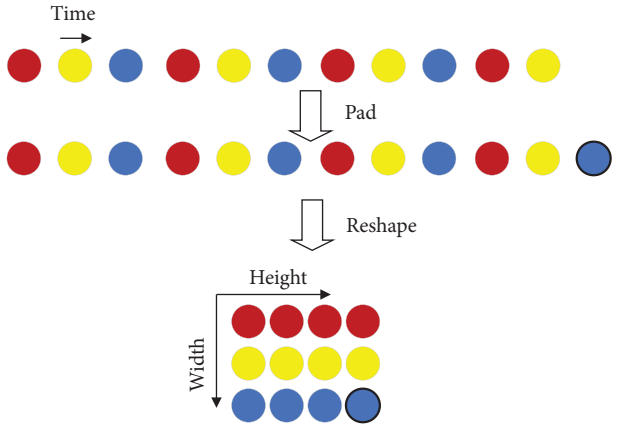


Figure 7: Schematic diagram of filling and reshaping in speech waveforms.

generator generates the speech according to the input conditions to deceive the discriminator, leading to an incorrect classification of type 1 in the generated speech. Eventually, the generator can achieve the effects of mixing the spurious with the genuine through the mutual game processing between the generator and the discriminator. The adversarial loss functions of the generator and discriminator are represented in equations (2) and (3), respectively.

$$\mathscr{L}_{\mathrm{Adv}}(G;D) = \mathbb{E}_s\left[(D(G(s)) - 1)^2\right], \tag{2}$$

$$\mathscr{L}_{\mathrm{Adv}}(D;G) = \mathbb{E}_{(x,s)}\left[(D(x) - 1)^2 + (D(G(s)))^2\right]. \tag{3}$$

In brief, MSD and MPD are described as discriminators, where x represents the real speech and s denotes the input condition (mel spectrum extracted from the corresponding real speech).

2.4.2. Feature Matching Loss. To improve the ability of the generator, the feature matching loss (FML) proposed in the MelGAN model is adopted. The FML improves the generator's forgery ability by comparing the difference between the real speech and the generated speech in the output features of each layer of the discriminator network. To measure this difference, the $L1$ distance is used. The feature matching loss function formula is mentioned in equation (4).

$$\mathscr{L}_{FM}(G;D) = \mathbb{E}_{(x,s)}\left[\sum_{i=1}^{T}\frac{1}{N_i}\left\|D^i(x) - D^i(G(s))\right\|_1\right], \tag{4}$$

where T represents the number of convolutional layers in the discriminator, and D^i and N_i indicate the features and number of features in the i-th layer of the discriminator, respectively.

2.4.3. Mel Spectrum Loss. By jointly optimizing the multiresolution spectral loss and the adversarial loss, parallel WaveGAN effectively models the time-frequency distribution of real speech waveforms. Similar to the multiresolution spectral loss, the HiFi-GAN adopts the Mel spectrum loss based on the characteristics of the human auditory system, thereby improving the perceptual quality of the generated speech. Specifically, the Mel spectrum loss is the $L1$ loss in

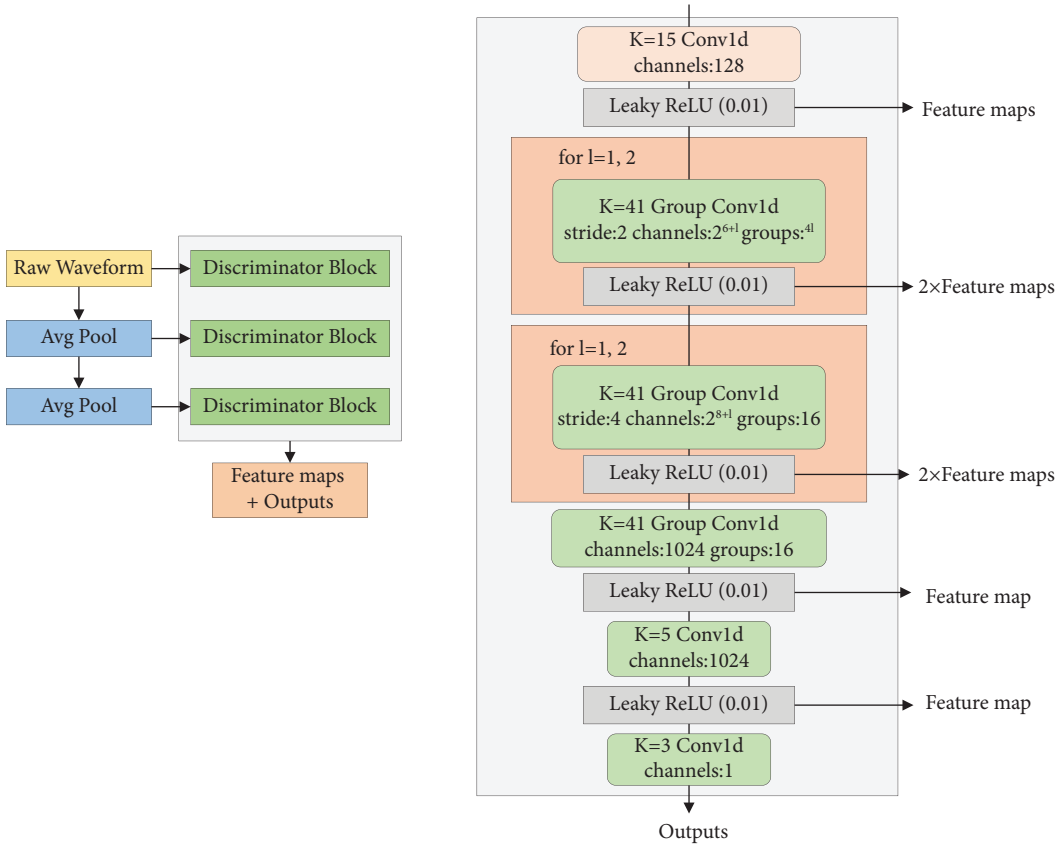


Figure 8: Schematic diagram of the multiscale discriminator.

the Mel spectrum between generated speech and real waveform, as illustrated in equation (5).

$$\mathscr{L}_{\text{Mel}}(G) = \mathbb{E}_{(x,s)}\left[\|\phi(x) - \phi(G(s))\|_1\right], \tag{5}$$

where $\phi(\cdot)$ represents the function of extracting the Mel spectrum from speech. It must be noted that the Mel spectrum extracted here is full-band (the lowest frequency is 0 Hz and the highest frequency is half of the speech sampling rate), which differs from the band-limited Mel spectrum as the input condition. This full-band Mel loss helps the model learn the full-band frequency information of the speech.

2.4.4. Total Loss. Feature matching loss and Mel spectrum loss are used as auxiliary losses to stabilize the model training and accelerate the convergence. Hence, the final loss functions used to train the generator and discriminator are mentioned in equations (6) and (7).

$$\mathscr{L}_G = \sum_{k=1}^{K}\left[\mathscr{L}_{\text{A dv}}(G; D_k) + \lambda\mathscr{L}_{FM}(G; D_k)\right] + \mu\mathscr{L}_{\text{Mel}}(G), \tag{6}$$

$$\mathscr{L}_D = \sum_{k=1}^{K}\mathscr{L}_{\text{A dv}}(D_k; G), \tag{7}$$

where D_k represents the kth subdiscriminator in MPD and MSD, λ and μ denote the hyperparameters used to control the proportion of each loss, and their values are, respectively, set to 2 and 45 in this experiment.

3. Experimental Setup

3.1. Corpus. The LibriSpeech [26] speech recognition corpus is used here as the training corpus, which was widely used in the public datasets. It contains 1000 h of audiobooks at 16 kHz with corresponding texts. The audio recordings are split and sorted into shorter segments of 35 s. Also, there are two clean training sets that contain 436 h of American English speech from 1172 speakers with widely varying tonal styles. These speakers are inconsistent between the training, development, and test sets. The THCHS-30 [27] Chinese Corpus is approximately 33.5 h long, with a total of 13,388 sentences recorded by 30 college students who speak fluent Mandarin, with an average length of 20 words and an average length of 9 s per sentence. All audios in the corpus correspond to text, Chinese characters are represented by pinyin, and "spring" is represented by "chun1 tian1" (1 represents the first sound), indicating that all models are trained on prestandardized data. For accelerating the convergence of the feature prediction network, Montreal Forced Aligner is used to enforce the alignment of the audio with the corresponding text, cut silent segments longer than 0.4 s, and resegmented the data into shorter utterances.

Table 2: Speaker encoder model parameters based on *x*-vector.

Initial learning rate	0.0001
Model embedding size	256
Model hidden layer size	256
Model layers	3
Speaker batch size	32
Number of utterances per speaker	10

Table 3: The key training parameters of the feature prediction network.

Dimensions of the speaker embedding vector	256
Silence duration (s)	0.4
Utterance duration (s)	16
Mel spectrum channel number	80
Initial learning rate	0.003
Final learning rate	0.00005
Spectral window length (ms)	50
Spectral window shift (ms)	12.5

3.2. Corpus Preprocessing. In LibriSpeech's training sets, one can notice ambient background noise and background noise when muted. To avoid extracting such silent segments from the complete utterance, the target spectrum is preprocessed by using spectral subtraction [28] based on voice activity detection (VAD) to identify and remove those long silences from the spectrum. For separating the silent and the non-silent data, webrtcvad is used as the python interface for VAD, which generates a binary flag to indicate if the voice segment is pronounced. Finally, to adjust the speaker's volume, the speech waveforms are normalized.

3.3. Parameter Settings. Experiments are performed on a single GPU and CPU (NVIDIA Tesla V100 GPU for training, Xeon(R) E5–2620 v4 2.10 GHz CPU, and NVIDIA GTX 1080Ti GPU for testing), and the network architecture of the model is built on PyTorch. Table 2 highlights the settings of the speaker encoder parameters based on the *x*-vector.

The key training parameters of the feature prediction network are listed in Table 3.

The model of the generator is trained with the AdamW optimizer, where $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1e^{-6}$. After each training epoch, the learning rate decays by a factor of 0.999. The batch length of each speech is set to 16384 samples, while the batch size is tuned to 12 audio samples. The key training parameters of the generator model are listed in Table 4.

For comparative experiments, WaveNet and WaveGlow vocoders are trained with the same training and model parameters as the original settings.

3.4. Performance Analysis. By visualizing the speaker embedding vectors, the relationship between the speaker embedding vectors can be monitored more intuitively. The naturalness of the final cloned speech is evaluated by the mean opinion score (MOS), while the similarity is computed

Table 4: The main training parameters of the generator.

Initial learning rate	0.0002
h_u	512
k_u	[16, 16, 4, 4]
k_r	[3, 7, 11]
d_r	[[[1, 1], [3, 1], [5, 1]], [[1, 1], [3, 1], [5, 1]], [[1, 1], [3, 1], [5, 1]]]

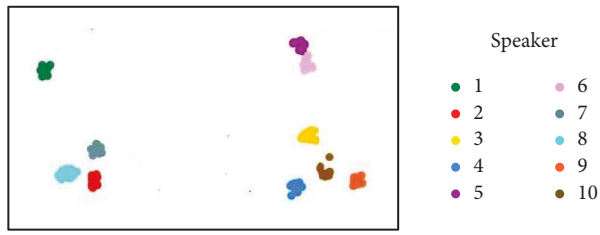


Figure 9: Embedding mapping based on *d*-vector.

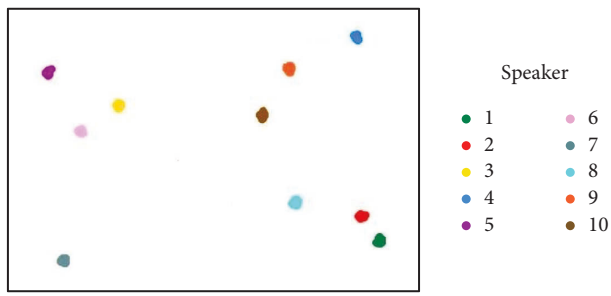


Figure 10: Embedding mapping based on *x*-vector.

by the similarity mean opinion score (SMOS). In the MOS test, 20 randomly selected samples are randomly selected as an evaluation set. A group of 20 listeners who were proficient in English listened through headphones and scored according to the quality of the samples. The MOS is based on the absolute category rating scale [29], with scores ranging from 1 to 5. The MOS scores are recorded with 95% confidence intervals (CI).

4. Results and Discussion

4.1. Embedding Vector Similarity. Using the output of the trained speaker encoder, the test-other dataset in LibriSpeech is tested, which contains 10 speakers and 10 utterances from every single speaker. The speaker embeddings are visualized by reducing the dimensionality, and a clustering algorithm [30] is used to map the embeddings of 100 utterances in a two-dimensional space. Figure 9 represents the embedding mapping of the d-vector method, and Figure 10 represents the embedding mapping of the x-vector method.

Different colors represent different speakers. It can be observed that the embedding vectors of the same speaker form the discourse clusters, and the embedding vectors of different speakers exhibit a certain distance. By comparing the two figures, it can be observed that the different speaker

Table 5: Parameters, inference speed, MOS, and PESQ scores of different vocoders.

Vocoder	MOS (CI)	PESQ	Parameters (M)	Speed on GPU	Speed on CPU
Ground truth	4.56 ± 0.08	4.48	—	—	—
WaveNet	3.97 ± 0.06	3.35	—	×0.002	—
WaveGlow	3.96 ± 0.07	3.19	—	×5.26	×0.13
HiFi-GAN	4.25 ± 0.07	3.63	13.94	×70.34	×2.42
Improved HiFi-GAN	4.38 ± 0.06	3.74	4.38	×78.67	×3.17

embeddings based on the *x*-vector are farther apart and easier to distinguish.

4.2. Speech Quality, Model Parameters, and Inference Speed Tests. To compare the speech quality of different vocoders, which are all trained until convergence, a perceptual evaluation of speech quality (PESQ) is used for objective evaluation, and MOS is used for subjective evaluation. Before the PESQ test, the number of sampling points is processed into an integer multiple of the frameshift by discarding the redundant silence segment at the end of each speech in the test set. Further, the real Mel spectrum from the processed test set is extracted and fed into the model to generate speech. Both the generated speech and its corresponding real speech are downsampled to 16 kHz and evaluated with PESQ using the pypesq library. Besides, the parameters of each vocoder model and the inference speed on GPU and CPU are also compared, respectively. The inference speed is measured by the reciprocal of the real time factor (RTF). RTF demonstrates the time required to generate a one-second speech. The model is considered to have a real-time capability if the time that it takes to generate a one-second waveform is less than or equal to one second. Therefore, the reciprocal value of RTF indicates that the model's inference speed is a multiple of real-time.

The results of different vocoders are listed in Table 5 for easy comparison of speech quality, inference speed, and parameters. Based on these results, it can be seen that the introduction of the improved HiFi-GAN significantly reduces the number of parameters, improves the inference speed of the model on GPU and CPU, and does not damage the quality of the speech generated by HiFi-GAN. The improved HiFi-GAN model not only reduces the number of parameters by about 68.58% but also slightly improves the voice quality (MOS increased by 0.13, PESQ increased by 0.11) when DSC and CMSC are combined. Besides, the inference speed on GPU and CPU has increased by about 11.84% and 30.99%, respectively. Compared with the flow-based WaveGlow and autoregression-based WaveNet, the improved HiFi-GAN is superior to them in all indicators.

4.3. Subjective Preference Evaluation. Subjective preference tests are performed in this section to evaluate the model's speech cloning effect. The testers in the subjective preference comparison test are all college students majoring in the corresponding language. The testers will select their favorite sample after listening to it. They can choose to have no preference if they have no inclination. Figure 11 depicts the test results.

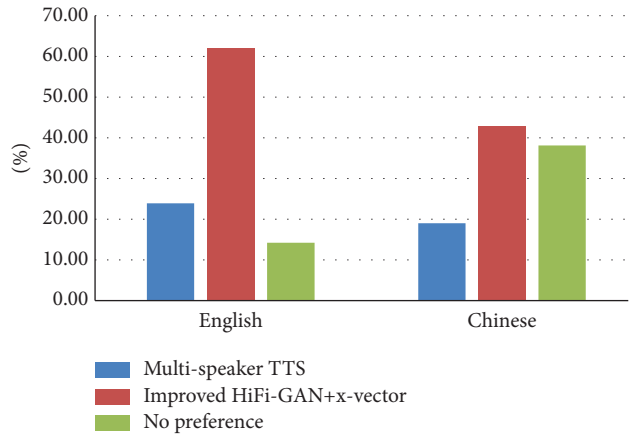


Figure 11: Scores of subjective preference evaluation.

The model enables voice cloning of different languages by adjusting the construction of character embedding. Figure 11 shows that the popularity of English voice cloning samples is greater than that of Chinese voice cloning samples. The testers prefer the voice cloning system based on the improved HiFi-GAN.

4.4. Speech Naturalness and Similarity. To verify the effectiveness of the improved HiFi-GAN model as a vocoder on the voice cloning task, the acoustic feature prediction model is trained with the same dataset as the front end. The dataset split used to train the acoustic feature prediction model is also consistent with the dataset split used to train the vocoder network so that exaggerating experimental results because of leaking test data during evaluation can be avoided. The Mel spectrum with target speaker features is transformed into time-domain speech waveforms through the vocoder. MOS and SMOS are used to evaluate the naturalness and similarity of the cloned speech. In the test, 30 real speech samples are randomly selected and divided into three groups according to the reference corpus. One group is present in the test set from the LibriSpeech dataset, and the second group is formed in the test set from the THchs-30 dataset. The third group is present in the test set from the VCTK dataset [31] in order to verify the universality of the improved vocoder. VCTK contains 44 h of clean speech from 109 speakers, and each speaker provides more than 400 utterances in British English.

In terms of speech naturalness, the experimental results are listed in Table 6. The *x*-vector-based methods outperform the d-vector-based methods on all the models, and the improved HiFi-GAN model combined with the *x*-vector

Table 6: MOS of cloning speech naturalness of different models.

Metric	Settings	LibriSpeech	VCTK	THchs-30
MOS (CI)	Multispeaker TTS	3.93 ± 0.06	3.57 ± 0.07	3.64 ± 0.05
	Multispeaker TTS + *x*-vector	4.02 ± 0.08	3.72 ± 0.09	3.78 ± 0.07
	WaveGlow + *d*-vector	3.85 ± 0.06	3.49 ± 0.08	3.47 ± 0.06
	WaveGlow + *x*-vector	3.93 ± 0.07	3.74 ± 0.08	3.69 ± 0.08
	HiFi-GAN + *d*-vector	4.21 ± 0.10	3.86 ± 0.06	3.92 ± 0.07
	HiFi-GAN + *x*-vector	4.30 ± 0.07	4.15 ± 0.07	4.13 ± 0.09
	Improved HiFi-GAN + *d*-vector	4.28 ± 0.09	4.06 ± 0.05	4.11 ± 0.04
	Improved HiFi-GAN + *x*-vector	4.36 ± 0.06	4.28 ± 0.08	4.28 ± 0.06

Table 7: SMOS of cloning speech similarity of different models.

Metric	Settings	LibriSpeech	VCTK	THchs-30
SMOS (CI)	Multispeaker TTS	3.56 ± 0.07	3.18 ± 0.06	3.25 ± 0.08
	Multispeaker TTS + *x*-vector	3.91 ± 0.06	3.44 ± 0.07	3.59 ± 0.06
	WaveGlow + *d*-vector	3.55 ± 0.09	3.11 ± 0.09	3.32 ± 0.07
	WaveGlow + *x*-vector	3.89 ± 0.08	3.47 ± 0.09	3.64 ± 0.05
	HiFi-GAN + *d*-vector	3.82 ± 0.05	3.38 ± 0.07	3.43 ± 0.09
	HiFi-GAN + *x*-vector	4.15 ± 0.07	3.61 ± 0.08	3.68 ± 0.08
	Improved HiFi-GAN + *d*-vector	3.99 ± 0.10	3.52 ± 0.06	3.61 ± 0.05
	Improved HiFi-GAN + *x*-vector	4.23 ± 0.06	3.80 ± 0.08	3.84 ± 0.07

embedding achieves the highest MOS on both datasets. If data augmentation is used, the quality of the speech synthesized by the voice cloning system may be better.

In terms of speech similarity, the experimental results are mentioned in Table 7. The *x*-vector-based methods perform much better than the *d*-vector-based methods on all the models, and the SMOS is higher, especially on the LibriSpeech dataset.

It can be observed from the abovementioned two tables that the x-vector is stronger than the d-vector in representing the target speaker, and the quality of cloned speech is significantly improved. The cloned speech effect referring to the LibriSpeech corpus is better than relying on the VCTK corpus. This may be due to the fact that the training is performed in American English, and the testing is done in British English. When the experimental results in the subjective preference evaluation are combined, it is clear that the quality of English speech synthesis is superior to that of Chinese speech synthesis, which may be due to the complexity of Chinese prosodic structure and the difficulty of feature expression. The improved HiFi-GAN model combined with x-vector embedding achieves the best results in both naturalness and similarity of cloning speech, suggesting that the improved HiFi-GAN model has good compatibility in voice cloning tasks. We plan to use the Mel spectrum output of the acoustic feature prediction network under the teacher forcing condition as the input of the back-end network to fine-tune the vocoder for narrowing the difference between the real Mel spectrum and the predicted Mel spectrum.

5. Conclusions

In this paper, a voice cloning method with fewer parameters, faster inference speed, and higher voice quality is proposed based on the multispeaker TTS model. First, to improve the similarity of cloning speech, the x-vector feature vector that can better represent the characteristics of the target speaker is extracted based on TDNN. Then, the HiFi-GAN vocoder is improved to effectively characterize the input Mel spectrum through a competitive multiscale convolution strategy, providing sufficient feature information for the subsequent network to generate a higher-quality speech signal. Finally, the model parameters are effectively reduced by the use of depth-wise separable convolution, and the inference speed is improved without degrading the quality of the generated speech. According to the experimental results, the method in this paper effectively reduces the parameters of the HiFi-GAN model and improves the generated speech quality (MOS increased by 0.13, PESQ increased by 0.11), and the model inference speed on GPU and CPU is increased by about 11.84% and 30.99%, respectively. This proves to be very meaningful for deploying the model to application scenarios with insufficient hardware conditions and limited memory and for improving the adaptability of the model. The improved HiFi-GAN model has remarkable performance and good compatibility on the voice cloning task and achieves the highest combined score combined with x-vector embedding in all the tests.

Data Availability

The data that support the findings of this study are available from from LibriSpeech corpus, THchs-30 corpus, and VCTK corpus, which are publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Conceptualization, methodology, software were done by Zeyu Qiu. Software was done by Jun Tang.Data curation was done by Yaxin Zhang. Formal analysis was done by Jiaxin Li.Visualization was done by Xishan Bai. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 62171470).

References

[1] S. Arik, J. Chen, K. Peng, W. Ping, and Y. Zhou, "Neural voice cloning with a few samples," in *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, pp. 10019–10029, NeurIPS, Montréal, Canada, December 2018.

[2] Z. Kons, S. Shechtman, A. Sorin, C. Rabinovitz, and R. Hoory, "High quality, lightweight and adaptable TTS using LPCNet," 2019, https://arxiv.org/abs/1905.00590.

[3] C. Li, X. Ma, B. Jiang et al., "Deep speaker: an end-to-end neural speaker embedding system," 2017, https://arxiv.org/abs/1705.02304.

[4] Y. Chen, Y. Assael, and B. Shillingford, "Sample efficient adaptive text-to-speech," 2018, https://arxiv.org/abs/1809.10460.

[5] H. B. Moss, V. Aggarwal, N. Prateek, J. González, and R. Barra-Chicote, "Boffin tts: few-shot speaker adaptation by bayesian optimization," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7639–7643, IEEE, Barcelona, Spain, May 2020.

[6] Z. Zhang, Q. Tian, H. Lu, L. Chen, and S. Liu, "Adadurian: few-shot adaptation for neural text-to-speech with durian," 2020, https://arxiv.org/abs/2005.05642.

[7] E. Cooper, C. I. Lai, Y. Yasuda, F. Fang, and X. Wang, "Zero-shot multi-speaker text-to-speech with state-of-the-art neural speaker embeddings," in *Proceedings of theICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6184–6188, IEEE, Barcelona, Spain, May 2020.

[8] L. Wan, Q. Wang, A. Papir, and I. L. Moreno, "Generalized end-to-end loss for speaker verification," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4879–4883, IEEE, Calgary, Canada, April 2018.

[9] J. Shen, R. Pang, R. J. Weiss, and N. Schuster, "Natural tts synthesis by conditioning wavenet on Mel spectrogram predictions," in *Proceedings of the 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 4779–4783, IEEE, Calgary, Canada, April 2018.

[10] R. Skerry-Ryan, E. Battenberg, Y. Xiao et al., "Towards end-to-end prosody transfer for expressive speech synthesis with tacotron," in *Proceedings of the 35th International Conference on Machine Learning*, pp. 4693–4702, PMLR, Stockholm, Sweden, July 2018.

[11] Y. Wang, D. Stanton, and Y. Zhang, "Style tokens: unsupervised style modeling, control and transfer in end-to-end speech synthesis," in *Proceedings of the 35th International Conference on Machine Learning*, pp. 5180–5189, PMLR, Stockholm, Sweden, July 2018.

[12] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio Speech and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[13] T. Kinnunen, L. Juvela, P. Alku, and J. Yamagishi, "Non-parallel voice conversion using i-vector PLDA: towards unifying speaker verification and transformation," in *Proceedings of the IEEE International Conference on Acoustics*, pp. 5535–5539, IEEE, New Orleans, LA, USA, June 2017.

[14] Q. Wang, C. Downey, L. Wan, P. A. Mansfield, and I. L. Moreno, "Speaker diarization with LSTM," in *Proceedings of the2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5239–5243, IEEE, Calgary, Canada, April 2018.

[15] Y. Jia, Y. Zhang, R. Weiss et al., "Transfer learning from speaker verification to multispeaker text-to-speech synthesis," in *Proceedings of the32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, pp. 4480–4490, NeurIPS, Montréal, Canada, December 2018.

[16] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, *Deep Neural Network Embeddings for Text-independent Speaker Verification*Interspeech, Stockholm, Sweden, 2017.

[17] A. Van Den Oord, S. Dieleman, H. Zen et al., "WaveNet: a generative model for raw audio," *SSW*, vol. 125, p. 2, 2016.

[18] R. Yamamoto, E. Song, and J. M. Kim, "Parallel WaveGAN: a fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in *Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6199–6203, IEEE, Barcelona, Spain, May 2020.

[19] K. Kumar, R. Kumar, and T. De Boissiere, "Melgan: generative adversarial networks for conditional waveform synthesis," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, pp. 1–12, NeurIPS, Vancouver, Canada, December 2019.

[20] J. Kong, J. Kim, J. Bae, and HiFi-Gan, "Generative adversarial networks for efficient and high fidelity speech synthesis," in *Proceedings of the34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, pp. 17022–17033, NeurIPS, Vancouver, Canada, December 2020.

[21] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, https://arxiv.org/abs/1406.1078.

[22] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*NeurIPS, New Orlean, Louisiana, 2014.

[23] A. G. Howard, M. Zhu, B. Chen et al., "Mobilenets: efficient convolutional neural networks for mobile vision applications," 2017, https://arxiv.org/abs/1704.04861.

[24] T. Salimans and D. P. Kingma, "Weight normalization: a simple reparameterization to accelerate training of deep neural networks," in *Proceedings of the 30th Conference on Neural Information []Processing Systems (NIPS 2016)*, pp. 901–909, NeurIPS, Barcelona, Spain, December 2016.

[25] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks [EB/OL]," 2018, https://arxiv.org/abs/1802.05957.

[26] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an ASR corpus based on public domain audio books," in *Proceedings of the 2015 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5206–5210, IEEE, South Brisbane, Australia, April 2015.

[27] D. Wang and X. Zhang, "Thchs-30: a free Chinese speech corpus," 2015, https://arxiv.org/abs/1512.0188.

[28] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[29] Itut Rec, *P. 800: Methods for Subjective Determination of Transmission Quality*, International Telecommunication Union, Geneva, Switzerland, 1996.

[30] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-SNE effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.

[31] C. Veaux, J. Yamagishi, and K. MacDonald, *CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit*, The Centre for Speech Technology Research (CSTR), The University of Edinburgh, Edinburgh, Scotland, 2017.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6174255, 18 pages
https://doi.org/10.1155/2022/6174255

Research Article

A Defect Detection Model for Industrial Products Based on Attention and Knowledge Distillation

Ze-Kai Zhang, Ming-Le Zhou, Rui Shao, Min Li, and Gang Li

Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250013, China

Correspondence should be addressed to Gang Li; lig@qlu.edu.cn

Received 13 July 2022; Accepted 14 September 2022; Published 10 October 2022

Academic Editor: Wenming Cao

Industrial quality detection is one of the important fields in machine vision. Big data analysis, the Internet of Things, edge computing, and other technologies are widely used in industrial quality detection. Studying an industrial detection algorithm that can be organically combined with the Internet of Things and edge computing is imminent. Deep learning methods in industrial quality detection have been widely proposed recently. However, due to the particularity of industrial scenarios, the existing deep learning-based general object detection methods have shortcomings in industrial applications. This study designs two isomorphic industrial detection models to solve these problems: T-model and S-model. Both proposed models combine swin-transformer with convolution in the backbone and design a residual fusion path. In the neck, this study designs a dual attention module to improve feature fusion. Second, this study presents a knowledge distiller based on the dual attention module to improve the detection accuracy of the lightweight S-model. According to the analysis of the experimental results on four public industrial defect detection datasets, the model in this study is more advantageous in industrial defect detection.

1. Introduction

Quality detection is an important task in the industrial production process, which is of great significance to protect the personal safety of users and avoid economic losses. In industrial quality detection, big data analysis technology, the Internet of Things, edge computing, and other technologies are widely used. Studying an industrial detection algorithm that can be organically combined with the Internet of Things and edge computing is imminent. Early machine vision algorithms used manual feature selection and trained classifiers to identify defect features. This approach relies too much on the robustness of the extracted features, resulting in much time-consuming development.

In recent years, convolutional neural networks (CNNs) have rapidly developed in image classification, object detection, and image segmentation. However, due to the particularity of industrial defect detection scenarios, deep learning has not been applied widely. As shown in Figure 1, the brightness, colour, object size, and background discrimination of industrial pictures differ from natural scenes. Specifically, the industrial surface defect detection datasets scale is relatively small, unlike large-scale general datasets such as ImageNet, PASCAL VOC2007/2012, and COCO in classical computer vision tasks. Moreover, the industrial datasets differ significantly from the public datasets in the number of samples, sample scale, and proportion of positive and negative samples. This difference leads to less application of general-purpose object detectors in industrial product surface defect detection. In addition, most industrial detection scenarios are offline and require a light model, while existing general object detection models are challenging to meet actual needs.

This study proposes a novel surface defect detection model for industrial products. The model consists of two submodels, T-model and S-model. T-model has a large depth and high detection accuracy and is suitable for scenarios with no obvious requirements for speed. S-model has a small depth and high speed and is more suitable for edge computing scenarios with high-speed requirements. Both models consist of a backbone, neck, and detector. In the backbone, this study combines CNN and swin-transformer.

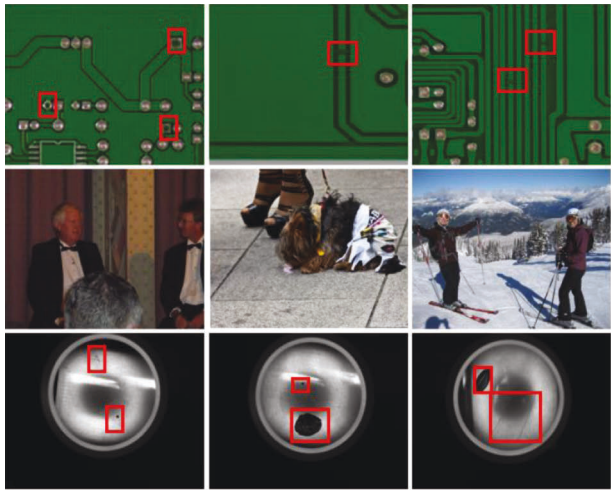


Figure 1: Comparison of industrial detection dataset with COCO dataset. The first row is the picture of the PCB defect dataset. The second row is the COCO dataset pictures. The third row is a picture of aluminium defects. It can be seen that in the field of industrial quality detection, there are many types of small objects, and the positive and negative samples are not balanced, making detection difficult.

CNN extracts local information in shallow features, swin-transformer extracts global information in deep layers, and each layer outputs weighted features through a unified residual path. In neck, this study designs a dual attention structure to focus on the features of object regions. In detector, this study designs multiple predictions heads on T-model and S-model.

In order to improve the detection effect of the lightweight S-model, a knowledge distiller is presented in this study. It contains the foreground attention-guided distillation of objects and global distillation. The knowledge distiller can significantly improve the detection accuracy of the S-model without extra overhead.

Overall, this study organically combines some existing techniques to create two detection models. An attention module is independently designed inside the model, and a knowledge distiller is designed to transfer the teacher network's knowledge for the lightweight network's guided training. The detection model designed in this study can not only be used for defect detection but also has a wide range of application scenarios in the direction of processing a large amount of image data in the cloud platform combined with the Internet of Things.

The main contributions of this study are listed as follows:

(1) This study proposes a novel model for surface defect detection of industrial products. Compared with the traditional CNN-based target detection methods, the model combines CNN and swin-transformer, which significantly improves the accuracy and can be used for application deployment in many image data analyses, edge computing, and other scenarios.

(2) This study designs two isomorphic object detection models. In backbone, this study extracts local and global important features by combining CNN and swin-transformer and designs a unified residual path to fuse features at different levels. In the neck, the attention module is designed to improve the effect of feature fusion. At the same time, this study designs the attention module to the decoupling detection head to improve detection accuracy.

(3) This study designs a novel attention-guided distillation strategy. The distiller uses the dual attention module to guide the generation of attention region features. The distiller transfers the knowledge of the T-model into the S-model, instructing the S-model to learn the T-model. The distiller in this study can obtain a lightweight, high-accuracy, and fast industrial quality detection model.

2. Related Work

2.1. Object Detection. In recent years, CNN-based object detection algorithms have been used in many domains. CNN-based object detection algorithms are generally divided into two categories: one stage and two stage. One-stage algorithms include YOLO [1–6], SSD [7], RetinaNet [8], and DSSD[9]. The methods are to directly divide the input image into multiple 1×1 grids, where each cell is responsible for detecting objects whose centre points fall within the grid. The methods significantly improve the detection speed, but the accuracy is slow. The two-stage algorithms include Rcnn [10], FastRcnn [11], FasterRcnn [12], and MaskRcnn [13]. These detection algorithms generate boxes via RPN, and the second-level detector uses boxes to conclude. The methods are better at detecting, but the speed is very low.

The one-stage object detector usually contains three parts: backbone, neck, and detector. The backbone is generally composed of multiple groups of convolutions for feature extraction. The well-known backbones include ResNet [14], ResNext [15], VGG [16], DenseNet [17], MobileNet [18], CSPDarkNet [19], and EfficentNet [20], etc. The neck fuses feature maps at different levels in the backbone to enhance the semantic and fine-grained features. Typical structures of neck mainly include FPN [21], PANet [22], Bi-FPN [20], etc. Two types of detectors commonly used in one-stage object detection are as follows: coupled and decoupled.

2.2. Attention. The attention mechanism is to make the network pay more attention to the area of the object and ignore the unimportant areas. Its essence is to use the relevant features to learn the weight distribution and then apply the learned weight to the original features. Attention makes the network pay more attention to the target object. Classical attention networks include SENet [23], SKNet [24], ResNext [15], CBAM [25], and self-attention [26].

2.3. Vision-Transformer. The transformer [26] is an attention-based encoder-decoder architecture in deep learning. Compared with CNN, vision transformer (Vit) [27] can obtain more refined global attention features and achieve good performance on multiple benchmarks such as

ImageNet, COCO, and ADE20k. However, it also has some drawbacks. First, the range of object scales for visual detection varies greatly, and the performance of the Vit [27] may not be optimal in different scenarios. Second, if the image resolution is high, transformer [26] requires much computation. Swin-transformer [28] solves this problem by shifting window partitions to calculate self-attention.

2.4. Knowledge Distillation. Knowledge distillation is a widely used method for model compression. Knowledge distillation is to transfer the knowledge of the T-model into S-model to improve the accuracy of lightweight models without adding extra computation. Knowledge distillation methods can be roughly categorized into response based [29–32], feature based [33, 34], and relation based [35]. Response-based methods use the output of the last layer of the teacher network to imitate the teacher's final prediction. This method is simple and efficient, but it relies on the output of the last layer and cannot make the student model obtain the supervision of the middle layer of the teacher model. Feature-based methods use the features in the middle layer of the teacher model to guide the student network to perform feature selection. However, two drawbacks need to be solved urgently. One is choosing the intermediate layer, and the other is matching the feature representation between the intermediate layer and the guiding layer if the layers' sizes are different. Relation-based methods take advantage of the inner product of the features between the two layers and employ the teacher structure as knowledge to guide the student model.

2.5. Application in Big Data, Industrial Internet of Things, Defect Detection, and Other Scenarios. Wang et al. proposed a deep learning model [36] combining GRU and LSTM and modeled the crack width of the dam, which can effectively predict the change of dam defects. Chen et al. proposed a training method for CNN and proposed a two-layer parallel training (BPT-CNN) architecture [37] in a distributed computing environment. BPT-CNN effectively improves the training performance of CNN, saves training time while maintaining accuracy, and has wide application fields.

In addition, there are many excellent survey proposed in the fields of big data, industrial Internet of things, etc. Pu et al. proposed an automatic fetal ultrasound standard plane recognition (FUSPR) [38] based on deep learning in an Industrial Internet of Things (IIoT) Environment. Cao et al. proposed a novel BERT-based deep space network (BDSTN) [39] to learn the demand pattern of taxis. Chen et al. combined CNN and LSTM to propose multiple closed spatiotemporal CNNs (MGSTC) [40] for traffic flow prediction; Wang et al. [41] proposed the application of big data technology to data mining, data analysis, and data sharing in large amounts of data, and to create huge economic benefits by using the potential value of data. Zhang et al. [41] summarized the existing blockchain-based systems and applications, which have broad application prospects in different data processing and transmission scenarios. Zhang et al. proposed a combined method of Weber local descriptor (WLD) and local binary pattern (LBP) for seam carving forgery detection [42]. To sum up, some detection methods can be combined with existing technologies such as CNN, not only in defect detection scenarios, but also in different application scenarios such as IoT and big data analysis.

3. Method

The flowchart of the proposed model is shown in Figure 2. The model consists of two submodels: T-model and S-model. In order to improve the detection effect of the lightweight S-model, this study designs a knowledge distiller and a dual attention module. By generating attention features from the pretrained T-model, the S-model training process is guided to learn the features from the T-model. In this way, the S-model reaches or even exceeds the detection effect of the T-model.

The structure of the T-model and S-model is shown in Figure 3. Both models propose a novel backbone combining convolutional layers with swin-transformer [28] to extract finer-grained image features. This study also adds weighted residual connection paths in the backbone. In the neck, a double-tower structure and dual attention modules are designed to improve the effect of feature fusion. Finally, multiple decoupled detectors are used to detect objects of different scales.

The main formula symbol table used in the rest of the method of this study is annotated as Table 1.

3.1. S-Model and T-Model

3.1.1. Backbone. Most general-purpose object detectors are based on CNN and employ large-scale convolution kernels in the first layer of the backbone to increase the receptive field. However, more detailed information will be lost with the increase of convolutional layers. Swin-transformer [28] uses window self-attention to significantly reduce computation load and extract small-scaled features. Therefore, this study designs a novel backbone based on swin-transformer [28].

As shown in Figure 3, our backbone includes block-B, block-S, SPP, and weighted residual fusion paths. The first layer selects a 3×3 convolution kernel to extract fine-scaled features. Block-B comprises three YOLOv5 C3, and three Conv cascaded. Block-S consists of swin-transformer [28], Conv cascade. The structure of the swin-transformer [28] is shown in Figure 4. In swin-transformer [28], the input features are divided into windows and encoded with relative position. The final output is obtained through multiple images down sampling (patch merging) and swin-block. In the steel surface classification experiment of swin-transformer [28], the effect of swin-transformer [28] with relative position encoding is better than that of swin-transformer [28] without it. Therefore, this study adds relative position encoding to block-S.

In order to further improve the detection of small objects, this study designs a unified weighted residual path (Figure 3) and performs fusion of features from the backbone. The fusion formula is as follows:

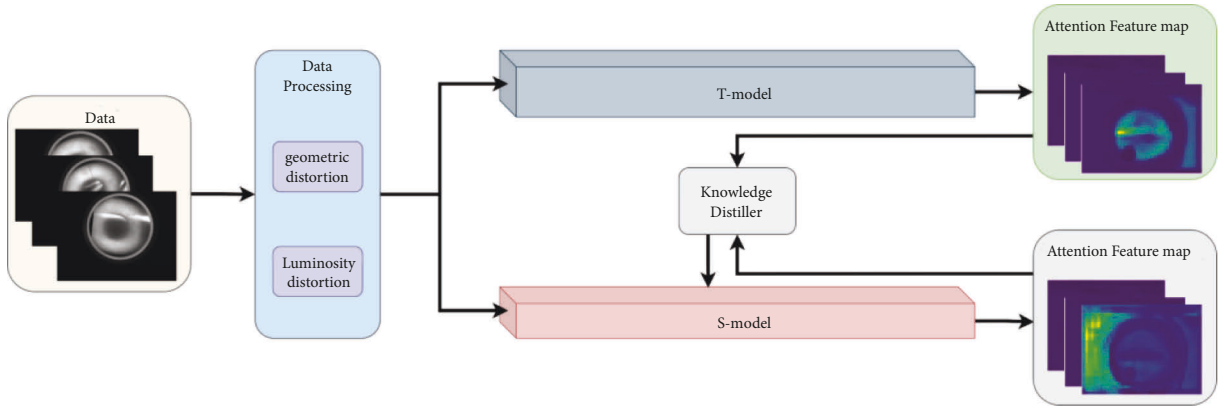


FIGURE 2: The flowchart of the model in this study. First, the dataset is expanded and enhanced by geometric and photometric transformations. Then, the data is input into the pretrained T-model and S-model, respectively, and the T-model output features guide the S-model training.

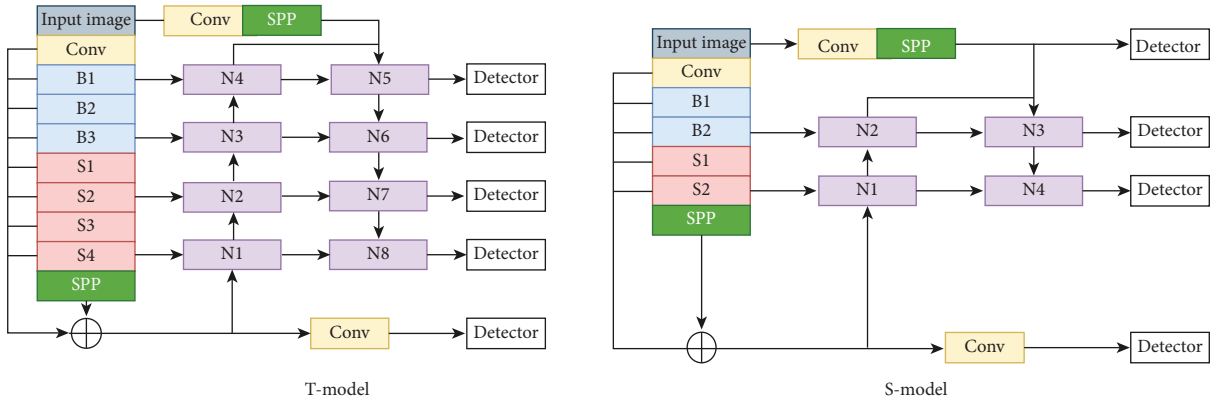


FIGURE 3: T-model and S-model. The two models are similar in structure, and the depth of the S-model network is much smaller than the T-model. The blue block-B is the Conv layer, the red block-S is the swin-transformer layer, the purple block-N is the feature fusion layer with the dual attention module added, and the detector is the decoupling detection head.

$$F = f\left(\sum_{i=1}^{n} Conv\left(conv\left(x_i\right)\right)\frac{expw_i}{\sum_1^n expw_j}\right), \tag{1}$$

where x_i is the original input feature and w is the adaptive learnable weight, $\sum_1^n w_j = 1$. $x_i$convolves with a 3×3 kernel to adjust the feature size, a 1×1 to integrate the channel, and then multiplies $expw_i / \sum_1^n expw_i$ before fusion.

3.1.2. Neck. The neck is designed to use better the features extracted by the backbone. In this study, the SPP structure is designed to enhance the invariance and robustness of image features. Figure 5 shows the structure of block-N in the top-down path and the bottom-up path in the neck. In block-N, this study designs a dual attention module, which will be introduced in detail in Section 3.2.

3.1.3. Decoupled Detector. In object detection, the role of the head in CNN is to regress the generated features to the bounding box and classified into some categories. Most methods use one head for classification and regression. This solution has limitations because classification relies more on fine-grained features, while regression relies more on semantic information to locate the object.

To stress the issue, YOLOX proposes the decoupled detector, which divides classification and regression into two types of problems. In this study, the decoupled detector is innovated (Figure 6), and a dual attention module is added to the classification branch. In our proposal, 6 decoupled detectors are used to detect features of the T-model, and four decoupled detectors to detect features of the S-model. The dense stacking of multiple decoupled detectors helps the models to detect objects of different scales.

3.2. Attention Module. Attention is usually embedded in CNNs and used to generate attention matrices to optimize features. CBAM [25] is a classic attention module that combines channel and spatial attention. Specifically, a $1 \times 1 \times C$ feature map will be obtained by pooling in channel attention, and a $H \times W \times 1$ feature map will be obtained by pooling in spatial attention. However, pooling leads to much information being lost. A mutual mapping between the three dimensions in a $C \times H \times W$ image is crucial to extracting attention.

Table 1: The main formula symbol used in the rest of the method.

Formula symbol annotation	Location
x_i: the original input feature w_i: the learnable weight of a single feature map w_j: the weight of any feature map to be accumulated.	Formula (1)
F: original input feature $M_c(F)$: channel attention $M_S(F)$: spatial attention F': the superposition of channel attention and spatial attention and multiplied by the original input F. $\otimes$: multiply by bit	Formulas (2)–(4)
$F^T_{i,j,k}$: the characteristics of teachers $F^S_{i,j,k}$: the characteristics of students f: converting $F^S_{i,j,k}$ to the same data dimension as $F^T_{i,j,k}$ H, W: specify the height and width of the feature C: channels	Formula (5)
x, y: the horizontal and vertical coordinates of the area object GT: the area position of the real frame	Formula (6)
M(F): the process of generating attention F: the original feature map T: distillation temperature, T = 20	Formulas (9) and (10)
L_{fg}: the distillation loss of foreground and background $Mask_{y,z}$: foreground mask $(1 - Mask_{y,z})$: background mask $ScaleMask_{y,z}$: scale mask $AttentionMask$: attention mask $\lvert F^T_{x,y,z} - f(F^S_{x,y,z})\rvert^2$: the difference between teacher feature and student feature	Formula (11)
$M(F^T_{x,y,z})$: the attention feature map of the teacher model $M(f(F^S_{x,y,z}))$: the attention feature map of the student model	Formula (13)
L_{fg}: the distillation loss of foreground and background L_A: attention mask loss L_G: distillation loss of the entire feature map QFL: QFocal loss L_{CIOU}: CIou loss $\alpha, \beta, \gamma, \delta, \varepsilon$: the weight parameters of the balance loss	Formula (17)
W: the updated weight parameter w_k: the weight parameter before the update L_i: the value of each loss n: the number of weight parameters	Formula (18)

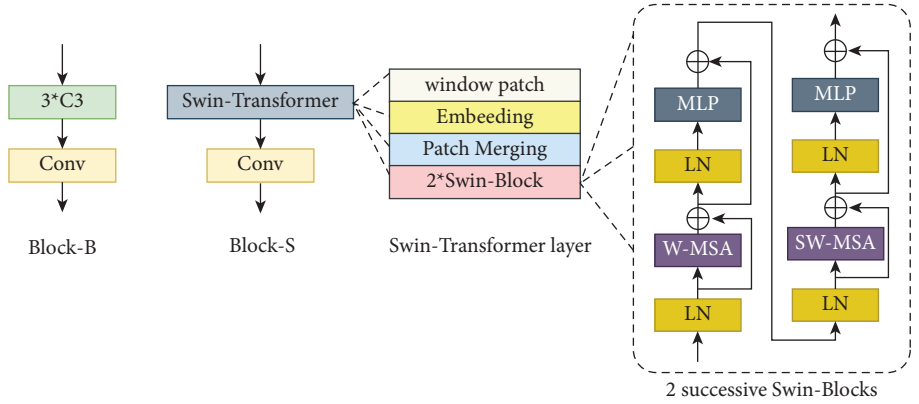


Figure 4: Structure of block-B and block-S.

The attention structure proposed in this study (Figure 7) includes channel and spatial attention. The resulting attention features are multiplied by input features as their features are combined. In terms of channel attention, this study first uses 1×1 convolution instead of pooling to achieve channel interaction and information integration. It then inputs the features into a two-layer neural network (MLP) to amplify the relationship between channels and

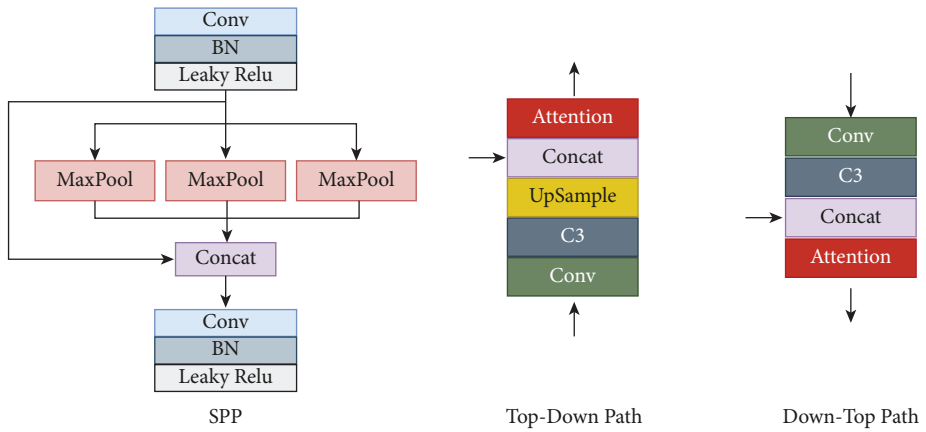


Figure 5: Block-N structure in neck, and block-N is reversed in the top-down and bottom-up paths.

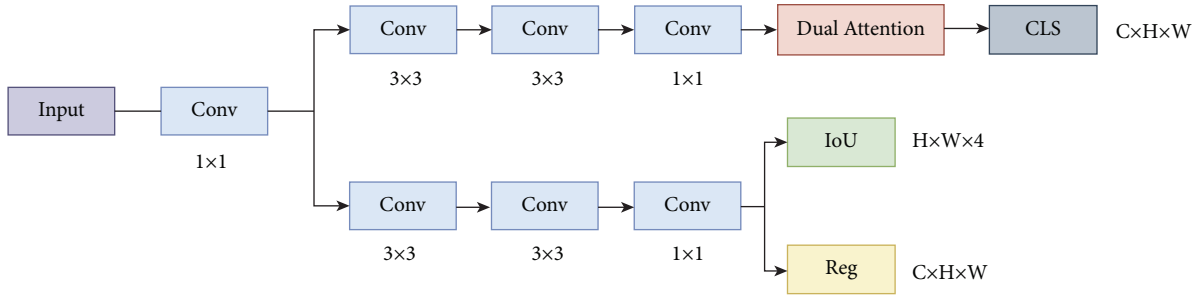


Figure 6: Decoupled detector.

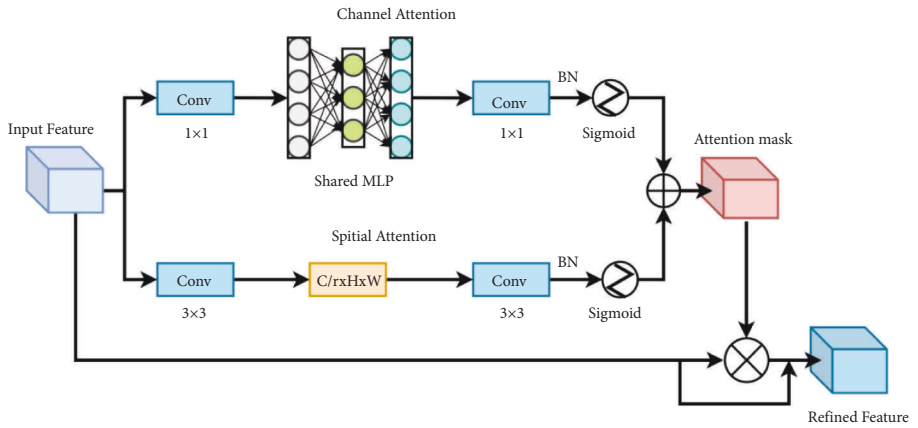


Figure 7: Attention flowchart. This study integrates spatial and channel attention and adds residual paths to them.

spatial dimensions. The number of neurons in the first layer in MLP is C/r (r is the reduction ratio, $r = 16$), the activation function is Relu, and the number of neurons in the second layer is C. In this study, the attention structure introduces BN to reduce the gradient dispersion and speed up the convergence speed. This study uses two dilated convolutions ($d = 4$, $r = 16$) identical to BAM [43] for spatial information integration to focus on spatial information.

The attention extraction formula in this study is as follows:

$$M_c(F) = sigmoid(BN(Conv(MLP(Conv(F))))), \quad (2)$$

$$M_S(F) = sigmoid(BN(Conv(Conv(F)))), \quad (3)$$

$$F' = (M_c(F) + M_s(F)) \otimes F, \quad (4)$$

where $M_c(F)$ denotes channel attention, $M_s(F)$ denotes spatial attention, and F' denotes the superposition of channel attention and spatial attention and multiplied by the original input F.

3.3. Knowledge Distillation Module. The lightweight model has speed and memory consumption advantages, but the detection effect is challenging to meet the requirements. Knowledge distillation is an effective method to improve the detection accuracy of small models. The general knowledge

distiller focuses on the extraction of the overall features. The formula is as follows:

$$L_D = \frac{1}{CHW}\sum_{i=1}^{C}\sum_{j=1}^{H}\sum_{k=1}^{W}\left|F_{i,j,k}^{T} - f\left(F_{i,j,k}^{S}\right)\right|^2, \tag{5}$$

where $FT/i, j, k$ and $FS/i, j, k$ denote the characteristics of teachers and students, respectively, and f denotes converting $FS/i, j, k$ to the same data dimension as $FT/i, j, k$. H, W specify the height and width of the feature, and C denotes the channel.

This study uses knowledge distillation to improve the S-model's performance. The image's foreground and background are distilled separately. The positive and negative samples are separated by separating the foreground and background, which solves the imbalance of positive and negative samples in the image. This study also designs an attention module to generate attention regions, forcing the student model to learn the vital features of the image.

In the distillation method of foreground and background, this study first sets the binary mask of foreground segmentation (the real position of the object frame) and sets objects within the ground truth box to 1 and objects outside the ground truth box to 0. It is expressed as follows:

$$Mask_{x,y} = \begin{cases} 1, & (x, y \in GT), \\ 0, & otherwise, \end{cases} \tag{6}$$

where x and y denote the horizontal and vertical coordinates of the area object and GT denotes the area position of the real frame. When the position of x, y falls in the real frame area, it is set to 1; otherwise, it is 0.

Due to the large-scale variation of objects in industrial detection datasets and the uneven distribution of positive and negative samples of objects, these will adversely affect the distillation effect. For this reason, this study uses the scaling mask to balance the object scale with reference to FGD [44] to solve the problem of an unbalanced object scale. The formula is as follows:

$$ScaleMask_{x,y} = \begin{cases} \dfrac{1}{H_{x,y} \times W_{x,y}}, & (x, y \in GT), \\ \dfrac{1}{\sum_{x=1}^{H}\sum_{y=1}^{W}\left(1 - Mask_{x,y}\right)}, & \end{cases} \tag{7}$$

where GT is the region of the ground truth. In this study, the scaling mask is used to normalize the foreground and background pixels. When there are two objects in an image, the bounding box of the large object may cover small object. In this study, when small objects and large objects are in a bounding box, the following formula is used:

$$Mask = \min\ (GT), \tag{8}$$

where GT represents the real box of the object. When a small object is surrounded by the box of a large object, the smallest bounding box is preferentially selected.

After separating the foreground and background, this study uses the designed dual attention module to generate the attention mask. The attention mask formula is as follows:

$$M(F) = \left(M_c(F) + M_s(F)\right) \otimes F, \tag{9}$$

$$AttentionMask(F) = softmax\left(\frac{M(F)}{T}\right), \tag{10}$$

where $M(F)$ denotes the process of generating attention, F denotes the original feature map, $M_c(F)$ denotes channel attention, $M_s(F)$ denotes spatial attention, T denotes distillation temperature, $T = 20$, and softmax is used to process features graph weights.

In the distillation process of foreground and background, binary mask, scale mask, and attention mask are used for attention-guided distillation. The loss function is as follows:

$$\begin{aligned} L_{fg} = &\sum_{x=1}^{C}\sum_{y=1}^{H}\sum_{z=1}^{W} Mask_{y,z} ScaleMask_{y,z} AttentionMask\left|F_{x,y,z}^{T} - f\left(F_{x,y,z}^{S}\right)\right|^2 \\ &+ \sum_{x=1}^{C}\sum_{y=1}^{H}\sum_{z=1}^{W}\left(1 - Mask_{x,y}\right) ScaleMask_{x,y} AttentionMask\left|F_{x,y,z}^{T} - f\left(F_{x,y,z}^{S}\right)\right|^2, \end{aligned} \tag{11}$$

where $L_{fg}, Mask_{y,z}, (1 - Mask_{y,z}), ScaleMask_{y,z}$, $AttentionMask$are the distillation loss of foreground and background, foreground mask, background mask, scale mask, attention mask, respectively, and $|F_{x,y,z}^{T} - f(F_{x,y,z}^{S})|^2$ is the difference between teacher feature and student feature.

In addition, this study uses the attention loss function to let the student model learn the attention mask of the teacher model. The loss formula is as follows:

$$L_A = SmoothL1\left(AttentionMask^T, AttentionMask^S\right), \tag{12}$$

where L_A is the attention mask loss function, $AttentionMask^T$ denotes the teacher's attention mask, and $AttentionMask^S$ denotes the student's attention mask.

In the feature distillation of the foreground, this study uses the designed attention module to distil the entire feature map. The loss formula is as follows:

$$L_G = \left| M\left(F^T_{x,y,z}\right) - M\left(f\left(F^S_{x,y,z}\right)\right)\right|^2, \quad (13)$$

where L_G is the distillation loss function of the entire feature map, $M(FT/x, y, z)$ is the attention feature map of the teacher model, and $M(f(FS/x, y, z))$ is the attention feature map of the student model.

3.4. Loss Function. In order to solve the unbalanced positive and negative samples in the PCB dataset, this study uses QFocal loss as the classification and confidence losses of the T-model and S-model. The formulas are as follows:

$$QFL(\sigma) = -\alpha_t \times |y - \sigma|^\beta \times [(1 - y)\log(1 - \sigma) + y log(\sigma)], \quad (14)$$

where y is the smoothed label in [0, 1] and σ is the prediction result. Focal loss introduces two factors α_t and $|y - \sigma|^\beta$, where $\alpha_t = y \times \alpha + (1 - y) \times (1 - \alpha)$ is used to balance positive and negative samples, and $|y - \sigma|^\beta$ is used to stress difficult detected samples. In addition, this study introduces CIoU Loss as the prediction box regression loss of T-model and S-model. The formula is as follows:

$$L_{CIOU} = 1 - IOU + R_{CIOU}\left(B_{pd}, B_{gt}\right), \quad (15)$$

where R_{CIOU} is the penalty term for the prediction box B_{pd} and the object box B_{gt}. CIoU loss considers the overlapping area, centre point distance, and aspect ratio in the prediction frame regression, which solves the problem of inconsistency between the real frame and the predicted frame during object detection. The normalized distance and penalty term between the centre points of the two bounding boxes are defined as follows:

$$R_{CIOU} = \frac{\rho^2\left(b, b_{gt}\right)}{C^2} + \alpha\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2, \quad (16)$$

where b, b_{gt} denote the centre points of B_{pd}, B_{gt} respectively, $\rho(x)$ is the Euclidean distance, and C is the diagonal length of the smallest bounding box covering these two boxes. α is a positive trade-off parameter. w, h are the width and height of the prediction box. In the selection of anchors, this study uses K-means to filter the anchors that meet the criteria.

In summary, the total loss function proposed in this study is as follows:

$$Total\ Loss = \alpha L_{fg} + \beta L_A + \gamma L_G + 2\delta QFL + \varepsilon L_{CIOU}, \quad (17)$$

where $L_{fg}, L_A, L_G, QFL, L_{CIOU}$ denote the distillation loss of foreground and background, attention mask loss, distillation loss of the entire feature map, QFocal loss, cIou loss, respectively, $\alpha, \beta, \gamma, \delta, \varepsilon$ are the weight parameters of the balance loss, respectively. In this study, an adaptive weight updater is designed to adjust the weight adaptively. The formula is as follows:

$$W = \frac{\exp(L_i)}{\sum_{i=1}^{n}\exp\ (L_i)} \times n \times w_k, \quad (18)$$

where W denotes the updated weight parameter, w_k denotes the weight parameter before the update, L_i denotes the value of each loss, and n denotes the number of weight parameters. In this study, each weight parameter is set to one before the training starts, and then in each training round, the weight parameter is updated according to the proportion of each loss value in the total loss. In this way, the loss with a large value will get a large weight in the next round of optimization, and the weight with a small value will be further reduced. At the end of the training, all tasks can be optimized almost simultaneously.

4. Experiments and Analysis

In this section, experiments are carried out on the PCB defect dataset, the NEU surface defect dataset, and the aluminium defect dataset.

4.1. Experimental Environment and Parameter Settings. This article implements the code in the PyTorch framework, version 1.9.0. CUDA version 11.4, cuDNN version 8.0. This study's model training and inference are performed on NVIDIA RTX 6000 × 1 and Intel i9-9900k@5 GHz × 1. Experimental platforms are GPU memory of 24 GB and CPU memory of 32 GB. The IDE used in the experiment is Pycharm 2019 Professional Edition.

In the training process, this study uses YOLOv5 as the baseline to build the T-model, uses the pretraining weights of YOLOv5 in the initial training of the T-model, and saves T-model weights after training. This study uses T-model for pretraining, loads the pretraining weights of the T-model to S-model for training, and uses knowledge distillation during the training process. Unless otherwise specified, the network in this study is trained with Adam for 300 iterations with an initial learning rate of 0.001, and the learning rate is adjusted using cosine annealing. This article uses a weight decay of 0.0001 and a momentum of 0.9. Also, the input image is resized to 640 × 640. The batch of the model in this study is eight during training, the batch is one during inferencing, and TensorRT is not used.

4.2. Object Detection and Evaluation Indicators. FPS [45]: in this study, the model inference is carried out under the same equipment conditions. The same size image is used to calculate FPS and evaluate the model's processing speed.

IOU [45]: object detection uses the IOU to calculate the degree of coincidence between the predicted box and the ground-truth box, which further measures the accuracy of detecting the corresponding object in a specific dataset.

mAP [45]: mAP is the sum of the average precision of all categories divided by the number of all categories. mAP@.5 is the model accuracy index when the IOU is 0.5. mAP@.5: .95 is obtained by calculating an mAP every 0.5 from IOU from 0.5 to 0.95 and finally averaging these maps.

4.3. Experiments on the Aluminium Defect Dataset. The aluminium defect dataset is the images of aluminium surface defects published by Baidu AI (Figure 8). It contains 412 images in total. There are four types of defects, i.e., Zhen_kong, ca_shang, zang_wu, and zhe_zhou, and one

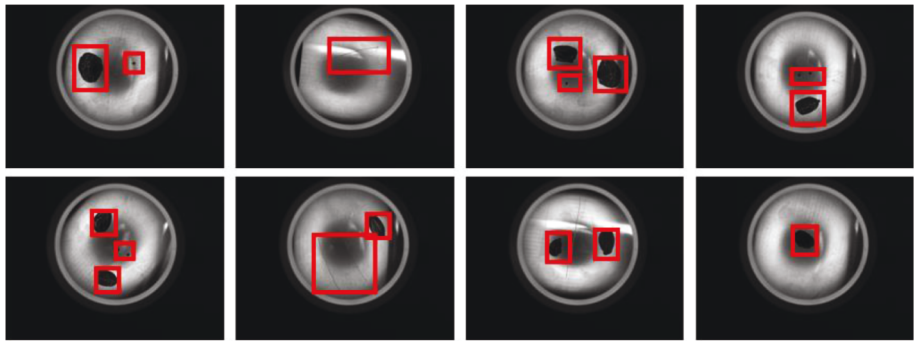


Figure 8: Example of aluminium defect.

image may contain different types of defects. In this study, the dataset is processed with photometric and geometric distortion. Specifically, this study brightens the picture and then flips and pans the picture to expand the dataset, increasing the number of pictures in the dataset to 1236. Then, the extended images are divided into a training set, validation set, and test set with a ratio of 8 : 1:1. It can be seen from Figure 8 that the surface defect scale of aluminium material varies greatly, which brings difficulties to the detection.

4.3.1. Comparative Experiments. Table 2 lists the comparison experiment results of aluminium defect datasets. It can be seen that the T-model of this study surpasses most classic object detection models, indicating that the model is more suitable for the field of industrial detection. T-model size is smaller than YOLOv5X, YOLOv4, YOLOR-P6, Faster-R-CNN, etc. The F1-score and mAP of the T-model are all in the leading position. T-model surpassed the newly proposed YOLOR 1.13% on mAP@.5, and F1-score exceeded 0.05. YOLOX has better detection performance than T-model, but T-model accuracy is close to YOLOX. However, due to the swin-transformer, the FPS advantage of the T-model is not apparent. In contrast, the S-model is lighter, the reasoning speed is fast, and the accuracy after knowledge distillation is close to the T-model.

In order to more intuitively show the detection effect of the T-model, Figure 9 shows the detection results of 16 pictures.

4.3.2. Ablation Experiments. Table 3 lists the ablation experiments performed with YOLOv5S as the baseline. In the ablation experiment of the backbone, this study uses the B4, S3, and SPP outputs in the backbone as the input of YOLOv5S neck for experimenting. The results in the second row show that when using the backbone of the T-model, mAP@.5 is 0.63% higher than that of YOLOv5S, and the other indicators are also slightly improved. On the neck ablation experiment, this study experiments with three C3s in the YOLOv5S backbone with one SPP output as the neck input. The third row shows that the detection effect is improved when introducing the neck in this study.

This paper also conducts ablation experiments with T-model as the baseline to verify the residual fusion path in the backbone and the down sampling path in neck. The ablation results are listed in Table 4. The first row shows the results without weighted fusion paths in the backbone and down sampling paths in neck. The second row shows that when adding the residual fusion path to the backbone, mAP@.5 increases by 0.58%, and the F1-score increases by 0.02. The third rows show that when introducing the down sampling path in neck, mAP@.5 is improved by 0.52%, and the F1-score is improved by 0.02. In general, the residual fusion path and the down sampling path are beneficial to industrial detection scenarios with drastic changes in scale.

4.3.3. Experiments on the Detector. This study examines several different detectors, including coupled detector, decoupled detector, and decoupled detector with added dual attention module. This study conducts experiments with T-model as the baseline. Specifically, this study experiments with these detectors on T-model, and the results are listed in Table 5. The experimental results in the second row show that the mAP@.5 and F1-score of decoupled detector are improved by 0.78% and 0.05, respectively, compared with coupled detector. The third line shows that the detection accuracy is further improved when adding the dual attention module to the classification branch of decoupled detector, indicating that the dual attention module improves the classification accuracy.

4.3.4. Experiments on the Dual Attention Module. This study analyses the features of CBAM. As shown in Figure 10, it is a heat map comparison between CBAM and the dual attention module. It can be seen that the dual attention module pays more attention to object areas such as wrinkles.

This study uses T-model as the baseline for experiments on the dual attention module (DA). The experimental results are listed in Table 6. It can be seen that the dual attention module effect surpasses CBAM.

4.3.5. Knowledge Distillation. This study conducts experiments related to knowledge distillation on lightweight models. The experiments are performed on isomorphism object detectors and heterogeneous object detectors, respectively. On isomorphic object detectors, this study experiments T-model and S-model, Efficientdet-d7, and Efficientdet-d3. This study experiments with faster-R-CNN-ResNet50, faster-R-CNN-

Table 2: Comparative experiment of the aluminium defect dataset.

Approach	Backbone	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1	Model size (M)	FPS
SSD	VGG-16	86.58	35.2	99.8	62.5	0.77	99.76	67
Faster-rcnn	ResNet-50	92.35	42.3	73.88	97.5	0.84	523	16
YOLOv3	DarkNet-53	95.9	54.4	94.7	95.7	0.95	234	19
YOLOv5S	CSPDarkNet	97.2	56.6	96.4	96.8	0.97	26	50
Efficientdet-d3	EfficientNet-B3	93.52	50.2	91.85	90.91	0.91	14.78	20
YOLOv5X	CSPDarkNet	94.3	54.9	95.1	94.7	0.95	167	21
Centrenet	Hourglass-104	60.3	38.6	51.09	60.61	0.55	124.61	48
Retinanet	ResNet-101	96.52	46.7	94.79	93.32	0.94	144.84	20
YOLOv4	CSPDarkNet-53	94.13	59.1	83.38	96.01	0.89	245.53	29
YOLOR-P6	CSPDarkNet	97	53.1	88.3	97.8	0.93	140.88	61
YOLOX-s	Darknet-53	99.2	58.2	98.91	99.3	0.99	34.2	41
T-model	Our backbone	98.53	58.8	98.85	98.38	0.98	81.6	19
S-model	Our backbone	97.82	55.86	97.83	96.87	0.97	25.1	36

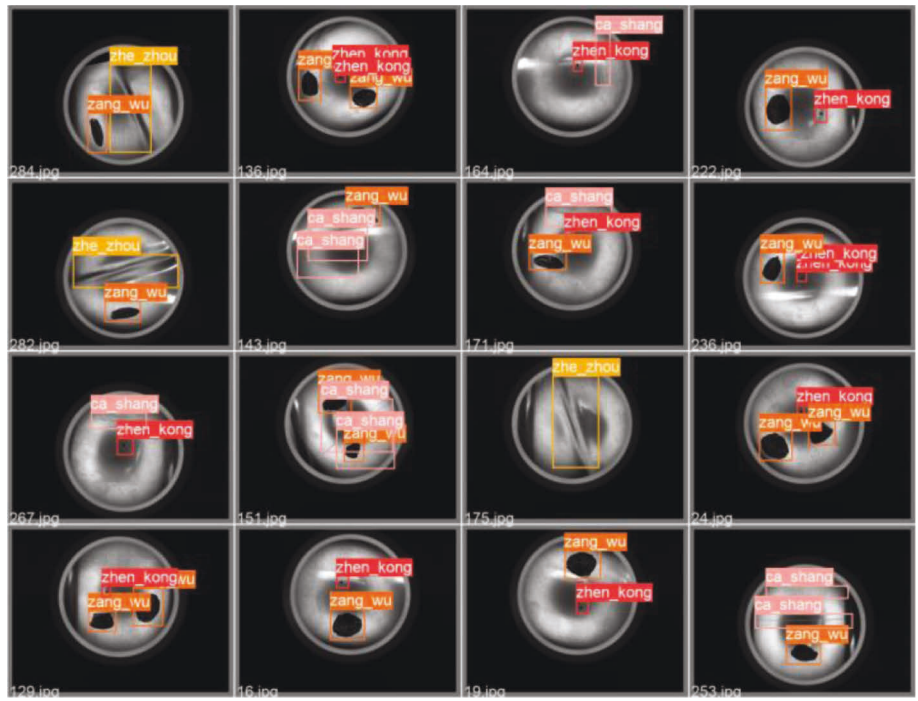


Figure 9: Effect of aluminium defect detection.

Table 3: Ablation experiment with YOLOv5S.

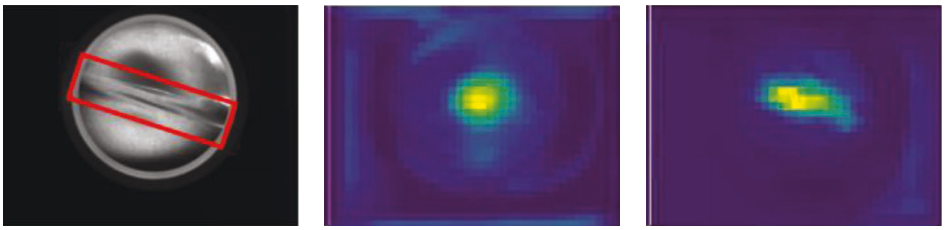
Approach and improvement	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
YOLOv5S	97.2	56.6	96.4	96.8	0.97
+Our backbone	97.83	56.94	97.84	96.87	0.97
+Our neck	97.71	57.13	97.1	97.38	0.98

Table 4: Ablation experiment with the T-model.

Approach and improvement	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
T-model	97.91	57.43	97.24	97.87	0.97
+Residual fusion path	98.02	57.57	98.16	98.47	0.99
+Residual fusion path + down sampling path	98.53	58.8	98.85	98.38	0.98

Table 5: Experiments on three detectors.

Detector	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
Coupled detector	97.24	57.05	86.4	97.31	0.93
Decoupled detector	98.02	57.7	98.16	98.47	0.98
Decoupled detector + attention	98.53	58.8	98.85	98.38	0.98



(a) (b) (c)

Figure 10: (a) Original image. (b) CBAM heatmap. (c) Dual attention module heatmap. The darker the yellow, the higher the attention.

Table 6: Comparison experiment between CBAM and dual attention module.

Approach	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
T-model without DA	97.52	57.2	96.75	97.87	0.97
+DA	98.53	58.8	98.85	98.38	0.98
+CBAM	96.68	53.15	95.75	96.83	0.96

Table 7: Knowledge distiller experiments.

Models	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)
S-model	95.76	54.23	95.82	95.25
S-model with KD	97.82	55.86	97.83	96.87
Faster-rcnn-VGG16	87.22	36.18	77.42	88.84
Faster-rcnn-VGG16 with KD	87.87	36.26	87.46	86.67
Efficientdet-d3	93.52	50.2	91.85	90.91
Efficientdet-d3 with KD	95.8	51.34	91.88	96.78
YOLOv4-tiny	84.43	35.79	80.46	86.01
YOLOv4-tiny with KD	84.94	35.6	80.44	86.9

VGG16, YOLOv4, and YOLOv4-tiny on heterogeneous object detectors. In addition, this study also experimented with output layer knowledge distillation on T-model and S-model as a comparative experiment. Table 7 lists the experimental results of various classic object detectors. It can be seen that knowledge distillation improves the detection performance of the S-model, especially when T-model and S-model are isomorphic. It proves that the knowledge distiller can improve the detection accuracy of small models without increasing the number of parameters.

4.4. Experiments on the PCB Defect Dataset. The PCB defect dataset has 1386 images, and its annotation files contain the object location information and classification. It contains six defects: missing hole, mouse bite, open circuit, spur, short, and spurious copper. Each image may contain multiple defective objects of the same type (missing holes in Figure 11). In the experiments in this study, 900 images of different defect types are selected for training (mainly minor defects that are difficult to identify). In our experiment, images and annotations are divided into the training set, validation set, and experiment set according to the ratio of 6 : 2 : 2.

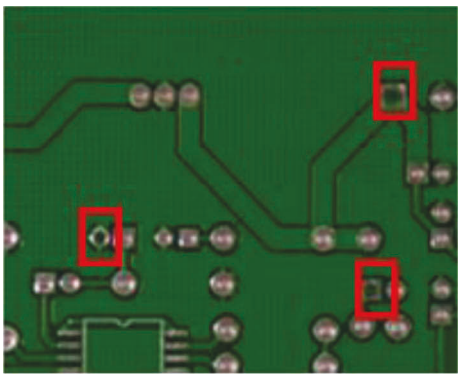


Figure 11: Missing hole.

This study conducts a comparative experiment on the PCB surface defect dataset, omitting FPS since the input size is still 640. It should be noted that the effect of the S-model is the result of distillation through the knowledge distiller. The

Table 8: Comparative experiment of PCB surface defects.

Approach	Backbone	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
SSD	VGG-16	17.1	10	19.93	54.64	0.29
Faster-rcnn	ResNet-50	78.33	50.24	67.74	74.51	0.7
YOLOv3	DarkNet-53	73.32	47.50	90.85	52.66	0.67
YOLOv5S	CSPDarkNet-53	91.2	51.8	81.82	92.12	0.86
Efficientdet-d3	EfficicientNet-B3	71.8	35	99.44	49.51	0.66
YOLOv5X	CSPDarkNet-53	91.8	50.7	95.2	91.3	0.93
Centrenet	Hourglass-104	43.4	14.53	43.79	52.2	0.48
Retinanet	ResNet-101	13.18	5	66.67	4.31	0.08
YOLOv4	CSPDarkNet-53	81.2	50.06	89.37	94.78	0.92
YOLOR-P6	CSPDarkNet	94.7	54.5	93.1	93.1	0.94
YOLOX-s	Darknet-53	95.84	62.03	87.16	93.73	0.91
T-model	Our backbone	94.79	58.02	88.14	95.65	0.91
S-model	Our backbone	92.51	56.98	83.79	92.33	0.88

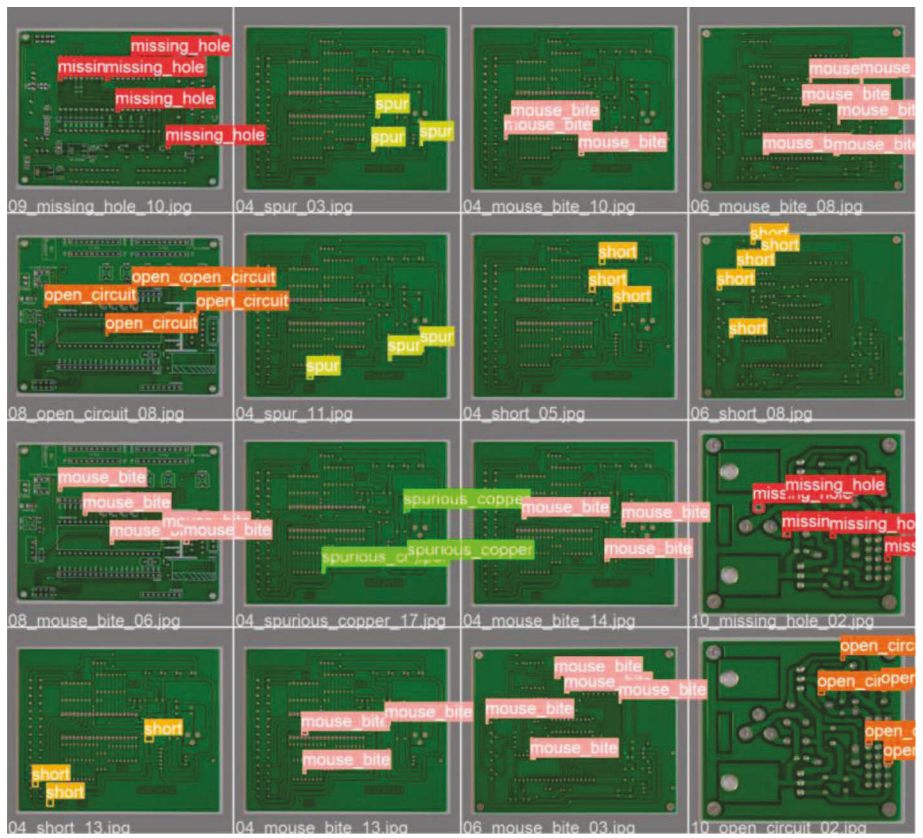


Figure 12: PCB defect dataset detection results.

comparison experiments are listed in Table 8. It can be seen that most of the classic object detection models are not effective for small object detection, and the T-model in this study shows a better performance.

In order to more intuitively show the detection effect of the T-model, Figure 12 shows the detection results of 16 pictures.

4.4.1. Ablation Experiment. Table 9 lists the ablation experiments performed in this study with YOLOv5S as the baseline. The experimental operations are consistent with the experiments on the aluminium defect dataset. The results in the second row show that after using the backbone of the T-model in this study, mAP@.5 is 3.68% higher than that of YOLOv5S indicating that the backbone in this study is beneficial for small object detection. The third line shows that the detection effect is improved when introducing the neck of the T-model. The dual attention module makes the network pay more attention to the object and improves the small object detection effect.

This study also uses T-model as the baseline to conduct ablation experiments to verify the residual fusion path in the backbone and the down sampling path in the neck. The

Table 9: Ablation experiment with YOLOv5S.

Approach and improvement	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
YOLOv5S	91.2	51.8	81.82	92.12	0.86
+Our backbone	91.98	52.48	96.05	92.46	0.94
+Our neck	92.05	53.09	96.62	93.6	0.95

Table 10: Ablation trials with T-model.

Approach and improvement	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
T-model	92.39	54.21	91.71	89.84	0.88
+Residual fusion path	93.65	56.44	89.94	94.43	0.92
+Residual fusion path + down sampling path	94.79	58.02	88.14	95.65	0.91

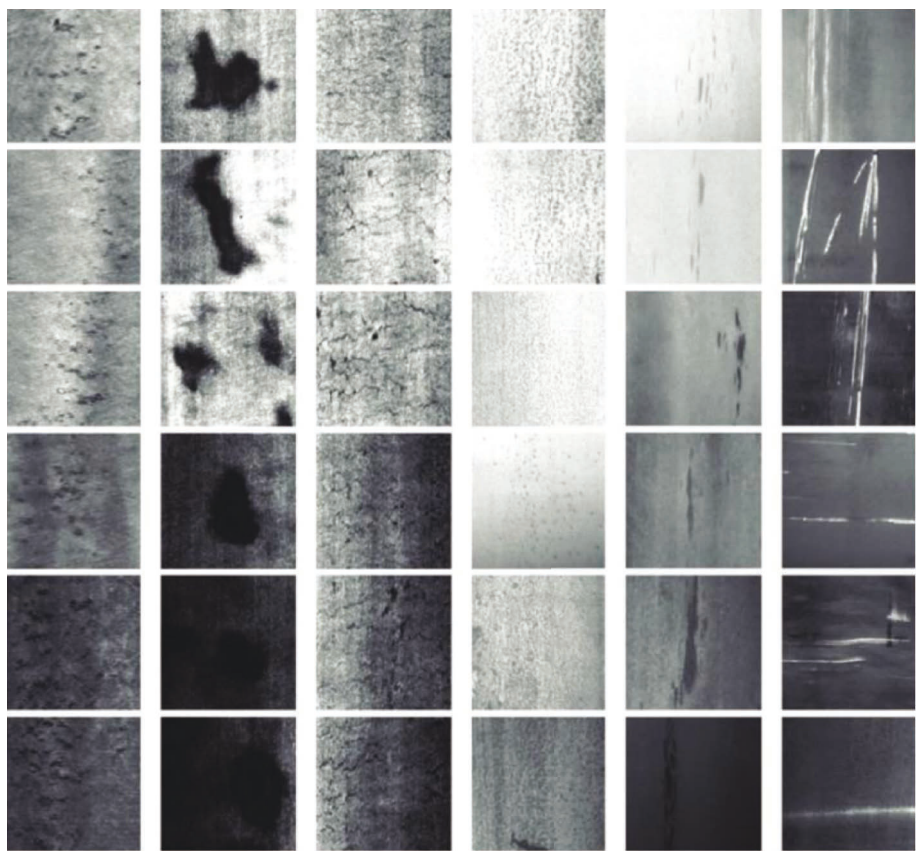


Figure 13: Example of surface defects in steel.

experimental operations are consistent with the experiments on the aluminium defect dataset. The ablation test results are listed in Table 10. The first row shows the results without weighted fusion paths in the backbone and down sampling paths in neck. The second row shows that when adding the residual fusion path to the backbone, the effect is significantly improved, indicating that the residual path can better fuse the fine-grained features in the backbone. The third row shows that by introducing the down sampling path in neck, mAP@.5 is improved by 2.4%, and the F1-score is improved by 0.03. It shows that the down sampling path is beneficial for small object detection because more fine-grained feature information is fused into neck. Overall, the residual fusion and the down sampling paths can improve the detection effect.

4.5. Experiments on the NEU Surface Defect Dataset. North-eastern university releases the NEU surface defect dataset. This dataset collects six typical defects on the surface of the hot-rolled strip: rolled-in scale, patches, crazing, pitted surface, inclusion, and scratches. Each image has several defects of the same type. The label file marks the category and specific location of the defective object.

Table 11: Comparative experiments on NEU defect datasets.

Approach	Backbone	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
SSD	VGG-16	66.73	26.6	87.27	34.52	0.5
Faster-rcnn	ResNet-50	76.95	39.9	44.14	87.86	0.59
YOLOv3	DarkNet-53	68.9	33.5	72.2	65.9	0.7
YOLOv5S	CSPDarkNet-53	70.9	35.6	78.4	64.3	0.35
Efficientdet-d3	EfficicientNet-B3	65.65	34	84.76	48.44	0.77
YOLOv5X	CSPDarkNet-53	72	37.7	66.8	72.5	0.69
Centrenet	Hourglass-104	39.24	13.5	55.5	18.2	0.29
Retinanet	ResNet-101	66.54	31.9	81.73	44.62	0.57
YOLOv4	CSPDarkNet-53	68.73	32.9	95.6	40.54	0.56
YOLOR-P6	CSPDarkNet	76.52	38.3	50.86	83.37	0.63
YOLOX-s	Darknet-53	79.38	40.23	53.73	84.88	0.66
T-model	Our backbone	75.56	39.68	53.45	83.16	0.65
S-model	Our backbone	74.41	39.2	49.05	82.28	0.62

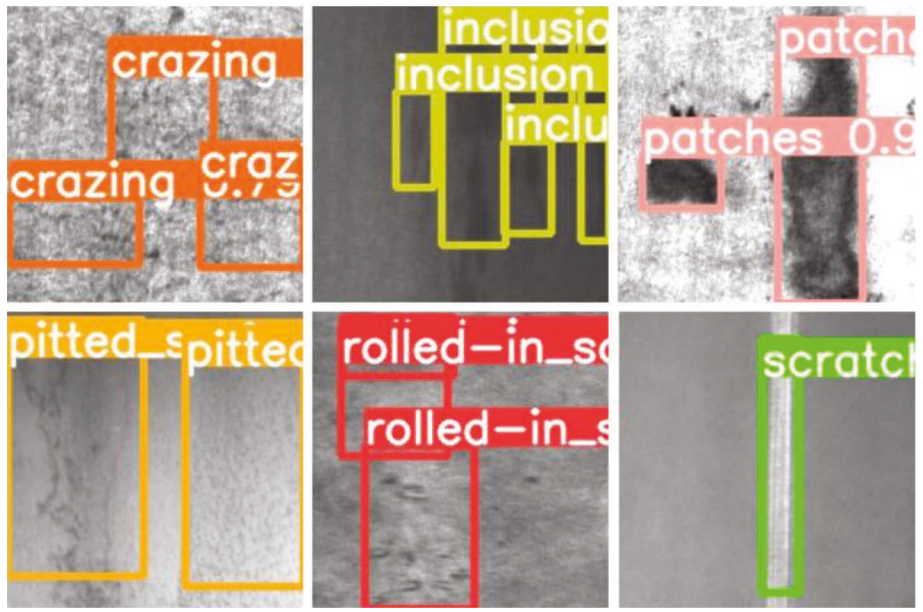


Figure 14: Steel defect detection effect.

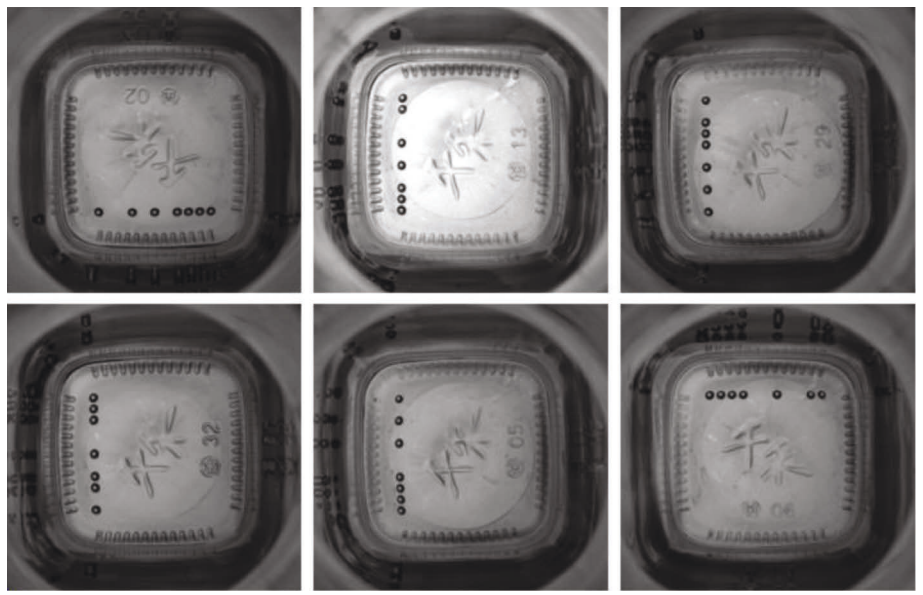


Figure 15: Example of bottom mould point data.

Table 12: Comparative experiments on NEU defect datasets.

Approach	Backbone	mAP@.5 (%)	mAP@.5-.95 (%)	Precision (%)	Recall (%)	F1
SSD	VGG-16	70.44	32.52	78.94	49.48	0.5
Faster-rcnn	ResNet-50	83.62	38.73	74.89	75.4	0.59
YOLOv3	DarkNet-53	91.37	40.66	84.39	92.33	0.7
YOLOv5S	CSPDarkNet-53	93.21	62.1	85.19	91.44	0.35
Efficientdet-D3	EfficicientNet-B3	86.39	55.67	67.49	89.9	0.77
YOLOv5X	CSPDarkNet-53	98.4	64.8	91.47	99.04	0.69
Centrenet	Hourglass-104	68.51	43.18	84.14	48.74	0.29
Retinanet	ResNet-101	80.79	49.89	66.28	77.56	0.57
YOLOv4	CSPDarkNet-53	96.1	50.06	89.37	94.78	0.56
YOLOR-P6	CSPDarkNet	99.3	59.9	53.5	99	0.63
YOLOX-s	Darknet-53	99.55	70.68	80.32	99.9	0.66
T-model	Our backbone	99.36	69.53	81.1	99.5	0.89
S-model	Our backbone	98.1	65.2	79.44	98.91	0.89

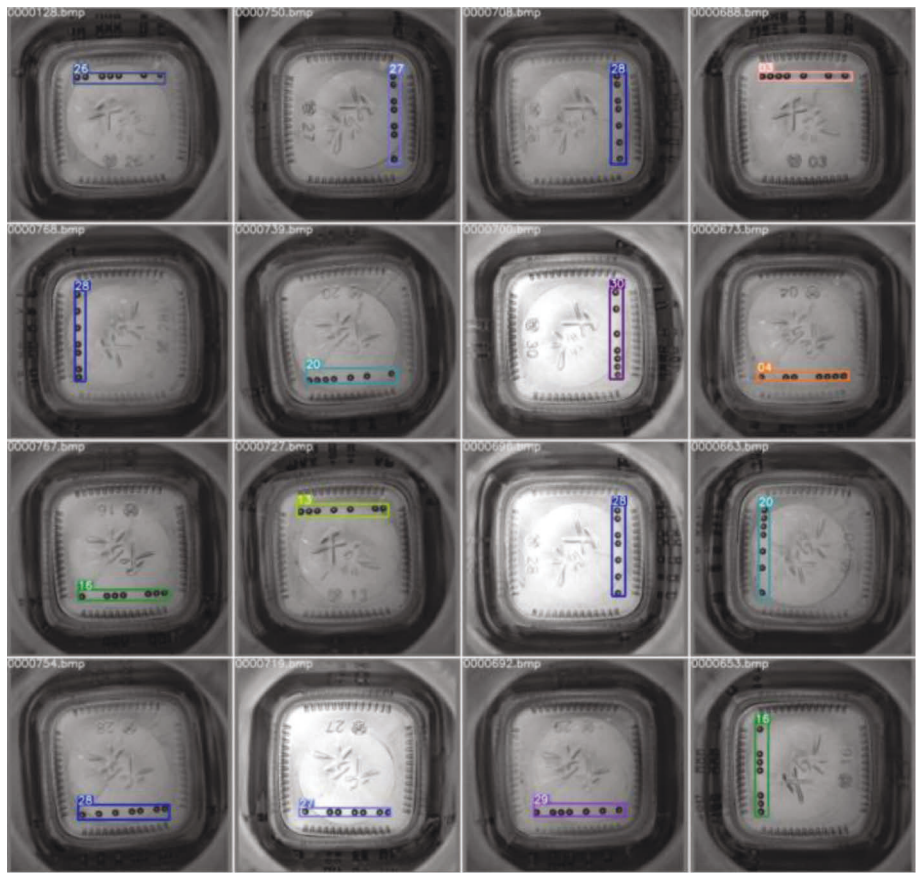


Figure 16: Glass bottle bottom mould point recognition effect.

Figure 13 shows different kinds of steel defects. The categories' defects have significant differences in appearance. Such scratches (last column) could be horizontal, vertical, slanted, etc. At the same time, each category has similarities in defects, such as rolled-in scale, crazing, and pitted surface. In addition, due to the influence of lighting and specific materials, the grey value of each category of defect images will also change. Object detection in the NEU surface defect dataset contains three difficulties: intraclass defects have significant appearance differences, interclass defects have similar aspects, and defect images are affected by changes in lighting and specific materials. According to the observation of the dataset, some cracks are concentrated in one direction because the steel is slender, and scratches are in any direction. Cracks and scratches are very similar, so scratches in specific directions are easily identified as cracks. Therefore, this study increases the data enhancement of cracks, rotates, splices crack pictures, and reduces the recognition error rate.

In order to show the generalization ability of this model in industrial detection, this study conducts a comparative

experiment on the NEU surface defect database, where the S-model is the result of knowledge distillation. The comparative test results are listed in Table 11. It can be seen that the T-model and S-model still have better performance, indicating that the object detection model of this study has a particular generalization ability in multiple industrial detection datasets. Figure 14 shows the different steel defect recognition results; it can be seen that all the defects are accurately identified.

4.6. Experiments on Glass Bottle Bottom Mould Point Dataset. This is a dataset for recognising mould point sequences at the bottom of glass bottles, with different permutations representing different product lot numbers. The mould point identification on the bottom of glass bottles is mainly used to locate the batch of glass products online to trace the product. The dataset contains 900 glass bottle bottom mould point images, each with a native resolution of 800×780 and 18 types. Each image has a corresponding label file. In this study, the label files correspond to the images one by one, and the training set, test set, and validation set are divided according to the ratio of $6:2:2$.

Figure 15 shows the images of different glass bottle bottom mould points, and the high similarity between objects makes detection difficult. In order to demonstrate the generalization ability of the model in industrial detection, this study conducts a comparative experiment on the glass bottle bottom mould point database, where the S-model results from knowledge distillation. The comparison test results are listed in Table 12. It can be seen that T-model and S-model still have better performance, indicating that the object detection model in this study has a specific generalization ability on multiple industrial detection datasets. Figure 16 shows the recognition effect of 16 different model point images. It can be seen that all model point objects are accurately recognized and positioned.

5. Conclusions

This study proposes a deep learning model for industrial quality detection. The model consists of T-model and S-model, which aims to meet detection tasks under different conditions. The model uses a combination of swin-transformer and convolution to extract the global information of the image. A dual attention module is designed to improve the neck's attention to important areas of the image, thereby improving the detection effect of the model. This study also designs a knowledge distiller using a dual attention module to improve the detection effect of the S-model. Finally, this study designs an adaptive loss weight updater to adjust the loss weights automatically. The experimental results show that the T-model in this study has high accuracy and is suitable for online data processing in scenarios such as IoT intelligent computing and big data analysis. The S-model in this study is fast and suitable for use in scenarios such as edge computing. In general, the model in this study can meet the needs of different scenarios and achieve a balance between accuracy and speed.

Data Availability

Aluminium defect dataset can be obtained from https://aistudio.baidu.com/aistudio/datasetdetail/13564 PCB defect dataset can be obtained from https://robotics.pkusz.edu.cn/resources/dataset/PCB/ and NEU surface defect dataset can be obtained from http://faculty.neu.edu.cn/me/songkc/Vision-based_SIS_Steel.html.

Disclosure

Zekai Zhang, Mingle Zhou, and Rui Shao are the co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was completed with the technical support of Key R&D Plan of Shandong Province (Soft Science Project) (2021RZA01016) and Plan of Youth Innovation Team Development of Colleges and Universities in Shandong Province (SD2019-161), Jinan.

References

[1] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas NV USA, June 2016.

[2] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, HI USA, July 2017.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, http://arxiv.org/abs/1804.02767.

[4] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, http://arxiv.org/abs/2004.10934.

[5] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," 2021, http://arxiv.org/abs/2107.08430.

[6] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You only look one-level feature," in *Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13039–13048, Virtual, June 2021.

[7] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot MultiBox detector," in *Proceedings of the 2016 Computer Vision 14th European Conference (ECCV)*, pp. 21–37, Amsterdam The Netherlands, October 2016.

[8] T. Y. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, Venice Italy, October 2017.

[9] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional Single Shot Detector," 2017, http://arxiv.org/abs/1701.06659.

[10] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on*

Computer Vision and Pattern Recognition (CVPR), pp. 580–587, Columbus OH USA, June 2014.

[11] R. B. Girshick, "Fast R-CNN," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago Chile, December 2015.

[12] S. Ren, K. He, R. B. Girshick, J. Sun, and R.-C. N. N. Faster, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[13] K. He, G. Gkioxari, P. Dollár, R. B. Girshick, and R.-C. N. N. Mask, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, Venice Italy, October 2017.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas NV USA, June 2016.

[15] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5987–5995, Honolulu, HI USA, July 2017.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of 2015 the 3rd International Conference on Learning Representations*, San Diego CA, USA, May 2015.

[17] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2261–2269, Honolulu HI USA, July 2017.

[18] A. G. Howard, M. Zhu, B. Chen et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017, http://arxiv.org/abs/1704.04861.

[19] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: a new backbone that can enhance learning capability of CNN," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1571–1580, Seattle WA USA, June 2020.

[20] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the Proceedings of 2019 the 36th International Conference on Machine Learning (ICML)*, p. 61056114, Long Beach California USA, June 2019.

[21] T. Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, Honolulu HI USA, July 2017.

[22] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8759–8768, Salt Lake City UT USA, June 2018.

[23] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, 2020.

[24] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 510–519, Long Beach CA USA, June 2019.

[25] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional Block Attention module," in *Proceedings of the 15th European Conference Computer Vision (ECCV)*, pp. 3–19, Munich Germany, September 2018.

[26] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, Long Beach CA USA, December 2017.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," in *Proceedings of the 9th International Conference on Learning Representations (ICLR 2021)*, Virtual Event Austria, May 2021.

[28] Z. Liu, Y. Lin, Y. Cao et al., "Swin transformer: hierarchical vision transformer using shifted windows," in *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002, Montreal QC Canada, October 2021.

[29] G. E. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," 2015, http://arxiv.org/abs/1503.02531.

[30] J. H. Cho and B. Hariharan, "On the efficacy of knowledge distillation," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4793–4801, Seoul Korea (South), October 2019.

[31] C. Yang, L. Xie, S. Qiao, and A. L. Yuille, "Training deep neural networks in generations: a more tolerant teacher educates better students," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*, pp. 5628–5635, Honolulu Hawaii, USA, January 2019.

[32] M. Phuong and C. Lampert, "Distillation-based training for multi-exit architectures," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1355–1364, Seoul Korea (South), October 2019.

[33] J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, "Self-speculation of clinical features based on knowledge distillation for accurate ocular disease classification," *Biomedical Signal Processing and Control*, vol. 67, Article ID 102491.

[34] M. Sepahvand and F. Abdali-Mohammadi, "Overcoming limitation of dissociation between MD and MI classifications of breast cancer histopathological images through a novel decomposed feature-based knowledge distillation method," *Computers in Biology and Medicine*, vol. 145, Article ID 105413.

[35] F. Tung and G. Mori, "Similarity-preserving knowledge distillation," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1365–1374, Seoul Korea (South), October 2019.

[36] J. Wang, Y. Zou, P. Lei, S. Sherratt, and L. Wang, "Research on recurrent neural network based crack opening prediction of concrete dam," *Journal of Internet Technology*, vol. 21, pp. 1151–1160, 2020.

[37] J. Chen, K. Li, K. Bilal, X. Zhou, K. Li, and P. S. Yu, "A Bi-layered parallel training architecture for large-scale convolutional neural networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 5, pp. 965–976, 2019.

[38] B. Pu, K. Li, S. Li, and N. Zhu, "Automatic fetal ultrasound standard plane recognition based on deep learning and IIoT," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7771–7780, 2021.

[39] D. Cao, K. Zeng, J. Wang et al., "BERT-based deep spatial-temporal network for taxi demand prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9442–9454, 2022.

[40] C. Chen, K. Li, S. G. Teo, X. Zou, K. Li, and Z. Zeng, "Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 4, pp. 1–23, 2020.

[41] J. Zhang, S. Zhong, T. Wang, H. C. Chao, and J. Wang, "Blockchain-based systems and applications: a survey," *Journal of Internet Technology*, vol. 21, no. 1, pp. 1–14, 2020.
[42] D. Zhang, Q. Li, G. Yang, L. Li, and X. Sun, "Detection of image seam carving by using weber local descriptor and local binary patterns," *Journal of Information Security and Applications*, vol. 36, pp. 135–144, 2017.
[43] J. Park, S. Woo, J. Y. Lee, and I. S. Kweon, "BAM: bottleneck attention module," in *Proceedings of the British Machine Vision Conference 2018 (BMVC)*, pp. 1–14, Newcastle UK, September 2018.
[44] Z. Yang, Z. Li, X. Jiang et al., "Focal and Global Knowledge Distillation for Detectors," 2021, https://arxiv.org/abs/2111.11837.
[45] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics Now*, vol. 10, no. 3, p. 279, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2605140, 20 pages
https://doi.org/10.1155/2022/2605140

Research Article

Ship Target Detection in Optical Remote Sensing Images Based on Multiscale Feature Enhancement

Liming Zhou[1,2] **Yahui Li**[1,2,3] **Xiaohan Rao**[1,2] **Cheng Liu**[1,2] **Xianyu Zuo**[1,2] **and Yang Liu**[1,2,3]

[1]*Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng, Henan, China*
[2]*School of Computer and Information Engineering, Henan University, Kaifeng, Henan, China*
[3]*Henan Province Engineering Research Center of Spatial Information Processing and Shenzhen Research Institute, Henan University, Kaifeng 475004, China*

Correspondence should be addressed to Cheng Liu; liucheng@henu.edu.cn

Received 1 July 2022; Revised 15 August 2022; Accepted 13 September 2022; Published 6 October 2022

Academic Editor: Wenming Cao

Due to the multiscale characteristics of ship targets in ORSIs (optical remote sensing images), ship target detection in ORSIs based on depth learning is still facing great challenges. Aiming at the low accuracy of multiscale ship target detection in ORSIs, this paper proposes a ship target detection algorithm based on multiscale feature enhancement based on YOLO v4. Firstly, an improved mixed convolution is introduced into the IRes (inverted residual block) to form an MIRes (mixed inverted residual block). The MIRes are used to replace the Res (residual block) in the deep CSP module of the backbone network to enhance the multiscale feature extraction capability of the backbone network. Secondly, for different scale feature maps' perception fields, feature information, and the scale of the detected objects, the multiscale feature enhancement modules—SFEM (small scale feature enhancement module) and MFEM (middle scale feature enhancement module)—are proposed to enhance the feature information of the middle- and low-level feature maps, respectively, and then the enhanced feature maps are sent to the detection head for detection. Finally, experiments were implemented on the LEVIR-ship dataset and the NWPU VHR-10 dataset. The accuracy of the proposed algorithm in ship target detection reached 79.55% and 90.70%, respectively, which is improved by 3.25% and 3.56% compared with YOLO v4.

1. Introduction

Target detection is important in military and civilian fields and has a wide range of application scenarios. Traditional target detection algorithms are mostly based on sliding windows and artificial feature extraction. Although it has achieved good results, there are still a series of deficiencies. Firstly, the method based on sliding window has high running cost and time complexity. Secondly, the robustness of manually designed features is poor [1–3]. Upon the development of deep learning, target detection methods based on deep learning have gradually replaced traditional methods. Many representative natural image target detection algorithms have been proposed and widely used, such as R-CNN [4], fast R-CNN [5], faster R-CNN [6], SSD [7], and YOLO [8–10] series.

As an important technology of ocean monitoring, ship target detection is significant in national security and maritime transportation safety. With the development of aerospace technology, ORSIs data are increasing. Therefore, more and more scholars try to apply the natural image target detection algorithm to ship detection in ORSIs. However, unlike natural images, ship detection in ORSIs is more difficult. As shown in Figure 1, the large field of view and the small ship target in Figure 1(a) cause a huge difference in the ratio of foreground pixels and background pixels of the ship target, which increases the difficulty of detection. In Figure 1(b), the texture and clarity of the shadow-affected

(a) (b) (c)

Figure 1: Display of ship target images in ORSIs. (a) Small-scale ship target. (b) Ship target obscured by clouds. (c) Ship targets against complex backgrounds.

ships are poor in the case of cloud and shadow occlusion. The small, medium, and large ships in Figure 1(c) have poor contrast between the foreground and background due to the influence of the wave background and wake [11]. Therefore, ship target detection in ORSIs still faces great challenges.

In ORSIs, large field of view, low resolution, and complex background make the detection precision of large, medium, and small ships generally low. Therefore, how to improve the detection precision of multiscale ship targets has become a research hotspot of scholars. Zhou et al. [12] proposed the MSSDNet (multiscale ship detection network). Firstly, CSPMRes2 (cross stage partial network with modified Res2Net) is used as the basic module of the backbone network for multiscale feature extraction. Secondly, FC-FPN (feature pyramid network with fusion coefficients) is used for multiscale feature adaptive fusion. The multiscale ship detection accuracy of this method has reached an advanced level. However, compared with the baseline network, the network introduces a large number of parameters and computation, the reasoning speed is only half of the baseline network, and the model size is about 1.7 times of the baseline network. Zhou et al. [13] proposed a ship object detection method based on feature enhancement. It uses the EIRM (elastic inception residual module) module to perform feature enhancement on the low- and middle-level feature maps. EIRM module extracts the feature information of multiscale ship targets through scaling strategy and uses SGPANet for further extraction and fusion of multiscale features. This method only adds a small amount of parameters relative to the baseline network. Wang et al. [14] proposed the FSoD-Net (full-scale object detection network). Firstly, a wider backbone network is proposed for feature extraction. For different scale feature maps, it adopts different feature extraction modules and regression layers. This method has achieved advanced performance in full-scale target detection in ORSIs.

According to the above analysis, aiming at the problem that it is difficult to detect multiscale ship targets in ORSIs, this paper proposes a ship target detection method based on multiscale feature enhancement based on YOLO v4 [15]. Experiments on the LEVIR-ship and NWPU VHR-10 datasets show that the proposed method achieves better results than YOLO v4. The main contributions of this paper are summarized as follows:

(1) The MIRes are proposed and applied to the deep layers of the backbone network. Firstly, to enhance the feature extraction ability of the backbone network, the Res in the deep CSP module of the backbone network is replaced by the IRes. Secondly, to extract the multiscale feature information of the ship target, the mixed convolution is improved and the depthwise separable convolution of the IRes is replaced by the modified mixed convolution.

(2) Two multiscale feature enhancement modules, SFEM and MFEM, are proposed for different size feature maps to enhance the features of low- and middle-level feature maps, respectively. SFEM and MFEM act on feature maps of 52×52 and 26×26 scales, respectively, and use atrous convolutions with different dilation rates to enhance the receptive field of feature maps while extracting information.

(3) Experiments on the LEVIR-ship dataset show that the proposed method achieves higher accuracy in the detection of multiscale ships in ORSIs. Simultaneous ablation experiments demonstrate the benefit of all partial improvements in this paper. The extended experiments on the NWPU VHR-10 dataset show that the method in this paper can achieve good results on different datasets and different categories.

This paper is organized as follows: Section 2 reviews some related work. Section 3 introduces the proposed method in detail. Section 4 is the experimental part, including comparative experiments and ablation experiments. Section 5 gives the conclusion.

2. Related Work

2.1. One-Stage Detection Algorithm Backbone Network. With the development of convolutional neural network classification techniques, scholars have proposed deeper

networks for classification tasks, such as ResNet50 [16], Darknet53 [10], and ResNet-101 [16]. These deep networks have powerful feature extraction capabilities, so they are widely used in the backbone network of target detection. Although the deep convolutional neural network has strong robustness, a large number of convolutions will cause the loss of small target pixels and the destruction of texture structure, so its applicability for small target detection is poor.

Qi et al. [17] used the YOLO v3 framework with auxiliary networks for object detection. The auxiliary network is lighter than the backbone network and has fewer convolutions, so it can retain more location information. The backbone network and auxiliary network extract features at the same time, and the location information extracted from the auxiliary network is transmitted to the backbone network. The combination of the deep semantic information extracted by the backbone network and the location information extracted by the auxiliary network greatly improves the detection accuracy of the network. Qing et al. [18] used the improved RepVgg as the backbone network for feature extraction. The backbone network uses multibranch structure for feature extraction during training, and each convolution is followed by a residual connection to accelerate the convergence of the network. At the same time, considering multiple branches will increase inference time, so only single-branch structure is used when inference. Then, the neck network uses the improved FPN [19] for feature fusion and finally uses four scales for detection, which achieves good detection results. Xu et al. [20] used a densely connected network (DenseNet) to enhance the feature extraction ability of YOLO v3 backbone network for poor detection accuracy in ORSIs object detection. Then, some of the residual blocks in the third and fourth residual units of the Darknet53 backbone network are replaced by dense connection blocks. Finally, four detection heads are used for detection, and the detection accuracy is significantly improved.

Sandler et al. [21] explored the impact of bottleneck structure on accuracy and parameters, proposed IRes (inverted residual block), and built a lightweight MobileNet V2 network based on the IRes. The structure of the IRes makes the feature extraction part have a wider channel and can extract richer feature information. Meanwhile, the introduction of depthwise separable convolution makes the module lightweight. Although depthwise separable convolutions tend to cause the problem of accuracy degradation, the wider structure makes up for this shortcoming. Tan et al. [22] discussed the effect of different convolution kernel sizes on network performance, proposed mixed convolution, and built MixNet using mixed convolution for natural image target detection. Different from traditional convolution, mixed convolution divides the input channels into different groups, then each group corresponds to a depthwise separable convolution with different kernel sizes, and finally fuses the outputs of each convolution. Mixed convolution introduces different convolution operations in a single convolution, so it can extract feature information of different scales from a single feature map while reducing part of the amount of computation and parameters.

2.2. Feature Enhancement Module. To solve the problem that multiscale targets are difficult to detect ORSIs, researchers have applied feature pyramids such as FPN [19] and PANet [23] for multiscale detection. Zhang et al. [24] found that there was a problem with the simple fusion of feature pyramids, so they proposed a multilevel feature pyramid. Firstly, they used CAFUS (content-aware feature upsampling) instead of upsampling to solve the fusion problem of feature maps of different scales. Then, FEM (feature enhancement module) is proposed to enhance the fused feature map. Wang et al. [25] proposed SE-SSD (spatial enhanced single shot multi box detector). Firstly, the spatial information is enhanced by increasing the number of image channels. Secondly, the output of the backbone network is modified, and a pooling operation is removed. Finally, a context feature enhancement module is designed to act on the middle and low layer feature maps to enhance the receptive field of the middle and low layer feature maps. Wang et al. [26] proposed an improved YOLO v3 algorithm for multiscale object detection. Dense connection blocks are used instead of residual units in Darknet53 to enhance the feature extraction capability of the backbone network. The feature enhancement module is proposed to act on the middle- and low-level feature maps to enhance the location information contained in the shallow feature maps. Finally, the FPN is improved. This method significantly improves the detection accuracy of small targets in ORSIs.

Liu et al. [27] proposed a lightweight RFB inspired by the human receptive field. RFB mainly includes two parts: multibranch structure and atrous convolution [28]. Each branch structure contains convolution kernels of different sizes and atrous convolutions with different expansion rates. The original RFB structure mainly consists of three branches, each of which contains a convolution kernel of 1×1, 3×3, and 5×5, corresponding to dilated convolutions with expansion rates of 1, 3, and 5, respectively. The convolution kernels of different sizes can better simulate the receptive field and extract feature information, which is significantly better than the convolution kernel of fixed size. Atrous convolution can capture feature information in larger regions without increasing parameters.

Although the above methods have achieved good detection accuracy, Chen et al. [29] pointed out that the key to the feature fusion pyramid is to divide and conquer, so it only uses a single-layer feature map for detection. However, single-layer feature maps have poor detection accuracy for small objects. Therefore, this paper retains PANet for multiscale feature map fusion and then uses different modules for feature enhancement of different scale feature maps to improve the accuracy of multiscale ship detection.

3. Methodology

In this section, the algorithm proposed in this paper is introduced in detail through four sections: Section 3.1; overall network structure; Section 3.2;backbone with MIRes; Section 3.3;PANet with multiscale feature enhancement; and Section 3.4;LOSS function.

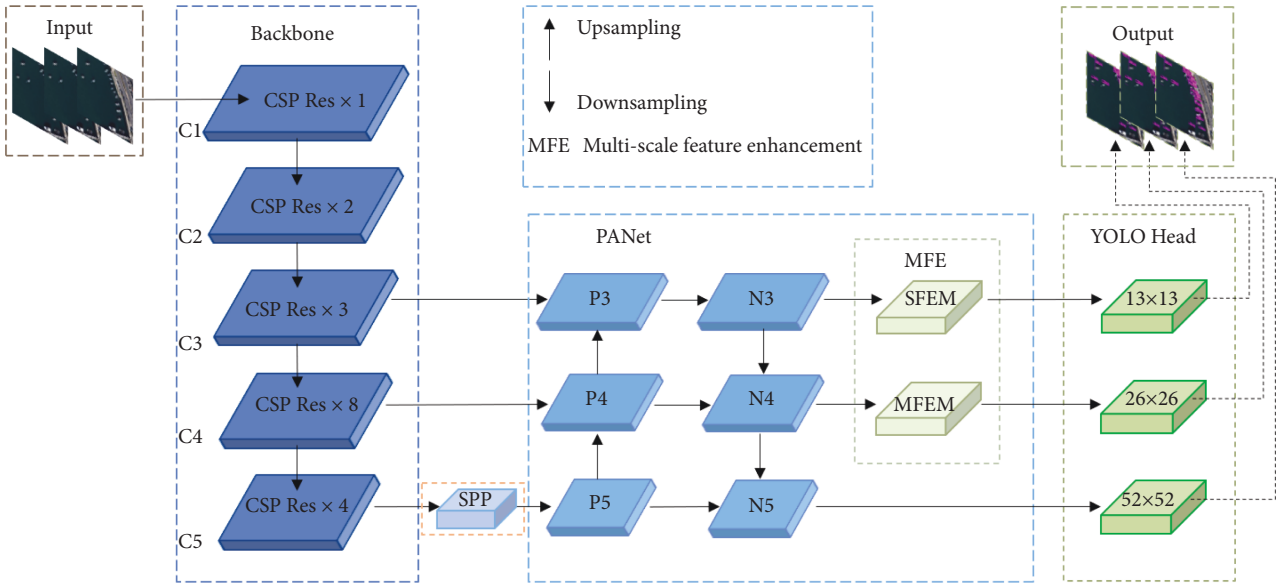


Figure 2: Overall network structure diagram.

Input: $T = \{T_1 \cdots T_n\}$, T is the da taset images
Begin: $T_1 \cdots T_n = 416 \times 416$, resize input image to 416×416.
$C = \{C_1 \cdots C_X\}, 1 \le x \le 5$.$C$ is the output characteristic diagram of each stage of the backbone network.
for C_X in C *do*:
if $x \le 2$
continue.
end
else if: $2 < x \le 4$
PANet←C_x**.** Send the middle and low-level feature map extracted from the backbone network into PANet for feature fusion.
end
else:
$C_x = SPP(C_x)$, PANet←C_x**.** After the C_5 feature map passes through the SPP module, it is sent to PANet for feature fusion.
end
$P_s = SFEM(P_s)$,P_m = MFEM(P_m). Feature enhancement of middle and low-level feature maps P_s and P_m after PANet fusion.
end
Output: detect(P_s, P_m, P_l)**.** The enhanced middle and low-level feature map P_s and P_m and high-level feature map P_l are input into the detection head for detection, and finally the detection results are output.

Algorithm 1: Multiscale feature enhanced ship target detection.

3.1. Overall Network Structure. The overall network structure of the algorithm in this paper is shown in Figure 2. The network mainly consists of four parts: the CSP Darknet53 backbone network is used for feature extraction, the PANet neck network is used for feature fusion, MSFE is used for multiscale feature enhancement, and the YOLO Head is used for detection. The backbone network contains five CSP modules (C1–C5). C1–C4 contain different numbers of residual blocks (1,2,8,8), and C5 contains four MIRes. Among them, the 13×13 and 26×26 feature maps extracted by the third and fourth CSP modules are directly used as the input of PANet. The feature map of size 52×52 extracted by the fifth CSP module first passes through the SPP module and then is input into PANet for feature fusion. PANet fuses feature maps of different scales through upsampling and downsampling. After the feature maps are fused by PANet, the 13×13 and 26×26 scale feature maps are sent to the detection head together with the 52×52 scale feature maps through the multiscale feature enhancement modules—SFEM and MFEM—for detection. Algorithm 1 describes the basis idea of multiscale feature enhanced ship target detection.

Figure 3 is the structure of the SPP module. In the SPP module, the input feature map first goes through three different sizes of max pooling layers (5, 9, 13), and then the output of each pooling layer is fused with the input feature map. The SPP module was proposed for input image resizing [30]. In this paper, it is mainly used to increase the reception range of backbone features and separate context features.

Figure 4 shows the CSP module structure. In the CSP structure, the input feature map is sent to two branches of

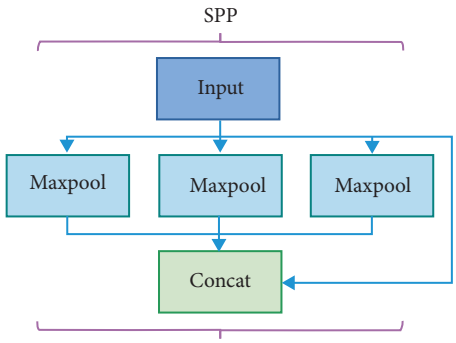


Figure 3: SPP module structure diagram.

different depths, and the branch containing Res × N is responsible for feature extraction, and then directly fused with another branch. The CSP module can improve the accuracy of the model while reducing weight.

3.2. Backbone with MIRes. The amount of feature information is closely related to the accuracy of object detection, and the C5 feature map of the CSP Darknet53 backbone network contains generous contextual feature information [31]. Therefore, this paper considers to enrich the feature information of C5 by improving the feature extraction capability of the fifth CSP module to improve the accuracy of multiscale ship target detection. At present, many researchers improve the feature extraction ability of the network by increasing the depth, but the increase of the depth is always accompanied by a surge in the amount of computation. Therefore, in order not to increase the redundant computation, this paper considers increasing the width of the network instead of the depth.

The wider network structure of the IRes enables it to have strong feature extraction capabilities, and the presence of depthwise separable convolutions makes it lightweight enough. However, for multiscale ship detection, it still lacks the ability to extract multiscale features. Therefore, this paper proposes MIRes, which introduces an improved mixed convolution into the IRes. The MIRes are used to replace the Res in the fifth CSP module, which greatly improves the multiscale feature extraction capability of the backbone network with a small increase in the amounts of parameters and computation.

The MIRes structure is shown in Figure 5, where MC is a mixed convolution and c is the channels number. MIRes adopts the IRes structure, and the number of network channels is expanded by 6 times compared with Res. Channel extended mixed convolution is used for multiscale feature extraction. Table 1 shows the network architecture and parameters of Res and MIRes.

For the same input, we can use a simple formula to show the difference in the output feature maps of mixed convolution and depthwise separable convolution. In this case, the input feature map size h (height) and w (width) are equivalent and the output feature map size is the same. Assuming that a depthwise separable convolution is$W^{(k,k,c,m)}$, $Y^{(h,w,c\bullet m)}$ is the output tensor. Then, each output feature map can be represented by the following formula (1). Different from the depthwise separable convolution, if the mixed convolution input channels are divided into g groups, the number of input and output channels is equal. Similarly, the convolution kernels are also divided into g groups. Then, the output for the t-th group can be expressed in formula (2). The total output of the mixed convolution can be represented in formula (3).1

$$Y_{x,y,z} = \sum_{-\frac{k}{2}\le i\le\frac{k}{2},-\frac{k}{2}\le j\le\frac{k}{2}} X_{x+i,y+j,z/m}\bullet W_{i,j,z}\forall_z = 1,\cdots,m\bullet c, \tag{1}$$

$$\hat{Y}^t_{x,y,z} = \sum_{-\frac{k_t}{2}\le i\le\frac{k_t}{2},-\frac{k_t}{2}\le j\le\frac{k_t}{2}} \hat{X}^t_{x+i,y+j,z/m}\bullet\hat{W}^t_{i,j,z}\forall_z = 1,\cdots,m\bullet c_t, \tag{2}$$

$$Y_{x,y,z_0} = \mathrm{Concat}\left(\hat{Y}^1_{x,y,z_1},\cdots\hat{Y}^g_{x,y,z_g}\right). \tag{3}$$

In formula (1), $k \times k$ is the size of the convolution kernel, c is the input channel, and m is the channel multiplier. In formula (2) $\langle\hat{X}^{(h,w,c_1)},\cdots\hat{X}^{(h,w,c_g)}\rangle$ is the mixed convolution channel of group g, and $\langle\widehat{W}^{(k_1,k_1,c_1,m)},\cdots\widehat{W}^{(k_g,k_g,c_g,m)}\rangle$ is the depthwise separable convolution channel of group g. In formula (3), $z_0 = z_1, + \cdots + z_g = m\bullet c$。

ORSIs have a large field of view, and most ship targets have small foreground pixels. Large kernel convolution can easily cause the loss of target pixels and bring a lot of computation. Therefore, we have improved the kernel size of the mixed convolution and changed the way the number of channels is divided. Figure 6 shows the proposed mixed convolution structure. Firstly, we removed the 9×9 convolution to avoid the problem of information loss during the convolution process. Secondly, a 1×1 convolution is added, which further reduces the amount of computation and parameters on the basis of retaining more detailed feature information. Finally, we abandon the equal distribution and choose the exponential distribution for channel partitioning of convolutions with different kernel sizes. The latter retains more low-latitude feature information while realizing multiscale feature extraction. Compared with the method of using equal division between channels in natural images, the performance of exponential division in ORSIs is higher. The exponential channel division method is shown as follows:

$$C_x = \begin{cases} 2^{-x}, 0 < x < i-1, \\ 2^{-x+1}, x = i, \end{cases} \tag{4}$$

where i is the number of convolution kernels and C_x is the number of channels of the x-th convolution. According to the parameters in Table 1 and formula (3), the output feature map of the proposed mixed convolution can be expressed by the following formula:

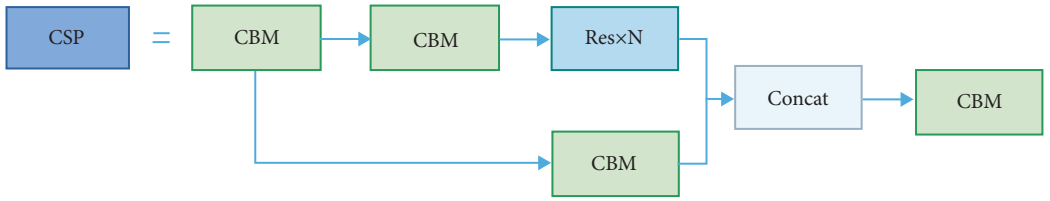


Figure 4: CSP module structure diagram.

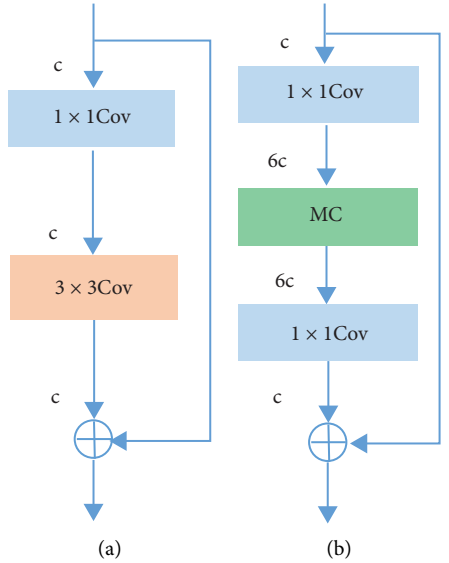


Figure 5: Comparison of (a) Res and (b) MIRes.

Table 1: Res and MIRes network architecture and parameters.

	Res	MIRes
Input	$13 \times 13 \times 512$	$13 \times 13 \times 512$
Operation	$\begin{bmatrix} 1 \times 1\text{Conv} \times 512 \\ 3 \times 3\text{Conv} \times 512 \end{bmatrix}$	$\begin{bmatrix} 1 \times 1\text{Conv} \times 3072 \\ \begin{bmatrix} 1 \times 1\text{Dwise} \times 1536 \\ 3 \times 3\text{Dwise} \times 768 \\ 5 \times 5\text{Dwise} \times 384 \\ 7 \times 7\text{Dwise} \times 384 \end{bmatrix} \\ 1 \times 1\text{Conv} \times 512 \end{bmatrix}$
Output	$13 \times 13 \times 512$	$13 \times 13 \times 512$

$13 \times 13 \times 512$ represents the width, height, and number of channels of the feature map.

$$Y_{13,13,3072} = \text{Concat}\left(\hat{Y}^{1}_{1,1,1536}, \hat{Y}^{2}_{3,3,768}, \hat{Y}^{3}_{5,5,384}, \hat{Y}^{4}_{7,7,384}\right). \quad (5)$$

MIRes has a wider network structure and introduces mixed convolution, so the multiscale feature extraction capability of the network is greatly improved At the expense of a small amounts of parameters (Parameter 9M) and computation (BFLOPS 0.759 G), MIRes greatly enhance the multiscale feature extraction capability of the backbone network, and the detection accuracy of multiscale ship targets is significantly improved (2.45%).

3.3. PANet with Multiscale Feature Enhancement. PANet is the mainstream solution for multiscale target detection. It is based on a rule: shallow feature maps (C3) contain higher resolution and more location information, and deep feature maps (C5) have larger receptive field and more semantic information [32]. The shallow receptive field is smaller, and its location information is more beneficial to target localization. The deep receptive field is larger, which contains more semantic information, which is beneficial to the classification of objects. Take Figure 7 as an example. The shallow network (C3) contains more local information, that is, fine-grained information, and the receptive field at this

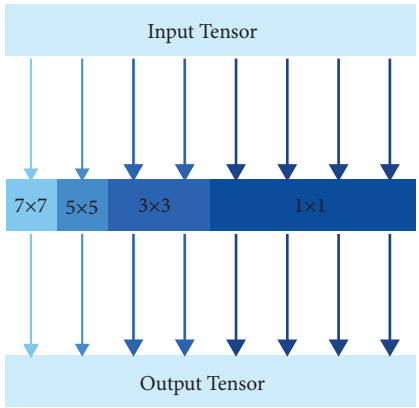


Figure 6: The proposed mixed convolutional structure diagram.

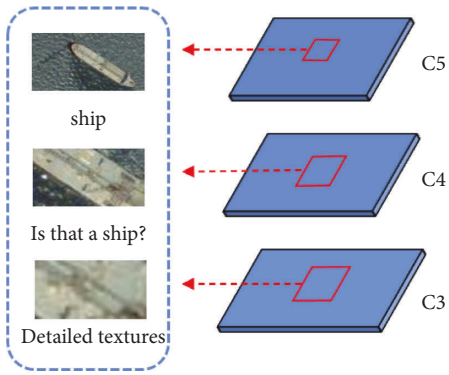


Figure 7: Example diagram of semantic information. The shallow feature maps are usually some corners and so on. The middle layer is part of the object. The deep layer is usually a whole object with rich semantic information.

time is relatively small. Therefore, the local information of the feature map obtained by the shallow network is relatively rich, the resolution of the feature map at this level is relatively high, and the receptive field of a single pixel is relatively small, which can capture more location information. As the number of downsampling or convolution increases, the deep feature (C5) gradually increases the receptive field and the overlapping area between the receptive fields. The information represented by the pixels at this time is the information of a region, and the feature information obtained is the feature information between this region or adjacent regions, which is relatively not fine-grained and low in resolution, but rich in semantic information [33, 34]. The feature information extracted by MIRes first passes through the SPP module and then is input into PANet for fusion. In PANet, the deep feature map is fused with the upper layer feature map in a ratio of 1 to 1 after upsampling, so that the deep semantic information is transferred to the shallow feature map. The feature map of the shallow layer is fused with the feature map of the next layer in a ratio of 1 after downsampling, so that the position information of the shallow layer is transmitted to the deep layer. This fusion mode can improve the accuracy of multiscale target detection, but for small- and medium-sized ship targets, this simple fusion strategy may not be able to adapt to the multiscale characteristics of ship targets in remote sensing images. Therefore, this paper further improves the accuracy of multiscale ship target detection by enhancing the feature information of the middle- and low-level feature maps.

The receptive field mechanism of RFB can significantly enhance the receptive field and feature information of feature maps. However, there are still problems when RFB is directly applied to ship target detection in ORSIs. The large convolution cores and large expansion rates in RFBs may cause target miss and misdetection for small- and medium-sized ships. Therefore, for the appeal problems, this paper proposes multiscale feature enhancement modules: SFEM and MFEM. For low-level feature maps, SFEM introduces a multibranch structure and uses ordinary convolution and atrous convolution with an expansion rate of 1 for feature extraction. For the midlevel feature maps, feature enhancement is performed using MFEM containing atrous convolutions with large expansion rates. Considering the parameters of the network at the same time, different from other deep feature enhancement modules, we abandon the deeper network structure and use a wider network for multiscale feature extraction. Finally, considering that the input and output of the SFEM and MFEM modules do not involve channel transformation, the 1×1 convolution used to adjust the dimension.

For the low-level feature map, it is mainly responsible for the detection of small-scale ship targets. For small targets, too large receptive field will introduce a large amount of background, resulting in the decline of detection accuracy. Therefore, in this paper, the atrous convolution with a large expansion rate is replaced by an ordinary 3×3 convolution, and only ordinary convolution and atrous convolution with an expansion rate of 1 are used to extract features. At the same time, the 5×5 large kernel convolution branch is removed and a 3×3 convolution branch is added. Figure 8 shows the structure of SFEM. R represents the dilation rate of the atrous convolution, and shortcut represents the residual connection. The input feature map goes through four-branch structure for feature extraction. Each branch contains the same input, and the feature information extracted by each branch is fused and added to the input feature map. Compared with the RFB module, SFEM can enhance the shallow position information while reducing the introduction of noise, effectively improving the accuracy of ship detection.

For the midlevel feature map, it needs appropriate receptive field for the detection of medium-scale ship targets. Therefore, we use a combination of standard convolution and atrous convolution to extract features. Its structure is similar to SFEM. On the basis of retaining the multi-branch structure, atrous convolution is used to enhance the receptive field and capture a wider range of feature information. Figure 9 shows the structure of MFEM. R is the dilation rate of atrous convolution, and shortcut is the residual connection. MFEM retains the network structure of SFEM and the atrous convolution with large expansion rate in RFB, which can enhance the receptive field of the effective area while maintaining the size of the feature map.

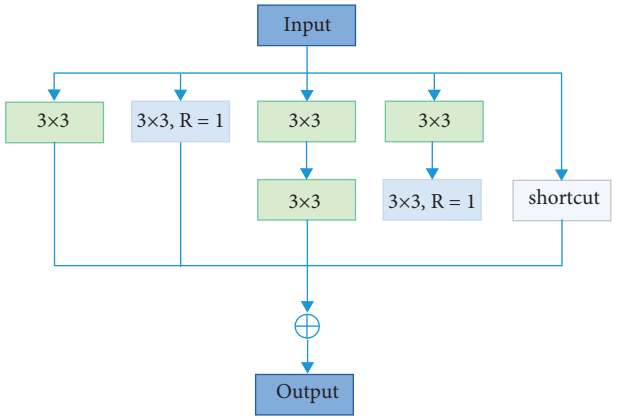


Figure 8: SFEM structure diagram.

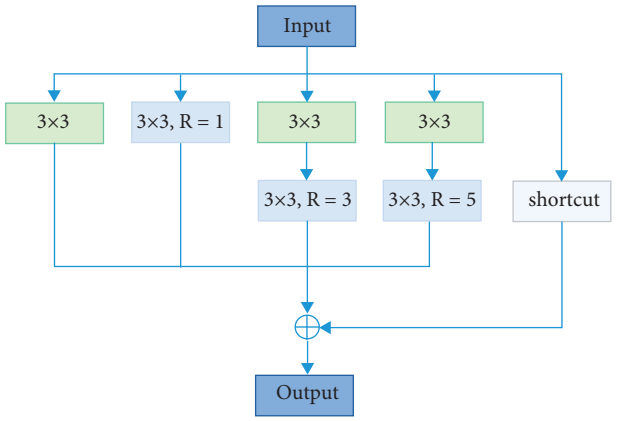


Figure 9: MFEM structure diagram.

SFEM and MFEM can enhance the feature description of multiscale ship targets and improve the accuracy of ship detection. At the same time, it can achieve better results than some more advanced enhancement modules such as RFB and EIRM [13]. In Section 4, a series of comparative experiments are implemented to verify the effectiveness of the module.

3.4. Loss Function. The feature map after feature enhancement is predicted at the detection head. The three prediction scales of the algorithm in this paper are $13 \times 13, 26 \times 26, and\ 52 \times 52$. During prediction, the feature map is divided into $S \times S$ grids, and each grid will contain multiple prediction boxes. The final result is obtained by calculating the joint intersection of IOU [35] and then filtering with nonmaximum suppression (NMS). In this paper, the CIOU [36] loss function is used to calculate the bounding box regression loss, and the loss functions such as GIOU [37], DIOU, and IOU are compared.

The IOU loss function is usually used to calculate the bounding box regression loss, the intersection-over-union (IOU). Assuming that the predicted box is defined asB_{pre} and the ground-truth box is defined asB_{gt}, the IoU loss function (7) is as follows:

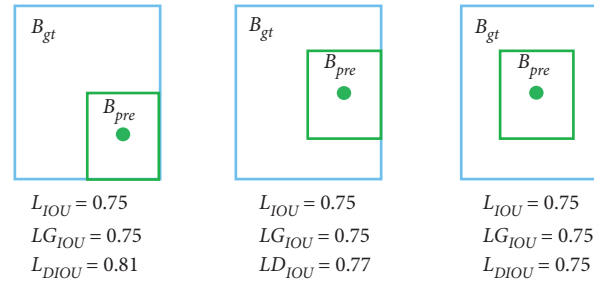


Figure 10: Loss function limits. Green is the predicted box. Blue is the true box.

$$IOU = \frac{\left|B_{pre} \cap B_{gt}\right|}{\left|B_{pre} \cup B_{gt}\right|}, \tag{6}$$

$$L_{IOU} = 1 - \frac{\left|B_{pre} \cap B_{gt}\right|}{\left|B_{pre} \cup B_{gt}\right|}. \tag{7}$$

However, as shown in Figure 10, there are still problems with the IOU loss function. Firstly, when B_{pre} and B_{gt}do not intersect, IOU=0. Secondly, when the twoB_{pre} are not the same, the two IOUs are equal.

To solve the problem of IOU=0 when the predicted frame and the real frame do not intersect, GIOU introduces the minimum bounding box based on the predicted frame and the real frame. Assuming that the predicted frame is B_{pre}, the real frame isB_{gt}, and A_c is the minimum bounding box, GIOU can be expressed by

$$G_{IOU} = IOU - \frac{\left|A_c - B_{pre} \cap B_{gt}\right|}{\left|A_c\right|}. \tag{8}$$

However, when the predicted box and the ground-truth box input contain relationship, GIOU will degenerate into IOU, which still has the problem of equal IOU. DIOU is based on IOU and GIOU, taking into account the distance between the ground-truth box and the predicted box, and can guide the direction movement of the prediction box when the two do not intersect. The DIOU loss function formula is as follows:

$$DIOU = IOU - \frac{\rho^2\left(b, b_{gt}\right)}{c^2}. \tag{9}$$

In formula (9), b and b_{gt} are the center points of the prediction madness and the real box, respectively,ρrepresents the Euclidean distance between the two center points, and c is the diagonal length of the minimum bounding box between the real box and the prediction box. The DIOU parameters are shown in Figure 11. According to Figure 9, it can be seen that when the target box completely includes the prediction box, GIOU will completely degenerate into IOU. Therefore, GIOU cannot converge well in advanced algorithms, thus affecting the detection accuracy. And, DIOU adds a penalty

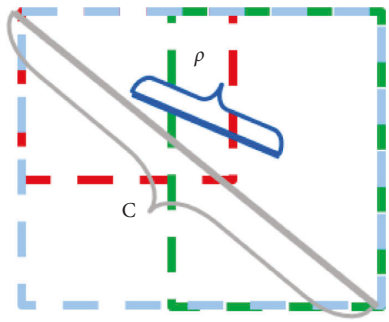


Figure 11: Schematic diagram of DIOU. c is the diagonal of the minimum bounding box of the real box and the predicted box, and ρ is the distance between the center points of the two.

term$\rho^2(b, b_{gt})/c^2$on the basis of GIOU. It can minimize the normalized distance between the center points of the two bounding boxes, while the diagonal is unchanged, so that it converges faster than GIOU.

Three important factors for bounding box regression loss are overlap area, center point distance, and aspect ratio. DIOU ignores the key factor of aspect ratio. CIOU considers the aspect ratio of the predicted box and the ground-truth box, which can be well adapted to the aspect ratio characteristics of the ship target. The CIOU formula is as follows:

$$CIOU = \frac{\rho^2\left(b, b_{gt}\right)}{c^2} + \alpha v,$$

$$\alpha = \frac{v}{(1 - IOU) + v}, \tag{10}$$

$$v = \frac{4}{\pi^2}\left(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h}\right)^2.$$

In the above formulas,$\alpha$is a positive parameter. $v$means consistency of aspect ratio.w^{gt} and h^{gt} are the width and height of the real box, and w and h are the width and height of the predicted box.

4. Experiments and Results

To verify the proposed method, a series of experimental comparisons are implemented in this section. In Sections 4.1, 4.2, and 4.3, the dataset, experimental details, and evaluation indicators are introduced. In Section 4.4, a comparative experiment is implemented to compare with other algorithms and the baseline network. Section 4.5 is the ablation experiment. Section 4.6 is an extended experiment.

4.1. Dataset. The LEVIR dataset contains a large number of background images (that is, images that do not contain objects), so we segment the LEVIR dataset to form the LEVIR-ship dataset. The LEVIR-ship dataset only retains images containing ship targets for ablation experiments and comparative experiments to verify the superiority of the proposed algorithm in multiscale ship target detection. The NWPU VHR-10 dataset has various categories, the target scale changes greatly, and the target distribution is relatively concentrated, which is very challenging. Therefore, we conduct extended experiments on the NWPU VHR-10 dataset to verify the applicability and scalability of the proposed algorithm for different datasets and different targets. We will give a detailed introduction to the LEVIR-ship and NWPU VHR-10 datasets, respectively.

The LEVIR [38] dataset contains a total of 21,952 images with a resolution of 600×800 pixels. There are three categories: aircrafts, ships, and storage tanks. We separate out the ship categories and remove images that do not contain objects to form the LEVIR-ship dataset. LEVIR-ship has a total of 1494 images and 3025 ship targets, and each image contains at least one ship target. According to the original partitioning method, the training set contains 876 images of 1790 ship targets, and the test set contains 618 images of 1235 ship targets. The target location distribution and scale distribution in the LEVIR-ship dataset are shown in Figure 12. In Figure 12(a), the horizontal axis and the vertical axis are the positions of the target in the width and height of the image, respectively, the horizontal axis in Figure 12(b) is the target pixel, and the vertical axis is the number of targets. It is obvious that the target location distribution and scale distribution in the lever ship data set are relatively uniform.

The NWPU VHR-10 [39] dataset contains a total of 800 high-resolution images and ten classes of objects. In this paper, 150 background images without targets are deleted, and 650 images with targets are reserved for training and testing. The training set and test set are randomly divided in a ratio of 5 to 5. Figure 13 shows the target location distribution and scale distribution in the NWPU VHR-10 dataset. In Figure 13(a), the horizontal axis and the vertical axis are the positions of the target in the width and height of the images, respectively, the horizontal axis in Figure 13(b) is the target pixel, and the vertical axis is the number of targets. Most of the objects in the NWPU VHR-10 dataset are gathered in the central area of the images, and there are few small-scale objects, mostly medium-scale targets.

4.2. Implementation Details. Before training, the images are resized to 416×416 size and then sent to the network for training. The training has a total of 4000 iterations. The training batch size is set to 64, and the initial learning rate is 0.0001. When iterating to 3200 and 3600 times, the learning rate is reduced by one tenth, respectively. The training of the network was performed on an RTX3060 GPU, and the training was accelerated using CUDA 11.3 and cuDNN 8.05.

The algorithm in this paper performs location and category regression based on predefined anchor boxes on $S \times S$ grids of each image. Therefore, the generation of anchor boxes has a great impact on the performance of the network. The K-means++ algorithm can generate better anchor boxes to accelerate the convergence of the network. In order to speed up the convergence of the network, this paper uses the K-means++ clustering algorithm to obtain the size of the prior box. The clustering process of the K-means++ algorithm is shown in Algorithm 2. The size of

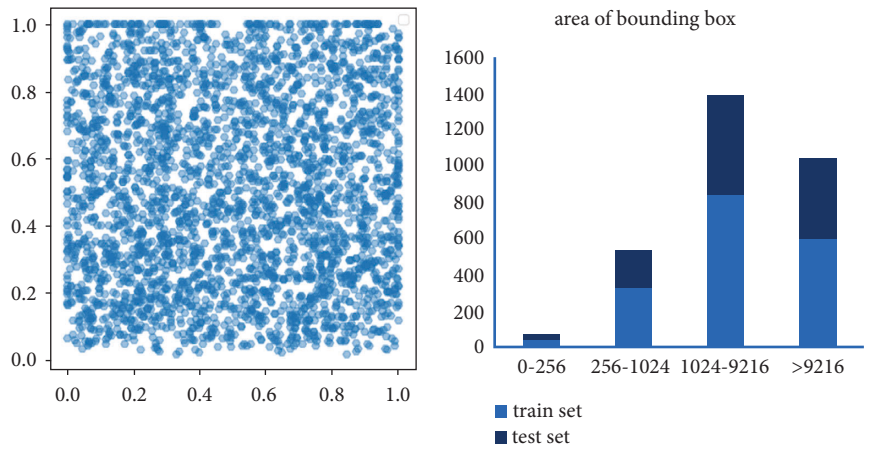


Figure 12: Distribution of object locations and scales in the LEVIR-ship dataset. (a) Target location distribution. (b) Target scale distribution.

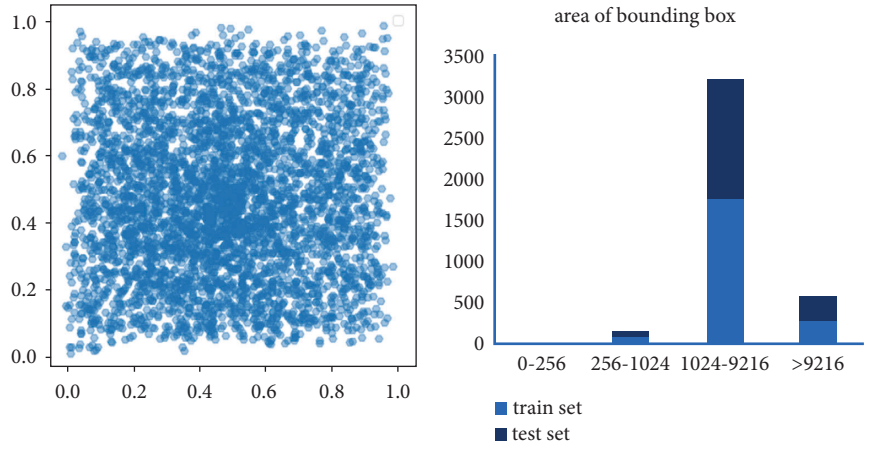


Figure 13: Distribution of object locations and scales in the NWPU VHR-10 dataset. (a) Target location distribution. (b) Target scale distribution.

Table 2: The priori box size for the LEVIR-ship dataset and NWPU VHR-10 dataset.

Feature map	Receptive field	Detection object	Anchor box	
			LEVIR-ship	NWPU VHR-10
13 × 13	Large	Large	(77, 102), (104,138), (152,202)	(36, 27), (57,65), (130,121)
26 × 26	Medium	Medium	(31, 41), (45,60), (61, 81)	(22, 13), (16,21), (18, 39)
52 × 52	Small	Small	(8, 11), (13, 17), (21, 28)	(6, 8), (15, 8), (10, 16)

the prior box is shown in Table 2. Each prediction scale sets three prior frames, which are, respectively, suitable for the detection of large, medium, and small objects.

4.3. Evaluation Criteria. In this paper, the general evaluation methods map and FPs are used to evaluate the algorithm. The mAP is a general evaluation criterion in object detection, which is determined by P (precision) and R (recall). The formulas for calculating P and R are as follows:

$$P = \frac{TP}{TP + FP}, \tag{11}$$

$$R = \frac{TP}{TP + FN}. \tag{12}$$

In formula (13) and formula (14), TP represents positive samples, that is, the number of ships detected; FP represents the negative sample, that is, the number of ships incorrectly

Table 3: the LEVIR-ship dataset with the experimental results.

Method	mAP (%)	FPS (s)
LARGE-RAM [38]	60.80	0.2
Faster RCNN [41]	79.0	6.4
EIRM-SGPANet [13]	77.82	51.7
YOLO v4	76.30	61.8
Ours	79.55	56.2

Input: $P = \{P_1 \cdots P_n\}$, P is the set of center points.
Step 1: randomly select a sample from the dataset as the initial cluster center C_1.
Step 2: first calculate the shortest distance $D(x)$, between each sample and the current existing cluster center, then calculate the probability of each sample being selected as the next cluster center $D(x)^2/\sum_{x\in X} D(x)^2$Finally, select the next cluster center according to the roulette method.
Step 3: repeat Setp2 until K cluster centers are selected.
Step 4: for each sample X_i in the dataset, calculate its distance to K cluster centers and classify it into the class corresponding to the cluster center with the smallest distance.
Step 5: recalculate its cluster center $C_i = 1/|C_i|\sum_{x\in C_i} X$ for each category C_i.
Step 6: repeat Step 4 and Step 5 until the position of the cluster center no longer changes.
Output: K cluster centers $P = \{P_1 \cdots P_k\}$.

Algorithm 2: K-means++ algorithm

detected; FN stands for false samples, that is, the number of ships missed [40]. According to P and R, the calculation of mAP can be expressed as follows:

$$mAP = \frac{\sum_{i=1}^{N_{cls}} \int P_i(R_i) dR_i}{N_{cls}}. \tag{13}$$

In the above formula, N_{cls} represents the number of categories in the dataset.P_i represents the precision of the i-th class, and R_i represents the recall of the i-th class. In this paper, FPS is used to evaluate the inference speed of the algorithm. The FPS formula is

$$FPS = \text{frameNumelapsedTime}. \tag{14}$$

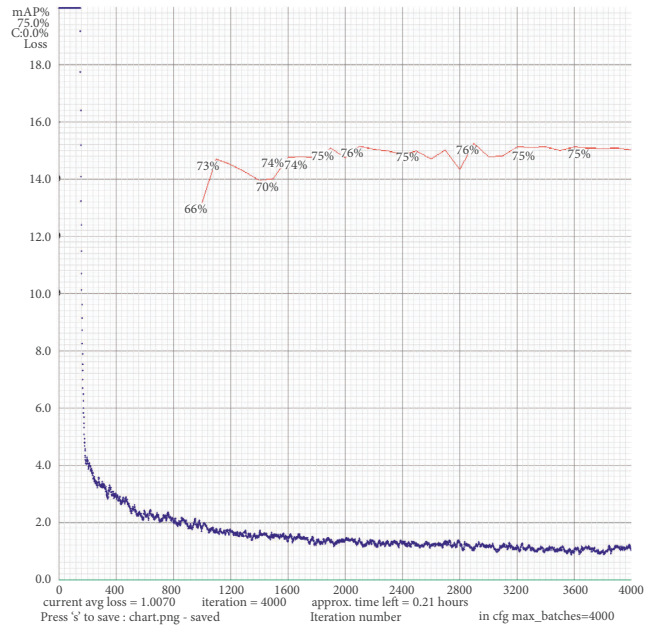
In the above formula, frameNum is the number of images, and elapsedTime is the inference time. The higher FPS indicates that the model inferences faster.

4.4. Comparative Experiments. In this section, the comparison experiment is divided into two parts. In Section 4.4.1, the proposed algorithm is compared with other methods and baseline network, and the advantages of this algorithm are analyzed. In Section 4.4.2, the feasibility of this improvement is compared. The improvement feasibility comparison mainly includes the feasibility of mixed convolution improvement, the comparison of Res, and the comparison of feature enhancement modules (SFEM and MFEM) with RFB and EIRM.

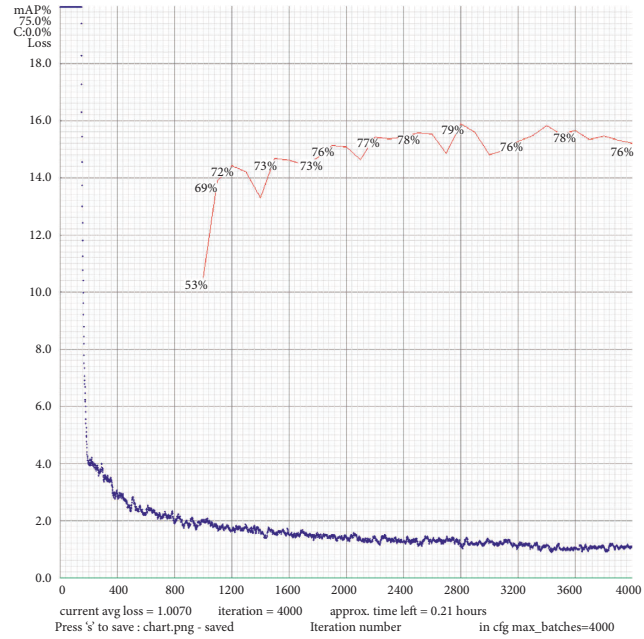
4.4.1. Comparison with Other Methods and Baseline Network. According to the experimental parameter settings in Section 4.2, a comparative experiment was conducted on the lever ship data set, and the results are shown in Table 3. First, compared with the baseline network, the improvement of the algorithm in this paper has achieved better results. Although the detection speed (FPS) decreased slightly, the accuracy of multiscale ship detection was improved by 3.25%. Secondly, compared with other two-stage algorithms, the mAP and FPS of this algorithm have reached the highest level. In general, the accuracy of the proposed algorithm has been greatly improved compared with the baseline network, and the speed still has great advantages compared with the two-stage algorithm, meeting the requirements of real-time detection ($FPS \geq 30$).

Figure 14 is a comparison chart of the loss function curve between YOLO v4 and the algorithm in this paper. The blue is the loss value curve, and the red is the mAP curve. The loss curve represents the difference between the predicted and actual results. According to the loss curve and mAP curve, it is obvious that the proposed algorithm converges faster than YOLO v4, and the loss value is lower and the accuracy is higher. The model in this paper is significantly better than the YOLO v4 network in terms of detection accuracy.

Figure 15 is a comparison chart of the detection results between the algorithm in this paper and the YOLO v4 algorithm. In order to make the detection results more representative, we selected a total of six groups of images for comparison. Among them, Figures 15 (a)–15 (g) is the original image, Figures 15 (h)–15 (n) is the detection result of YOLO v4, and Figures 15 (o)–15 (u) is the detection result of the proposed algorithm. In Figure 14, the images Figures 15 (a)–15 (c) contain small, medium, and large-scale ship targets; in Figure , the ship targets have long wakes and complex wave backgrounds; in Figure 15 (f), the ship targets are occluded by shadows; in Figure 15 (g), the ship target is affected by strong light and the contrast between light and dark is strong. According to the



(a)

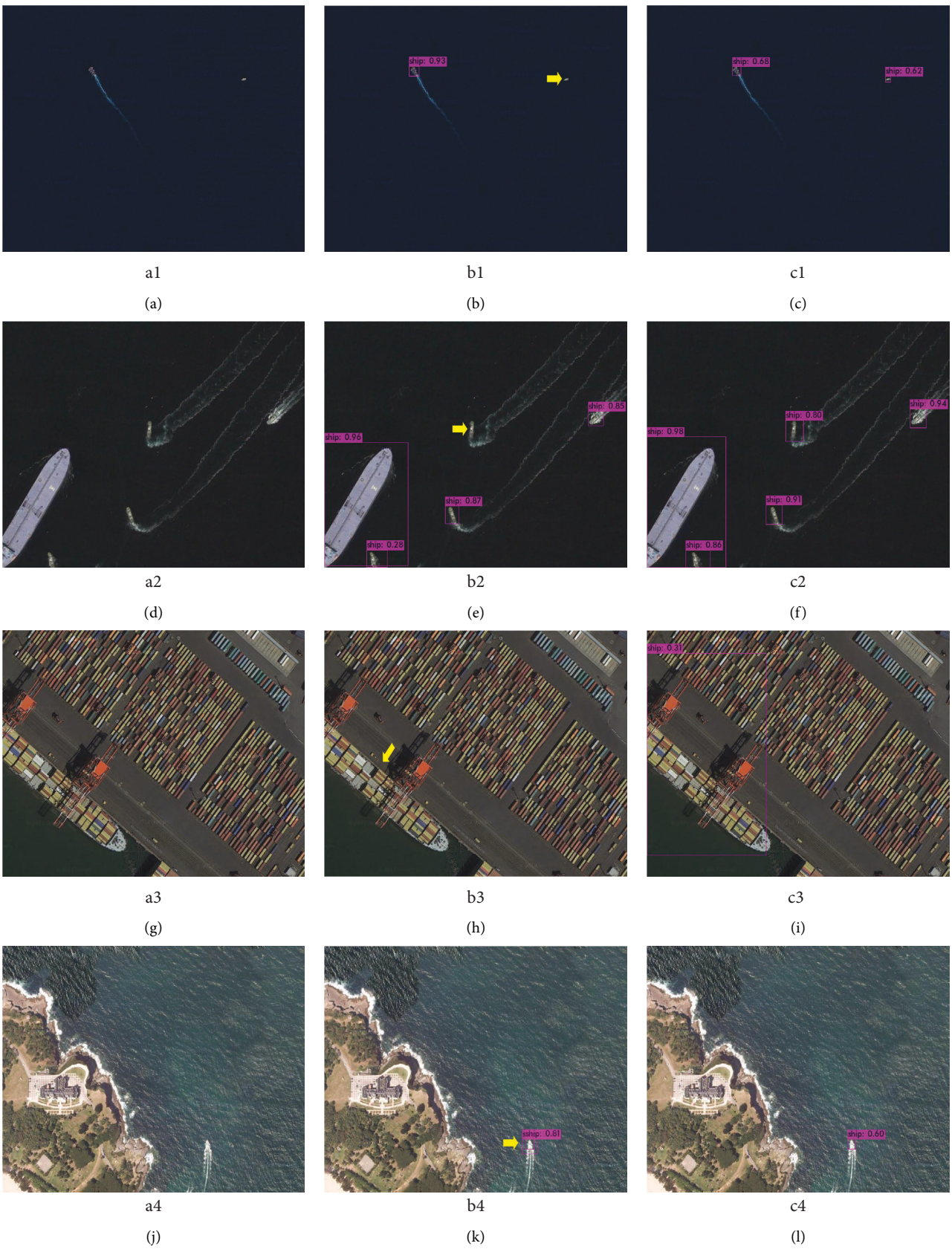


(b)

Figure 14: Loss curve comparison chart. (a) YOLO v4 loss curve. (b) The loss curve of the algorithm in this paper.

characteristics of the above images and the detection results of Figures 15 (h)–15 (n), it can be seen that the feature extraction ability and the receptive field of the feature map of YOLO v4 algorithm are obviously insufficient, which is easy to cause missed detection and false detection of multiscale ships. According to Figures 15 (o)–15 (u), it is obvious that the proposed algorithm has high detection accuracy in multiscale ship detection and can detect large, medium and small-scale ship targets well. At the same time, under the influence of wake, wave background, light, and shadow, the algorithm in this paper can well suppress the interference of background factors and accurately detect ship targets of various scales.

4.4.2. Comparison of the Baseline Network and the Proposed Method. In order to verify the influence of the size of the convolution kernel and the channel allocation method on the detection accuracy, in this paper, different convolution kernels and assignment methods are used to perform



a1 b1 c1

(a) (b) (c)

a2 b2 c2

(d) (e) (f)

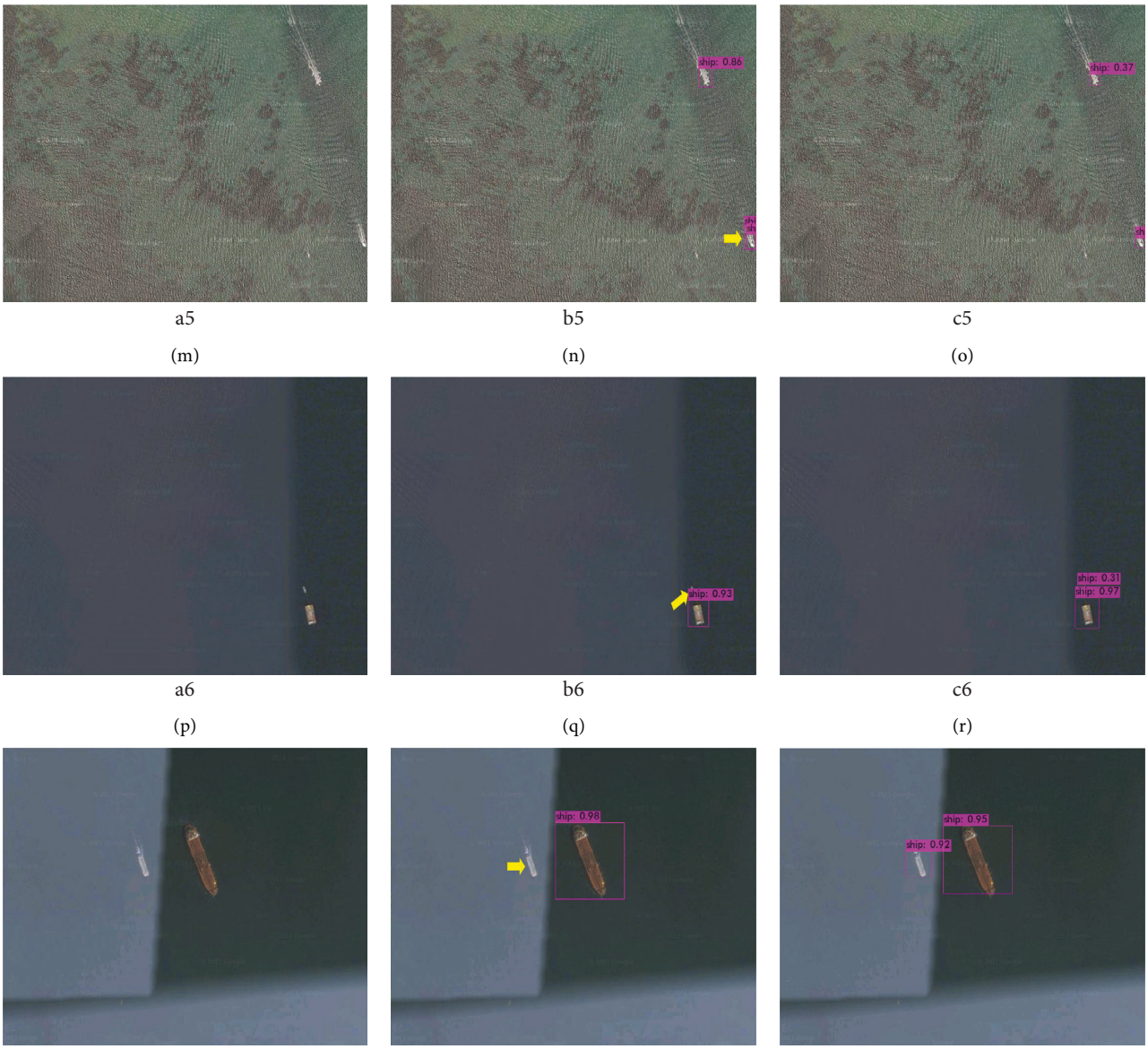
a3 b3 c3

(g) (h) (i)

a4 b4 c4

(j) (k) (l)

Figure 15: Continued.



a5 (m) b5 (n) c5 (o)

a6 (p) b6 (q) c6 (r)

a7 (s) b7 (t) c7 (u)

Figure 15: Comparison of test results. The yellow arrow points to the targets of missed detection and false detection.

Table 4: Comparison of convolution kernel size and distribution method of mixed convolution.

Method	mAP (%)	Parameter (M)	BFLOPS (G)	FPS (s)
MIRes-3579	77.78	254	60.442	56.2
MIRes-1357	77.8	253	60.359	61.8
MIRes-3579exp	76.9	253	60.384	61.8
MIRes-1357exp (ours)	78.75	253	60.322	61.8

ablation experiments. The experimental results are shown in Table 4. 1357exp means that the convolution kernels are 1, 3, 5, and 7, and the channels are divided in the exponential form. 1357 means that the convolution kernels are 1, 3, 5, and 7, and the channels are divided equally. At the same time, we added two evaluation indicators—Parameter and BFLOPS. Parameter represents the size of the model, that is, the amount of parameters, and BFLOPS represents the amount of model calculation. According to the results, it can be seen that removing the large kernel convolution and

Table 5: Comparison of different residual blocks.

Method	mAP (%)	Parameter (M)	BFLOPS (G)	FPS (s)
Res (YOLO v4)	76.3	244	59.563	61.8
IRes	77.2	252	60.309	61.8
MIRes (ours)	78.75	253	60.322	61.8

Table 6: Comparison of feature enhancement modules.

Feature enhancement	mAP (%)	Parameter (M)	BFLOPS (G)	FPS (s)
S-RFB	77.73	254	61.873	56.2
S-EIRM	77.95	257	64.354	56.2
SFEM	79.21	247	64.791	61.8
M-RFB	78.26	257	61.872	56.2
M-EIRM	78.16	272	64.265	51.7
MFEM	78.55	267	65.55	56.2
S-RFB, M-RFB	76.52	258	63.423	56.2
S-EIRM, M-EIRM	76.79	276	68.297	51.7
SFEM, MFEM	79.55	271	70.778	56.2

replacing it with a 1×1 convolution can bring a little improvement in accuracy, while the amounts of parameters and calculation are also slightly reduced. In the comparison between the exponential division and the equal division, the exponential division achieves higher accuracy. Therefore, for the multiscale ship detection of ORSIs, due to the complexity of ORSIs and the multiscale characteristics of ship targets, compared with large kernel convolution, 1×1 convolution may be more efficient. The comparison results verify that the improvement of the mixed convolution in this paper enables it to achieve excellent results in multiscale ship detection in ORSIs.

To verify the superiority of the proposed MIRes, the influence of various residual blocks on the detection accuracy is compared in this paper. Table 5 shows the comparison results of various residual blocks. It is obvious that although IRes has better feature extraction ability, its accuracy has been improved compared to Res. However, the MIRes proposed in this paper achieves better results, and the increase in the amounts of parameters and computation is almost negligible. Compared with Res and IRes, the introduction of mixed convolution enables MIRes to have stronger multiscale feature extraction capabilities, so it can achieve better detection accuracy in multiscale ship target detection in ORSIs.

Table 6 shows the comparison results of the multiscale feature enhancement module and other feature enhancement modules. For fairness, RFB, EIRM [13], and the multiscale feature enhancement module are added to the same location on the basis of the improved backbone network. SFEM and MFEM in this paper act on the middle and low-level feature maps, respectively. Therefore, two RFBs and two EIRMs were added in the experiment to act on the middle and low-level feature maps. S-RFB represents adding RFB modules to the low-level feature map branch, and M-RFB represents adding RFB modules to the midlevel feature map. S-EIRM means adding EIRM to the low-level feature map branch, and M-EIRM means adding EIRM to the middle-level feature map. According to the results, it is obvious that, in the low-level feature map, the atrous convolution of the RFB with a large expansion rate may introduce a large amount of background, which will cause a decrease in accuracy. The scaling strategy of EIRM is also less accurate. SFEM only uses a combination of atrous convolution with an expansion rate of 1 and ordinary convolution, which enhances the feature description while suppressing the interference of background noise, and improves the accuracy of ship detection. For midlevel feature maps, MFEM also achieves better results than RFB and EIRM. Finally, the proposed multiscale feature enhancement module achieves higher accuracy than EIRM and RFB. Therefore, the strategy of using different enhancement modules for feature maps of different scales can achieve better results than the fixed feature enhancement strategy.

In order to verify the impact of the IOU loss function on the detection accuracy, this paper uses GIOU, DIOU, and CIOU to conduct comparative experiments. The experimental results are shown in Table 7. It is obvious that, in the algorithm of this paper, CIOU achieves the best effect compared to GIOU and DIOU. CIOU introduces the strategy of anchor box aspect ratio, which enables it to achieve higher accuracy than DIOU.

4.5. Ablation Experiments. To verify the effectiveness of each module, ablation experiments were implemented on the LEVIR-ship dataset. Table 8 is the results of the ablation experiments. It is obvious that MIRes and the multiscale feature enhancement modules SFEM and MFEM proposed in this paper can effectively improve the detection accuracy of the baseline network YOLO v4. This proves that the improvements proposed in this paper can effectively improve the accuracy of multiscale ship detection in ORSIs. MIRes used improved mixed convolution and wider network structure, which greatly enhances the multiscale feature extraction capability of the backbone network. MFEM and SFEM act on the middle- and low-level feature maps to

Table 7: Comparison of loss functions.

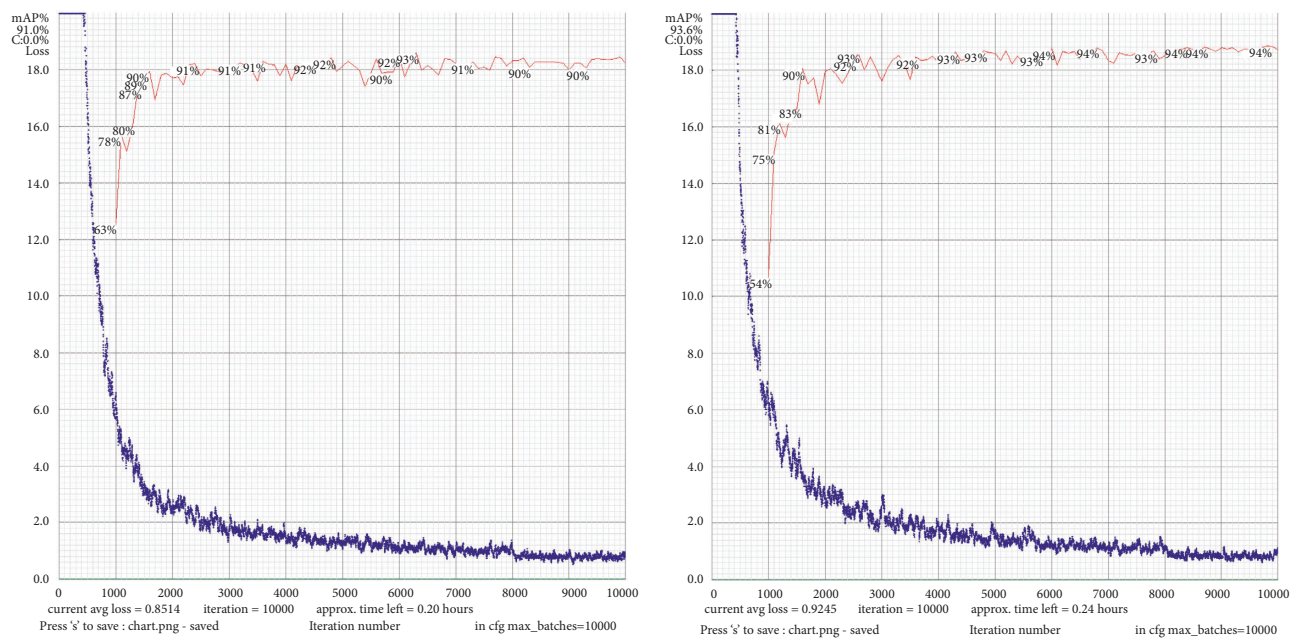
Loss function	mAP (%)	FPS (s)
GIOU	77.69	56.2
DIOU	77.56	56.2
CIOU	79.55	56.2

Table 8: Ablation experiment.

MIRes	SFEM	MFEM	mAP (%)	Parameter (M)	BFLOPS (G)	FPS (s)
Baseline			76.3	244	59.563	61.8
√			78.75	253	60.322	61.8
	√		78.84	247	64.791	61.8
		√	77.18	258	64.791	56.2
√	√		79.21	256	65.55	61.8
√		√	78.55	267	65.55	56.2
	√	√	77.65	262	70.019	56.2
√	√	√	79.55	271	70.778	56.2

Table 9: Extended experimental results of NWPU VHR-10 dataset.

Method	AI (%)	SH (%)	ST (%)	BD (%)	TC (%)	BC (%)	GTF (%)	HB (%)	BR (%)	VH (%)	mAP (%)
FCOS [42]	90.47	73.72	90.36	98.94	89.38	80.82	96.74	87.91	61.92	88.16	85.84
SHDET [43]	100	81.36	90.90	98.66	90.84	82.57	98.68	91.11	76.43	89.82	90.04
FMSSD [44]	99.70	89.90	90.30	98.20	86.00	96.80	99.60	75.60	80.10	88.20	90.40
MRFF-RCA [40]	99.50	88.40	90.20	98.70	89.20	95.40	99.20	89.60	82.20	92.90	92.50
YOLO v4	99.97	87.16	98.54	97.38	97.73	94.54	99.65	82.67	77.61	93.42	92.87
Ours	99.93	90.72	98.45	97.67	99.66	96.45	99.80	83.55	83.64	91.69	94.17



(a) (b)

Figure 16: Comparison of loss curves. (a) YOLO v4 loss curve. (b) The proposed algorithm loss curve.

Figure 17: Continued.

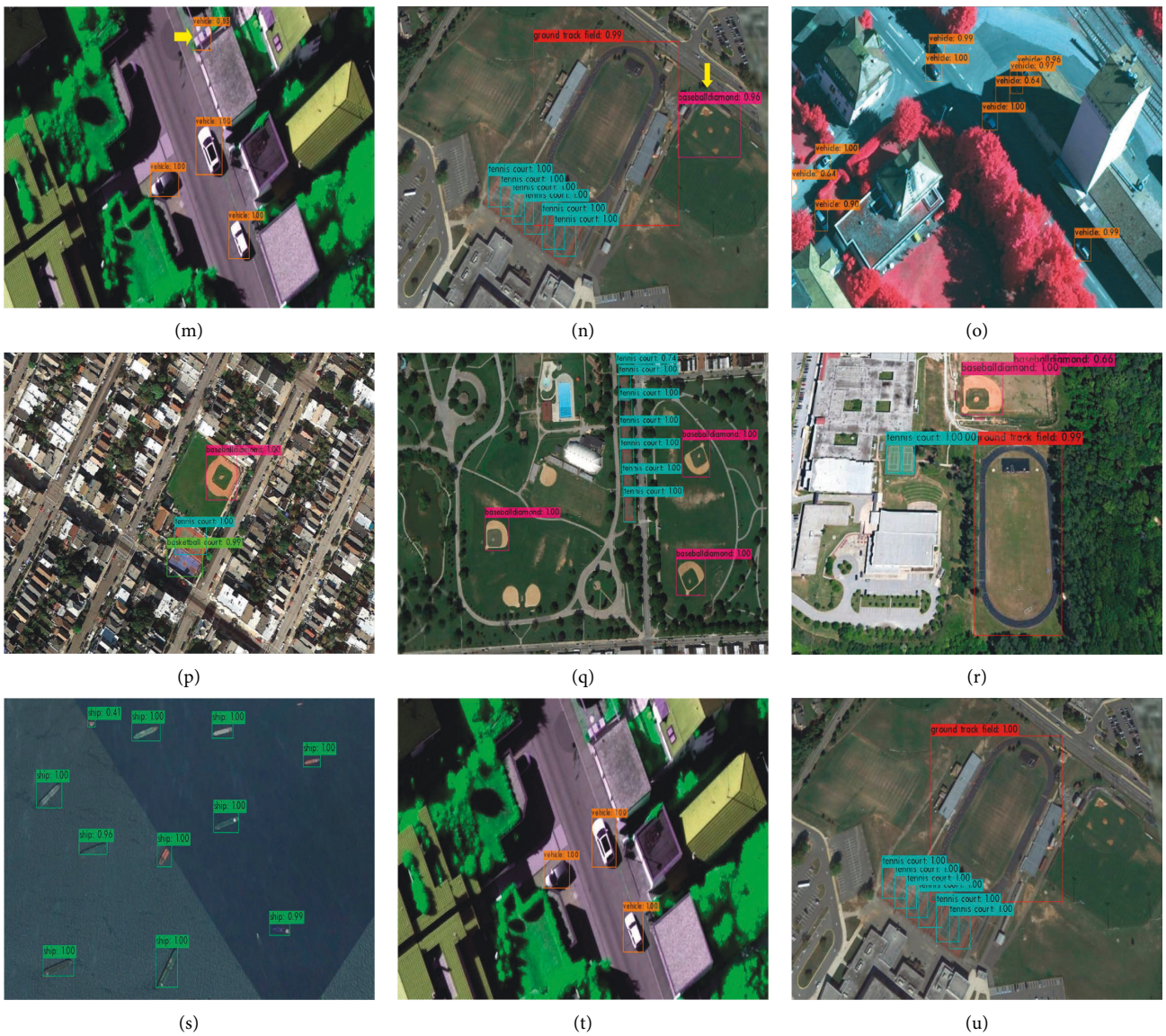


Figure 17: Comparison of detection results in the NWPU VHR-10 dataset. The yellow arrow points to the targets of missed detection and false detection.

further enhance the feature description of ship targets, thereby improving the accuracy of ship detection. At the same time, SFEM, MFEM, and MIRes were added, and the method in this paper finally reached mAP of 79.55%. Compared with the baseline network, it has obtained a great improvement.

4.6. Extend Experiment. To further verify the effectiveness and scalability of our algorithm on other categories, this paper implements extended experiments on the NWPU VHR-10 dataset. The experimental results are shown in Table 9. AI, SH, ST, BD, TC, BC, GTF, HB, BR, and VH represent airplane, ship, storage tank, baseball diamond, tennis court, basketball court, ground track field, harbor, bridge, and vehicle, respectively. It can be seen from the results that the accuracy of the proposed algorithm for the overall mAP and ship category has reached the highest level. At the same time, it is extended to other categories, such as BD, TC, BC, GTF, HB, and BR. The algorithm in this paper is improved compared to YOLO v4.

Figure 16 is the comparison of the loss curve between the algorithm in this paper and the YOLO v4 algorithm on the NWPU VHR-10 dataset. Obviously, although the loss curve amplitude of the algorithm in this paper is large, the detection accuracy is steadily improved, while the loss value converges. The YOLO v4 algorithm converges faster and the curve is smooth, but its detection accuracy is low.

Figure 17 is the comparison of some detection results between the proposed algorithm and the YOLO v4 algorithm on the NWPU VHR-10 dataset. In order to reflect the scalability of the proposed algorithm in other categories, this paper selects images that contain different targets and are difficult to detect. The detection results contain a total of

seven groups of images. Figures 17(a)–17(g) are the original images, Figures 17(h)–17(n) are the YOLO v4 detection results, and Figures 17(o)–17(u) are the detection results of the algorithm in this paper. Among them, some targets in Figures 17(a) and 17(b) are occluded by shadows, some targets in Figures 17(c) and 17(d) are not completely intercepted, and the colors of targets and backgrounds in Figures 17(e) and 17(f) are similar. Targets in Figure 17(g) are larger. According to the detection results of Figures 17(h)–17(n) and Figures 17(o)–17(u), it can be seen that YOLO v4 is prone to missed detection and false detection when detecting shadow-occluded targets, intercepted targets, and targets with similar background color and texture. The proposed algorithm can extract more multiscale feature information, and the multiscale feature enhancement module can further enhance the feature description and receptive field of the target and suppress the interference of background information. It has better detection results for multiscale targets in complex backgrounds.

5. Conclusion

To improve the accuracy of multiscale ship target detection in ORSIs, the multiscale feature enhancement ship target detection algorithm based on the one-stage algorithm YOLO v4 is proposed in this paper. Firstly, in order to improve the multiscale feature extraction capability of the backbone network, this paper improved the mixed convolution and proposed MIRes based on the improved mixed convolution. The MIRes are used to replace the Res in the deep CSP module. The wider network structure of the MIRes and the multikernel mixed convolution greatly enhance the feature extraction capability of the backbone network and the feature map receptive field. Secondly, the multiscale feature enhancement modules SFEM and MFEM are proposed, which act on the middle and low-level feature maps to enhance the feature map receptive field and feature information. Finally, comparative and extended experiments are implemented on the LEVIR-ship dataset and the NWPU VHR-10 dataset. The experimental results on the LEVIR-ship dataset show that the proposed algorithm achieves a 3.25% improvement compared to the baseline network and meets the requirements of implementation detection. At the same time, compared with the current relatively excellent feature enhancement modules, the multiscale feature enhancement module in this paper has achieved the best results. The experimental results on the NWPU VHR-10 dataset show that the proposed algorithm achieves a 3.56% improvement in the ship category compared to the baseline network. At the same time, it also achieved good results in the categories of base diamond, tennis court, basketball court, ground track field, harbor, and bridge.

Data Availability

The research data come from the network public dataset. The NWPU VHR-10 dataset can be downloaded at https://www.heywhale.com/mw/dataset/5e9d2c33ebb37f002c618636. The LEVIR dataset can be downloaded at http://levir.buaa.edu.cn/Code.htm. The LEVIR-ship dataset and experiment results can be downloaded at https://pan.baidu.com/s/18n1KIsUOQiCGbPwDDPme6g?pwd=1111.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Basic Research Program of China (Grant no. 2019YFE0126600); the Major Project of Science and Technology of Henan Province (Grant no. 201400210300); the Key Scientific and Technological Project of Henan Province (Grant no. 212102210496); the Key Research and Promotion Projects of Henan Province (Grant nos. 212102210393 and 202102110121); Kaifeng Science and Technology Development Plan (Grant no. 2002001); National Natural Science Foundation of China (Grant no. 62176087); and Shenzhen Science and Technology Innovation Commission (SZSTI), Shenzhen Virtual University Park (SZVUP) Special Fund Project (Grant no. 2021Szvup032).

References

[1] C. Y. Li Kequan, J. Liu, and X. Mu, "Survey of deep learning-based object detection algorithms," *Computer Engineering*, vol. 7, 2021.

[2] W. S. Bao, "Survey of object detection algorithm based on deep learning," *Transducer and Microsystem Technologies*, vol. 41, no. 4, 2022.

[3] S. H. Song Zhina and Y. Li, "A survey on ship detection technology in high-resolution optical remote sensing images," *Geomatics and Information Science of Wuhan University*, vol. 16, no. 11, pp. 1703–1715, 2021.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*, IEEE Computer Society, Manhattan, New York, U.S, 2013.

[5] R. Girshick, *Fast R-CNN*, Computer Science, Manhattan, New York, U.S, 2015.

[6] S. Ren, K. He, R. Girshick, J. Sun, and R.-C. N. N. Faster, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[7] W. Liu, D. Anguelov, D. Erhan et al., *SSD: Single Shot MultiBox Detector*, Springer, Berlin/Heidelberg, Germany, 2016.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *You Only Look once: Unified, Real-Time Object Detection*, IEEE, Manhattan, New York, U.S, 2016.

[9] J. Redmon and A. Farhadi, *YOLO9000: Better, Faster, Stronger*, pp. 6517–6525, IEEE, Manhattan, New York, U.S, 2017.

[10] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, https://arxiv.org/abs/1804.02767, 2018.

[11] L. Tian, Y. Cao, B. He, Y. Zhang, C. He, and D. Li, "Image enhancement driven by object characteristics and dense feature Reuse network for ship target detection in remote sensing imagery," *Remote Sensing*, vol. 13, no. 7, p. 1327, 2021.

[12] K. Zhou, M. Zhang, H. Wang, and J. Tan, "Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion," *Remote Sensing*, vol. 14, no. 3, p. 755, 2022.

[13] L. Zhou, Y. Li, X. Rao et al., "Feature enhancement-based ship target detection method in optical remote sensing images," *Electronics*, vol. 11, no. 4, p. 634, 2022.

[14] G. Wang, Y. Zhuang, H. Chen et al., "FSoD-Net: full-scale object detection from optical remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[15] A. Bochkovskiy, C. Y. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, https://arxiv.org/abs/2004.10934.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, June 2016.

[17] Z. Qu, F. Zhu, and C. Qi, "Remote sensing image target detection: improvement of the YOLOv3 model with auxiliary networks,," *Remote Sensing*, vol. 13, no. 19, p. 3908.

[18] Y. Qing, W. Liu, L. Feng, and W. Gao, "Improved YOLO network for free-angle remote sensing target detection," vol. 13, no. 11, p. 2717, 2021.

[19] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, 2017.

[20] D. Xu and Y. Wu, "Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection," *Sensors*, vol. 20, no. 15, p. 4276, 2020.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, 2018.

[22] M. Tan and Q. V. Le, "Mixnet: Mixed depthwise convolutional kernels," 2019, https://arxiv.org/abs/1907.09595.

[23] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, Salt Lake City, UT, USA, June 2018.

[24] K. Zhang and H. Shen, "Multi-stage feature enhancement pyramid network for detecting objects in optical remote sensing images," *Remote Sensing*, vol. 14, no. 3, p. 579, 2022.

[25] G. Wang, Y. Zhuang, and Z. Wang, "Spatial enhanced-ssd for multiclass object detection in remote sensing images," in *Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 318–321, Yokohama, Japan, August 2019.

[26] D. L. Wang, W. B. Du, and Y. T. Liu, "Remote sensing images detection based on dense connection and feature enhancement," *Computer Engineering*, vol. 48, no. 6, pp. 251–256, 2021.

[27] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 385–400, 2018.

[28] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[29] Q. Chen, Y. Wang, and T. Yang, "You only look one-level feature," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13039–13048, Nashville, TN, USA, June 2021.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[31] Q. Chen, Y. Wang, T. Yang, X. Zhang, J. Cheng, and J. Sun, "You Only Look One-Level Feature," 2021, https://arxiv.org/abs/2103.09460.

[32] B. He, Q. Zhang, M. Tong, and C. He, "An anchor-free method based on adaptive feature encoding and Gaussian-guided sampling optimization for ship detection in SAR imagery," *Remote Sensing*, vol. 14, no. 7, p. 1738, 2022.

[33] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian, "Picking deep filter responses for fine-grained image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1134–1142, Las Vegas, NV, USA, June 2016.

[34] Y. Yao, G. Cheng, and X. Xie, "Optical remote sensing image object detection founded by multi-resolution function fusion," *Journal of Remote Sensing*, vol. 25, no. 5, pp. 1124–1137, 2021.

[35] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 658–666, Long Beach, CA, USA, June 2019.

[36] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: faster and better learning for bounding box regression," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.

[37] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 2019.

[38] Z. Zou and Z. Shi, "Random access memories: a new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1100–1111, 2018.

[39] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 98, pp. 119–132, 2014.

[40] J. Liu, D. Yang, and F. Hu, "Multiscale object detection in remote sensing images combined with multi-receptive-field features and relation-connected attention," *Remote Sensing*, vol. 14, no. 2, p. 427, 2022.

[41] R. Dong, D. Xu, J. Zhao, L. Jiao, and J. An, "Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8534–8545, 2019.

[42] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: fully convolutional one-stage object detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9627–9636, Seoul, Republic of Korea, November 2019.

[43] D. Zhu, S. Xia, J. Zhao et al., "Spatial hierarchy perception and hard samples metric learning for high-resolution remote sensing image object detection," *Applied Intelligence*, vol. 52, no. 3, pp. 3193–3208, 2022.

[44] P. Wang, X. Sun, W. Diao, and K. Fu, "FMSSD: feature-merged single-shot detection for multiscale objects in large-scale remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3377–3390, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9825827, 1 page
https://doi.org/10.1155/2023/9825827

Retraction

Retracted: Research on Application of Ecological Sports Innovation in Efficient Development Based on DCN Deep Learning

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Hu, "Research on Application of Ecological Sports Innovation in Efficient Development Based on DCN Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9586509, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9586509, 8 pages
https://doi.org/10.1155/2022/9586509

Research Article

Research on Application of Ecological Sports Innovation in Efficient Development Based on DCN Deep Learning

Xueyan Hu

School of Xi'an Physical Education University, Xi'an 710068, China

Correspondence should be addressed to Xueyan Hu; 106036@tea.xaipe.edu.cn

Received 7 July 2022; Revised 17 August 2022; Accepted 26 August 2022; Published 4 October 2022

Academic Editor: Wenming Cao

With the continuous improvement of social and economic level, the relationship between human and nature is deteriorating. The ecological concept has been attached importance, so the concept of ecological sports has been born. For physical education, if the ecological physical education teaching resources can be effectively developed, it will be a lifelong benefit for students and teachers. Based on the DCN deep learning method, this paper makes an innovative research on ecological sports, aiming to promote the efficient development of sports, train the samples according to the physical quality of different individuals, get the most suitable sports events, and then promote the development of physical education.

1. Introduction

In the 1960s, the global environment was facing a crisis, and people's living environment was gradually deteriorating. The concept of ecological civilization also began to involve all levels of human social life, which is the future development direction of sports. However, the physical education classroom is the main channel for cultivating students' awareness of ecological sports, and it is also the main center of school physical education. At the same time, the emergence of ecological sports has also brought new challenges to school sports. Therefore, contemporary physical education workers have a new goal: to find an ecological sports curriculum with harmonious development of "environment, one person, one sports." The emergence and development of higher education, to a certain extent, reflects the level of social development in the corresponding stage. Colleges and universities are not isolated existence, they reflect various ecological backgrounds, and colleges themselves have different ecological factors. From an ecological point of view, the sustainable development of sports mainly refers to the balance between the sports system and society, people, and nature. Research on the development of sports ecology has become a common topic in the field of sports theory. The school sports in the system began to undertake the educational and teaching tasks and goals of sports ecological awareness. Ecological civilization constructs the theoretical cornerstone of scientific and reasonable physical education. At the same time, it also provides a different perspective for the reform and development of physical education courses in colleges and universities in my country [1, 2].

In recent years, the reform and development of physical education in colleges and universities in my country have been mixed. On the one hand, the reforms in recent years have had certain effects. For example, more optional courses have been added. Students can choose the best matching physical education courses according to their physical fitness, environment, and interests, highlighting diversity and liberalization. Secondly, physical education courses began to gradually increase the allocation of class hours for practical courses, so that students can have a relaxed and happy body. On the other hand, there are still many problems with the current physical education courses in colleges and universities. The problems of backward physical education teaching concepts and strong utilitarianism have not been fundamentally solved. But it did not come to fruition. The university stage is an important period that affects the socialization of individual students and the formation of their personalities, so more attention should be paid to the development of self-physical education and self-health of

business school students, as well as to the development of the awareness and ability of lifelong health and lifelong physical education of college students [3]. Meanwhile, the evaluation system for physical education teaching in domestic colleges and universities is influenced by the traditional educational ideas of the past and has limitations. On the one hand, many physical education teachers attend classes just to simply complete their teaching tasks. On the other hand, many students are not actively participating in physical education courses just to get the credits required for graduation. This almost utilitarian phenomenon is also a common problem in current colleges and universities, which is not only detrimental to the long-term development of physical education courses in colleges and universities but also detrimental to the development of students' physical quality W and their health. Therefore, the teaching in colleges and universities, students should be given more attention and strengthen the cultivation of correct values and worldviews; therefore, the mainstream sports concept nowadays should be developed toward the mode of mutual integration with ecology [4].

The physical education curriculum in higher education is closely related to ecological and natural civilization. The prerequisite for the smooth development of university physical education courses is to have a good campus cultural atmosphere and a good ecological civilization environment for sports. Each university has its characteristics, so the ecological environment on campus is also determined by the characteristics of the campus, each with its characteristics [5]. This paper is to study the future development direction of college physical education courses from the perspective of ecological sports and combine college physical education courses with the ecological environment, to put forward reasonable suggestions for the development of ecological sports in China's colleges and universities, mainly using the DCN deep learning method to promote the development of sports and to select the most suitable sports items by testing the physical quality of students. In turn, physical education is targeted to complete physical education.

In the 1970s, Brown Brenner [6] of Cornell University continued to pursue the ecological route, trying to establish an "ecology of human development," and in the 1990s, Wilson [7] explored and analyzed ecological physical education. The British scholar Eggleston [8], with his unique style, found that ecology is mainly concerned with the behavior and way of life of the organism, understanding the connection between it and the surrounding ecological environment, with the main purpose of scientific research on the distribution of educational resources. Smith and Williams [9] conducted a more in-depth study and analysis of the ecological problems of physical education that emerged from microclassroom ecology and macro-educational, cultural, and ecological crises, and the concept of ecological sport has gradually penetrated the physical education curriculum and become one of the main objects of research.

Through the analysis of the research status at home and abroad, it is found that the researchers all hold different viewpoints. However, the scientific research on the microscopic and macroscopic educational ecology in my country has become an important content of educational research in colleges and universities: "The basic spirit of synthesis, connection, balance." Therefore, this paper proposes a practical method for sports innovation based on the concept of ecological development and the computational theory of sports deep learning, aiming to accurately guide people to the most suitable sports events and better promote the healthy development of people's bodies.

2. Overview of Ecological Sports Research

2.1. Importance of Ecological Sports Resource Development. Eco-physical education, as the name implies, is the organic integration of physical culture and ecological environment, to achieve common development. Such a form of physical education curriculum not only meets the current theme of the times but also innovates and develops based on the original physical education, which can better promote students' physical exercise and also better promote students' physical culture development. As education reform continues to progress, physical education that helps students release stress is especially important as they face the pressure brought by academics in various subjects. Ecological physical education can be more effective in improving students' physical skills and can also enhance their mental health. The ecological physical education program focuses on exercising students' physical fitness and improving their physical ability so that students will have the appropriate survival skills and good adaptability when they enter society in the future, which will benefit their development [6–9].

In daily ecological physical education, teachers should lead students to establish the correct concept of physical education, make them realize that "health comes first," and continuously innovate the teaching classroom along this line to cultivate sports that suit students' interests. We should also follow this route to create innovative teaching classes and develop sports that suit students' interests. This can be done in several ways, such as teaching content, curriculum organization and implementation, curriculum evaluation, and curriculum objectives, to bring students a healthy ecological physical education with reasonable teaching strategies. In addition, while organizing fitness activities for students, teachers should also transmit the concept of environmental protection to students, and in the long run, students will naturally form a sense of ecological protection, which is a very meaningful thing for our living ecological environment.

2.2. Eco-Physical Education Innovation Path

2.2.1. Establishing the Correct Ideological Awareness. Teachers should first make students pay attention to ecological physical education in their minds. In this current ecological physical education environment, teachers should cultivate students' awareness of ecological physical education, to establish a new concept of the harmonious development of human beings, sports, and nature. Deepen students' understanding of the concept of ecological physical education, so that students can establish the correct concept of sports thinking and improve the lack of students'

cognition of ecological physical education. With the development of the times, it is more important for teachers to keep pace with the times and actively carry out ecological physical education courses to change the traditional physical education mode and concept in students' minds, and to study further in their spare time to enhance their professional skills and constantly update their ecological physical education ideology [10, 11].

2.2.2. School Teachers and Students Promote Ecological Sports as a Trinity. In traditional physical education, teachers always mechanically impart knowledge and theory to students, and the imitation of movements is too homogeneous, which makes students gradually lose interest in learning physical education courses. In addition, the lack of interaction between teachers and students in the process of physical education is not conducive to the establishment of emotions between teachers and students, and the lack of emotional communication between teachers and students also makes teachers unable to understand the actual situation of students in the first time. Moreover, teachers should make continuous exploration and innovation in the way of ecological physical education. For example, schools can broadcast knowledge about eco-physical education through campus broadcasts at regular intervals, so that students can learn anytime and anywhere during recess activities and will also exchange and discuss with each other and their classmates, thus enhancing their understanding and knowledge of the ecological curriculum of physical education [12–15]. Teachers can also assign learning tasks for students to set up blackboards about ecological sports knowledge based on their understanding of the ecological sports curriculum concept, and in the process of collecting and integrating ecological sports knowledge, students can also gain a deeper understanding of the ecological sports teaching concept utilizing information technology. This process expands students' knowledge horizons and develops their creative thinking and hands-on skills. The school should also actively hold an ecological sports culture festival and encourage every student to actively participate in it to deepen their knowledge of the ecological sports curriculum.

2.2.3. Scene Breakthrough. Scene is in the classroom teaching, the teacher simulates the teaching link and presents a number of designed scenes. In this process, the teachers showed different picture forms through passionate language and body expression to impress the judges. As such in track-and-field relay running teaching, teachers in the teaching process, first of all, give students to explain how to hand over the baton, the teacher through the language of explanation at the same time with their action demonstration, to give the judges that teachers present a clear and clear action intuitive sense; next, the teacher through the organization of different forms of relay running forms, such as chase running relay, short distance running relay, graphics running relay, and other forms, and the teacher can present these several field schematics through the form of board book when simulating the explanation to give the judges a clear feeling at a glance, and the judges' teachers will also feel that the teaching organization form of the class teacher is relatively new and unique, a bright feeling in front of their eyes, and it can also show the actual teaching ability level of the teacher in the real classroom teaching [16–19].

Moreover, the presentation of real scenes is also a kind of demonstration in the simulation classroom. In the simulated class, although there are no real students, the judges see real scenes, including real people and objects, which is a means and form to reflect the teaching ability of teachers. For example, in a lesson, the teacher needs to solve the key points of a lesson and effectively break through the difficult points, but how this difficult breakthrough is presented in the simulated class session, which requires the teacher to show the difficult breakthroughs bit by bit in a realistic way through the organization of different classroom practices [20–22]. Such as the combination of exercises in the articulation point teaching, such as hurdle running teaching, teachers know that the teaching of the starting leg and swing leg teaching is the focus of hurdle running teaching, many teachers in the class process are the starting leg teaching and swing leg teaching decomposition demonstration teaching, and in the teacher simulation class link, teachers can start the teaching of the starting leg and swing leg articulation point teaching through the form of simulation class to show. Of course, the articulation point teaching is the highlight of this lesson, so in the usual physical education teaching process, teachers should focus on teaching this articulation point, explain it through their language and body demonstration, and tell the articulation point teaching step by step through learning and practicing, to give the judges a fresh feeling, so that the judges can feel the different forms of teaching, and also let the judges through the process of not having a real classroom but the process of seeing the real scene without the real classroom [23–25].

In conclusion, among the different forms of proficiency assessment for teachers, the simulated classroom form is an effective means to show teachers' teaching ability without real classroom teaching, which is a form of effectively showing the "real" classroom teaching session through teachers' verbal explanation and body language. The emergence of this teaching style is a comprehensive examination of the teaching ability of physical education teachers. Therefore, in teaching, teachers have to show the virtual teaching classroom through various teaching methods, through the effective switching of scenarios, scenes, and real scenes, to present an extraordinary teaching effect.

3. DCN Deep Learning for Sports Innovation

3.1. DCN Model. DCN is a deformable convolution, which is based on the traditional CNN by adding a position offset to the sampled grid points on the input feature map (or input image), as shown in Figure 1.

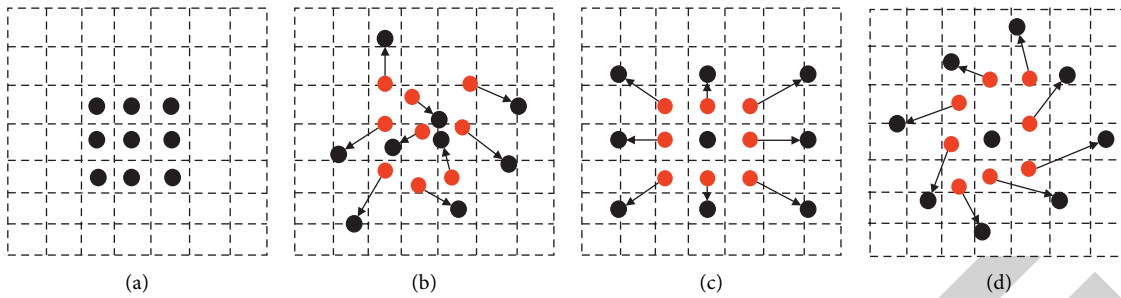


Figure 1: DCN location offset diagram.

Currently, various models of complex networks appear to be too simple in their construction methods and specific expressions compared to real networks, without considering the weights of the edges in the network and the directedness problem [26]. There is proposed a deterministic complex network model, the DCN model, and a weighting approach is used to allow the model to be better applied to real networks. In this paper, three edge weight assignment methods are used to assign values to the edges of the DCN model, respectively, to construct the weighted DCN model and derive the node strength formula of the weighted DCN model to obtain the optimal edge weight assignment method.

The DCN model obtains a deterministic complex network, or DCN for short, by n times of continuous growth, where the number of nodes in layer i is equation (1) and the total number of nodes is equation (2)as follows.

$$N(i) = 2^{i-1}, \tag{1}$$

$$N(n) = 2^{n+1} - 1. \tag{2}$$

The degree of each node in layer i is equation (3), the degree distribution is equation (4), the clustering coefficient is equation (5), and the average shortest path is equation (6), where equation (6) $i = 1, 2, 3, \ldots$

$$k = 4 \bullet 2^{n-i}, \tag{3}$$

$$P(k) = (k - i + 3)^{-1}, \tag{4}$$

$$c_{N \longrightarrow \infty} = 1, \tag{5}$$

$$L_{N \longrightarrow \infty} = 2. \tag{6}$$

Characteristics of point weight distribution of weighted DCN models with different assignment methods.

The weighted network can be represented by the weighted adjacency matrix equation (7), where $i = 1, 2, \ldots, N$, N is the size of the network, that is, the total number of nodes. The matrix element w_{ij} $(w_{ij} > 0)$ represents the weights of the edges connected by node i and node j. The degree distribution and node strength distribution are physical quantities that reflect the characteristics of the network. In this paper, we use three types of assignments to construct the weighted DCN model and analyze the distribution characteristics of point weights.

$$W = \left(w_{ij}\right). \tag{7}$$

3.2. Assignment Method with Constant Edge Weights. For the DCN model with $n = 5$, some parameters of this model are shown in Table 1. Define the edge weight between nodes with subordination in the DCN model with $n = 5$ as 1, the edge weight between nodes in the next generation as 2, the edge weight between nodes in the next two generations as 3, and the edge weight between nodes in the next three generations as 4, and the edge weight between nodes in the next four generations as 5.

From the generation process of the DCN model, it is known that the DCN model is very hierarchical, and the properties (degree, point power) of each node in the same layer are the same. The number of node layers is 1, 2, 3, 4, 5, 6, and the corresponding node strengths are 258, 99, 37, 16, 12, 15 from the defined equation (8) of point power.

The $n = 5$ weighted DCN model has a large point weight for each node and decreases with the number of layers, and the strength of the master node is the largest, which means that each node in the DCN is important and the master node plays an important role.

$$s_i = \sum_{j \in \Gamma(i)} w_{ij}. \tag{8}$$

The above is weighted according to the unequal weight of each edge, but of course, you can also set the weight of each edge to be the same according to the specific situation, such as all 1, and then calculate the node strength of the whole weighted DCN, which can be obtained from following equations.

$$s_i = \sum_{j \in \Gamma(i)} w_{ij} = k_i = 4 \bullet 2^{n-i} + i - 3, \tag{9}$$

$$P\left(s_i\right) = \frac{2^{i-1}}{\left(2^{n+1} - 1\right)}, \tag{10}$$

where s_i refers to the point weights and $P(s_i)$ refers to the node strengths. The node strength distribution of a deterministic weighted complex network with all edge weights of 1 is shown in Figure 2.

Table 1: Related parameters.

$n=5$	Degree of node k	Number of nodes $n(i)$
1	62	1
2	31	2
3	16	4
4	9	8
5	6	16
6	5	32

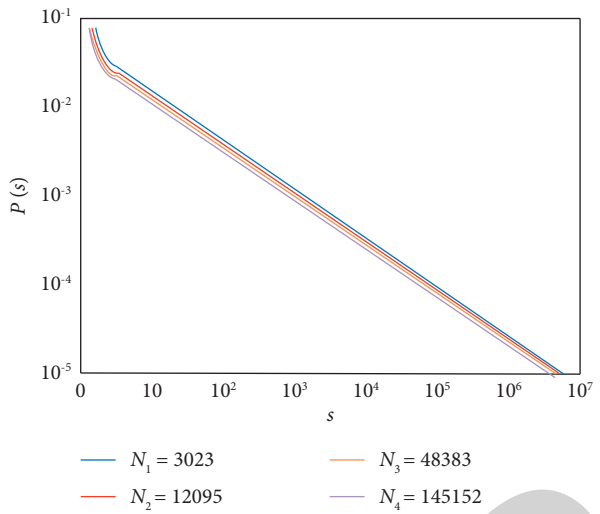


Figure 2: Point weight distribution of deterministic weighted complex networks with edge weight 1.

From Figure 2, the node strength distribution of the weighted DCN with constant edge weights obeys a power-law distribution in the range of $s > 20$ as shown in the following equation:

$$P(s) = s^{-\gamma}. \tag{11}$$

Therefore, the point weights of the weighted DCN with constant edge weights as a whole are consistent with the power-law distribution, which is consistent with the actual weighted network with a power-law distribution of point weights, so the weighted DCN model can describe the actual network more vividly.

3.3. The Side Weights Obey the Exponential Distribution of the Assignment. Assuming that the weights of edges in a weighted complex network obey an exponential distribution, that is, equation (12), the sample values obeying the exponential distribution are all greater than 0; that is, $w>0$ is consistent with the case where the edge weights are all greater than 0 in the actual network.

$$P(w) = \theta e^{-\theta w}. \tag{12}$$

The intensity distribution $P(s)$ of the weighted network nodes obeying the edge weights of the exponential distribution can be regarded as the probability density function of the sum distribution of the exponential distributions of the same parameters. For example, for 2 exponential distributions obeying parameters θ in equations (13) and (14), the sum distribution is equation (15), and then, the probability density function of Z is equation (16) from the convolution formula.

$$P_X(x) = \theta e^{-\theta x}, \tag{13}$$

$$P_Y(y) = \theta e^{-\theta y}, \tag{14}$$

$$Z_2 = P_X(x) + P_Y(y), \tag{15}$$

$$\begin{aligned} P_Z(z) &= \int_0^{\infty} P_X(x) P_Y(z-x) \mathrm{d}x \\ &= \int_0^{\infty} \theta e^{-\theta x} \theta e^{-\theta(z-x)} \mathrm{d}x = \theta^2 z e^{-\theta z}. \end{aligned} \tag{16}$$

By analogy, the probability density function of the sum distribution of the exponential distributions of the same parameters is

$$P_{ski}(s) = \frac{\theta^{k_i}}{(k_i - 1)!} s^{k_i - 1} e^{-\theta s}, \tag{17}$$

where $s > 0$, $\theta > 0$, $1 \le i \le n+1$, k_i is the degree of node i, and $k_i = 4 \bullet 2^{n-i} + i - 3$. From equation (17) and the intensity distribution of the weighted DCN nodes under each parameter, it can be seen that the intensity of node i obeys the Gamma distribution $s \sim Ga(K_i, \theta)$, so it does not have the power-law characteristic.

3.4. The Edge Weights Obey the Assignment of the Nodal Degree Product Function. Let the degree of node i and node j be k_i and k_j, respectively, then the edge weights connecting these 2 nodes satisfy the relation equation (18), where α can effectively regulate the intensity magnitude of the nodes. From the regularity and hierarchy of the DCN model, it is known that the point weights of any node in the same layer are the same. The point weight is defined by equation (19) [27, 28]as follows:

$$w_{ij} = \left(k_i, k_j\right)^{\alpha}, \tag{18}$$

$$s_i = \sum_{j \in \Gamma(i)} w_{ij} = \sum_{j \in \Gamma(i)} \left(k_i, k_j\right)^{\alpha}, \tag{19}$$

where $\Gamma(i)$ denotes the set of all nodes connected to node i. For some node m in layer i, the strength s_i of node m is calculated by summing the weights of all the edges connected to node m. Node m is directly connected to the nodes of immediate relatives from layer 1 to layer $i-1$, so we have

$$s_{i1} = w_{i,1} + w_{i,2} + \ldots + w_{i,i-1} = \sum_{j=1}^{i-1} \left(k_i, k_j\right)^{\alpha}, \tag{20}$$

where k_i and k_j denote the degrees of node i and node j, and W_{ij} denotes the edge weights connecting these 2 nodes.

All direct subrelative nodes from level $i+1$ below node m to level $n+1$ are directly connected to m with.

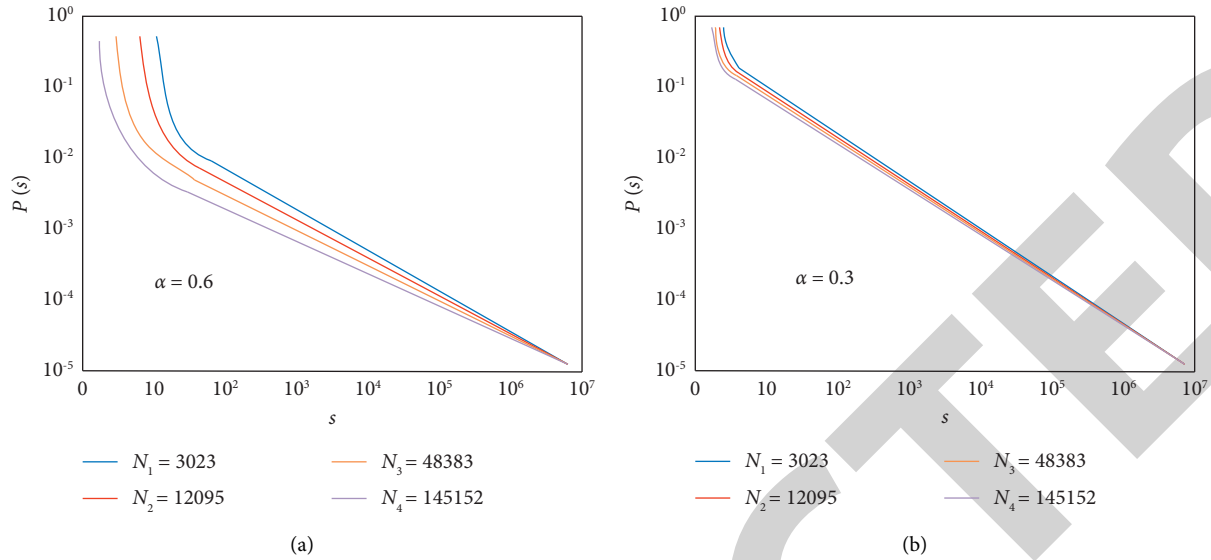


Figure 3: Point weight distribution of deterministic complex weighted networks with edge weights as degree product distribution about (a) $\alpha = 0.6$; (b) $\alpha = 0.03$.

$$s_{i2} = w_{i,i+1}\bullet 2 + w_{i,i+2}\bullet 4 + \ldots + w_{i,n+1}\bullet 2^{n+1-i} = \sum_{j=i+1}^{n+1}\left[w_{ij}\bullet 2^{j-i}\right] = \sum_{j=i+1}^{n+1}\left[\left(k_i, k_j\right)^{\alpha}\bullet 2^{j-i}\right]. \tag{21}$$

The degree n of any node in the i-th layer is $k_i = 4 * 2^{n-i} + i - 3$, the total number of nodes in the i-th layer is $N(i) = 2^{i+1}$, and the total number of nodes in the DCN is $N(i) = 2^{n+1} - 1$. Therefore, the strength of any node m in the i-th layer of the weighted DCN is

$$\begin{aligned} s_i &= s_{i1} + s_{i2} \\ &= \sum_{j=1}^{i-1}\left(k_i, k_j\right)^{\alpha} + \sum_{j=i+1}^{n+1}\left[\left(k_i, k_j\right)^{\alpha}\bullet 2^{j-i}\right] \\ &= \sum_{j=i}^{i-1}\left[\left(4\bullet 2^{n-1} + i - 3\right)\left(4\bullet 2^{n-j} + j - 3\right)\right]^{\alpha} \\ &\quad + \sum_{j=i+1}^{n+1}\left[\left(4\bullet 2^{n-i} + i - 3\right)\left(4\bullet 2^{n-j} + j - 3\right)\right]^{\alpha} 2^{j-i}, \end{aligned} \tag{22}$$

$$P\left(s_i\right) = \frac{2^{i-1}}{2^{n+1} - 1}. \tag{23}$$

3.5. Results and Analysis. Firstly, the nodes of the weighted DCN are analyzed and calculated, and the intensity distribution law is shown in Figure 3.

From Figure 3, it can be seen that when $s > 10^4$, the four-point power distribution curves in the double logarithmic axis show an obvious linear relationship, that is, the relationship between $P(s)$ and the power-law distribution of s is very well conformed. From Figure 3, it can be seen that the power-law relationship of the point-weight distribution is more obvious than Figure 3; that is, the smaller α is, the more the point-weight distribution tends to be a power-law distribution, so a suitable α value can be set to simulate it according to the actual weighted network.

Therefore, for the weighted DCNs whose edge weights obey the node-degree product function, the nodal intensity distribution as a whole obeys the power-law distribution of $P(s) \sim s^{-\gamma}$, which is consistent with the actual weighted networks.

When the DCN is assigned with the product by node degree, the edge weight formula is equation (24), where the parameter α can take different values according to the actual situation; that is, it represents different physical quality conditions of different people, to achieve the purpose of effectively regulating the intensity of their sport.

$$w_{ij} = \left(k_i k_j\right)^{\alpha}. \tag{24}$$

Further, this paper investigates how the point weight distribution pattern changes if the value of α is changed for a given weighted DCN in Mexico. Figure 4 shows the point weight distribution of the weighted DCN ($N = 3023$) with edge weights obeying the nodal degree product function under different α. From Figure 4, it can be seen that the four-point power distribution curves under the double logarithmic axis show an obvious linear relationship in a certain range; that is, the point power distribution obeys the power-law distribution in a certain range, and the smaller α is, the more obvious the power-law relationship is; moreover, the node strength of the weighted DCN increases rapidly when α gradually becomes larger so that a suitable α value can be set according to the requirements of the actual weighted network for simulation.

Thus, after assigning values to the edges of the DCN model employing nodal degree products, α represents the physical fitness detection, and some characteristics of the

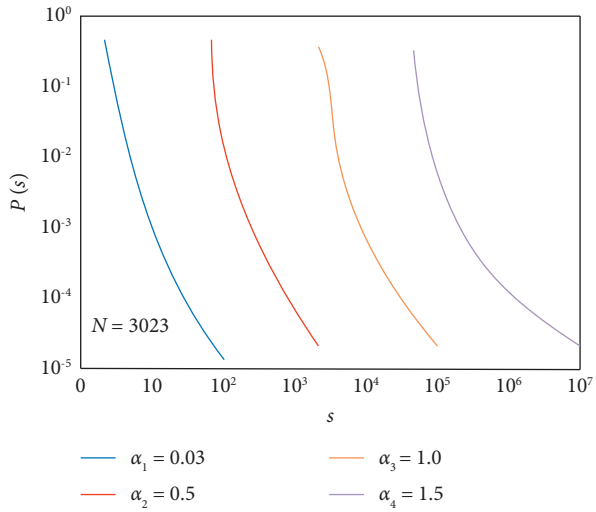


Figure 4: Weighted DCN point weight distribution under different α.

weighted DCN model were obtained through analytical calculations: (1) The nodal strength of the model is subject to a power-law distribution, and the power-law feature becomes more pronounced as the parameter α decreases; (2) for a weighted DCN model of a given size, the node strength increases rapidly as α becomes progressively larger, thus enabling different parameters to be set for simulating different actual weighted networks, which is convenient and flexible; and (3) the point power of the model is associated with the node degree, and $s(k)$ grows with k in the power-law form $s(k) \sim k^{\beta}$. The smaller the α is the more obvious the power-law relationship is, so that different α values can be set according to the requirements of the actual weighted network, and the size of the network node strength can be adjusted effectively so that the most suitable sports project can be selected.

4. Conclusion

With the continuous development of China's economy and society, the requirements for ecological sports have become higher and higher, and the sports system bred in the era of Wang et al. civilization can no longer fully adapt to the new requirements for sports in the era of ecological civilization in the twenty-first century. The ecological development of human sports is the inevitable trend of our sports development. Therefore, universities in China must speed up the reform and development of physical education courses in the direction of ecology to further promote the ecological development of physical education in China. However, teachers effectively innovate the ecological physical education classroom, which is conducive to the healthy development of students' body and mind, and students continuously enhance their own ecological and environmental awareness in the process of learning, which effectively realizes the concept of sustainable development of human and society and nature, and on this basis improves students' interest in sports, so that students develop the correct sports ideology and concept, thus promoting students' all-round development. Therefore, this paper investigates the distribution law of point weights of the weighted DCN model under three types of edge weight assignment and draws the following conclusions: (1) when the edge weight obeys the exponential distribution, the node strength of the weighted DCN model obeys the Gamma distribution, so in this case, it does not have the characteristics of power law; (2) when the edge weight is constant or the edge weight obeys the node degree product function, the point weight distribution of the weighted DCN model satisfies the power-law relationship. From the statistical analysis of the real weighted network, the statistical distribution $P(s)$ of the node weights has characteristics similar to the degree distribution; that is, it obeys the power-law distribution. These characteristics are reflected in the weighted DCN model, so the weighted DCN model can simulate the real world better than other networks; (3) from the results of studying the weighted DCN model, it can be seen that among the three edge weight assignment methods, the edge weight obeying the node degree product method is the best because this method is the most suitable for actual needs. The weighted model in this way not only has a power-law relationship between the point weight distribution $P(s)$ but also satisfies the power-law relationship between the point weight and the node degree, and can effectively adjust the parameter α of the point weight size according to actual needs, that is, different physical conditions. By changing the value of α, the network can be made closer to the actual network, to obtain a better effect of sports recommendation items, and to better realize the innovation and development of ecological sports.

Data Availability

The dataset used to support the findings of this study can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] D. Y. Trejos, J. C. Valverde, and E. Venturino, "Dynamics of infectious diseases: A review of the main biological aspects and their mathematical translation," *Applied Mathematics and Nonlinear Sciences*, vol. 7, no. 1, pp. 1–26, 2022.

[2] G. Lieblein and C. Francis, "Towards responsible action through agroecological education," *Italian Journal of Agronomy*, vol. 2, no. 2, p. 83, 2007.

[3] Rizhao and Shandong, "PRC. ON the ecological construction of sports tourism," *Journal of Qufu Normal University (Natural Science)*, vol. 28, no. 4, pp. 114–116, 2002.

[4] A. N. Akkılıc, T. A. Sulaiman, and H. Bulut, "Applications of the extended rational sine-cosine and sinh-cosh techniques to some nonlinear complex models arising in mathematical physics," *Applied Mathematics and Nonlinear Sciences*, vol. 6, no. 2, pp. 19–30, 2021.

[5] S. Loland, "Record sports: an ecological critique and a reconstruction," *Journal of the Philosophy of Sport*, vol. 28, no. 2, pp. 127–139, 2001.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9460985, 9 pages
https://doi.org/10.1155/2022/9460985

Research Article

Development of Network Security Based on the Neural Network PSD Algorithm

Jianxun Li, Song Ji, and Yiran Jiang

School of Information Science and Engineering, Baoding University of Technology, Baoding 071000, Hebei, China

Correspondence should be addressed to Song Ji; 13938757257@hait.edu.cn

Received 17 July 2022; Revised 15 August 2022; Accepted 12 September 2022; Published 30 September 2022

Academic Editor: Wenming CAO

The more frequent occurrence of network security incidents has an impact on network security. Through the research on network security situational awareness, this paper constructs a multilevel network security situation evaluation index system from various aspects and uses the Elman neural network optimized by the genetic algorithm to evaluate network security situation. Aiming at the disadvantage of subjective dependence in the traditional assignment method of basic probability assignment function, Elman neural network is used to obtain the basic probability assignment function to increase its objectivity, and it is optimized with the PSD algorithm. In addition, the neural network is further improved by the genetic algorithm. In the traditional D-S evidence theory, an evidence correction step is added to optimize the situation that the final judgment result is incorrect due to evidence conflict. Finally, the fusion rules of the D-S evidence theory are used to fuse the support degrees of the four first-level situations to different security levels to obtain the final network security situation assessment result. The results show that the prediction accuracy of the GA-Elman neural network model is as high as 80%, which is significantly higher than that of the traditional D-S model, indicating that the model proposed in this paper has improved the accuracy of the assessment and prediction results. In conclusion, this study provides feasible theoretical prediction guidance for the accurate assessment of network security posture, reveals the improvement ideas for network security development, and is of great significance for the maintenance of network environment security.

1. Introduction

The rapidly developing economy has provided development opportunities for the field, and network-related technologies have been widely used [1, 2]. In the long-term social development process, people find that the rational use of the Internet can improve the way of life and make people's activities more convenient. However, with the continuous expansion of the network application scope and the continuous increase in the frequency of use, network security incidents occur more and more frequently, which has an impact on network security. Due to the expansion of network usage, network security incidents occur frequently. The popularization of the network provides convenience to people's lives, but it also brings greater difficulty to network security maintenance [3, 4].

Network attack is a typical representative of network security incidents [5], and data leakage is one of the serious consequences of network attacks. The total number of data breaches in 2020 surpassed the previous maximum and reached a new high, becoming the highest number of breaches so far. The frequent occurrence of security incidents shows the importance of paying attention to the security of the network environment [6]. In 2014, the state made it clear that in order to ensure national security, the maintenance of network security as a part of it is also crucial. However, because the defense side of traditional network security defense technology does not really combine data, intelligence, and other information resources with network security, some single security technology is relatively backward, and it is necessary to achieve a high level of network security protection. It is not enough to use only

one-sided technologies such as firewall, intrusion detection, and access control [7]. The birth of network security situational awareness technology has solved the problems existing in traditional security technology. The emergence of this technology enables the protection terminal to take the relevant protection measures proactively while detecting whether a network security attack occurs or not and performs real-time feedback and operations [8, 9]. Compared with foreign countries, China's network security situational awareness research started relatively late and has gone through many stages. As China's research on SOC/SIEM and other types of products and technologies has been going on for many years, it has always been at the level of alerting security incidents and has not provided any guarantee for protecting network security. The emergence of big data technology has made a breakthrough for domestic researchers and realized the security analysis technology based on big data [10]. Through the research on network security situational awareness, academic personnel at home and abroad have achieved many remarkable results. In order to improve the accuracy of network security situation assessment, Siaterlis et al. [11] proposed a new network security situation assessment scheme by combining evidence theory with situation assessment. Dutt et al. [12] proposed an IBLT model based on tools to resist security attacks, which can identify some behaviors in the network environment, thereby effectively improving the accuracy of evaluation. Almeida et al. [13] connected modular components in a multilayered manner, proposing a situational awareness model based on EXEHDA-ISSA. Zhang and He [14] used big data analysis and combined forecasting methods to predict the security situation of smart grids; Liu [15] improved the ability of D-S evidence theory to confront conflict by proposing predictive models based on improved D-S.

Network security situational awareness is one of the hot research points in the current network security direction. The reason is that it can dynamically show the overall situation of network security by evaluating the security influencing factors at various levels and, at the same time, makes accurate predictions on the development trend of future network security situation [16]. This paper selects the method of combining spiking neural network (PSD) to evaluate and predict the network security situation. The optimized neural network assigns BPA to each feature, uses the PSD algorithm to process the input of relevant data, improves the parameter optimization ability of the system, and obtains more accurate results while improving the objectivity of BPA assignment. At the same time, this paper corrects the BPA by introducing the Pearson coefficient and the distance of evidence to solve the problem of wrong conclusions caused by conflicting evidence.

2. Establishment of the Theoretical Model

2.1. DS-GAElman Evidence Theory

2.1.1. D-S Evidence Theory Synthesis Rules. The specific formula of the synthesis rule of D-S evidence theory [17] is as follows:

(1) Two mass function synthesis rules are

$$m = m_1 \oplus m_2, \begin{cases} m(\varnothing) = 0, \\ m(A) = k^{-1}, \sum_{A_1 \cap A_2 = A} m_1(A_1) \cdot m_2(A_2), \end{cases} \tag{1}$$

where $k = 1 - \sum_{A_1 \cap A_2 + \varnothing} m_1(A_1) \cdot m_2(A_2) = \sum_{A_1 \cap A_2 + \varnothing} m_1(A_1) \cdot m_2(A_2)$.

(2) Multiple mass function synthesis rules are

$$m = m_1 \oplus m_2 \oplus \ldots m_n \begin{cases} m(\varnothing) = 0, \\ m(A) = k^{-1}, \sum_{A_1 \cap A_2 \cap \cdots A_n = A} m_1(A_1) \cdot m_2(A_2) \cdot \ldots \cdot m_n(A_n), \end{cases} \tag{2}$$

where $k = 1 - \sum_{A_1 \cap A_2 \cap \ldots A_n = \varnothing} m_1(A_1) \cdot m_2(A_2) \cdot \ldots \cdot m_n(A_n) = \sum_{A_1 \cap A_2 \cap \ldots A_n = \varnothing} m_1(A_1) \cdot m_2(A_2) \cdot \ldots \cdot m_n(A_n)$.

2.1.2. Improvement of D-S Evidence Theory. After a lot of research, researchers believe that D-S evidence theory can be selected as a solution [18] because the theory can describe uncertainty and has the ability to solve such problems. However, in the process of using evidence theory, researchers have found that there may be problems such as contradicting the actual results, unfairness, or even being rejected by one vote when fusing evidence. Among these emerging problems, the fusion results of highly conflicting evidences that contradict intuitive results have become the focus of researchers. When the D-S evidence theory fuses evidences with less conflict, the error between the final synthesis result and the correct result will be smaller, and when there is high conflict between the evidences, the synthesis result of the D-S evidence theory will have a larger deviation even against the correct result. In this regard, this paper calculates the credibility of evidence by introducing Pearson correlation and calculates the average evidence probability according to the credibility [19]. The evidence weight w is determined, and w is used as a correction coefficient to modify the weight of the original evidence body, so as to obtain the improved evidence BPA. Finally, all BPAs are synthesized using the basic synthesis rules to obtain the final result. The improved steps are as follows:

(1) Calculate the similarity of the Pearson correlation coefficient between each evidence:x and y are two variables, and assuming that there is a certain linear relationship between x and y, this correlation can be represented by the Pearson correlation coefficient P ranging from −1 to 1. When P is a number greater than 0, it means that any one of the variables in x and y increases with the increase of the other variable. This situation indicates that the relationship between x and y is a positive correlation. Conversely, when P is negative, it indicates that the relationship between x and y is linearly negative. When one variable in x and y increases, the other value decreases instead. If P

is equal to 0, it means that there is no linear correlation between x and y. The formula for the correlation coefficient is as follows:

$$P_{ij} = \frac{\operatorname{cov}(m_i, m_j)}{\sigma_{m_i}\sigma_{m_j}} = \frac{E\left(\left(m_i - \mu_{m_i}\right)\left(m_j - \mu_{m_j}\right)\right)}{\sqrt{E\left(m_i^2\right) - E^2\left(m_i\right)}\sqrt{E\left(m_j^2\right) - E^2\left(m_j\right)}}. \tag{3}$$

In the formula, $\mu_{m_j} = m_i - E(m_i)$, $\mu_{m_j} = m_j - E(m_j)$. The correlation coefficient of the above two evidences can be extended to $n \times n$ evidences, and the correlation matrix P_{ij} can be obtained as follows:

$$P_{ij} = \begin{bmatrix} p_{11} & p_{12} & \cdot & \cdot & p_{1n} \\ p_{21} & p_{22} & \cdot & \cdot & p_{2n} \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ p_{n1} & p_{n2} & \cdot & \cdot & p_{nn} \end{bmatrix}. \tag{4}$$

(2) Calculate the support degree of evidence m_i:

$$\sup_{m_i} = \sum_{j=1, j\neq i}^{n} P_{ij}\left(m_i, m_j\right). \tag{5}$$

(3) Calculate the credibility of the evidence:By analyzing the support formula of m_i, it can be seen that the greater the correlation between the two evidences, the higher the support, then the higher the credibility of m_i, so the credibility of m_i:

$$\operatorname{de} p_{m_i} = \frac{\sup_{m_i}}{\sum_{i=1}^{n} \sup_{m_i}}, \operatorname{de} p_{m_i} \in [0, 1]. \tag{6}$$

(4) Calculate the adaptive weight w_i separately for different hypothesis results:The smaller the d_i, the closer the distance to the average evidence probability, the greater the weight of the evidence. It can be known that if the distance is inversely proportional to the weight, then the probability weight w_i of each evidence is given as

$$w_i = \frac{1/d_i}{\sum_{i=1}^{n} 1/d_i}. \tag{7}$$

(5) Calculate the probability m_i' of each evidence separately:

$$m_i = w_i m_i. \tag{8}$$

(6) All m_i are fused sequentially using the synthesis rules, and after all the iterative syntheses are completed, the situation that meets the decision rules is selected as the final result of the problem.

2.2. Neural Network Model and Its Improvement. Feedforward neural networks are only affected by the input data at the current moment and do not care about the data at the previous moment [20, 21]. However, when solving practical problems, it will be found that there are intricate relationships among many data, including certain time series. If only the data at the current moment is considered, the final output may be biased, and the feedforward neural network does not have the ability to remember and cannot effectively deal with timing problems. Therefore, recurrent neural networks have been considered to solve this type of problem [22].

As the basic element of information transmission, Elman neural network was first proposed by Elman [23]. It adds an additional layer to the hidden layer to record the historical data of the hidden layer, that is, the context layer. Because of its dynamic and nonlinear characteristics, it is very sensitive to time series data, and the data have complex correlation. Elman neural network is selected as the core module to obtain BPA, and the PSD learning algorithm is used for deep learning and optimization [24].

2.2.1. Elman Neural Network Model. Compared with the ordinary neural network structure, the Elman network adds a new layer of undertaking layer, the implicit layer transmits the processed data to the undertaker layer, the undertaker layer memorizes the information transmitted by the implicit layer and combines the received data with the input layer of the next moment as the input of the implicit layer at the next moment [25, 26]. By storing it through the undertaker layer and outputting it to the hidden layer at the next moment, the neural network has the function of dynamic memory recognition of historical input data and enhances its ability to process dynamic information. The specific mathematical model is given as

$$\begin{cases} h(k) = g(w_3 \cdot q(k)), \\ q(k) = f[w_1 \cdot q_c(k) + w_2(u \cdot (k-1))], \\ q_c(k) = q(k-1), \end{cases} \tag{9}$$

where h is the output of the output layer, $g$is the transfer function of the output layer, w_3 is the weight of the data processed by the implicit layer in the data received by the output layer, q is the state of the implicit layer, and k is the current moment. In the second formula, f is the handling function of the implicit layer, and in most cases, Sigmoid is selected, and w_1 refers to the weight of the data processed by the implicit layer in all the received data of the undertaken layer, w_2 is the weight of the information received by the input layer to the implicit layer, and u is the input of the input layer. q_c in the second and third formulas refers to the state that is output for the undertaker layer, and $k-1$ in q indicates the previous moment.

2.2.2. Improvements to Elman Neural Networks. The traditional Elman neural network takes the gradient descent method as its weight update method, which is easy to fall into the local minimum value in the case of multiple local

minimum values in the error function; in this case, this method cannot guarantee that the optimal solution to be solved can be found, which will have a bad impact on the network training effect, so that the error of the final output result increases [27].

(1) Initialize the Elman neural network structure:Set related parameters including the number of neurons in each layer of the Elman neural network.

(2) Genetic algorithm parameter initialization:Set relevant parameters such as population size, probability of crossover and mutation, and target scalar function.

(3) Code to generate the initial population:The weights and thresholds of the Elman neural network are used as parameters to be optimized, and they are encoded as an individual, and multiple individuals form a population.

(4) Determination of fitness function:The genetic algorithm needs to assign the individual probability through a certain index and provides the basis for the genetic algorithm to survive the fittest through the size of the selected probability. The fitness function is used as an index to evaluate the pros and cons of chromosomes, and it needs to use a specific function as an evaluation criterion. Therefore, the fitness function needs to be used to calculate the fitness value of each individual, and the individual in the group can be distinguished from the bad by comparing the size of the function value. The fitness function f in this paper is

$$f = \frac{1}{\sum_{i=1}^{n} y_i' - y_i}. \tag{10}$$

In the formula, f is the fitness value, y_i' and y_i are the actual and expected weights and thresholds output by the ith output node, and n is the total number of network input samples.

(5) Perform the selection operation:According to the fitness function set in (d), it can be seen that the greater the fitness, the better the individual, and the better the genes of the individual in the next generation. Therefore, the probability of the individual being selected can be improved by increasing the selection probability. In this paper, the roulette method is used as the method of selection operation, and the probability of each individual being selected is

$$\mathrm{P}_i = \frac{f_i}{\sum_{i=1}^{n} f_i}, 1 \le i \le n. \tag{11}$$

The formula is to take the ratio of the individual fitness value to the sum of all individual fitness as the probability of the individual being selected.

(6) Obtain the optimization result: Determine whether the obtained results meet the training requirements and whether the termination conditions are met. If it is reached, pass the obtained optimal initial weights and thresholds to the Elman neural network, ending the iterative process. If the termination condition is not reached, the iterative evolution is continued until the condition is finally reached, then the iteration ends, and the result is obtained.

2.2.3. PSD Learning Algorithm. By analogy with the Widrow–Hoff (WH) learning rule, we can get

$$\Delta w_i = \alpha x_i (y_d - y_o), \tag{12}$$

where w_i is the weight of the corresponding quantity of the ith input, α is the learning rate, y_d is the expected sequence, y_o is the actual output sequence, and x_i is the input sequence. Because the output is actually a sequence containing pulse spikes, it is very difficult to obtain the derivation. In the PSD rule, a sharp pulse and a convolution kernel are convolved to obtain a derivable continuous value:

$$K\left(t - t^j\right) = V_0 \cdot \left(\exp\left(\frac{-\left(t - t^j\right)}{\tau_s} \right) - \exp\left(\frac{-\left(t - t^j\right)}{\tau_f} \right) \right). \tag{13}$$

3. Cybersecurity Situation Assessment

The selection of security situation indicators is the basis of the multilevel index system, and it is also the basis for the data sources and situation level standards of the security situational awareness process in the following sections. The results of the selection of security situation indicators directly affect the accuracy and rationality of subsequent security situation assessment and security situation prediction results. Because the network security environment is complex and diverse, certain selection rules need to be adopted when processing the data indicators of the network security situation. In general, the major functional modules of the network security system should be covered [28]. However, if too many indicators are selected, it is difficult to determine the importance and weight of each indicator. It will also increase the overall complexity of the situational awareness model and affect the situation assessment and prediction results. Therefore, it is necessary to obtain more comprehensive and accurate security situation indicators according to reasonable principles.

3.1. Construction of the Indicator System. According to the principle of hierarchy and classification, this paper proposes a relatively complete model of the first situational awareness index system, and according to the classification, the indicators are divided into four indicators: vulnerability, disaster tolerance, threat, and stability. The function of the first step is to select several main indicators, and these indicators need to have high independence and strong generality and use them as first-level indicators. Then, for each first-level indicator, a certain number of second-level indicators are selected as auxiliary indicators to describe the security situation. The network security situation assessment index system constructed is shown in Table 1.

Table 1: Network security situation assessment indicator system.

First-level indicators	Secondary indicators
Availability	Traffic change rate, average survival time of major devices, mean time between failures of major devices, and total subnet traffic
Vulnerability	Total number of security devices, total number of open ports on major devices, number and extent of network vulnerabilities, device system type, and memory capacity
Threatening	Type and number of alarms, type and severity of attacks, historical frequency of attack events, and distribution and size of threat data packets
Disaster tolerance	Network topology, network bandwidth, number of concurrent threads of subnet servers, operating system types of subnet devices, and types of services provided

3.2. Quantification of Indicators. As can be seen from the description in the previous section, there are many kinds of factors that affect the security of the network environment, and the raw data formats of these factors are different. In addition to differences in format, there are many other situations that make these data unsuitable for direct experimental manipulation since the vast majority of manipulations use numeric data. Therefore, we need to quantify the collected data on the basis of CVSS. Some of the formulas are mentioned in the next section.

3.2.1. Degree of Network Vulnerability. In the vast majority of security incidents, vulnerabilities in the network environment are the main reason for cyber-attacks. Vulnerabilities vary in severity and impact on the network. The quantitative formula for the degree of network vulnerability is mainly composed of the number and category of vulnerabilities. The formula is as follows:

$$\mathrm{b} = \frac{\sum_{k=1}^{n} \sum_{t=1}^{s} w_{k1} I_k N_{kl}}{B_{\mathrm{n}}}. \tag{14}$$

Among them, b represents the vulnerability level, B_{n} represents the total number of network vulnerabilities, n represents the number of devices, S represents how many different vulnerabilities exist, k represents the number of devices, I represents the number of vulnerabilities, and w_{k1} represents the weight of vulnerabilities, and this value is given a specific score by the CVSS standard, and $N_{kl}$is the number of certain vulnerabilities contained in a certain device.

3.2.2. Network Topology. When in a real network environment, the network topology will also have a certain impact on network security [29]. Since topology is used to determine how devices are placed, their security and degree of security are also key objectives to consider when measuring network security. The formula used is as follows:

$$\begin{aligned} \mathrm{topo} &= \sum_{k=1}^{n} TP_k, \\ TP_i &= \begin{cases} 1.0, \mathrm{n} \in [0, 4], \\ 0.6, n \in (4, 6], \\ 0.2, n \in (6, +\infty), \end{cases} \end{aligned} \tag{15}$$

where topo is the network topology quantization value, and TP_k refers to the number of nodes present in the environment measured by n in the fractional metric of the topology used by a network.

3.2.3. The Severity of the Attack. The destruction of the network environment after being attacked is a factor affecting network security. The broken situation after being attacked generally increases. At the same time, it is also necessary to consider that the degree of danger of different devices under the same attack is different, and the degree of damage to network security caused by different types of attacks is also different. Therefore, the quantification formula of attack severity is as follows:

$$A = \frac{\sum_{k=1}^{n} \sum_{i=1}^{s} r_k b_k Q_{kl}}{a_n}. \tag{16}$$

In the formula, n is the total number of devices, S is the number of types of attacks that are suffered, r_k refers to the level of a host receiving a certain attack, which is derived from the degree of possible impact of the attack, Q_{kl} refers to how many times the host has been attacked of a certain type, b_k refers to the degree of impact of a vulnerability, a_n represents the total number of types of attacks in a certain period of time, and A is a score that refers to the degree of damage caused by the attack.

3.2.4. The Degree of Occurrence of Attacks. The number of security incidents in a certain period of time will also affect the score of the entire situation indicator. If the number of security incidents occurs in a certain period of time, it means that the security of the environment is low. On the contrary, if the number of security incidents is small, the network security is relatively high. Therefore, the quantitative formula for the frequency of attack incidents is

$$f = \frac{N}{t}, \tag{17}$$

where f represents the occurrence of attack events, t is a fixed time period, and N represents the total number of attack events in the time period t.

3.3. Criteria for Classification of Safety Levels. The network situation is divided into five security levels, as shown in Table 2. Class I indicate that the current network security is high. At this time, the status of various indicators of the network has reached an excellent level, and it is in a state of continuous security and is under attack. Class II indicates that the network is currently in a relatively secure state, and the security posture of the network is in a good state, but compared to class I, the security state has been slightly affected. Class III means that the network situation is basically safe, and the network is in a medium security state. The network may have been attacked. Although the situation is basically safe, it is necessary to focus on some indicators in the poor state to prevent the network security state from deteriorating. Class IV indicates that the network is under moderate threat and may have suffered a certain degree of attack. Security personnel need to immediately look for indicators that are under moderate threat, and immediately take reasonable measures to make the network security state out of the threat state and restore security. Class V indicates that the system is extremely dangerous at this time, and the network security has been seriously affected. Emergency measures must be taken to eliminate the danger immediately.

Before the evaluation, it is necessary to process the data collected in the real network environment that may have some duplicate information and incomplete information. The processed dataset is divided based on a sliding window method, and a sample set is generated , some of which are used for neural network training, and the remaining part of the data is used as a test set for evaluation and testing, and the trained GA-Elman neural network is used to evaluate the model. BPA acquisition of indicator evidence is required to carry out the D-S evidence theory. The weight value of each evidence is revised, and then the updated basic probability distribution function is obtained, and finally, the evidence is iteratively fused through fusion rules. The final output value of the GA-Elman neural network model is the situational BPA. Therefore, this paper establishes the corresponding relationship between the situational level and the situational BPA, as shown in Table 3. The maximum confidence of the BPA is taken as the result, and the final decision of the network security posture assessment result is made according to the assessment level table and the correspondence table, and the assessment result is output.

3.4. Data Analysis of Network Security Situation Based on DS-GA-Elman. The experimental data in this paper are the public dataset NSL-KDD [30]. Most of the public datasets are real, and effective data are obtained by a well-known laboratory which had been collecting network traffic data for several months, and the dataset has been used by many researchers. Therefore, the comparison of experimental results using this dataset has certain comparability [31].

After the processing of the GA-Elman evaluation model, this paper obtains the support of each sample for different network security levels. Then uses the improved D-S evidence theory to fuse the BPA after correction and finally takes the fusion result with the largest value as the final evaluation result. According to the corresponding relationship between the situation level and situational BPA established in Table 3, the security level is evaluated. Situation states [O1, O2, O3, U] represent A-level security state, B-level security state, dangerous state, and unknown state, respectively. A-level security state only includes class I, and B-level security includes class II and III security levels. Hazardous states include class IV and class V safety levels. The final BPA value is shown in Table 4.

Table 2: Network security assessment level.

Security level	Description of network operation status
Class I (safety)	Network service is not affected
Class II (relatively safe)	Network service is slightly affected
Class III (basic safety)	Network services are moderately affected
Class IV (less safe)	Network services are severely affected
Class V (unsafe)	Network services are severely affected

Table 3: BPA value corresponding to the network security assessment level.

Security level	Situation BPA value
Class I (safety)	(0.8–1.0]
Class II (relatively safe)	(0.45–0.8]
Class III (basic safety)	(0–0.45]
Class IV (less safe)	(0.55–1.0]
Class V (unsafe)	(0–0.55]

Then, the final BPA results obtained by each sample are classified according to the network security classification standard, and the following conclusions are obtained (Table 5).

From Table 6, the error rate of using the traditional D-S evidence theory to evaluate the network security situation is as high as 30%. Compared with the D-S evidence theory, the number of errors in the situation assessment results of the Elman neural network increased by 10%, and the correct rate was only 60%. The results based on DS-GA-Elman neural network has an accuracy rate of 80% for the evaluation results of the test set, which is 10% higher than that of the traditional D-S evidence theory situation assessment results. The evaluation results of the Elman neural network model have a high accuracy rate. In order to verify the accuracy of the model evaluation, different samples and the number of samples were randomly selected many times for experiments. The comparison of the evaluation accuracy of each model is shown in Figure 1.

3.5. DEIPSO-GRU Prediction Model Results and Analysis. This paper compares the prediction results of the basic GRU and PSO-GRU-based prediction methods with the situation prediction results of the DEIPSO-GRU-based prediction method. The prediction results of multiple models are shown in Figure 2. It can be seen that the four forecasting methods show roughly the same trend of change. The maximum predicted value was achieved in sample 9.

Table 4: Evidence of cybersecurity posture based on BPA values.

Sample number	O_1	O_2	O_3	U
4	0.023	0.098	0.821	0.031
15	0.241	0.398	0.162	0.148
26	0.346	0.443	0.119	0.144
42	0.279	0.398	0.251	0.039
65	0.109	0.406	0.239	0.261
74	0.152	0.374	0.279	0.163
81	0.056	0.178	0.706	0.037
98	0.197	0.218	0.499	0.042
88	0.118	0.235	0.507	0.139
93	0.326	0.102	0.428	0.118

Table 5: Evaluation results of D-S, Elman, GA-Elman, and DS-GA-Elman.

Samples	D-S	Elman	GA-Elman	DS-GA-Elman	Desired security level
4	VI	IV	IV	V	IV
15	III	III	III	III	III
26	III	II	III	III	III
42	III	II	II	II	II
65	III	II	III	III	III
74	IV	III	II	III	III
81	IV	IV	IV	IV	IV
98	V	V	IV	IV	V
88	V	IV	V	V	V
93	V	IV	V	IV	IV

Table 6: Test results of various evaluation algorithms.

Algorithm category	Correct number	Correct rate (%)
D-S	7	70
Elman	7	60
GA-Elman	6	70
DS-GA-Elman	8	80

For predicting the prediction results of each model, the relative error formula $RE = |y_i - y_i'|/y_i' \times 100\%$ and the average absolute percentage error $\text{MAPE} = \sqrt{\sum_{i=1}^{n} ((y_i' - y_i)/y_i')^2}/n$ are used for error comparison and analysis, where y_i' represents the true situation value, y_i represents the output value of the model, n is the number of prediction samples, and the lesser the MAPE is, the more accurate the prediction results are. The average percentage error values are shown in Table 7. The RE value comparison is shown in Figure 3.

As can be seen from the prediction results in Figure 3, the GRU showed higher predictive values than the other two prediction models. In particular, at sample 5, it reaches a maximum of 102. The DEIPSO-GRU security situation prediction model is basically consistent with the trend of the real situation. Only a few predicted values differ from the actual situation values, and most of the predicted situation values and the situation change curves are consistent with the actual situation value change trends. However, some of the prediction results of the basic GRU prediction model and the

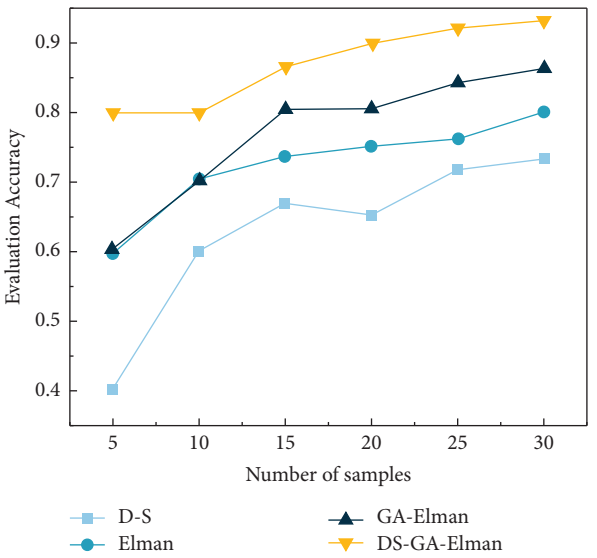


Figure 1: Comparison of accuracy of each model.

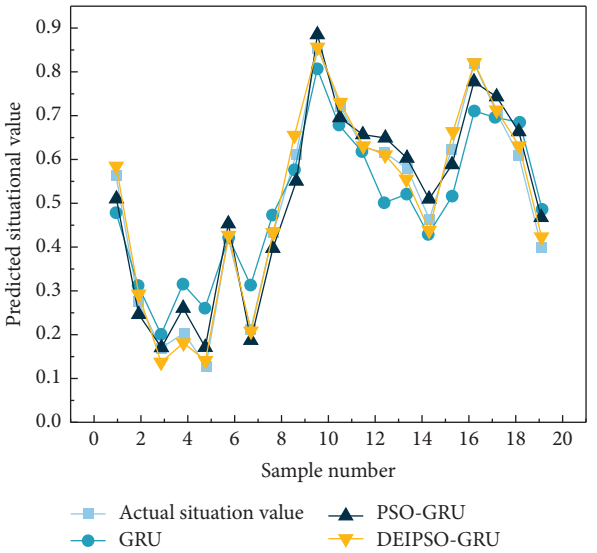


Figure 2: Multimodel prediction of situational value.

Table 7: MAPE value corresponding to GRU, PSP-GRU, and DEIPSO-GRU algorithm.

Algorithm category	MAPE value	Prediction correctness
GRU	0.042	Lower
PSP-GRU	0.023	Medium
DEIPSO-GRU	0.014	Higher

traditional PSO optimized GRU prediction model has a large deviation from the actual situation value. From the figure, there is a large error with the actual situation value. At the same time, in Figure 3, the error between the prediction results of the DEIPSO-GRU model and the actual situation value is relatively small, and it is relatively lower than the other two prediction models in most cases. Compared with the traditional GRU structure prediction, the accuracy of the GRU prediction model optimized by the PSO algorithm has

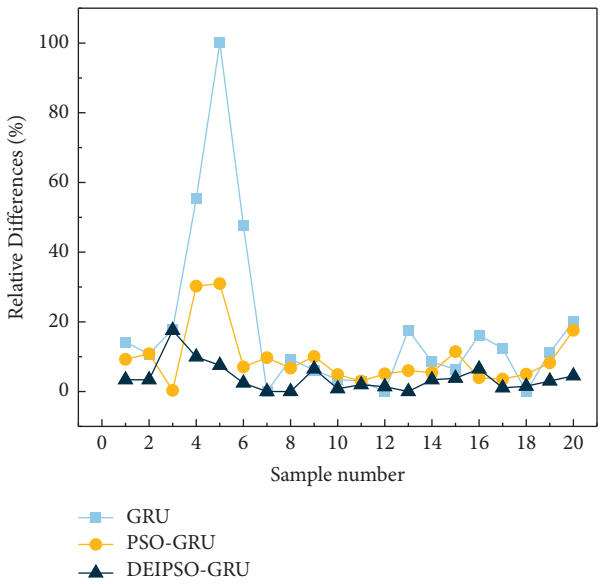


Figure 3: Relative error comparison between GRU, PSO-GRU, and DEIPSO-GRU.

been improved, but there are still many prediction deviations. The predicted MAPE values of this model are relatively low, but still fall short. However, the prediction model based on DEIPSO-GRU neural network has a relatively stable performance and the lowest MAPE value when making predictions. Therefore, the network security situation prediction model based on the DEIPSO-GRU structure is feasible and effective, and the prediction accuracy rate is improved.

4. Conclusion

This paper proposes a network security situation assessment model based on DS-GA-Elman neural network. The advantage of neural network is used to obtain the basic probability distribution function required in the synthesis rule of D-S evidence theory, and the disadvantage of the subjective dependence of traditional BPA is removed. The specific conclusions are as follows:

(1) The Elman neural network improved by the genetic algorithm is used to obtain BPA, and the optimization method of GA is beneficial to improve the accuracy of BPA. The D-S evidence theory is improved, and the Pearson correlation similarity and evidence distance theory are used to revise the BPA, and then the evidence is fused according to the fusion rules, which effectively prevents the occurrence of inaccurate assessment decisions caused by evidence conflicts.

(2) Propose a network security assessment model based on DS-GA-Elman. Through the analysis of the NSL-KDD dataset, the situation elements are extracted from the data according to the index system, the data are processed through quantitative formulas and normalization operations, and the dataset is divided, and the network security situation assessment model proposed in this paper is used to test the test. The final evaluation results have an accuracy rate of up to 80%.

(3) Compared with the evaluation results of other models, the results of the network security situation evaluation model are better. In addition, this paper proposes a prediction model based on DEIPSO-GRU to predict the network security situation, and the predicted results are basically consistent with the actual situation value. The results are compared with the prediction results of other models, and it is verified that the network security situation prediction model proposed in this paper has a higher prediction accuracy.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Wang, "Applying Internet information technology combined with deep learning to tourism collaborative recommendation system," *PLoS One*, vol. 15, no. 12, Article ID e0240656, 2020.

[2] J. F. Miao, Z. S. Wang, X. Ning, X. Nan, C. Weiwei, and L. Ruimin, "Practical and secure multifactor authentication protocol for autonomous vehicles in 5G," *Software: Practice and Experience*, 2022.

[3] D. Zhao and J. Liu, "Study on network security situation awareness based on particle swarm optimization algorithm," *Computers & Industrial Engineering*, vol. 125, pp. 764–775, 2018.

[4] J. Zhang, "Construction and exploration of virtual simulation experimental teaching platform for network security and computer technology," *Journal of Physics: Conference Series*, vol. 2173, no. 1, Article ID 012013, 2022.

[5] N. Hoque, M. H. Bhuyan, R. C. Baishya et al., "Network attacks," *Journal of Network and Computer Applications*, vol. 40, pp. 307–324, 2014.

[6] D. Nie and C. Xu, "Non-GAAP earnings quality in firms with data breach incident," *Asian Review of Accounting*, vol. 29, no. 3, pp. 383–398, 2021.

[7] L. I. Xian-Zong and S. S. School, "Research on computer network security defense technology," *Computer Knowledge and Technology*, 2015.

[8] H. Tianfield, "Cyber security situational awareness; proceedings of the IEEE international conference on Internet of things," in *Proceedings of the Farmaco*, Chengdu, China, December 2016.

[9] Y. Zahri, A. H. R. Susanty, and A. Mustaffa, "Cyber security situational awareness among students: a case study in Malaysia," *International Scholarly and Scientific Research & Innovation*, vol. 11, no. 7, 2017.

[10] W. U. Ya-Kun, H. X. Guo, X. M. Wang et al., "Research overview on big data technology," *Journal of Liaoning University (Philosophy and Social Science Edition)*, vol. 4, no. 63, pp. 3–76, 2015.

[11] C. Siaterlis, V. Maglaris, and G. Bella, "One step ahead to multisensor data fusion for DDoS detection," *Journal of Computer Security*, vol. 13, no. 5, pp. 779–806, 2005.
[12] V. Dutt, Y. S. Ahn, and C. Gonzalez, "Cyber situation awareness: modeling detection of cyber attacks with instance-based learning theory," *Human Factors*, vol. 55, no. 3, pp. 605–618, 2013.
[13] R. B. Almeida, V. Covalski, R. Machado et al., "A hierarchical architectural model for network security exploring situational awareness," *Farmaco*, in *proceedings of the 34th ACM/SIGAPP Symposium*, Limassol, Cyprus, April 2019.
[14] D. Zhang and Q. He, "Security situation awareness method for smart grid," *International Core Journal of Engineering*, vol. 6, no. 5, pp. 49–55, 2020.
[15] D. Liu, "Prediction of network security based on DS evidence theory," *ETRI Journal*, vol. 42, no. 5, pp. 799–804, 2020.
[16] H. U. Jianping, H. Hao, H. E. Qi et al., "Discussion on the construction of cyber security situational awareness platform in health care industry," *Chinese Journal of Health Informatics and Management*, 2019.
[17] C. Xie, J. Bai, Z. Wen, G. Lu, and H. Wang, "Lightning risk assessment of transmission lines based on D-S theory of evidence and entropy-weighted grey correlation analysis; proceedings of the Energy Internet & Energy System Integration," in *Proceedings of the Farmaco*, Beijing, China, November 2017.
[18] Q. Zhao, S. Wang, K. Wang, and B. Huang, "Multi-objective optimal allocation of distributed generations under uncertainty based on D-S evidence theory and affine arithmetic," *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 70–82, 2019.
[19] A. Ly, M. Marsman, and E. J. Wagenmakers, "Analytic posteriors for Pearson's correlation coefficient," *Statistica Neerlandica*, vol. 72, no. 1, pp. 4–13, 2018.
[20] Y. Pan and H. Yu, "Biomimetic hybrid feedback feedforward neural-network learning control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 6, pp. 1481–1487, 2017.
[21] J. Hu, J. Zhang, C. Zhang, and J. Wang, "A new deep neural network based on a stack of single-hidden-layer feedforward neural networks with randomly fixed hidden neurons," *Neurocomputing*, vol. 171, no. C, pp. 63–72, 2016.
[22] K. Gregor, I. Danihelka, A. Graves et al., "DRAW: a recurrent neural network for image generation," *Computer Science*, pp. 1462–1471, 2015.
[23] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.
[24] G. Ren, Y. Cao, S. Wen, T. Huang, and Z. Zeng, "A modified elman neural network with a new learning rate scheme," *Neurocomputing*, vol. 286, pp. 11–18, 2018.
[25] W. Z. Sun and J. S. Wang, "Elman neural network soft-sensor model of conversion velocity in polymerization process optimized by chaos whale optimization algorithm," *IEEE Access*, vol. 5, pp. 13062–13076, 2017.
[26] W. J. Shan, "Digital streaming media distribution and transmission process optimisation based on adaptive recurrent neural network," *Connection Science*, vol. 34, no. 1, pp. 1169–1180, 2022.
[27] W. Jia, D. Zhao, Y. Zheng, and S. Hou, "A novel optimized GA–Elman neural network algorithm," *Neural Computing & Applications*, vol. 31, no. 2, pp. 449–459, 2019.
[28] X. Ning, P. Duan, W. Li, and S Zhang, "Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer," *IEEE Signal Processing Letters*, vol. 27, pp. 1944–1948, 2020.
[29] A. Ledwoch, H. Yasarcan, and A. Brintrup, "The moderating impact of supply network topology on the effectiveness of risk management," *International Journal of Production Economics*, vol. 197, pp. 13–26, 2018.
[30] A. Kumarshrivas and A. Kumar Dewangan, "An ensemble model for classification of attacks with feature selection based on KDD99 and NSL-KDD data set," *International Journal of Computer Application*, vol. 99, no. 15, pp. 8–13, 2014.
[31] C. Wang, X. Wang, J. Zhang et al., "Uncertainty estimation for stereo matching based on evidential deep learning," *Pattern Recognition*, vol. 124, Article ID 108498, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9803857, 1 page
https://doi.org/10.1155/2023/9803857

Retraction

Retracted: Deep Learning of Subject Context in Ideological and Political Class Based on Recursive Neural Network

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] T. Jiang and X. Gao, "Deep Learning of Subject Context in Ideological and Political Class Based on Recursive Neural Network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8437548, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8437548, 8 pages
https://doi.org/10.1155/2022/8437548

Research Article

Deep Learning of Subject Context in Ideological and Political Class Based on Recursive Neural Network

Tingting Jiang[1] **and Xiang Gao**[2]

[1]*School of Marxism, Guangxi University of Chinese Medicine, Nanning 530200, China*
[2]*School of Public Health and Management, Guangxi University of Chinese Medicine, Nanning 530200, China*

Correspondence should be addressed to Xiang Gao; gx.ability@gmail.com

Received 24 June 2022; Revised 8 August 2022; Accepted 29 August 2022; Published 30 September 2022

Academic Editor: Wenming Cao

Ideological and political education is the most important way to cultivate students' humanistic qualities, which can directly determine the development of other qualities. However, at present, the direction of ideological and political innovation in higher vocational colleges is vague. In response to this problem, this study proposes a model based on HS-EEMD-RNN. First, the ensemble empirical mode decomposition (EEMD) method is used to decompose the measured values, and then the recurrent neural network (RNN) is used to train each component and the remaining items. Finally, through the mapping relationship obtained by the model, the response prediction value of each component and the remaining items can be obtained. In the RNN training process, the harmony search (HS) algorithm is introduced to optimize it, and the noise is systematically denoised. Perturbation is used to obtain the optimal solution, thereby optimizing the weight and threshold of the RNN and improving the robustness of the model. The study found that, compared with EEMD-RNN, HS-EEMD-RNN has a better effect, because HS can effectively improve the training and fitting accuracy. The fitting accuracy of the HS-EEMD-RNN model after HS optimization is 0.9918. From this conclusion, the fitting accuracy of the HS-EEMD-RNN model is significantly higher than that of the EEMD-RNN model. In addition, four factors, career development, curriculum construction, community activities, and government support, have obvious influences on ideological and political classrooms in technical colleges. The use of recurrent neural networks in the research direction of deep and innovative research on the subject context of ideological and political classrooms can significantly improve the prediction accuracy of its development direction.

1. Introduction

At present, the most fundamental task is to improve the overall quality of students, which is also the goal of social development. To cultivate socialist successors, schools, society, and families must cooperate closely to carry out student education from the basics and improve students' comprehensive quality. In education, ideological and political courses can play an extremely important role. Therefore, colleges and universities should take responsibility and regard cultivating students' correct thinking as one of the most important educational tasks [1]. The education of ideological and political courses is the most important way to cultivate students' humanistic qualities. Ideological and political education is the most important way to cultivate students' humanistic qualities. Quality in the broad sense mainly refers to material and spirit and is a quality both inside and outside, which contains extremely rich contents and can be divided into political and other aspects of quality [2]. Ideological quality plays a decisive role, so the most fundamental goal of school education is to cultivate talents with all-around development of morality, intelligence, physique, aesthetics, and labor and must have certain consciousness and scientific and technological knowledge. Ideological and political quality can directly determine the development of other qualities. No matter what kind of career students will engage in in the future, they must establish the correct three views. Only students with good ideology and morality can guide the correct development of other qualities and provide a steady stream of spiritual power

[3]. Therefore, innovative research in the ideological and political classroom is indispensable.

Foreign countries have civic education to carry out ideological domination of citizens and the corresponding expression of civic education discourse [4]. After the twentieth century, ideological and political education gradually formed two schools of universities based on historical research, namely, structuralism and functionalism [5]. Saussure founded thestructuralism school of ideological and political thinking, definingideological and political thinking as a science based on symbols and meanings.They advocated the study of the framework outline, practical operation, andsignificance of ideological and political thinking in daily life[6]. The functionalist ideological and political school advocates that more attention should be paid to the connotation of language itself, rather than to the external form and logical structure of ideological and political thinking. The ideological and political system can choose semantics independently, and the connotation of language is also transmitted [7]. Since the middle of the twentieth century, western discourse studies have gradually expanded from ideological and political studies to several related disciplines, such as philosophy, sociology, history, and literature. For example, Wittgenstein, a British philosopher, studied the relationship between ideology and politics and the external world, while Derrida, a French philosopher, believed that, to understand the specific practice of society, it is necessary to analyze ideology and politics [8]. After the 1960s, preliminary discussions on ideological and political studies began to appear in the field of sociology [9]. By comparing the studies of sociology by famous scholars, we can see their efforts in ideological and political studies, such as those of Habermas and Foucault. Habermas elaborated on the four principles of ideological and political discourse, which laid a foundation for the development of ideological and political studies. Foucault believed that discourse "covers various forms of formal or informal verbal interaction and written texts." "Discourse is power, and people give themselves to power through discourse" [10]. To solve the problem of the nonstationarity of signals, Huang et al. [11] proposed the empirical mode decomposition (EMD) method in 1998. However, the signal envelope of this algorithm is prone to shape distortion and causes an endpoint effect, which makes the decomposition of each component inaccurate. Some researchers [12] proposed EEMD, which can effectively solve the shortcomings of the above EMD algorithm. EEMD algorithm has been widely used. For example, some researchers used EEMD-RBF neural network to build a model of ideological and political direction [13]. Some scholars realized the extraction of key factors of ideological and political education through the EEMD algorithm [14]. Based on EEMD-JADE, some researchers have better isolated the environmental factors [15]. To improve training fitting and prediction accuracy, the recurrent neural networks (RNN) algorithm is introduced. RNN is a kind of structure-optimized neural network, which introduces a cyclic structure in the hidden layer and carries out internal connections among the hidden elements. During training, information can be propagated forward or backward in the model, which can improve the training efficiency and accuracy. At present, there are a lot of research achievements on the application of this algorithm. For example, researchers predicted the thematic context of ideological and political education through RNN and its fusion method [16]. Some researchers constructed a combination algorithm of CNN and RNN to explore continuous identity authentication in free text keystroke mode [17]. Some researchers also studied the necessity [18]. In the subject innovation of ideological and political class, the predicted value contains the prediction error and system noise, in order to further improve the precision of training fitting and the robustness of the algorithm.

In conclusion, it is difficult for a single algorithm model to solve complex problems, while, for the innovation of ideological and political classroom in higher vocational colleges affected by complex conditions, which is random and nonlinear, it is more difficult to achieve high precision prediction [19, 20]. Therefore, to overcome the shortcomings of the single algorithm, such as easily falling into the local minimum of the error function, slow convergence speed, and excessive convergence, this paper first uses the EEMD algorithm to smooth the displacement data. Then, using the strong nonlinear mapping ability of the RNN neural network, the harmony search algorithm was used to denoise and optimize it, and the advantages of each algorithm were integrated to improve the displacement prediction accuracy and model calculation efficiency, and the HS-EEMD-RNN ideological and political classroom prediction model was constructed. Recurrent neural networks have significantly improved the prediction accuracy of innovative directions in ideological and political classrooms and improved the use of computer network algorithms for practical applications and other such research.

2. Principles and Methods

2.1. Ensemble Empirical Mode Decomposition (EEMD). The EEMD algorithm adds enough normal distributed white noise into the original time series [21] and meanwhile performs EMD decomposition on the new time series. Then, the arithmetic average of each decomposition quantity is taking advantage of the property that the mean value of white noise is zero, and the IMF component and the remaining term R of EEMD decomposition can be obtained [22]. The decomposition steps are as follows:

(1) Considering that there are some problems in ideological and political prediction, the prediction is not accurate. In this study, we first add equal-length Gaussian white noise N times to the original time series to obtain several new time series:

$$y_i(t) = y(t) + n_i(t), \tag{1}$$

where $y_i(t)$ is the time series after adding white noise for the i-th time, $y(t)$ is the original time series, and $n_i(t)$ is normally distributed white noise signal.

(2) Then, the EMD algorithm is used to decompose the new time series with white noise, and the eigenmodal function and a residual term can be obtained:

$$y_i(t) = \sum_{i=1}^{J} w_{ij}(t) + r_i(t), \tag{2}$$

where $w_{ij}(t)$ is the j-th eigenmode function component obtained after adding white noise for the i-th time and $r_i(t)$ is the residual term.

(3) Then N groups of intrinsic components and residual term $r(t)$ are averaged to obtain the final IMF component and residual quantity $R(t)$.

$$\begin{aligned} y_i(t) &= \sum_{i=1}^{J} w_{ij}(t) + r_i(t), \\ R(t) &= \frac{1}{N}\sum_{i=1}^{N} r_i(t), \end{aligned} \tag{3}$$

where $IMF_j(t)$ denotes the final decomposition component of EEMD and $R(t)$ is the final surplus.

(4) Finally, the original time series signal can be decomposed into

$$y(t) = \sum_{j=1}^{J} IMF_j(t) + R(t). \tag{4}$$

2.2. RNN Model. The advantage of RNN lies in the one-time activation of different neuralunits, that is, the output content of hidden layer nodes at the previous timecan not only reflect the time, but also be used as the input of the next outputlayer. At the same time, after processing the information at this moment, itcan lead to the hidden layer node at the next time point [23]. Compared with traditional neural networks, RNN can improve the convergence speed and training accuracy in the process of training fitting and predicting displacement [23].

RNN is a recursive process in the time dimension. The results of the previous time have a direct impact on the training for the next time. However, because the weights at each time step are the same, different units at the same time will share the same group weights [24]. For a sequence x of length T, the size of the input layer of the RNN is I, the size of the hidden layer is H, the size of the output layer is K, and the dimensions of the three weight matrices U, V, and W can be obtained:

$$\begin{cases} U \in R^{l\times H} \\ W \in R^{H\times H}, \\ V \in R^{H\times K} \end{cases} \tag{5}$$

where x^t is the input of the t-th neuron in the sequence, a^t is the input of the t-th hidden layer, and b^t is the nonlinear activation of a^t, that is, the output of the neural network. Therefore, a^t is determined by the input layer and the output b^{t-1}neuron of the previously hidden layer:

$$\begin{aligned} a_h^t &= \sum_i w_{ih} x_i^t + \sum_{h'} w_{h'h} b_{h'}^{t-1}, \\ b_h^t &= f\left(a_h^t\right). \end{aligned} \tag{6}$$

The innovation sequence of the ideological and political classroom starts from $t = 1$, so $b_0 = 0$, and then the information of the hidden layer is transmitted to the output layer, and the output result of the output layer is

$$\begin{aligned} a_k^t &= \sum_h w_{hk} b_{h,}^t \\ y_k^t &= \frac{e^{a_k^t}}{\sum_j e^{a_j^t}}. \end{aligned} \tag{7}$$

At time t, the residual of the ordinary neural network is $\delta_k^t = y_k^t - z_k^t$, Since the hidden layer of RNN needs to accept the signal of the hiddenlayer at the previous time in forward conduction and the feedback of the hiddenlayer at the next time in reverse conduction, the residual term of RNN outputlayer is as follows:

$$\delta_h^t = f'\left(a_h^t\right)\left(\sum_k \delta_k^t w_{hk} + \sum_{h'} \delta_{h'}^{t+1} w_{hh'}\right). \tag{8}$$

When the length of the displacement sequence is T, the residual δ^{T+1} is all 0, and the entire network has only one set of parameters U, V, and W, then the reciprocal of time t:

$$\begin{aligned} U\colon \frac{\partial O}{\partial w_{ih}} &= \frac{\partial O}{\partial a_h^t}\frac{\partial a_h^t}{\partial w_{ih}} = \delta_h^t x_i^t, \\ V\colon \frac{\partial O}{\partial w_{hk}} &= \frac{\partial O}{\partial a_k^t}\frac{\partial a_k^t}{\partial w_{hk}} = \delta_k^t b_h^t, \\ W\colon \frac{\partial \theta}{\partial w_{h'h}} &= \frac{\partial \theta}{\partial a_h^t}\frac{\partial a_h^t}{\partial w_{h'h}} = \delta_k^t b_{h'}^t. \end{aligned} \tag{9}$$

Then it can be written in the unified form (assuming $x_i^t = a_i^t = b_i^t$ for the input layer):

$$\frac{\partial O}{\partial w_{hij}} = \frac{\partial O}{\partial a_h^t}\frac{\partial a_j^t}{\partial w_{ij}} = \delta_j^t b_i^t. \tag{10}$$

Finally, for time $t = 1, 2, \ldots, T$, by summing the recursion of the RNN neural network, the derivative of the RNN network concerning the weight parameter can be obtained:

$$\frac{\partial O}{\partial w_{ij}} = \frac{\partial O}{\partial a_h^t}\frac{\partial a_j^t}{\partial w_{ij}} = \sum_t \delta_j^t b_i^t. \tag{11}$$

2.3. Harmony Search Algorithm. For the displacement components obtained by EEMD decomposition, the high-frequency components are mainly caused by monitoring error noise, especially the IMF_1 component, so it is necessary

to systematically denoise them [25]. The basic idea of the harmony search algorithm is a metaheuristic optimization algorithm that imitates the harmony tuning process played by a band. The implementation process of optimizing the RNN neural network is as follows [26]:

Step 1: Initialize the relevant algorithm parameters.

HMS determines the global search capability of the HS algorithm, but an excessively large HMS value will affect the speed of convergence to the optimal solution. A large HMCR value is beneficial to the local search of the algorithm. The specific values of algorithm parameters can be determined by simulation experiments [27].

Step 2: Define the objective function and initialize the harmony memory.

Define the objective function: the average relative error F between the expected output (measured value) and the network output (fitted value), where the smaller the F value, the better the degree of optimization.

$$F = \frac{1}{H}\sum_{i=1}^{H}\frac{\left|y'(i) - y(i)\right|}{\left|y(i)\right|}, \tag{12}$$

where H is the total number; $y'(i)$ and y(i) are the measured and fitted values of the i-th sample, respectively.

Construct HMS random initial harmonics $x^1, x^2, \ldots, x^{HMS}$ to add a harmonic memory bank and the harmony x_i in the memory; the bank corresponds to a set of weights and thresholds of the RNN neural network. The harmony memory bank is

$$\begin{aligned} HM &= \begin{bmatrix} x^1 & f\left(x^1\right) \\ x^2 & f\left(x^2\right) \\ \vdots & \vdots \\ x^{HMS} & f\left(x^{HMS}\right) \end{bmatrix} \\ &= \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_N^1 & f\left(x^1\right) \\ x_1^2 & x_2^2 & \cdots & x_N^2 & f\left(x^2\right) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^{HMS} & x_N^{HMS} & \cdots & x_N^{HMS} & f\left(x^{HMS}\right) \end{bmatrix}, \end{aligned} \tag{13}$$

where x_j^i is the j-th weight or threshold of the i-th group of harmony in the RNN network and $f(x^i)$ denotes the i-th group of objective function values, that is, the average relative error between the i-th group of network outputs and the expected output.

Step 3: Create new harmonies.

To create new harmony is to construct new solution vectors. According to the harmony learning memory bank, tune tuning, randomly generate the tone $y'(j = 1, 2, \ldots, N)$ of the new harmony:

$$x_j' = \begin{cases} x_j' \in \left(x_j^1, x_j^2, \cdots, x_j^{HMS}\right) & r_1 < HMCR \\ x^L + r_1\left(x^U - X^L\right) & r_1 \geqslant HMCR \end{cases}, \tag{14}$$

where r_1 represents random value within [0, 1]. $x_j' \epsilon (x_j^1, x_j^2, \ldots, x_j^{HMS})$ is the randomly selected component in the j-th column HMS; Z represents the upper and lower limits of the JTH column. x^U and x^L are the upper and lower limits of the value of the j-th column components. Equation (14) shows that, for any new tone, it will be selected from the memory library with the probability of HMCR, and the value will be randomly selected from its corresponding value interval with the probability of 1-HMCR [23]. If x_j' is the solution component in the harmony memory, it is fine-tuned as

$$x_j' = \begin{cases} x_j' + r_3 bw & r_2 < PAR \\ x_j' & r_2 \geqslant PAR \end{cases}, \tag{15}$$

where r_1 and r_2 are distributed random values in [0, 1] and [−1, 1], respectively.

Step 4: Update the memory bank.

Step 5: End the loop.

Repeat steps 3 and 4 until the termination criterion is met or the maximum number of iterations is reached.

2.4. Modeling Process. Based on the advantages of the above algorithms, the specific modeling process of the EEMD-RNN combination model based on HS optimization is as follows:

(1) Determine the research measurement points of the influencing factors of the ideological and political classroom theme, decompose the measurement values through EEMD, and obtain several groups of IMF components and residual items R.

(2) The RNN neural network algorithm is used to train and fit the IMF component and the residual item R. The input is the environmental quantity and the time-dependent variable, and the output is the key influencing factor. During the training process, the thresholds and weights of the RNN are optimized through the harmony search (HS) algorithm until the optimal solution is obtained.

(3) Calculate the IMF component and residual item R corresponding to the prediction set by training the optimal fitting mapping relationship.

(4) Equal weight summation of the predicted values of each component displacement IMF and the remaining term R is the final prediction result.

2.5. Model Accuracy Assessment. To evaluate prediction accuracy, MAE, RMSE, and MAPE are applied to evaluate the accuracy [28], which are expressed as

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right|,$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2}, \quad (16)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right|,$$

where Y_t is the measured value of the influencing factors of the ideological and political classroom, $\hat{Y}_t$ is the calculation amount of the model, n is the total amount of monitoring data, and t is the time corresponding to the monitoring data.

3. Experimental Procedure and Result Analysis

3.1. Data and Factor Selection. The scientific advancement of ideological and political course teaching activities in technical colleges requires the improvement of course design as support. It is necessary to focus on teaching design and actively explore the construction of a scientific teaching planning system to ensure that ideological and political courses can be displayed to the greatest extent. The advantages of teaching design have significantly improved the quality of talent training in technical schools. This study mainly selects relevant classrooms and students and teachers as the research objects, selects the experimental samples as the actual values of the experiment, and selects the influencing factors as career development, curriculum construction, community activities, government support, family education, teacher-student environment, and other factors. Seven schools in a city were selected as the experimental data source.

3.2. Test Analysis. There are two EEMD decomposition parameters: the standard deviation *Nstd* of white Gaussian noise is generally 0.01~0.4, and the number of noise additions NE is generally 50 or 100. To obtain the best decomposition result, the comprehensive *Nstd* is 0.01, and the NE is 100. The EEMD decomposition results of the training set of the model are shown in Figure 1, in which 7 groups of IMF components and 1 group of residual *R*, IMF_1, IMF_2, and IMF_3 components are high-frequency components, and the remaining IMF components are low-frequency components. For high-frequency components, it is mainly caused by monitoring error noise, especially the IMF1 component. Therefore, systematic denoising is necessary to improve model accuracy.

To reduce the interference of noise errors in the measured values to the training results, in the process of optimizing the RNN training process of the harmony search algorithm, the size of the harmony memory bank *HMS* is 100, and the search range is [−1, 1]. The retention probability HMCR of the harmony memory is 0.85, the pitch fine-tuning probability PAR is 0.2, and the fine-tuning step size *bw* is 0.2. After the training of the HS optimized RNN, the

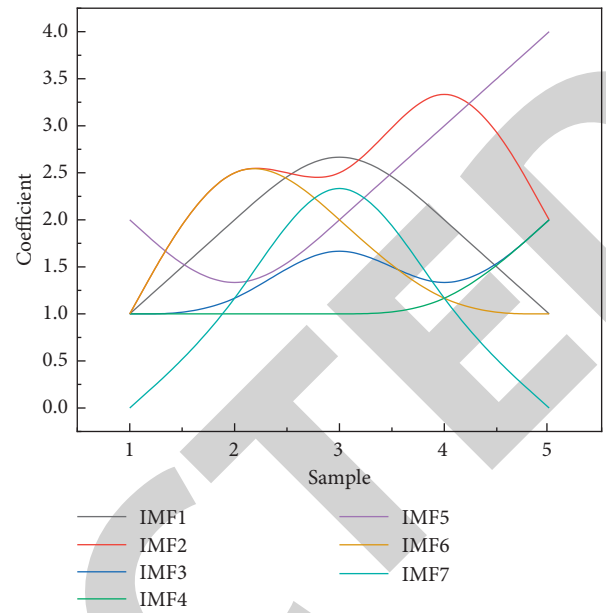


Figure 1: EEMD decomposition results.

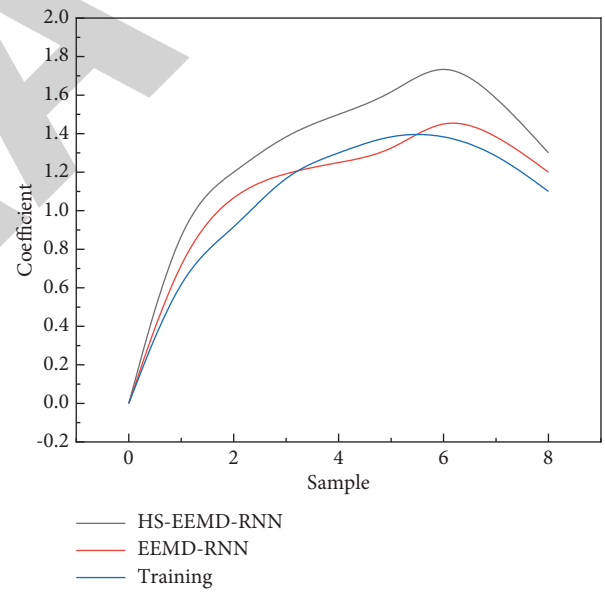


Figure 2: Fitting results between two models and actual values.

Table 1: Comparison of fitting accuracy of the two models.

Models	R	MAE	RMSE	MAPE
EEMD-RNN	0.9865	0.197	0.207	0.107
HS-EEMD-RNN	0.9918	0.074	0.086	0.041

fitting values of each component *IMF* and the residual R are equally weighted to obtain the fitting value of the displacement. The fitting result is shown in Figure 2.

At the same time, compared with the fitting results of the EEMD-RNN model, the fitting accuracy is demonstrated in Table 1. The EEMD-RNN model alone has a fitting accuracy

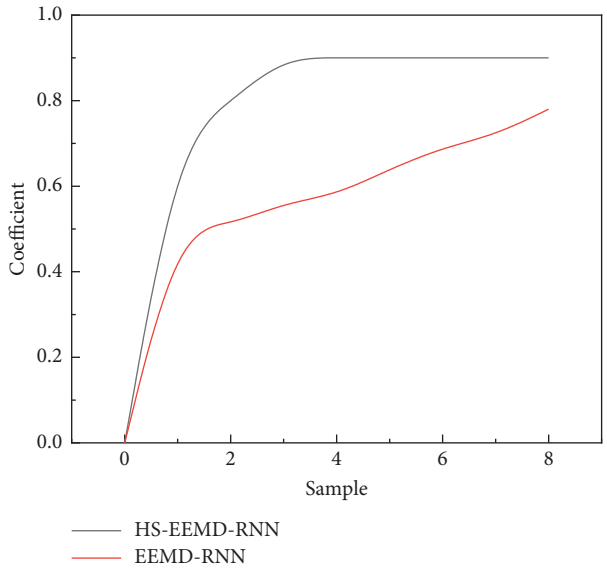


Figure 3: Comparison of prediction accuracy between EEMD-RNN model and HS-EEMD-RNN model.

Table 2: Comparison of prediction error between EEMD-RNN model and HS-EEMD-RNN mode.

Models	MAE	RMSE	MAPE
EEMD-RNN	0.221	0.250	0.123
HS-EEMD-RNN	0.088	0.101	0.048

of 0.9865, while the HS-EEMD-RNN model optimized by the HA algorithm has a fitting accuracy of 0.9918, and the fitting accuracy of the HS-EEMD-RNN is significantly higher.

The network trained by HS-EEMD-RNN is used to predict the displacement IMF of each component and the remaining term R, and the equal weight summation is the final dam displacement prediction result. At the same time, the displacement prediction value of EEMD-RNN is calculated, and the displacement prediction result is shown in Figure 3. Using MAE and RMSE as well as MAPE, three indicators are employed to evaluate the predictions of the two models, which are demonstrated in Table 2. The comparison results reveal that the HS-EEMD-RNN is better than the EEMD-RNN, which can meet the prediction requirements of the ideological and political classroom.

Among them, the sudden jump point in the change process is generally the focus of ideological and political classroom prediction, and the overfitting of the previous model and the large prediction error occurred more at the sudden jump point, so the EX5 displacement process line was selected to have the largest jump degree. The measured values of the seven factors are compared with the predicted values of the model to better evaluate the prediction accuracy of the model. It can be observed from Table 3 that the average absolute percentage error of HS-EEMD-RNN is 8.91%, which is significantly lower than 27.17% of EEMD-RNN.

Table 3: Comparison of predicted values.

Samples	The predicted value affects the ratio		RMSE	
	HS-EEMD-RNN	EEMD-RNN	HS-EEMD-RNN	EEMD-RNN
1	0.4	0.2	0.34	4.10
2	0.7	0.6	8.20	23.82
3	0.6	0.4	1.87	6.22
4	0.6	0.5	3.76	8.17
5	0.6	0.4	1.12	15.36
6	0.6	0.3	3.63	8.79
7	0.4	0.2	16.71	42.22
Ava	—	—	8.91	27.17

The error analysis between the predicted values obtained from the proposed HS-EEMD-RNN model and the actual value predictions shows that four factors, career development, curriculum building, club activities, and government support, have a more significant impact on the civic science classroom in technical colleges. Therefore, innovations are made based on such factors, as shown in Section 3.3.

3.3. Innovation Direction Prediction Results. In the basis of clear teaching design, to improve thecomprehensive effect of ideological and political course teaching activities intechnical schools, teachers should pay attention to choosing different entry points to reform and innovate theteaching mode and teaching a path and play therole of ideological and political teaching. In addition, students should begiven active guidance from multiple perspectives, so that they can activelyparticipate in the study and practice of ideological and political knowledge,so as to strengthen their professional quality and effectively ensure thatideological and political teaching activities achieve remarkable results.

(1) Based on career development, strengthen professional ethics education

Good career development is a key concern of technical school students. In the process of carrying out ideological and political education forstudents, teachers should take students' professional growth and careerdevelopment as the basis, effectively carry out ideological and politicaleducation for students' vocational ability and ensure that students'professional ability and professional level are significantly improved. At thesame time, it is necessary to strengthen the professional ethics of students toensure that the teaching effect and the comprehensive quality of personneltraining work can be maximally promoted.

(2) Build quality courses and guide students to explore

There is a close connection between the efficiency of technical schools and the innovation of curriculum in the process of implementing ideological education and guidance to students. Combined with the practical needs of talent training in technical schools, improve the construction of ideological and political

courses to ensure that students can implement active and effective teaching guidance, so that the level of ideological awareness of students in technical schools can be continuously improved, and the training of talents in technical schools can be improved. Teachers of ideological and political courses in technical schools can systematically plan the needs of students' career development. Then, according to the personal growth needs and career development needs of students, explore the construction of basic curriculum systems such as employment and entrepreneurship guidance courses, legal basic knowledge courses, mental health education courses, current affairs, and politics courses, and create excellent ideological and political courses to improve education effect. At the same time, based on the construction of quality courses, the school should combine the application of information technology to build a platform, and political education can be provided from multiple perspectives. The development of class teaching activities provides rich talent support; effective guarantees are further optimized, as well as systematic advancement of talent training.

(3) Carefully plan community activities and build an experience platform

In the practice of talent training in technical colleges, it is necessary not only to construct excellent courses from the perspective of theoretical education and guide students to systematically explore ideological and political content but also to consciously build experiential learning for students. The platform guides students to conduct a diversified exploration of the content of ideological and political courses. Positive educational guidance can be implemented for students. The goal becomes achieving all-around development and training students to become political education talents widely recognized by society.

(4) Strengthen government support and build an ideological education platform

In the process of reforming and innovating, the leadership of schools and the support of enterprises are important, so that the students of technical colleges can enjoy themselves. Professional ability, moral cultivation, and so forth have been comprehensively improved to optimize the talent cultivation work of technical schools as a whole. The development of effective teaching activities needs to be supported by a sound policy system. Only local governments increase their support for technical schools and combine the needs of technical talents in the region to build a comprehensive and flexible system, adapt to the principles and policies of the talent development direction of technical schools, and further optimize the job training work and ideological and moral education work of technical school students to create a positive and healthy talent training environment in the educational practice of technical schools. Gradually improve and then optimize the overall efficiency of the construction of high-quality technical personnel in technical colleges.

4. Conclusion

To better determine the influencing factors of the theme of ideological and political classrooms, this paper uses environmental variables and time-sensitive variables as input, and the output is the key influencing factor, which is used to build the HS-EEMD-RNN model, so it can efficiently determine the research points of ideological and political classrooms. Through experimental analysis, the model based on HS-EEMD-RNN achieves better prediction accuracy than that based on EEMD-RNN. The experimental results show that some influencing factors selected in this paper, including career development, curriculum construction, community activities, government support, family education, and teacher-student environment, have played a key role in the ideological and political classroom. Good prediction accuracy can clarify teaching design and improve ideological and political classroom teaching activities in technical schools. At the same time, students can more actively participate in the learning and practice of ideological and political knowledge, strengthen their professional quality, and effectively ensure that ideological and political teaching activities can achieve significant results.

Based on the HS-EEMD-RNN model, this study analyzes the influencing factors of ideological and political education in higher vocational colleges and predicts the innovation direction of ideological and political classrooms. The study finds that HS-EEMD-RNN has better effects than EEMD-RNN, and the specific conclusions are as follows:

(1) The HS can elevate and prove the training and fitting accuracy. The fitting accuracy of the EEMD-RNN model alone is 0.9865, while the fitting accuracy of the HS-EEMD-RNN model optimized by the HS algorithm is 0.9918. The fitting accuracy is significantly higher.

(2) The average absolute percentage error of the HS-EEMD-RNN model is 8.91%, which is significantly lower than 27.17% of the EEMD-RNN model. The prediction accuracy of the HS-EEMD-RNN is better than that of the EEMD-RNN, meeting the forecasting requirements of ideological and political classrooms.

(3) The influence of career development, curriculum construction, community activities, and government support on ideological and political classes in technical colleges is obvious; that is, the innovation direction should be based on career development, building excellent courses, elaborately planning community activities, and strengthening government support.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5490779, 10 pages
https://doi.org/10.1155/2022/5490779

Research Article

The Application of Computer Intelligence in the Cyber-Physical Business System Integration in Network Security

Shi Lin,[1,2,3] **Ma Yang**,[1] **Yan Lu**,[1] **and Liquan Chen**[1]

[1]*School of Cyber Science and Engineering, Southeast University, Nanjing 210096, Jiangsu, China*
[2]*College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, Jiangsu, China*
[3]*Jiangsu Financial Information Management Center, Nanjing 210024, Jiangsu, China*

Correspondence should be addressed to Shi Lin; shiling717@nuaa.edu.cn

Received 16 June 2022; Revised 19 July 2022; Accepted 1 August 2022; Published 26 September 2022

Academic Editor: Wenming CAO

In order to address the false alarm detection problem caused by the inability to identify the transgression scene pages in the process of horizontal transgression detection, this study proposes a deep learning-based LSTM-AutoEncoder unsupervised prediction model. The model uses long short-term memory network to build AutoEncoder, extracts text features of page response data of horizontal transgression scenario, and reconstructs text features to restore. Meanwhile, it counts the error between the restored result and the original page response, judges whether the detection result of horizontal transgression is false alarm according to the error threshold of unknown page, and tests the effectiveness of the model effect under real business data by comparing it with other two algorithms, one-class SVM and AutoEncoder, which provides security for enterprise network business. The results show that the LSTM-AutoEncoder model achieves a more balanced index in terms of accuracy, precision, recall, and $F1$-score in the case of MAE, which is 0.3% more and 0.2% more than the case of MSE in terms of recall and accuracy. It is concluded that the LSTM-AutoEncoder model is more in line with the real business requirements, and the simple model architecture selected for this study can reduce the complexity of the model, speed up the prediction time of the model in the application phase, and improve the performance of the detection software. This indicates that this study has some application prospects in network security.

1. Introduction

Web security issues are closely related to our lives, among which transgression vulnerabilities are the most common. Transgression vulnerability is one of the business logic vulnerabilities, which is very common in current web applications [1, 2]. Horizontal override is one of the override loopholes, which means that users with the same authority can query, modify, delete, and add other people's information by modifying parameter variables to achieve illegal behavior [3, 4]. When a hacker exploits an override vulnerability, it is difficult to be monitored and processed by defense software because the attack behavior is no difference from normal user requests and does not contain sensitive characters and special characters [5]. In recent years, with the penetration of information technology into all aspects of people's lives, information security has begun to be frequently mentioned and concerned, especially in the modern society with highly developed computer technology and network technology. The development of all walks of life is closely related to information security, so issues involved in it must be paid enough attention to [6]. The problem of leaks caused by computer network security issues has also given people a warning. Network security problems can easily lead to personal information leakage and damage to reputation. To avoid such problems, people have put forward higher requirements for the use of computers and began to pay close attention to computer use and network security, so as to ensure the function of computer services [7]. Therefore, it is virtually important to detect the horizontal override vulnerabilities and repair the defects in time to avoid losses.

A lot of research work has been carried out [8–12]. In the actual business production process of an enterprise, it is not

only necessary to enhance the security development awareness and code specifications of code developers, but also necessary to have good security detection for the upcoming web applications. For the detection of privilege escalation vulnerabilities, there are the following related studies. Sun et al. proposed a method of using static inspection to detect logical vulnerabilities in web applications and constructed sitemaps of visitors with different privileges through source code to determine whether there is unauthorized behavior [13]. Monshizadeh et al. proposed the MACE tool to detect privilege escalation vulnerabilities in large code bases, which found many serious and unknown vulnerabilities and achieved good results. Deepa et al. adopted a black-box method to identify logic flaws and vulnerabilities and constructed finite state machines by modeling expected behaviors [14]. Song et al. obtained the accessible links of the target website by means of URL link crawling, used different user permissions to access the linked pages, and parsed the web page responses to determine whether there are horizontal and vertical unauthorized vulnerabilities [15]. Ma et al. proposed a new permission control vulnerability detection method, that is, by establishing a five-layer model combined with authorization context information to detect whether the target system has permission control vulnerabilities [16]. Although there are many methods for unauthorized detection, the following problems exist in real business scenarios. First, the method of source code auditing is extremely expensive and requires relatively strong security knowledge personnel to spend a lot of time on detection. At the same time, due to the short iteration cycle and the huge amount of product code, it is difficult for security experts to take into account all the vulnerability points in web applications [17]. Secondly, the current automation level over-authority detection tools and methods often have the problem of false positives, which increases the cost of the verification of the detection results [18]. Although there have been many combined applications of artificial intelligence and network security [19–21], and the recognition of horizontal unauthorized scene pages has been realized, it is prone to the problem of false positives.

To address this problem, this study proposes a horizontal transgression detection model based on LSTM-AutoEncoder, which applies deep learning transgression scenario page recognition to transgression detection. And at the same time, this paper well solves the problem of false positives of horizontal transgression detection, ensures timely detection and repair of horizontal transgression vulnerabilities, as well as significantly improves the accuracy of horizontal transgression detection and reduces the work cost of manual review.

2. Horizontal Override Detection Based on LSTM-AutoEncoder

2.1. LSTM. The long short-term memory (LSTM) network was proposed by Hochreiter et al. [22, 23], using the control gate mechanism. The LSTM structure is shown in Figure 1, which is composed of multiple units. At present, the long short-term memory network has been applied in many fields. When this algorithm is used to process data, the data are generally divided into training values and observed values, the training values are used for network training, and the observed values are compared with the predicted values. For data with simple variation laws, this method has higher prediction accuracy [24].

In the LSTM structure diagram, h_t is the state output of the network at time t, and X_t is the input vector at time t. The state flow formula of the LSTM model is as follows.

The first is the forgetting stage, which will select historical information to be forgotten, f_t is the output state of the forget gate at time t, σ is the logistic sigmoid function, W_f is the weight matrix, and b_c is the bias vector:

$$f_t = \delta\left(W_f \cdot \left(X_t, h_{t-1}\right) + b_c\right) f. \tag{1}$$

Then, the state of the input gate is updated, which mainly includes the relevant factors of unauthorized identification. i_t is the output state of the input gate at time t, b_i is the bias vector, $\widetilde{C}_t$ is a new candidate value vector created by tanh, W_c is the weight matrix, and C_t is the memory cell state at time t:

$$i_t = \delta\left(W_i \cdot \left(X_t, h_{t-1}\right) + b_i\right). \tag{2}$$

Candidate value vector formula:

$$\widetilde{C}_t = \tanh\left(W_c \cdot \left(X_t, h_{t-1}\right) + b_c\right). \tag{3}$$

Memory cell state formula:

$$C_t = i_t \cdot \widetilde{C}_t + f_t \cdot C_{t-1}. \tag{4}$$

Finally, the result of the current state is output by the output gate, o_t is the output state of the output gate, W_o and U_o are the weight matrix, and b_o is the bias vector:

$$\begin{aligned} o_t &= \sigma\left(W_o X_t + U_o h_{t-1} + b_o\right), \\ h_t &= O_t \cdot \tanh\left(C_t\right). \end{aligned} \tag{5}$$

When LSTM's ability is adopted to process long-order data, training and processing the response data of the unauthorized scene web page can well ensure the contextual validity of the unauthorized scene data.

2.2. LSTM-AutoEncoder. AutoEncoder usually consists of an encoder and a decoder [25]. Hinton et al. proposed a prototype of auto-encoding to test their Boltzmann machine learning algorithm [26]. First, the web page response data x is input. Then, the encoding stage encodes the input response data to the hidden layer h through the mapping function f: $h = f(x)$, and the decoding stage passes the mapping function g: $x' = g$ in the decoding stage (h decodes the h of the hidden layer into new web page response data). The reconstruction error of the encoder is

$$\min \|x - g(f(x))\|_2^2. \tag{6}$$

By trying to use the AutoEncoder to convert the web page response data into text vector into a new response text vector, and by comparing the difference between the original

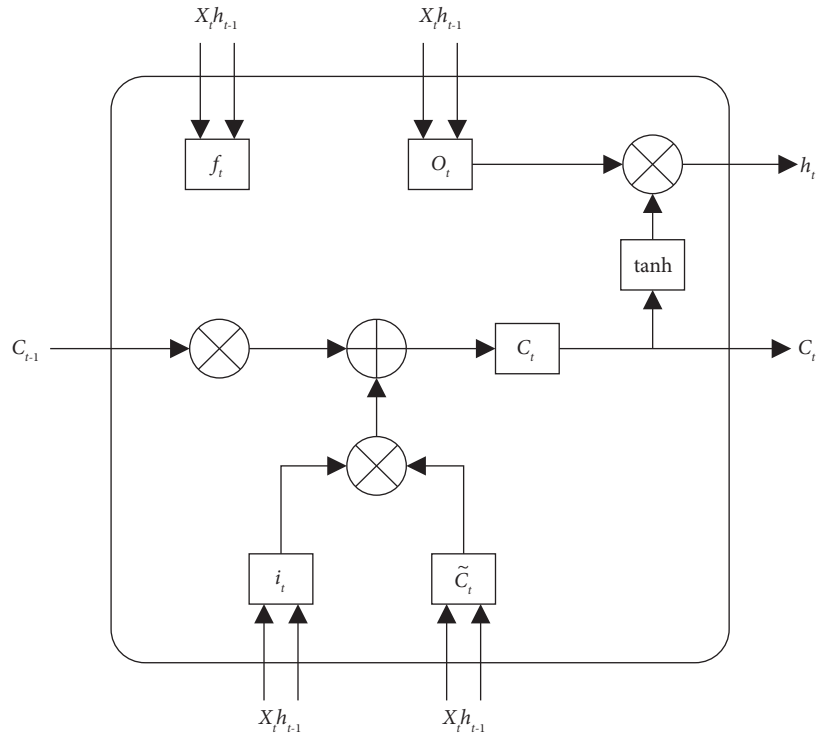


Figure 1: LSTM structure diagram.

text vector and the newly generated text vector, the category of the web page response data is judged.

LSTM-AutoEncoder was first proposed by Srivastava et al. to learn representations for video sequences. Currently, LSTM-AutoEncoder is widely used for extreme temporal prediction and anomaly detection in text sequences [27, 28]. The web page response of the horizontal unauthorized scene page is a text sequence, in which the unauthorized related scene page can be considered as a normal page, and the non-unauthorized related page can be considered as an abnormal page. Modeling based on this design idea can realize the detection of unauthorized scene pages.

The evaluation indicators of the LSTM-AutoEncoder model are mean squared error (MSE), mean absolute error (MAE), and commonly used evaluation indicators in machine learning, which contain precision, accuracy, recall, and $F1$-score [29]. The calculation formulas of MAE and MSE are as follows, where $X_{\text{prediction}}$ is the predicted value each time, X_{real} is the actual value, N is the total number of times, and t, i is the current number of times:

$$\text{MSE} = \frac{1}{N}\sum_{t=1}^{N}\left(X_{\text{prediction},t} - X_{\text{real},t}\right)^2, \tag{7}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(X_{\text{prediction},i} - X_{\text{real},,i}\right). \tag{8}$$

The calculation formulas of precision, accuracy, recall, and $F1$-score are as follows, where TP is a true positive example, FP is a false positive example, TN is a true negative example, and FN is a false negative example:

$$\begin{aligned}\text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}},\\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}.\end{aligned} \tag{9}$$

The formula for precision and $F1$-score is as follows:

$$\begin{aligned}\text{accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}},\\ F1 - \text{score} &= 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.\end{aligned} \tag{10}$$

The input of the LSTM-AutoEncoder model is the vectorized horizontal unauthorized scene page text, and the tokenizer module of the Keras framework is used for text preprocessing. First, the tokenizer instance object is built and passed in the stop character through the filters parameter, in which the maximum reserved phrase is set to 20, 000. Then, the training data for fitting are used to obtain a dataset text dictionary that can convert text into sequences. When new data need to be predicted, this dictionary can be used for conversion. Finally, the converted text sequence is deformed into a three-dimensional array required by the input of the LSTM layer, which is used as the training data to input the model.

Table 1 shows the architecture code of LSTM-AutoEncoder. The encoder and decoder are composed of two LSTM units, respectively. The RepeatVector layer is used to change the data dimension, and the TimeDistributed layer and the dense layer are used to deal with the many-to-many situation of the data decoding dimension.

Table 1: LSTM-AutoEncoder code architecture diagram.

Inputs = input (shape = (train_data.shape [1]), train_data.shape [2])
# ENCODER
$L1$ = LSTM (64, activations = relu) (inputs)
$L2$ = RepeatVector (train_data.shape [1]) (L1)
DECODER
$L3$ = LSTM (64, activations = relu, return_sequences = True) (L2)
Outputs = TimeDistributed (dense (train_data.shape [2])) (L3)
LSTM-AE = Model (inputs = inputs, outputs = outputs)
Inputs = input (shape= (train_data.shape [1]), train_data.shape [2])
# ENCODER
$L1$ = LSTM (64, activations = relu) (inputs)

2.3. Horizontal Override Detection Based on LSTM-AutoEncoder. In web applications, horizontal override occurs among users with the same privileges [30]. Its main feature is that users with the same authority can forge other users' behaviors by modifying parameters under the user's authority, and can view and use other users' data. The flowchart of LSTM-AutoEncoder-based horizontal override detection among users with the same privileges is as follows.

As can be seen from Figure 2, the detailed description of the horizontal unauthorized detection process based on LSTM-AutoEncoder is as follows.

The first step is to use user A's cookie and user B's cookie to visit the target website, respectively, use crawler technology to automatically crawl target page links and web page responses by carrying user cookies, and build a set of response pages for users A and B.

The second step is to deduplicate the pages of users A and B according to URL links and page responses. Page deduplication mainly removes the pages that both users A and B can access under the same permissions as the static pages of the web pages that are being accessed and the shared pages of the web pages. By comparing the URL links, request body similarity and web page similarity of users A and B, the pages in which the URL link is the same and the page similarity and request body similarity are 100% are identified as duplicate pages. Through the above method, the set of deduplicated pages of users A and B is obtained.

The third step is to use the deduplication page sets of users A and B to request each other. Concretely, the cookie of user A is used to request the URL link in the page collection of user B, and at the same time the parameters in the URL are replaced with the request parameters of user B, so as to get the page response of user A requesting user B. Then, the similarity between the page response of user A and the page response of user B is compared under the same link. When the similarity is greater than the set threshold (in the actual test process, the threshold is set to 98.4% according to the results of multiple tests), it is considered that this link is unauthorized. Similarly, user B's cookie is used to request user A's URL link to obtain another set of unauthorized links. The link set after deduplication of the two sets of links is the link set that is preliminarily judged to have an unauthorized vulnerability.

In the process of automatic level override detection, there are often a lot of noise data (such as UUID, timestamp, and other data dynamically generated by JS) in web pages. The degree is less than 100%, has not been deduplicated, and enables in the third step of the mutual request detection; the page similarity of the response page after the request of users A and B is higher than the threshold, thereby resulting in a false positive for the horizontal unauthorized detection result. The horizontal unauthorized vulnerability only occurs in the horizontal unauthorized scene page, so the problem of horizontal unauthorized false positive can be solved by constructing a deep learning model to detect whether the result page is an unauthorized scene page.

The fourth step is to perform data preprocessing on the response page data in the horizontal unauthorized detection result and then perform the unauthorized scene page detection through the LSTM-AutoEncoder model. When the model determines that the page type is an unauthorized scene page, it is concluded that the page has a horizontal unauthorized vulnerability. Otherwise, it is considered that the horizontal unauthorized link determined in the third step does not belong to the horizontal unauthorized scene page, and there is a false positive.

3. Test Analysis

3.1. Experimental Data and Experimental Environment. LSTM-AutoEncoder belongs to the unsupervised single-classification model, so only the unauthorized scene pages need to be collected during the model training stage. The number of pages without business data interaction in the web application website is much larger than the number of pages with user business data interaction. Therefore, in the model testing stage, the non-unauthorized scene pages used to test the performance of the model are randomly selected from the website pages without business interaction in the training set website. In order to fit the real business scenario, the experimental data collected in this paper come from a large number of web application websites such as Internet forums and e-commerce, which involves pages of unauthorized scenarios such as personal information, order payment, inquiry, invoice, complaint, and suggestion. Furthermore, the corresponding type of website directory is found through the e-commerce index website and forum index website and the links are crawled. Then, the website is registered and by the selenium automated testing tool, the script set is used to obtain the web page response of the unauthorized scene page rendered by the browser after registration. Meanwhile, the identification and labeling of the unauthorized scene page are carried out by the security service personnel, the data collection process consumes a total of 60 man-days, and Jieba is used to segment the collected web page responses. Considering that in existing mainstream web development frameworks such as React, Vue, and other web page source codes dynamically generated by JavaScript, English character tags are often randomly generated and contain very few data features, English data are discarded during word segmentation, and stop words are removed according to the Chinese stop words table and the

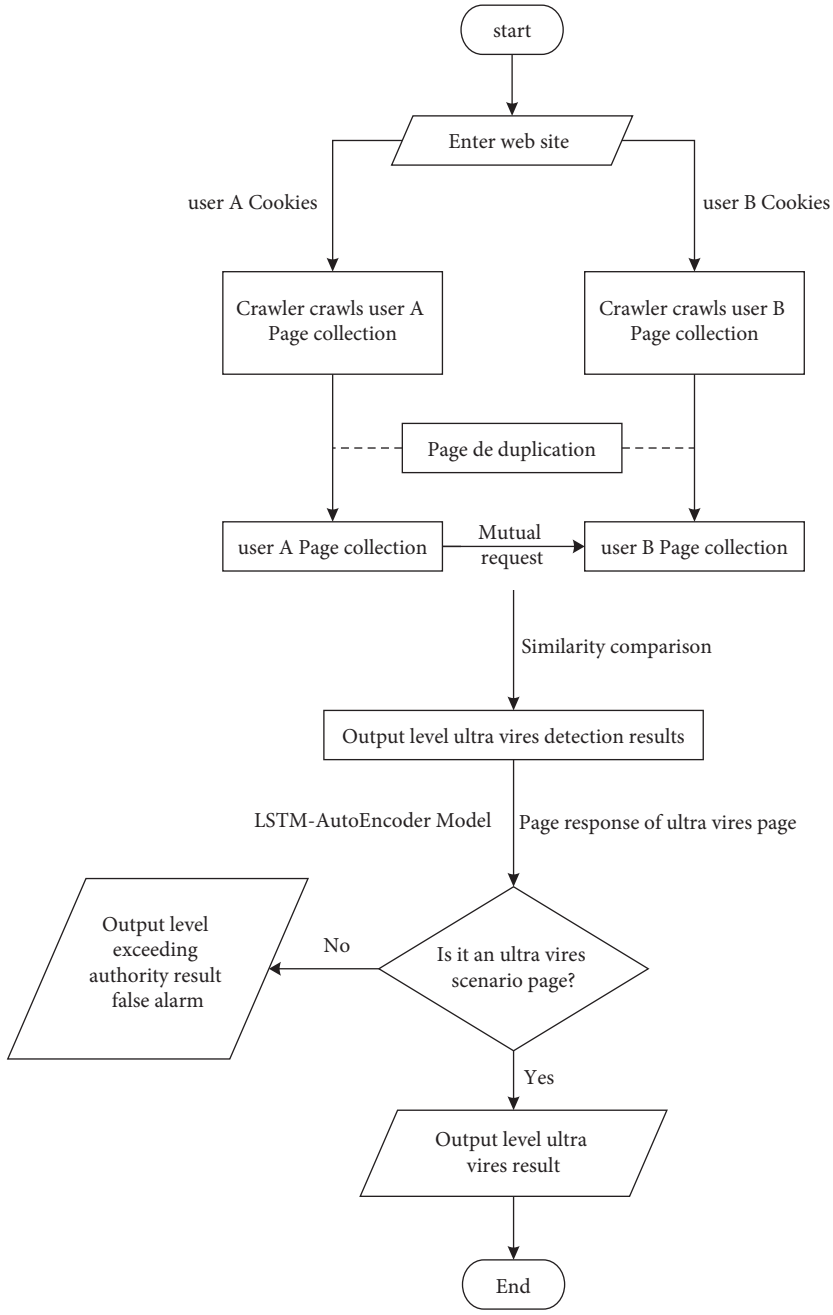


Figure 2: Flowchart of horizontal override detection based on LSTM-AutoEncoder.

place name information table. After cleaning, a total of 12,000 unauthorized scene page data for model training and 1,200 nonauthorized scene page data for testing model performance are collected. The data splitting function train_test_split in the scikit-learn framework is used to divide the data into training set, test set, and validation set according to the ratio of 0.8:0.1:0.1, and the sequence length of each sample data is unified to 1000 through the pad_-sequence function of the Keras module. The dataset statistics are shown in Table 2.

Because theresponse text of the web page is too long, Table 3 lists the first 15 texts of some dataset examples in the training set.

The experimentalprocess is the same as the detection flowchart given in Figure 3. Firstly, theautomatic horizontal unauthorized detection is used to perform horizontal unauthorized detection on the target website, and then the web page response of the target link corresponding to the detection result judged to be unauthorized is input into the LSTM-AutoEncoder model to detect the unauthorized scene page. If the output of the LSTM-AutoEncoder model is yes, the web page corresponding to the link is an unauthorized scene page, and it is determined that the link exists as a horizontal unauthorized scene, whereas if the detection result is no, it means that the link is not an unauthorized scene page, and the unauthorized result is a false positive.

Table 2: Dataset statistics.

Data type (number)	Training set	Test set	Validation set	Statistics
Unauthorized scenario page	9600	1200	1200	12000
Nonurban scenario page	0	0	1200	1200

Table 3: Examples of training set datasets.

Number	Example
1	Connect the poster QR code to log in using the Weibo account to access the poster account password verification code and change in dynamic prompt guarantee must
2	Screenshot of the client's wonderful download, watch video, homepage channel, special event reminder, homepage account setting recommendation
3	Car rental global-car rental home car store activities car rental-login registration-hello-order assets-account logout
4	Enter the mobile phone number-retrieve the user name-home member login service-retrieve the user name-mobile phone number verification code-now available
5	Free registration, complete information, and login problems
6	The user logs in to the homepage of Super Comics-update ranking-search and read-clear record-login synchronization-read click
7	Account-personal center-home interactive-home topics-Q&A center-index market-market data-announcement home
8	Merchant center, return to the homepage, welcome to the settings menu, release open store settings, profile account name, contact

The experimental environment is as follows: LSTM-AutoEncoder model uses Python 3.6 as the development language, Keras == 2.1.2 as the high-level packaging framework, and tensorflow-gpu == 1.10.0 as the deep learning framework, and the CPU is Intel i7-7700K, GPU model training for NVIDIA 2080 Ti Windows machines.

3.2. Experimental Method. The following methods are used for the judgment of how to use the LSTM-AutoEncoder model for horizontal unauthorized scene pages.

First, model training is performed using MSE and MAE as loss functions, respectively. After continuously adjusting the unit size, batch size, activation function, and round epoch, the optimal model with units = 64, batch size = 32, epoch = 50, and activation function of ReLU is obtained. Figures 3 and 4 show, respectively, MSE and MAE LOSS descent plots of the best model.

It can be seen from Figure 3 that the MSE of the training value is significantly higher than the test value at the first 10 training times, and then, with the increase of training round, the MSE of the training value is significantly lower than that of the test value and finally remains around 0.2, which shows that with the progress of the experiment, the MSE error of the test is gradually reduced, and the test prediction results are more and more accurate. In addition, Figure 4 shows the decrease of the model's mean absolute error loss. In this study, with the increase of training times, the MAE of both the training value and the test value shows a downward trend, and after 40 times of training, the MAE of the test value is higher than that of the training value, which indicates that when the number of training times reaches a certain number, the accuracy will reach the best accuracy, and the data prediction is most feasible at this time.

Secondly, the best models of MSE and MAE are adopted to predict the original training set, respectively, and the calculation formulas of MSE and MAE (Equations (7) and (8)) are used to obtain the MSE and MAE distribution

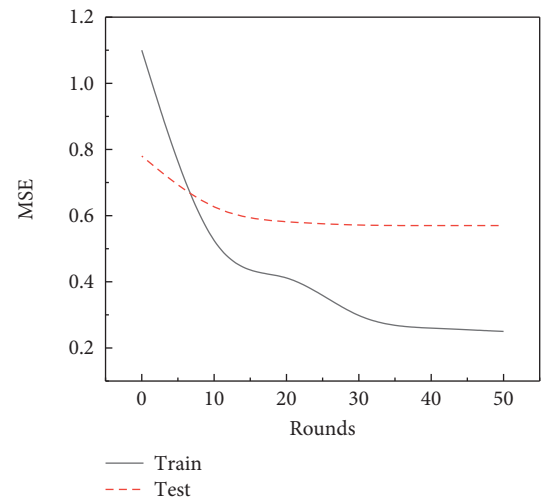


Figure 3: Model MSE loss drop chart.

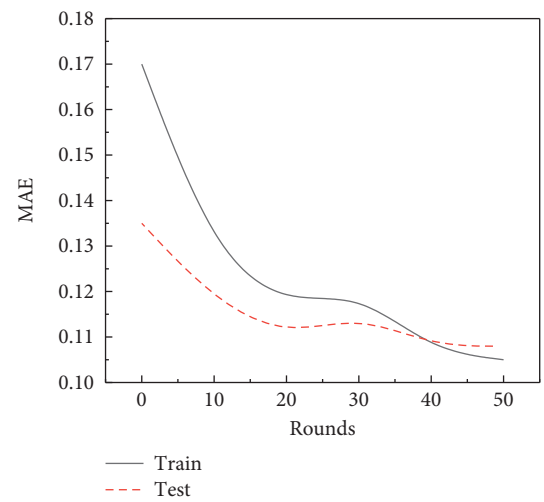


Figure 4: Model MAE drop graph.

intervals between the predicted value and the original value. The distribution interval determines the threshold for the restoration error. The distribution interval of MSE and MAE is shown in Figures 5 and 6, the threshold of MSE is 1.75, and the threshold of MAE is 0.65. As can be seen from Figure 5, when the MSE error is within 0.5, the number proportion is the highest, which indicates that the prediction effect is better, whereas when the number of MSE is above 0.5, the number proportion is less, which indicates that the model has higher prediction accuracy. Compared with MSE, the distribution range of MAE is also mainly within 0.5, of which the number of data within 0.2 accounts for the highest proportion, followed by the part with MAE of 0.4. The data plot of MAE and MSE shows that the model has better prediction effect.

Finally, the MSE and MAE distributions of the validation set and negative samples are calculated, and according to the threshold, the TP, TN, FP, and FN values of the model in the case of MSE and MAE can be calculated. Assuming that the threshold is k, test is the test data, error is the error data, and sum is the sum of the data; the calculation method is as follows:

$$\begin{aligned} TP &= \text{sum}(\text{test} < k), \\ TN &= \text{sum}(\text{error} > k), \\ FP &= \text{sum}(\text{error} < k), \\ FN &= \text{sum}(\text{test} > k). \end{aligned} \tag{11}$$

The ROC curve is usually used to evaluate the quality of a classifier. The abscissa is TPR, and the ordinate is FPR, which, respectively, represent the probability of the classifier classifying positive examples into pairs and the probability of the classifier wrongly classifying negative examples. Moreover, AUC represents the area under the ROC curve, and the closer the AUC is to 100%, the higher the predicted value is. The precision curve of the image and the ROC curve can be obtained from the TP, TN, FP, and FN values, as shown in Figure 7.

In order to verify the effectiveness of this paper, the one-class SVM model and the AutoEncoder model are trained and compared with the same data preprocessing method and the same training data. Among them, the one-class SVM model uses the preset model and default parameters in the scikit-learn module for training, the AutoEncoder model uses a double-layer fully connected layer as the encoder and decoder, and the unit parameter is 32 for training. The comparison results of the three models for the validation set are shown in Tables 4 to 6.

According to the experimental results, in the one-class model, the precision of the model reaches up to 0.974, indicating that the one-class model has a high recognition rate for unauthorized scene pages in the dataset. However, the recall rate of the model is only 0.473, indicating that for many non-unauthorized scene web pages, the model identifies them as unauthorized scene pages. Therefore, in the actual business process, the one-class SVM model cannot solve the problem of false positives of horizontal overreach.

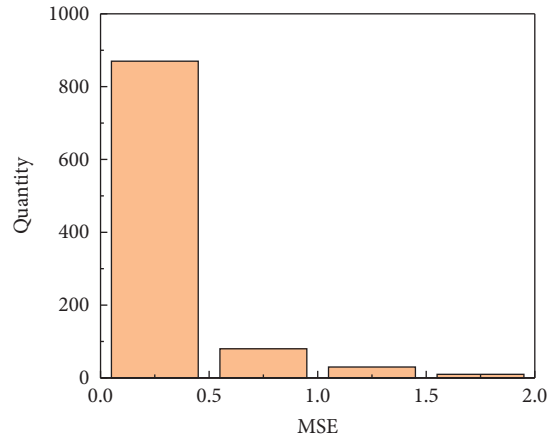


Figure 5: Distribution of MSE.

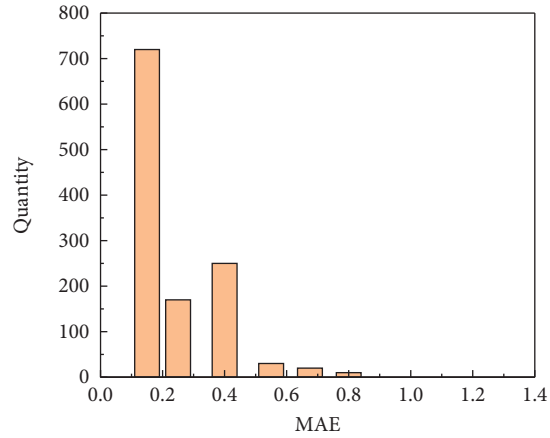


Figure 6: Distribution of MAE.

The recall rates of the AutoEncoder model in the case of MAE and MSE are 0.988 and 0.994, indicating that the case of the correct horizontal unauthorized scene page occupies a high proportion to all the horizontal unauthorized samples, but the average precision rate of such model in the case of MAE and MSE is lower by about 3% compared to that of LSTM-AutoEncoder indicating that the proportion of pages predicted by the model as horizontal unauthorized scenes is only 0.89%. Therefore, the overall performance of the AutoEncoder model is weaker than the LSTM-AutoEncoder model. The $F1$-score often represents the harmonic average of the precision rate and the recall rate, which is consistent with the business requirements for page detection in horizontal override scenarios. It can be seen from Table 3 that the accuracy, precision, recall, and $F1$-score of the LSTM-AutoEncoder model in the case of MAE have reached a relatively balanced index. Compared with the case of MSE, the $F1$-score of MAE is 0.3% more, and the accuracy rate of MAE is increased by 0.2%, which is more in line with real business needs. Therefore, the LSTM-AutoEncoder model trained with MAE as the loss function is used as the detection model of the online environment to realize the detection of horizontal unauthorized scene pages.

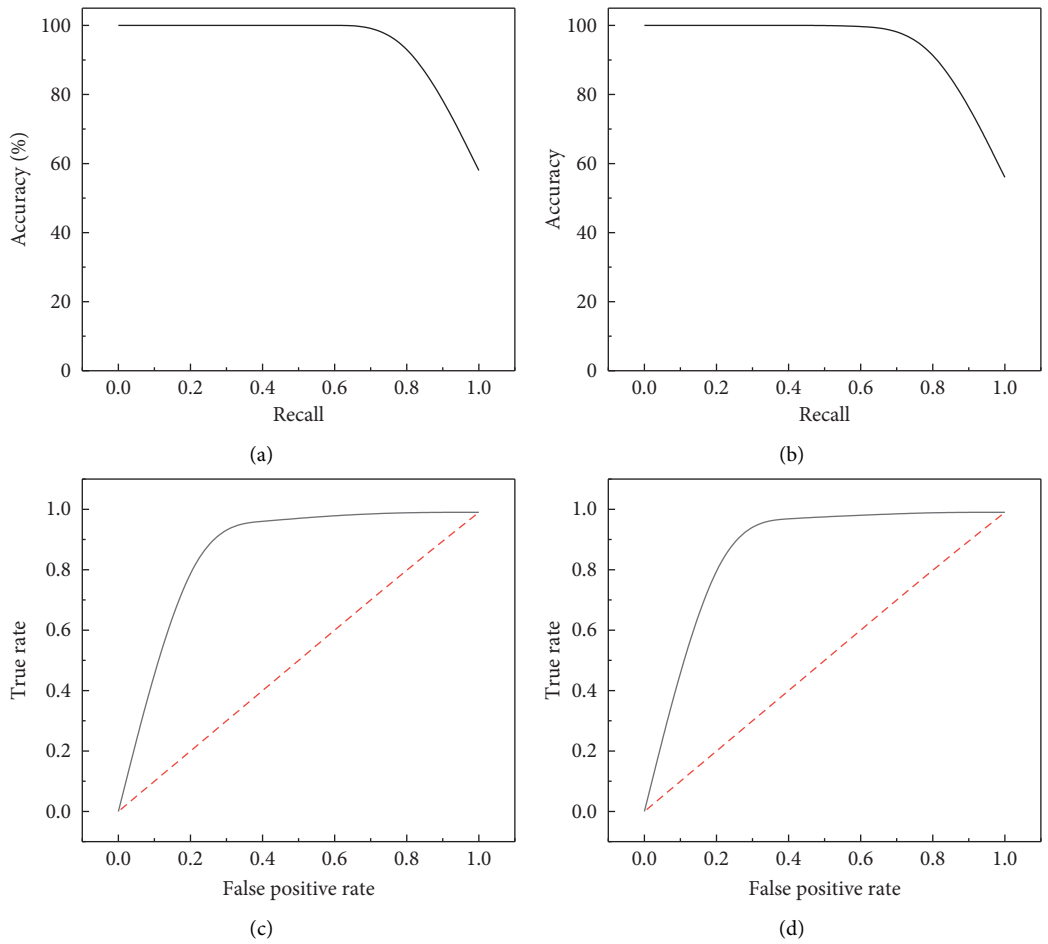


Figure 7: Accuracy and true rate graph, (a) MSE precision-recall plot; (b) MAE precision-recall plot; (c) MAE subject characteristic curve; (d) MSE subject characteristic curve.

Table 4: One-class SVM prediction results.

Precision	Accuracy	Recall	*F*1-score
0.974	0.730	0.473	0.636

Table 5: AutoEncoder prediction results.

Loss function	Precision	Accuracy	Recall	*F*1-score
MSE	0.895	0.940	0.994	0.942
MAE	0.897	0.940	0.988	0.940

Table 6: LSTM-AutoEncoder prediction results.

Loss function	Precision	Accuracy	Recall	*F*1-score
MSE	0.931	0.942	0.955	0.942
MAE	0.920	0.944	0.971	0.945

According to the characteristics of the LSTM network, its feature extraction ability for the contextual web page response text sequence is stronger, which makes the comprehensive performance of the model more excellent. In addition, the multi-layer LSTM codec has been tried in the experiment, but the accuracy of the model is only increased by 0.001%, and even the accuracy of the four-layer network has dropped. The recognition of scene pages brings better results. In the actual business inspection scenario, the efficiency of the model is the key to the rapid operation of the detection system. Blindly pursuing the accuracy rate and abandoning the performance parameters must be avoided in the development process. After the time-consuming test of the prediction module, the prediction time of the model is 0.04 s, which meets the business needs of calling a large number of model predictions. After weighing the pros and cons, choosing the simple model architecture as shown in Figure 3 can reduce the complexity of the model, accelerate the prediction time of the model in the application stage, and improve the performance of the detection software.

4. Conclusion

Horizontal transgression vulnerability is particularly important to the business data security of companies and users. In this paper, we use the sample data of web application transgression scenarios and train the LSTM-AutoEncoder model to identify the pages of transgression scenarios, which solves the problem of false positives in traditional transgression detection and improves the accuracy of transgression detection. The experimental data show that the

LSTM-AutoEncoder model has certain accuracy advantages over the traditional one-class SVM model and AutoEncoder model under certain amount of data scale and also has great advantages in processing web text sequences with contextual relationships, which is as follows [31]:

(1) In the one-class model, the model reaches a very high precision of 0.974 and has a high recognition rate for unauthorized scene pages in the dataset. However, the recall rate of the model is only 0.473, which makes the non-unauthorized scene web page be recognized by the model as an unauthorized scene page.

(2) The recall rates of the AutoEncoder model in the case of MAE and MSE are 0.988 and 0.994, respectively, and the cases that are correctly judged as horizontal unauthorized scene pages occupy a high proportion to all the horizontal unauthorized samples. However, the average precision rate of such model in the case of MAE and MSE is lower by about 3% compared to that of LSTM-AutoEncoder.

(3) In the case of MAE, the precision, accuracy, recall, and $F1$-score of the LSTM-AutoEncoder model have reached a relatively balanced index. Compared with the case of MSE, its $F1$-score is 0.3% higher and the accuracy rate is 0.2% higher, which is more in line with real business needs.

(4) After the time-consuming test of the prediction module, the prediction time of the model is 0.04 s, which meets the business needs of calling a large number of model predictions. After weighing the pros and cons, choosing the simple model architecture as shown in Figure 3 can reduce the complexity of the model, accelerate the prediction time of the model in the application stage, and improve the performance of the detection software.

The application of AI and network security is at a relatively early stage of development, and this paper provides a good application case. However, the experimental data in this paper also have certain shortcomings. In future research, it is necessary to extend the collection of training data to more types of pages of transgression scenarios in a large number of web applications on the Internet, so as to improve the generalization ability of the model and make it possible to test more transgression scenarios.

Data Availability

The dataset can be accessed upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China, "Joint Research of IoT Security System and Key Technologies Based on Quantum Key," under project number 2020YFE0200600.

References

[1] P. A. Karger and R. R. Schell, "Multics Security Evaluation: Vulnerability Analysis," in *Proceedings of the 18th Annual Computer Security Applications Conference, 2002. Proceedings*, December 2002.

[2] A. Almuhaideb and S. Saeed, "A process-based approach to abet accreditation: a case study of a cybersecurity and digital forensics program," *Journal of Information Systems Education*, vol. 32, no. 2, pp. 119–133, 2021.

[3] K. H. Kim, *Analysis of Security Vulnerability in Cooperative Communication Networks*, 2011.

[4] L. Li, Y. Li, Y. Liu et al., "Preparation of a novel activated carbon from cassava sludge for the high-efficiency adsorption of hexavalent chromium in potable water: adsorption performance and mechanism insight," *Water*, vol. 13, no. 24, p. 3602, 2021.

[5] M.-a c o T. C. S. L. P. Laverdiere-papineau, *Security Vulnerability Correction*, US, 2016.

[6] M. Prakash, M. Behera, and C. Security, 2020.

[7] T. E. Gasiba, U. Lechner, M. Pinto-Albuquerque, and A. Porwal, "Cybersecurity Awareness Platform with Virtual Coach and Automated Challenge Assessment," *Lecture Notes in Computer Science*, vol. 12501, pp. 67–83, 2021.

[8] A. Krishnamurthy, A. Mettler, and D. Wagner, "Fine-grained privilege separation for web applications," in *Proceedings of the 19th International Conference on World Wide Web, WWW*, pp. 551–560, Raleigh, North Carolina, USA, Apri 2010.

[9] S. Nagpure and S. Kurkure, "Vulnerability Assessment and Penetration Testing of Web Application," in *Proceedings of the 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, Pune, India, August 2017.

[10] M. Monshizadeh, P. Naldurg, and V. N. Venkatakrishnan, "MACE: Detecting privilege escalation vulnerabilities in web applications," *Bone*, vol. 47, pp. 690–701, 2014.

[11] T. Markmann, D. Gessner, D. Westhoff, and QuantDroid, "Quantitative approach towards mitigating privilege escalation on Android," in *Proceedings of the IEEE International Conference on Communications*, Budapest, Hungary, June 2013.

[12] K. Mollus, D. Westhoff, and T. Markmann, "Curtailing Privilege Escalation Attacks over Asynchronous Channels on Android," in *Proceedings of the International Conference on Innovations for Community Services*, Reims, France, June 2014.

[13] F. Sun, X. Liang, and Z. Su, *Static Detection of Access Control Vulnerabilities in Web Applications*, Shanxi Architecture, Shanxi Province, north China, 2011.

[14] G. Deepa, P. S. Thilagam, A. Praseed, and A. R. Pais, "DetLogic: a black-box approach for detecting logic vulnerabilities in web applications," *Journal of Network and Computer Applications*, vol. 109, pp. 89–109, 2018.

[15] H. G. Song, Y. Kim, and K. G. Doh, *Automatic Detection of Access Control Vulnerabilities in Web Applications by URL Crawling and Forced Browsing*.

[16] L. Ma, Y. Yan, and H. Xie, "A new approach for detecting access control vulnerabilities," in *Proceedings of the 2019 7th International Conference on Information, Communication and Networks (ICICN)*, Macao, China, April 2019.

[17] K. T. Mursi, B. Thapaliya, and Y. Zhuang, "A hybrid-optimizer-enhanced neural network method for the security vulnerability study of multiplexer arbiter PUFs," *Journal of*

Physics: Conference Series, vol. 1729, no. 1, Article ID 012010, 2021.

[18] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to mixed-integer optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 36–48, 2021.

[19] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[20] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, "Intrusion detection by machine learning: a review," *Expert Systems with Applications*, vol. 36, no. 10, pp. 11994–12000, 2009.

[21] O. Yavanoglu and M. Aydos, "A review on cyber security datasets for machine learning algorithms," in *Proceedings of the IEEE International Conference on Big Data*, Boston, MA, USA, December 2017.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] N. Srivastava, E. Mansimov, and R. Salakhutdinov, *Unsupervised Learning of Video Representations Using LSTMs*, JMLR.org, 2015.

[24] A. F. Gers, J. Schmidhuber, and F. Cummins, *Learning to Forget: Continual Prediction with LSTM*, in *Proceedings of the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99*, Edinburgh, USA, September 2000.

[25] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 12, pp. 3371–3408, 2010.

[26] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithm for Boltzmann machines," *Cognitive Science*, vol. 9, no. 1, pp. 147–169, 2010.

[27] V. Wan, Y. Agiomyrgiannakis, H. Silen, and J. Vit, "Google's Next-Generation Real-Time Unit-Selection Synthesizer Using Sequence-To-Sequence LSTM-Based Autoencoders," in *Proceedings of the Interspeech*, August 2017.

[28] C. Fang, Y. Tao, J. Wang et al., "Mapping relation of leakage currents of polluted insulators and discharge arc area," *Frontiers in Energy Research*, vol. 9, p. 694, 2021.

[29] H. I. W Noguchi and M. Yamamoto, "Proposing Multimodal Integration Model Using LSTM and Autoencoder," in *Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies (formerly BIONETICS)*, December 2016.

[30] P. Liu, X. Sun, Y. Han, Z. He, W. Zhang, and C. Wu, "Arrhythmia classification of LSTM autoencoder based on time series anomaly detection," *Biomedical Signal Processing and Control*, vol. 71, Article ID 103228, 2022.

[31] J. Kang, C. S. Kim, J. W. Kang, and J. Gwak, "Anomaly detection of the brake operating unit on metro vehicles using a one-class LSTM autoencoder," *Applied Sciences*, vol. 11, no. 19, p. 9290, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9784806, 1 page
https://doi.org/10.1155/2023/9784806

Retraction

Retracted: Strategies for Ideological and Political Education in Colleges and Universities Based on Deep Learning

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Y. Sun, "Strategies for Ideological and Political Education in Colleges and Universities Based on Deep Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5322677, 9 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5322677, 9 pages
https://doi.org/10.1155/2022/5322677

Research Article

Strategies for Ideological and Political Education in Colleges and Universities Based on Deep Learning

Ying Sun

Jilin Animation Institute, Changchun, Jilin 130012, China

Correspondence should be addressed to Ying Sun; 3180100350@caa.edu.cn

Received 18 July 2022; Revised 19 August 2022; Accepted 24 August 2022; Published 26 September 2022

Academic Editor: Man Fai Leung

Ideological and political education in colleges and universities is routinely burdened with the job of building morality and cultivating people, which is related to the cultivation of college students' ideals and beliefs, spiritual pursuits, and political literacy. Based on self-determination theory (SDT), this paper modeled different learning motivations in the early stage of ideological and political courses and analyzed the learning motivation of different student groups combining the Gaussian mixture model (GMM) and stacked autoencoder (SAE). Meanwhile, the study in this paper compared the participation characteristics of different learning motivation clusters, the differences between the ideological and political course performances of students with different learning motivations, and the potential link between learning motivation and learners' educational level. The experimental results show that students with extrinsic motivation will have better performance in the courses. The strength of extrinsic motivation is positively correlated with students' academic performance, and 70% of students with intrinsic motivation achieve excellent results. In addition, the χ^2 test result of the two courses selected is 6.442, which confirms the effectiveness of the clustering model proposed in this paper from the side and provides effective theoretical support for the implementation and reform of ideological and political education strategies.

1. Introduction

As information technology increasingly modernizes, tremendous changes have taken place in people's lives, and various fields such as politics, economy, and culture have all taken the express train of the information age to move forward. In recent years, with the gradual entry of information technology into the field of education, more and more education researchers have devoted themselves to the practical exploration of the deep integration of information technology and subject teaching [1]. The integration of information technology into the ideological and political teaching of colleges and universities will inevitably bring new opportunities and challenges to the education system [2].

At present, online education has become a hot issue and the optimization of online education service quality has gradually become common demand of social life. E-learning has gained enormous popularity due to the accessibility of various e-learning platforms such as Moodle, Blackboard, MOOC, and so on [3]. Learning management systems such as MOOC and Moodle are widely used in enterprises, universities, and educational institutions, which offer certain advantages to the teaching of modern education. Online education is a computer-based learning environment in which students can freely choose learning materials and control their own learning pace, and students' ongoing learning behavior will be recorded in log files without interrupting their learning process. Through in-depth analysis of log files, students' learning behaviors can be studied qualitatively, which can provide useful information for teachers to improve instructional design. Many studies have shown that online learning has a significant impact on students' autonomous learning and knowledge acquisition [4]. For example, Liu and Yuan found that students can only learn effectively when they are highly motivated in networking activities. Different learning motivations produce different learning outcomes [5]. Firat et al. used the IMeL questionnaire method to determine the level of intrinsic

motivation of open and distance education students [6, 7], and Dunn used the questionnaire method to classify learning motivation, which is used as an engagement indicator to predict students' final performance at the same time [8, 9]. However, not every student can benefit from an online learning model even if they enroll in the same courses, as learning outcomes are greatly influenced by students' self-regulated learning ability and motivation. In other words, learning motivation drives learning behavior, and learning behavior is an external manifestation of learning motivation and an important basis for evaluating the state of learning motivation. Differences in students' learning motivations will directly differentiate learning strategies selected. So, a high interest may be required for students' initial learning. With interest, students' curiosity will be stimulated, and students' motivation to learn will be enhanced. Different learning participants have different learning motivations and goals; the motivation may not be just to get better grades or certificates, it may be that they are interested in a particular chapter or forum, which will support them to communicate with more like-minded scholars. Therefore, it is urgent to better understand the inner relationship between students' learning behavior, final performance, and learning motivation. To gain a more objective understanding of the relationship between learning motivation, engagement, and final performance, different learning motivations can be measured by analyzing online learning behavior data. Currently, online learning is often lacking in guiding students' learning depth. Improving the depth of learning depends on the user experience on the online learning end of the learners and the reinforcement of learning motivation. However, it is not easy to evaluate students' learning motivation and identify their differences in detail. In the traditional offline classroom environment, teachers can relatively easily grasp the specific situation of students' learning motivation through face-to-face interaction, while in the online learning environment, teachers must obtain this information through the information recorded by the learning platform. In addition, students with different educational backgrounds are able to enroll in the same course according to their own will, which leads to a huge number of participants in a given course, so it is indispensable to develop a method that can automatically detect various motivation types from large datasets.

Although there have been some previous studies on learning motivation [10, 11], most of them are based on questionnaires, self-reports and interviews, and other data for analysis, and few use data collected from online learning platforms, which makes the analysis process very labor-intensive, and the final analysis results are too subjective. Even though there are some studies on online learning behavior, these studies are based on the analysis of students' overall performance in the whole course semester, and the final research results cannot be fed back to the current course in time. According to the above problems, this paper attempts to process a large amount of information (demographic information and clickstream data) generated in the ideological and political education of colleges and universities through the method of machine learning (deep learning) and combines corresponding models and theories (self-determination theory, Gaussian mixture model, and deep coding automatic model) to clarify the relationship between learning motivation and external performance of students' ideological and political education. At the same time, it also provides relevant constructive suggestions for the implementation and reform of ideological and political education strategies.

2. Models and Methods

2.1. Self-Determination Theory. Self-determination theory (SDT), one of the most comprehensive and empirically supported theories about motivation [12], describes the learning motivation in the online learning environment in detail. Deci and Ryan defined SDT as an experience-derived theory of human motivation and personality in social settings that distinguishes motivation from autonomy and control. Related studies have shown that SDT has made an important contribution to the determination of students' motivation in an online environment [13].

2.2. The Basic Principle of Gaussian Mixture Model Algorithm

2.2.1. Related Concepts of Gaussian Mixture Model. The Gaussian mixture model (GMM) is a model composed of multiple Gaussian distributions, and the basic Gaussian mixture model adopts the method of linear superposition [14]. It assumes that the dataset consists of multiple potential clusters that conform to a Gaussian distribution, and that the final stacking result is the distribution presented. Although in general, the same dataset may contain different types of distributions, considering that the model is represented by a mixture of multidimensional Gaussian probability models, it can still fit any type of probability distribution.

Supposing that a random variable X is a mixture of M Gaussian distribution models, the Gaussian mixture model can be represented as follows:

$$P(x \mid \theta) = \sum_{m=1}^{M} \alpha_m \phi(x \mid \theta_m), \tag{1}$$

where α_m represents the weight of the mth Gaussian distribution in the Gaussian mixture model. This parameter is generally not given and satisfies the following formula:

$$\sum_{m=1}^{M} \alpha_m = 1, \ (\alpha_m \geq 0). \tag{2}$$

Although it is generally impossible to know the components occupied by each Gaussian distribution in the mixture model, considering that the Gaussian mixture model is formed by the linear superposition of m Gaussian distributions, the sum of the weights of its constituent members is defined as 1. In formula (1), $\phi(x \mid \theta_m)$ is the probability density function of the mth submodel in the mixture model, Gaussian density function is the most used function, and θ_m can be represented by $\theta_m = (\mu_m, \sum m)$, so $\phi(x \mid \theta_m)$can be expressed by

$$\phi(x \mid \theta_m) = \frac{1}{\sqrt{2\pi\sum_m}} \exp\left\{-\frac{(x-\mu_m)^2}{2\sum_m}\right\}, \quad (3)$$

where μ_m represents the mean of the mth Gaussian distribution component and $\sum m$ represents the covariance matrix of the mth Gaussian distribution component.

Given that a Gaussian model can be determined by the probability density function and the parameter set θ, the parameter calculation of the hybrid model proposed in this paper uses the log-likelihood function of the maximization model to find the optimal sample parameters that make the cluster close to the actual distribution of the sample as far as possible, so as to obtain better performance clustering results. Equation (4) is a likelihood estimation function that calculates the key parameters of the sample dataset.

$$L(\theta) = L(X_1, X_2, \cdots X_m; \theta) = \prod_{i=1}^{m} P(X_i, \theta). \quad (4)$$

where $X_1, X_2, \cdots X_m$ represents the data samples of the model, $P(x, \theta)$ represents the probability density function of the model, and θ represents a vector consisting of one target parameter or multiple parameters to be estimated.

Considering the hidden variables in the sample set, the process of directly using the maximum likelihood estimation method to estimate the model parameters will be quite tedious, and the optimization process will be quite long [15]. Therefore, the implicit variables need to be represented first. After determining the hidden variables, the maximum value of the likelihood function is obtained and after taking the logarithm of formula (4), formula (5) is the result obtained.

$$L^{(1)}(\theta) = \sum_{i=1}^{m} \log p(x_i, \theta) = \sum_{i=1}^{m} \log \sum_{Z_i} p(x_i, z_i, \theta). \quad (5)$$

2.2.2. Expectation-Maximization Algorithm. The expectation-maximization (EM) algorithm was first proposed by Dempster et al. EM algorithms are widely used in many algorithms in machine learning [16, 17] such as the parameter estimation in the K mean, the support vector machine (SVM) [18], the GMM, the hidden Markov model (HMM) [19], and the subject generation model (latent Dirichlet allocation, LDA) [20]. The EM algorithm refers to the iterative solution of some target parameters using a strategy of great likelihood estimation in all datasets, including hidden variables. The iteration of the EM algorithm is mainly done by two steps: the expectation Step and the maximization Step. The expectation step of the expectation-maximization algorithm is to calculate the expectation of the model according to the hidden state of the model, compute the Gaussian distribution of the guessed hidden data, and then fix the model parameters and use the maximum likelihood estimation to calculate the complete data including the observed data and the hidden data in sequence, which will finally obtain the parameters of the Gaussian mixture model. By that analogy, M steps are performed. Then, through iterating E step and M step, that is, adjusting the model according to the parameters (E step) and then adjusting the parameters according to the model (M step), the E step and the M step are alternately performed until the parameters of the solved Gaussian mixture model do not change substantially. Meanwhile, the algorithm achieves convergence and obtains the optimal expectation of the Gaussian mixture model, the covariance matrix, and the weights of each Gaussian distribution. The expected value of the log-likelihood function of the mixed model can be described by the initial values of the model parameters that have been selected, and the specific definition is

$$E_Q\left[\log p(\theta \mid Y, Q) \mid \theta^{(i)}, Y\right] = \int \log\left[p(\theta \mid Y, Q)\right] p\left(Q \mid \theta^{(i)}, Y\right) dQ, \quad (6)$$

where Q represents the implicit data that cannot be observed and $\theta^{(i)}$ represents the posterior standard deviation after the $i+1$th iteration. The conditional expectation probability of the joint distribution of the mixture model can be expressed by

$$L(\theta, \theta_i) = \sum^{m} \sum P\left(z_i \mid x_i, \theta_j\right) \log P(x_i, z_i \mid \theta). \quad (7)$$

The constraint of the maximum value of the log-likelihood function parameter under the conditional probability can be expressed by

$$\theta^{j+1} = \arg\max_{\theta} L\left(\theta, \theta^{j}\right). \quad (8)$$

Continue to iterate the above E step and M step and end the iteration when $\theta^{(i)}$ and $\theta^{(i+1)}$ are infinitely close.

2.3. Deep Autoencoding Model

2.3.1. Autoencoder. Autoencoder network belongs to a kind of neural network structure under unsupervised learning classification, which usually consists of three layers [21]. The ideal goal of an autoencoder network is to reproduce the original input data. Similar to the Seq2Seq model in natural language processing, an autoencoder network usually consists of two parts: an encoder and a decoder. Among them, the goal of the encoding network is to convert the relatively high-dimensional original input data into an encoding vector in a low-dimensional space, while the role of the decoding network is the opposite, which is to restore the low-dimensional vector to the data representation of the high-dimensional space [22].

The encoder consists of an encoding function f_θ, and for each input data x, the encoding function can be expressed by

$$h = f_\theta(x) = f_\theta(\omega x + b), \quad (9)$$

where h is an intermediate vector obtained after the input data x are encoded by the encoder in the hidden layer, which can also be called an encoding vector. In the same idea, the decoding network is composed of a decoding function $g_{\theta'}$ that is, the decoder, which remaps the intermediate vector h from the low-dimensional space to the high-dimensional space, as shown in the following formula:

$$\widehat{x} = g_{\theta'}(h) = g_{\theta'}(\omega' h + b'). \tag{10}$$

As shown in formula (11), the parameters θ and θ' in the encoding stage and the decoding stage are obtained by continuously adjusting and updating to minimize the reconstruction error. . The essential principle of the autoencoder is to minimize the reconstruction error$L(x, \widehat{x})$of the samples for M training sample data sets, which is often used to measure the difference between the input data x and the output data $\widehat{x}$. To sum up, the specific definition is shown in the following formula:

$$\phi_{AB}(\theta, \theta') = \frac{1}{M} \sum_{m=1}^{M} L(x, g_{\theta'}(f_{\theta}(x^m))). \tag{11}$$

Both the encoder and the decoder are implemented by nonlinear mapping, and their specific definition is shown in equations (12) and (13), where s_f represents the activation function of the encoding stage and s_g represents the activation function of the decoding stage. The parameter set of the encoder can be represented as $\theta = \{W, b\}$, and the parameter set of the decoder can be represented as $\theta' = \{W^T, b\}$, where b and d both represent bias vectors and W and W^T represent weight matrices.

$$f_{\theta}(x) = s_f(Wx + b), \tag{12}$$

$$g_{\theta'}(x) = s_g\left(W^T x + d\right), \tag{13}$$

$$L(x, \widehat{x}) = \|x - \widehat{x}\|^2. \tag{14}$$

2.3.2. Deep Feature Learning. The deep learning network model is to input the original input data into a neural network with multiple hidden layers. After the nonlinear operation of multiple hidden layers in the middle, the final output of the hidden layer is the same as the input data through the deep network model. The deeper it learns, the deeper the features are abstracted [23]. However, some datasets do not have initial labels, and deep feature learning is divided into three categories based on whether the initial labels are involved in the entire network training process, namely, supervised feature learning, semi-supervised feature learning, and unsupervised feature learning. Among them, supervised feature learning can be called classification. Semi-supervised feature learning is between supervised feature learning and unsupervised feature learning, referring to the existence of labeled data and unlabeled data in the training data. Unsupervised feature learning is also called clustering [24].

2.4. Experimental Data

2.4.1. Data Selection. The datasets used in this study were collected and curated by researchers at the Open University and relevant content and materials for courses were delivered through a virtual learning environment (VLE). Hereby, we extracted two different courses A and B on ideological and political education from this platform for research.

2.4.2. Data Preprocessing. The original clickstream data of students collected from the Open University Learning Analysis Dataset (OULAD) were used as features for analysis. Since there are some redundant features in the original feature set, we removed these features, filled in the missing values, and analyzed certain features in sequence. Then, the existing features were summed and averaged to characterize the available learning into four categories: demographics, interaction behavior, registration information, and evaluation information. In terms of data selection, we selected two representative courses in the Open University in 2018 and 2019. The selected course presents more than 300 students with at least two lesson videos in the course and a large number of students in the course who failed their final grades. For all courses that meet these criteria, we have selected courses A and B, with a total of 846 students in the selected courses. Finally, we anonymized the data in accordance with the ethical and privacy requirements applied in the Open University.

3. Experimental Results and Analysis

3.1. Analysis of Different Learning Motivation Clusters and Learning Behavior Characteristics. This article used two courses, A and B, to examine the association between motivation, student engagement, and final performance. After preprocessing the data, a total of 744 students enrolled in these two courses, with 363 enrolling in course A and 381 enrolling in course B. In order to determine the motivation level of students as early as possible based on the interaction data and demographic information of the virtual learning environment, we used the interaction data before the first job submission. In Table 1, we calculated the distribution of final grades for students in both courses who did not submit their first assignment.

According to Table 1, if a student does not submit their first assignment, the probability of failing the course is greater than 90% (withdrawing is also considered a failure), making first assignment submission an important factor in final grade prediction. However, considering that students' learning motivation is largely determined by their demographic information, simply submitting the first assignment is not sufficient to understand the different behavioral patterns of students for the purpose of predicting learning motivation, and their behavioral data and learning styles recorded in the online learning platform system shall prevail. To this end, we integrated and analyzed students' behavioral data, demographic information, assessment quiz information, and course registration information recorded by the virtual learning environment before submitting the first assignment, such as Homepage Clicks, Average Clicks, Forum Clicks, and so on. This article kept all the details of this information in the virtual learning environment. Table 2 shows the characteristics and associated descriptions used in this study.

This section examines the quantitative statistics of relevant features in each of the different learning motivation

Table 1: Final performance distribution for students who did not submit the first assignment.

Course	Withdrawn	Fail	Pass	Prob of failure
A	272	71	17	0.95
B	266	80	33	0.91

Table 2: Characteristics and description for learning analysis.

Learning characteristics	Signalment
Age	Learner gender
Highest education	Learner age
Region	The living areas of a learner when taking the course
Forum clicks	Number of times learner visits forums
Resource clicks	Number of times learner visits lecture notes and slides
Homepage clicks	Number of times learner visits the first page of a lesson
Average clicks	Average number of times learner visits different resources
Average delayed days	Average number of days delayed in submitting assignments
Final performance	Learner final performance
Activity type	This includes watching course videos, browsing websites, or quizzes
Date registration	The exact time students register for the course online
Assessment weight	Represents the weight of each assessment score in the student's final score
Assessment type	It could be a daily quiz, a midterm exam, or a final exam

classes and compares these feature values in the same group. This article described the level of participation of each variable in different categories by mean (Mean) as well as standard deviation (Std). Table 3 describes the differences in indicators of different characteristics of each group. As you can see from the table, Avg delay represents the level of procrastination of students in different motivation groups, and the results show that students in the IMFS and UMS groups are more likely to delay submitting assignments than the remaining two groups. The larger values for the standard deviation of the two groups of students (Std = 12.11 and Std = 10.37) indicate that the two groups of students are weak in time management and control of learning systems. Conversely, students in the IMPS and EMS groups have nearly the same standard deviations, which are (Std = 8.91 and Std = 8.23), indicating that students in the two groups always submit assignments at similar times or study at a fixed rate. In the first job, the average scores for IMPS group, IMFS group, EMS group, and UMS group are Mean = {70.55, 65.72, 79.91, 71.87}, respectively. It can be found that students in the UMS group have a higher average score than students in the IMFS group, indicating that students in the UMS group work harder than students in the IMFS group from the beginning of the course to the submission of the first assignment, with higher grade point averages and higher assignment scores. However, it is also concluded that students in the UMS group are more likely to delay submitting assignments, and as course content becomes more complex, their interest in learning is more likely to be affected. For students in the EMS group, they scored the highest average on their first assignment and studied at a relatively stable pace during this time. These quantitative and other data analysis results are consistent with our previous analytical modeling of learning motivation. Compared with using deep learning algorithms to analyze multidimensional clickstream data and cluster different learning motivations, it is more explanatory to analyze different learning motivation groups from statistical dimensions.

3.2. Experimental Results and Analysis of Learning Motivation Clustering Model Based on GMM and SAE. In this section, we compared the relationship between the clustering results of learning motivation and students' final grades to verify our clustering results, and the final clustering results of different learning motivations are shown in Figure 1. The relevant clustering results are usually hierarchical or partitioned, and partitioning makes it easier to show each cluster in a 2D plot rather than a dendrogram, so this paper used the partition method to display the final clustering results. These components are calculated based on the maximum possible variance of the variables used in order to show as many defects as possible in the data.

It can be seen from Figure 1 that the clustering results of the cluster scatter diagram of learning motivation and students' final grades are roughly divided into three groups, and there may be a small amount of overlap between these clusters. Among them, the black clustering can be interpreted as a class of groups with extrinsic learning motivation. These students have strong learning goals and objectives, and most of them are rewarded or certified for high scores. In addition, the extrinsically motivated group is the largest group, and most participants in this group completed the course, with most of their final grades tending to be above 80 and very few students failing their course. This is consistent with the definition of extrinsic learning motivation as explained. In contrast, most of the students in the red clustering have less than 60 points in their academic performance, and the number is smaller than that of the black class. This may be due to the fact that some of this group of students do not have final grades, and the ultimate goal of them is not to get high grades or to get an award

Table 3: Characteristics and comparison of participation degree of different learning motivation groups.

Learning motivation	Avg clicks Mean/std	Resource clicks Mean/std	Score of the first assignment Mean/std	Avg delay Mean/std
IMPS	49.22/44.01	40.44/35.35	70.55/8.96	0.37/8.88
IMFS	25.56/19.78	19.87/13.61	65.72/13.09	3.43/12.11
EMS	87.81/65.98	80.21/59.42	79.91/7.89	2.44/8.13
UMS	86.12/59.98	80.36/60.87	71.87/10.94	8.23/10.37

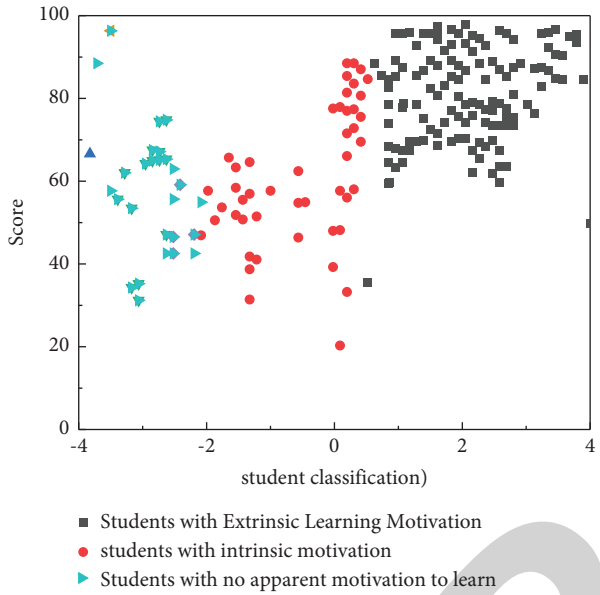


Figure 1: The relationship between clustering results and students' final scores.

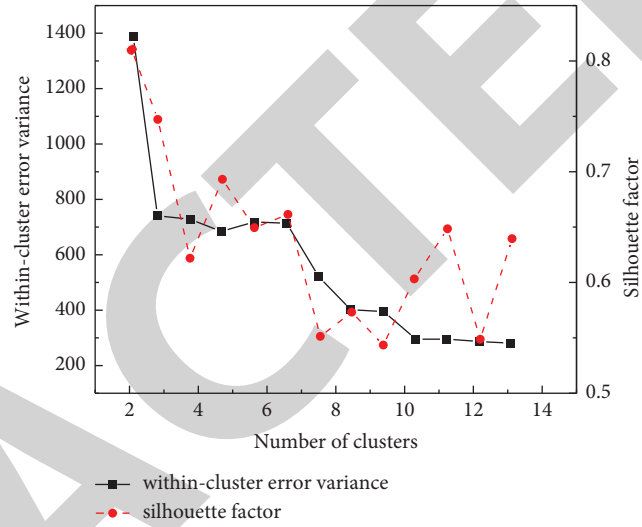


Figure 2: The influence of intra-cluster error variance (SSE) and silhouette coefficient on the optimal number of clusters.

certificate, whereas they may be interested in a certain knowledge point or want to discuss and exchange with other scholars in the forum of the learning website. We call them intrinsically motivated students. The learning behavior of this group of students is often driven by intrinsic motivations such as emotion, interest, or temporary curiosity and is rarely affected by external motivations. Most of the students with cyan dots in the figure have no grades and no obvious distribution rules. We classify them as students with other motivations or no obvious motivations. This group of participants is less active on online learning platforms. Only a small number of students' final results are counted, and the attrition rate is very high, which is the smallest group among the three groups.

3.3. Influence of Intra-Cluster Error Variance and Silhouette Coefficient on Model Performance. Since the learning features analyzed do not belong to the previously defined categories, this study used internal evaluation indicators to analyze the performance of the clustering model. In order to determine the optimal number of clusters for the clustering results, this section uses the error variance within the cluster and the silhouette coefficient to verify the reliability and validity of the clustering results, and the results are shown in Figure 2. The intra-cluster error variance (SSE) is used as the performance index [25]. The smaller the index value is, the higher the convergence degree of each cluster is. However, the value of intra-cluster error variance (SSE) is not as small as possible because there is an extreme case that all sample points are regarded as a cluster, in which the error variance within the cluster is 0, and the final classification effect is obviously not achieved. Therefore, it is necessary to seek a balance between the number of clusters and the variance of intra-cluster error variance (SSE). The elbow method solves this balance problem [26]. Assuming that there is an initial value K, we defined it as the largest possible number of clusters and then incremented the number of clusters from 1. In fact, the data have an underlying pattern, that is, there is a real optimal number of clusters, and when the number of clusters set by the model continues to approach this value, the intra-cluster error variance (SSE) will decrease rapidly. However, when the number of clusters set exceeds the actual optimal number of clusters, the rate of decrease of the intra-cluster error variance will become slow. By analyzing the changes in the value of the descending derivative, the final optimal cluster can be determined.

In general, the higher the average silhouette coefficient, the better the quality of the clusters. From the average silhouette coefficient in the above figure, the optimal number of clusters should be 2 clusters, and the value of the average silhouette coefficient is the highest at this time [27, 28]. However, according to the elbow rule, when $k = 2$, the correlation curve of the intra-cluster error variance (SEE) is not a rapidly decreasing trend. In other words, it is not an inflection point when $k = 2$. But when $k = 3$, the value of

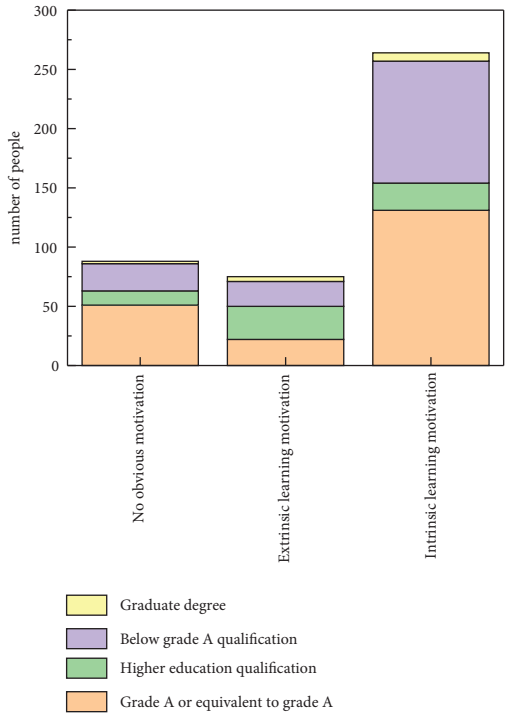


Figure 3: The distribution of educational level in different learning motivation groups.

intra-cluster error variance (SEE) will drop sharply, which is an inflection point, and the value of the average profile coefficient is also relatively high at this time, only slightly lower than the value at $k = 2$. Therefore, when comprehensively analyzing the average silhouette coefficient and intra-cluster error variance, $k = 3$ is the best choice for the number of clusters. The final clustering results obtained in this paper are consistent with the type of motivation defined by SDT theory, which indirectly proves the rationality and interpretability of this method. The influence of intra-cluster error variance (SSE) and silhouette coefficient on the optimal number of clusters is shown in Figure 2.

3.4. The Correlation Experiment between Learning Motivation and Educational Level. Except for the above, this paper also analyzes the relationship between different educational levels and learning motivation. In this dataset, the educational attainment of the participants is broadly classified into 4 categories: "A level or equivalent," "higher education (HE) qualification," "below A level," and "postgraduate qualification." Figure 3 shows the distribution of students with different educational levels in different motivation groups. As can be seen from the graph, students with "A grade or equivalent" and "below A grade" educational levels make up about 70% of the student body. In the group of students with extrinsic motivation, the distribution of students of the four education levels is relatively even. In the group of students without obvious motivation (no motivation), students with "higher education (HE) qualification" and "postgraduate qualification" educational levels account for a small proportion of the total students, which is in line with the fact that most of the students with higher educational levels will have strong learning motivation or learning goals.

Table 4: Chi-squared test for comparing different educational levels with different motivation groups in courses A and B.

Course	χ^2 test	df	P value
A	6.442	6	0.371
B	6.442	6	0.349

Table 4 illustrates the chi-square test for different educational levels for different motivation groups, and the χ^2 statistic tests the difference between the expected and true distribution of different educational levels for different motivation groups. Based on this, we used a table to calculate the P value, with $P < 0.05$ indicating that the null hypothesis is rejected, and there is a significant correlation between educational background and learning motivation. The parameter df refers to degrees of freedom, which represents the number of independent values that vary in the calculation. Because both course A and course B have 4 different educational levels and three different sets of motivations for learning, the degree of freedom is 6. According to Table 4, the P values for both courses are greater than 0.05. Therefore, it is concluded that there is no significant relationship between educational attainment and learning motivation, accepting the null hypothesis.

4. Conclusion

By understanding the motivation for learning, it can help the reform of the ideological and political education system and methods of colleges and universities, and also help the teachers to better understand the learning progress and situation of the learners, so as to timely prescribe the right medicine according to the specific problems and intelligently intervene in different learners. On the basis of SDT, this paper analyzed and modeled different learning motivations in the early stage of the course. Meanwhile, the study in this paper compared the engagement characteristics of different learning motivation clusters, studied the differences of course performance between student groups with different learning motivations and the potential link between learning motivation and learners' educational attainment, and clarified the relationship between different student groups inside and outside the school and demographic information such as age distribution. The main conclusions are as follows:

(1) The study proves that students' learning behavior is an important indicator to identify students' different learning motivations, and different learning motivations are significantly related to students' final ideological and political performance. Most of the students with intrinsic learning motivation have higher ideological and political academic performance, and some students with unclear motivation also have high academic performance. This may be related to other learning data not recorded outside the online learning platform, which can be further analyzed in future studies.

(2) This paper proposes a deep clustering algorithm combining GMM and SAE based on the demographic information and clickstream data about learning behavior recorded in the Open University virtual learning environment. The optimal number of clusters is determined by two internal evaluation indexes of clustering experiments, that is, silhouette coefficient and intra-cluster error variance. The maximum value of the silhouette coefficient is 0.82, and the intra-cluster error variance is less than 1220 under the number of clusters from 2 to 14, with the minimum value of 198.

(3) The deep clustering model proposed in this paper is compared with the clustering models in other papers, and a comparative study is carried out from multiple perspectives. The results show that the model greatly improves the accuracy of clustering results through in-depth analysis of clickstream data, and the statistical test values of χ^2 in A and B courses are both 6.442, and the P values are only 0.371 and 0.349, respectively, which are greater than 0.05. The results further show that the model can more accurately analyze the learning motivation of different student groups and make up for the shortcomings in other clustering algorithms.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] Y. Shi, M. Wang, and T. Zhang, "Considerations of the subject teaching method applying in the information technology course [M]," *Advanced Research On Computer Education, Simulation And Modeling*, pp. 385–391, Springer, Berlin, Germany, 2011.

[2] J. Lin and I. Department, *Innovation of Ideological and Political Teaching in the New Perspective of "We Should Adjust According to the Situation and Times"*, Journal of Hubei Correspondence University, Wuhan, China, 2018.

[3] D. R. Garrison and T. Anderson, "E-Learning in the 21st century: a framework for research and practice [M]," *A Framework for Research and Practice*, Routledge, England, UK, 2003.

[4] V. Kavitha and R. Lohani, "A critical study on the use of artificial intelligence, e-Learning technology and tools to enhance the learners experience," *Cluster Computing*, vol. 22, no. S3, pp. S6985–S6989, 2019.

[5] T. J. Dunn and M. Kennedy, "Technology Enhanced Learning in higher education; motivations, engagement and academic achievement," *Computers & Education*, vol. 137, no. AUG, p. 104, 2019.

[6] X. Xu, J. Wang, and H. Peng, "Prediction of academic performance associated with Internet usage behaviors using machine learning algorithms [J]," *Computers in Human Behavior*, vol. 98, 2019.

[7] D. Filvà, M. A. Forment, and F. GarcíA-PeñALVO, "Clickstream for learning analytics to assess students' behavior with Scratch [J]," *Future Generation Computer Systems*, vol. 93, 2019.

[8] M. Firat, H. Kilinc, and T. V. Yuzer, "Level of intrinsic motivation of distance education students in e-learning environments," *Journal of Computer Assisted Learning*, vol. 34, no. 1, pp. 63–70, 2018.

[9] P. Barba, G. E. Kennedy, and M. D. Ainley, "The role of students' motivation and participation in predicting performance in a MOOC [J]," *Journal of Computer Assisted Learning*, vol. 32, no. 3, pp. 218–231, 2016.

[10] K. M. Law, V. C. Lee, and Y. T. Yu, "Learning motivation in e-learning facilitated computer programming courses," *Computers & Education*, vol. 55, no. 1, pp. 218–228, 2010.

[11] E. W. Cheng and D. C. Ho, "The influence of job and career attitudes on learning motivation and transfer," *Career Development International*, vol. 6, no. 1, pp. 20–28, 2001.

[12] M. Standage, J. L. Duda, and N. Ntoumanis, "A test of self-determination theory in school physical education," *British Journal of Educational Psychology*, vol. 75, no. 3, pp. 411–433, 2005.

[13] J. Park, M. Lee, and H. J. Chang, "Symmetric graph convolutional autoencoder for unsupervised graph representation learning," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, November 2020.

[14] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004 ICPR 2004*, Washington, DC, USA, August 2004.

[15] P. Beerli and J. Felsenstein, "Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach," *Proceedings of the National Academy of Sciences*, vol. 98, no. 8, pp. 4563–4568, 2001.

[16] I. Chowdhury and Y. Watanobe, "Classification of online judge programmers based on Rule extraction from self organizing feature map," in *Proceedings of the 9th IEEE International Conference on Awareness Science and Technology (iCAST)*, Fukuoka, Japan, September 2018.

[17] M. C. Goulden, E. Gronda, and Y. Yang, "CCVis: visual analytics of student online learning behaviors using course clickstream data [J]," *Electronic Imaging*, 2019.

[18] C. Saunders, M. O. Stitson, and J. Weston, "Support vector machine [J]," *Computer intelligence*, vol. 1, no. 4, pp. 1–28, 2002.

[19] N. Arica and F. T. Yarman-Vural, "A shape descriptor based on circular hidden Markov mode," in *Proceedings of the 15th International Conference on Pattern Recognition ICPR-2000*, Barcelona, Spain, September 2002.

[20] M. Hoffman, D. M. Blei, and F. R. Bach, "Online learning for latent dirichlet allocation," in *Proceedings of the International Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 2010.

[21] S. Shi and Q. Tong, "Research on Optimization Algorithm of auto-encoding neural network applied to rolling bearing fault diagnosis," *Journal of Physics: Conference Series*, vol. 1871, no. 1, Article ID 012078, 2021.

[22] Y. Wu, M. Zeng, and Z. Fei, "KAICD: a knowledge attention-based deep learning framework for automatic ICD coding [J]," *Neurocomputing*, vol. 469, 2020.

[23] S. Guan, M. Lei, and H. Lu, "A steel surface defect recognition algorithm based on improved deep learning network model

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 7536330, 8 pages
https://doi.org/10.1155/2022/7536330

Research Article

A Retrospective Analysis on the Effects and Complications of Endoscope-Assisted Transoral Approach and Lateral Cervical Approach in the Resection of Parapharyngeal Space Tumors

Danni Guo, Changling Sun, Xiao Yin, Hongyang Qu, Bingyi Dai, Lin Hu, Chen Zhou, and Xiaodong Du

Department of Otorhinolaryngology-Head & Neck Surgery, Affiliated Hospital of Jiangnan University, Wuxi 214122, China

Correspondence should be addressed to Xiaodong Du; 9812015213@jiangnan.edu.cn

Received 10 June 2022; Revised 11 July 2022; Accepted 18 August 2022; Published 23 September 2022

Academic Editor: Wenming Cao

Objective. To observe and compare the effects and complications of endoscope-assisted transoral approach and lateral cervical approach in the resection of parapharyngeal space (PSS) tumors. *Methods*. From January 2013 to September 2021, 69 patients with parapharyngeal space tumors in the Affiliated Hospital of Jiangnan University were divided into the control group ($n = 37$) and the observation group ($n = 32$) according to the mode of operation. The tumors in the parapharyngeal space were resected by the lateral cervical approach in the control group, and the tumors in the parapharyngeal space were removed by endoscopy-assisted transoral approach in the observation group. The general clinical data and operation conditions of the two groups, including operative blood loss, operation time, drainage volume and drainage time, hospital stay, perioperative pain degree, and tumor resection rate were collected and analyzed statistically. The patients were followed up for 6 months, and the complications of the two groups were recorded. *Results*. Compared with the control group, the operation time in the observation group was significantly shorter, and the amount of intraoperative bleeding in the observation group was significantly less than that in the control group, and the difference was statistically significant ($P < 0.05$). The postoperative drainage was less and the hospital stay in the observation group was shorter than that in the control group, and the difference was statistically significant ($P < 0.05$). There was no significant difference in tumor resection rate between the two groups. The visual analog scale (VAS) score on the 1st and 3rd day after operation in the observation group was lower than that in the control group. After treatment, some patients in the two groups had complications such as nerve injury, dysphagia, difficulty in mouth opening, massive hemorrhage, and parotid fistula. The total incidence of complications in the observation group was lower than that in the control group, and the difference was statistically significant ($P < 0.05$). *Conclusions*. The effect of the endoscope-assisted transoral approach is similar to that of the lateral cervical approach in the resection of tumors in parapharyngeal space, but the endoscope-assisted transoral approach has shorter operation time, less intraoperative bleeding, and less postoperative drainage. The indwelling time and hospital stay of the drainage device were shorter than those of the patients with transcervical approach, and the perioperative stress response of patients with endoscope-assisted transoral approach is mild, which is beneficial to the physical and mental recovery of the patients.

1. Introduction

Parapharyngeal space (PSS) tumors account for about 0.5% of head and neck tumors and generally occur in adults, 30–50 years old account for about 50%, occur in childhood is very rare, female incidence is more than male [1]. There are many kinds of tumors in PSS, with more than 70 kinds, but benign tumors account for 70–80%, malignant tumors account for 20–30%, including primary tumors, direct spread tumors, and distant metastatic tumors. It is recognized that the more common PSS tumors are pleomorphic adenoma, schwannoma, and paraganglioma [2, 3]. Considering the concealment of PSS, it is often difficult to detect PSS tumors in the early stage. It is generally believed that PSS tumor volume larger than 2.5 cm will lead to clinical symptoms, such as upper neck mass, lateral pharyngeal wall or glossopalatine arch surface eminence, throat foreign body sensation, sore throat, cough, dysphagia, tinnitus, snoring,

and even dyspnea [4]. With the growth of the tumor, it will also lead to corresponding neurological dysfunction, characterized by hoarseness, ambiguous speech, limited mouth opening, Horner syndrome, and so on, which seriously affect the life of the patients [4]. The diagnosis of PSS tumors needs imaging examination for preoperative evaluation, according to the typical imaging features of the tumor to make a higher accuracy of preoperative diagnosis. The clinical diagnosis has a high coincidence rate. Ultrasound and CT-guided fine biopsy can make a preliminary judgment of the tumor before open operation and choose the appropriate mode of operation [5].

With regard to PSS tumors, complete surgical resection is the optimal choice for the therapeutic strategy [6]. Due to the diversity of tumor types and the complexity of anatomical structure in PSS, it is very important to choose a reasonable surgical approach in order to avoid neurovascular damage and even functional disorders [6]. In the continuous exploration of many clinicians and scholars, there are more surgical approaches, including lateral cervical incision, cervicomandibular approach, transcervical-parotid approach, transcaliber approach, and so on. In recent years, with the gradual development of endoscopic technology, clinicians have also investigated the feasibility of endoscopic-assisted transcaliber approach in the treatment of PSS tumors [7, 8]. With the further research of technology, endoscope-assisted minimally invasive surgery can remove many large and deep tumors in PSS. However, all kinds of surgical approaches have their advantages and disadvantages. According to the application experience of clinicians at present, it is acknowledged that the external cervical approach can fully expose the surgical field, and the important blood vessels and nerves can be protected to a certain extent, but due to the large scope of the operation and great trauma to the tissue, such as removal of the submandibular gland and amputation of the mandible in part of the operation, there are many postoperative complications, and there are scars on the face, influencing the appearance [9]. The trauma of the transcaliber approach is less than that of the external cervical approach, and there is no scar on the face, but due to the lack of surgical field, it has a certain blindness in separating the tumor and its surrounding tissue, and the scope of application is relatively limited. Traditionally, it is considered to be only suitable for relatively small tumors near the pharyngeal cavity [10]. Endoscope-assisted transoral minimally invasive treatment of PSS tumors avoid mandibular amputation, important neurovascular injury, and facial scar [11]. However, there are few clinical comparative studies on cervical incision and endoscope-assisted transcaliber approach in the treatment of PSS tumors. In order to provide references and suggestions for the selection of surgical approaches for PSS tumors, this study selected the complete clinical data of patients with PSS tumors and established relevant inclusion and exclusion criteria to exclude other interference factors other than the influence of surgery, so as to make the comparison more reasonable. Based on the above, the purpose of this article was to observe and compare the efficacy and complications of endoscopic-assisted transoral and lateral cervical approaches in parapharyngeal space (PSS) tumor resection.

2. Patients and Methods

2.1. Clinical Information. From January 2013 to September 2021, 69 patients with parapharyngeal space tumors in the Affiliated Hospital of Jiangnan University were divided into the control group ($n = 37$) and the observation group ($n = 32$). All patients were diagnosed based on imaging manifestations and pathological findings. There were 18 males and 19 females in the control group. The patients in the control group ranged in age from 15 to 76 years, and the average age was (48.05± 15.62) years. Their course of disease ranged from 2 months to 36 months with an average of (12.11 ± 9.27) months. In the observation group, there were 14 males and 18 females. The patients in the observation ranged in age from 22 to 73 years, and the average age was (51.78 ± 13.89) years. The course of disease ranged from 1 months to 24 months with an average of (10.06 ± 6.54) months. The main clinical symptoms of the patients were neck swelling and pain, sleep snoring, hoarseness, foreign body sensation in the throat, dysphagia, inarticulate speech, tinnitus, and tightness, and even some cases had symptoms of holding breath and dyspnea in the lateral position, and a few patients had no obvious clinical symptoms. A mass was found inadvertently or during physical examination. Main signs: upper neck or submandibular mass; soft palate eminence of gloss palatine arch; protuberance of lateral wall of pharynx; and all patients underwent double diagnostic examination of oral cavity and neck. Auxiliary examinations included neck CT and/or MRI, polysomnography, and esophageal barium meal radiography, all of which were diagnosed as parapharyngeal space tumors before operation. Inclusion criteria (1) patients with parapharyngeal space tumors were diagnosed by cervical CT and/or MRI, polysomnography, esophageal barium meal radiography, and pathological examination; (2) it accords with the indication of surgical treatment; and (3) the patients and their families informed consent and signed the informed consent form. Exclusion criteria: (1) patients with other neurological diseases, nasopharynx, oral, esophageal tumors, and cognitive impairment that may lead to difficulty in mouth opening; (2) patients with severe malnutrition, malignant tumors, systemic immune system diseases, or major organ dysfunction and elderly patients; (3) patients with severe heart, kidney, liver, and other important organ dysfunction; and (4) do not want to participate in this study.

2.2. Treatment Method. The tumors in the parapharyngeal space were resected by the lateral cervical approach in the control group, and the tumors in the PSS were resected by endoscope-assisted transoral approach in the observation group.

Lateral cervical approach: behind the cervicomaxillary angle, from the top of the mastoid, down along the anterior edge of the sternocleidomastoid muscle to the level of the great angle of the hyoid, cut open the skin and subcutaneous

tissue. Lifted the platysma myocutaneous flap, exposed the inferior pole of the parotid gland and the submandibular gland, pulled the sternocleidomastoid muscle backward, exposed the posterior abdomen of the digastric muscle, carefully dissected the internal carotid artery, external carotid artery, internal jugular vein, and vagus nerve; dissected the marginal branch of the facial nerve at the inferior margin of the mandible and the facial artery; separated the ancient inferior nerve from the deep surface of the digastric muscle; and dissected and protected the accessory nerve at 1/3 on the posterior edge of the sternocleidomastoid muscle. The tumor could be fully exposed by cutting off the styloid hyoid muscle and the posterior abdomen of the digastric muscle. If the tumor was located in the upper part of the PSS near the skull base, part of the mandibular angle bone should be removed (generally, the width of the resection area should not exceed 1/2 of the width of the mandible and the length should not exceed 3 cm) to expose the tumor completely. If it is close to the tumor capsule and separated carefully, the tumor can be removed sharply. After resection, the operative cavity was washed with iodophor and normal saline, negative pressure drainage was placed, and the incision was sutured layer by layer. The drainage tube stayed for two to three days. This surgical method is suitable for the resection of most PSS tumors, including some limited malignant tumors.

The operation of endoscope-assisted transoral approach: before operation, patients were treated with shoulder pads in supine position, head tilted back, and Davis mouth opener was placed through mouth to expose oropharynx. The front end of the red children's catheter was placed into the bottom of the nose through the anterior nostril, reaching back to the posterior nostril to the nasopharynx, down to the mouth, pulling the catheter out of the mouth, knotting and fixing both ends of the catheter at the upper lip, so as to pulled the soft palate forward to better expose the field of operation. Under nasal endoscope, at the tonsillar fossa of the lateral wall of the pharynx, a low-temperature plasma scalpel (Arthrocare Corporation EIC 5874-01, USA) was used to make a longitudinal incision from top to bottom. Incision of mucous membrane and submucous tissue (if the tonsil was too large and affected the visual field of operation, tonsillectomy should be performed first), exposed the superior constrictor of pharynx, dissected and separated the superior constrictor of pharynx, reached the surface of the tumor, closed to the tumor and separated the tumor capsule with plasma scalpel (COBLATE Level 7), the separation sequence should be paid attention to (see Figure 1). In order to avoid damaging the important blood vessels and nerves around the tumor due to blocking the visual field, we started from the inner side of the tumor, then turned to the front of the tumor, then separated the upper and lower parts of the tumor, and finally separated the outer side and back of the tumor. The separation should be gentle, avoid violence, and be carefully separated along the gap between the tumor capsule and the surrounding tissue. If adhesion between the capsule and surrounding tissue is found during separation, carefully separate it with vascular forceps, cut it open with a plasma scalpel under endoscope, and be operated by two surgeons. If necessary, the assistant could hold the

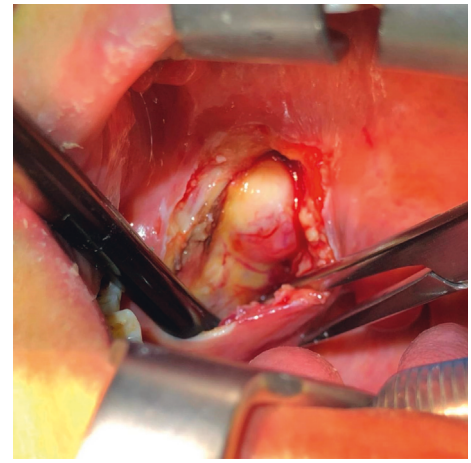


Figure 1: Surgical field of direct view in the endoscope-assisted transoral approach.

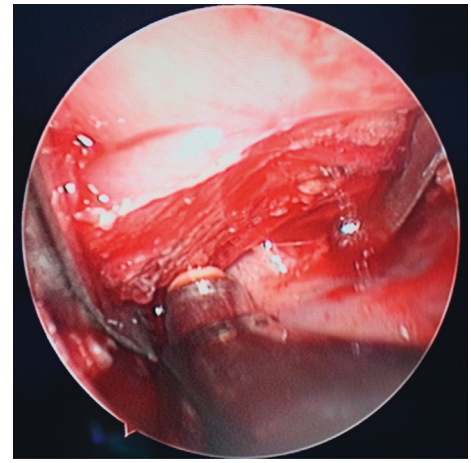


Figure 2: Surgical field of endoscopy in the endoscope-assisted transoral approach.

endoscope, and the operator could operate with both hands. During the operation, the assistant held the suction device to suck the field bleeding in time to ensure the clarity of the operation field under the microscope (see Figure 2). If the amount of bleeding was more, you could use gauze to stop the bleeding first, and after the tumor was completely removed, carefully looked for the bleeding point and stopped the bleeding with plasma scalpel (COAG Level 3). If the size of the tumor was too large to be completely removed, the tumor could also be removed step by step. After tumor removal, we carefully checked whether there was bleeding in the operative cavity, rinsed the operative cavity alternately with iodophor and normal saline, and sutured the muscle and the mucous membrane of the pharyngeal wall layer by layer. The iodoform gauze was placed in the operative cavity according to the intraoperative conditions to drain the exudate from the operative cavity, stayed for two to three days. After operation, cephalosporins and tinidazole were given intravenously to prevent infection and, if necessary, corticosteroids were used intravenously to reduce edema in the operative area.

Surgical instruments and equipment: endoscopy and monitoring using KarlStorz company of Germany nasal endoscope products (0°mirror, diameter 4.0 mm, video surveillance system); low-temperature plasma scalpel (Arthrocare Corporation EIC 5874-01, USA); ultrasonic knife (Johnson Company HAR9F, USA); unipolar needle electric knife; and conventional surgical instruments.

2.3. Observation Index. (1) Operating time, (2) the amount of intraoperative bleeding, (3) postoperative drainage and drainage time, (4) instruct the patient to spit out oral secretions to estimate the amount of bleeding, (5) the amount of fluid and bleeding in the external cervical drainage tube, (6) the time from operation to discharge, and (7) visual analog score (VAS) was used to evaluate the pain degree of perioperative patients [12]. Using the 10 cm swimming scale, the 0 end indicates no pain, the 10 end indicates the most severe pain, and the score range is 0–10 points. The higher the score, the more severe the pain (8) tumor resection rate and (9) complications occurred within 6 months after operation.

2.4. Statistical Analysis. SPSS 23.0 statistical software was adopted to process the data. The measurement data were presented as $(\overline{x} \pm s)$. The group design t-test was adopted for the comparison, and the analysis of variance was adopted for the comparison between multiple groups. Dunnett's t-test was adopted for comparison with the control group. The counting data were presented in the number of cases and the percentage. χ^2 test was adopted for comparison between groups, and bilateral test was employed for all statistical tests.

3. Results

3.1. Comparison of Postoperative Pathological Results between the Two Groups. The pathological results showed that among the 32 patients in the observation group, there were 8 cases (25.00%) of nerve schwannoma, 6 cases (18.75%) of pleomorphic adenoma, 4 cases (12.50%) of branchial cleft cyst, 4 cases (12.50%) of elongated styloid process, 2 cases (6.25%) of hemangioma, 1 case (3.13%) of Warthin tumor, 1 case (3.13%) of paraganglioma, 5 cases (15.63%) of squamous cell carcinoma, and 1 case (3.13%) of synovial sarcoma; among the 37 patients in the control group, there were 13 cases (35.14%) of nerve schwannoma, 10 cases (27.03%) of branchial cleft cyst, 5 cases (13.51%) of pleomorphic adenoma, 2 cases (5.41%) of lipoma, 1 cases (2.70%) of hemangioma, 1 case (2.70%) of Warthin tumor, 1 case (2.70%) of paraganglioma, 1 case (2.70%) of lymphangioma, 1 cases (2.70%) of squamous cell carcinoma, 1 case (2.70%) of adenocarcinoma, and 1 case (2.70%) of pleomorphic liposarcoma.

The minimum diameter of the tumor in the observation group was 1.0 cm, the maximum diameter was 9.8 cm, and the average diameter was (3.86 ± 1.19) cm . In the control group, the minimum diameter was 1.2 cm, the maximum diameter was 8.0 cm, and the average diameter was (4.25 ± 1.37) cm. There was no significant difference in the above-mentioned indexes between the two groups ($P > 0.05$).

3.2. The Time of Operation and the Amount of Intraoperative Blood Loss between the Two Groups. Compared with the control group, the operation time of the observation group was significantly shorter and the amount of intraoperative blood loss was significantly less than that of the control group ($P < 0.05$). The results are shown in Table 1.

3.3. Comparison of Postoperative Drainage and Hospital Stay. Compared with the control group, the posterior drainage in the observation group was less, and the hospitalization time in the observation group was shorter than that in the control group, and the difference was statistically significant ($P < 0.05$). The results are shown in Table 2.

3.4. Comparison of Tumor Resection and Perioperative Pain. There was no significant difference in tumor resection rate between the two groups. The VAS score of the observation group was lower than that of the control group at 1st day and 3rd day after operation, and the difference was statistically significant ($P < 0.05$). The results are shown in Table 3.

3.5. The Incidence of Complications after Surgical Treatment between the Two Groups. After treatment, complications such as nerve injury, dysphagia, difficulty in mouth opening, massive hemorrhage, and parotid fistula occurred in both groups. The total incidence of complications in the observation group was significantly lower than that in the control group, and the difference was statistically significant ($P < 0.05$). The results are shown in Table 4.

4. Discussion

The position of the PSS is deep, from the base of the skull to the hyoid plane, which is the potential space of the head and neck with complex anatomical relationship [12]. PSS tumors are relatively rare, accounting for about 0.5% of head and neck tumors, including benign tumors (80%) and malignant tumors (20%). The most common is pleomorphic adenoma originating from the parotid gland, followed by neurogenic tumors (including schwannoma, neurofibroma, and paraganglioma) [13]. The early clinical symptoms of PSS tumor are not obvious and are easy to be ignored. Because of its deep location, most patients will be inadvertently found during a health examination. When the diameter of PSS tumor is larger than 3 cm, pharyngeal mass, neck mass, pain, foreign body sensation, dysphagia, and even dysphagia may occur due to tumor invasion of pharyngeal cavity or adjacent tissues [14]. Horner syndrome, which occurs when sympathetic nerves are compressed, is typically characterized by myosis, ptosis, and decreased sweat. Local expansion of the tumor can compress the sympathetic nerve, causing snoring, dyspnea, and other compression symptoms [15, 16]. The previous study unveiled that most patients had no obvious

Table 1: Comparison of operation time and intraoperative blood loss ($\overline{x} \pm s$).

Group	N	Operation time (min)	Intraoperative bleeding volume (ml)
Control group	37	109.86 ± 33.95	162.43 ± 93.70
Observation group	32	91.72 ± 31.24	118.13 ± 78.46
t		2.296	2.110
P		0.025	0.039

Table 2: comparison of postoperative drainage and hospital stay ($\overline{x} \pm s$).

Group	N	Drainage volume (mL)	Hospitalization time (d)
Control group	37	24.05 ± 6.91	6.38 ± 1.54
Observation group	32	15.78 ± 6.43	5.50 ± 1.28
t		5.119	2.557
P		<0.01	0.013

Table 3: comparison of tumor resection and perioperative pain [n (%), $\overline{x} \pm s$].

Group	N	Tumor resection rate		VAS Scoring (Points)	
		Complete resection	Partial residue	1 day after operation	3 days after operation
Control group	37	34 (91.89)	3 (8.11)	3.54 ± 1.44	2.51 ± 1.13
Observation group	32	30 (93.75)	2 (6.25)	2.81 ± 1.49	1.97 ± 0.88
χ^2/t		0.212		2.066	2.189
P		0.644		0.043	0.032

Table 4: Comparison of complications after surgical treatment (N, %).

Group	N	Occurrence of complications					Total incidence rate
		Nerve injury	Dysphagia	Difficulty in opening mouth	Massive hemorrhage	Salivary gland fistula	
Control group	37	5 (13.51)	2 (5.41)	3 (8.11)	2 (5.41)	2 (5.41)	14 (37.84)
Observation group	32	2 (6.25)	1 (3.13)	1 (3.13)	1 (3.13)	0 (0.00)	5 (15.63)
χ^2							4.243
P							0.039

characteristic clinical manifestations, especially in the early stage of the disease. About 36% of the patients had ear tightness and earache, 13% of the patients had dysphagia, 11% of the patients had hearing loss, 10.5% of the patients had hoarseness, and about 6.4% of the patients had facial and lower jaw pain [17], suggesting that clinicians should carefully inquire the medical history of atypical clinical symptoms and make detailed examinations when receiving such patients, so as to avoid misdiagnosis and mistreatment.

Therefore, preoperative examination and evaluation is particularly indispensable for the therapeutics of PSS tumors. The more intuitive method in the diagnosis and examination of tumors in the PSS is to first use the method of joint diagnosis with both hands to understand the size, texture, location, and range of activity of the tumor, such as carotid body tumor can touch the sense of pulsation and have a certain degree of compression, therefore, the tumor can move in the anterior and posterior position but not up and down [18]. However, because of the deep location of PSS, it is difficult to accurately estimate the size of the tumor and its relationship with the surrounding tissue and important structures, so imaging examination is particularly important for the preoperative evaluation of PSS tumors. Through CT or MRI examination, we can understand the degree of tumor involvement and its relationship with the surrounding important tissues, such as the relationship with carotid sheath, skull base, and intracranial invasion, so as to help clinicians to judge in time and determine the operation plan [19, 20].

Surgical treatment still remains the optimal choice for PSS tumors [21]. Due to the various pathological types, deep anatomical location and complex surrounding structure of tumors in the PSS, the selection of appropriate surgical approach is the key to successful operation. In order to formulate the surgical approach before operation, we should focus on the size and nature of the tumor and its relationship with the surrounding important blood vessels and nerves and choose a surgical approach that can fully expose the operative field and remove the tumor completely. Meanwhile, it is the principle of surgical treatment of PSS tumors

that can avoid damage to important blood vessels and nerves and minimize postoperative scars and deformities.

Low-temperature plasma (LTP) technology has emerged in recent years, and its application in the neck has proven to be successful such as tonsillectomy and adenoidectomy. The low-temperature plasma scalpel is composed of a conductive medium (sodium) that generates a high concentration of plasma around the electrode, which is made up of highly ionized particles. Because the current does not generate a large amount of heat through the tissue and does not cause visual damage to the tissue, such particles have quite enough energy to detach the molecular bonds in the material, which in turn differs various molecules and reduces the tissue. At the same time, the characteristics of ultralow temperature and strong hemostatic effect of plasma can completely cut off the tumor tissue. Low-temperature plasma scalpel also has some disadvantages. First of all, it is difficult to decompose the levels and anatomical structures of some tissues during the operation, which will lead to tissue damage. Second, the temperature parameters during the operation are not easy to control, which may lead to vascular injury and tissue perforation, resulting in a series of complications. The incision is tiny and the ablation depth can be carefully regulated due to the extremely low working temperature, causing less injury to the surrounding skin during the procedure.

Transcervical approach is the most commonly used surgical approach for PSS tumors [22]. It is suitable for most patients with PSS tumors, with these advantages: (1) large operating space, directly entering the PSS, better exposing the important blood vessels and nerves in the PSS, and can avoid the injury of parotid gland and facial nerve; (2) can better protect the internal and external carotid arteries; (3) it is easy to separate the blood vessels at the edge of the tumor, and it is convenient to stop bleeding. However, the transcervical approach will bring greater surgical trauma to the patients, and the postoperative incision scar is larger, which will bring psychological burden to the patients. In addition, in order to completely remove the tumor during the operation, it is necessary to cut off part of the nerves and blood vessels, increasing the risk of facial nerve injury. Some scholars believe that the lateral cervical approach is limited in the treatment of tumors and angiogenic tumors located in the medial or upper part of PSS [22].

The resection of tumors in PSS by transoral approach was first proposed by Goodwin and Chandler [23]. The advantages of this approach are direct surgical approach, simple operation, less surgical trauma, early oral feeding, no surgical scar in the neck, and shorter hospitalization time [24]. However, in the past, since the surgical approach was operated in the oral cavity, restricted by lighting equipment and other surgical instruments, the operative field was narrow, it was difficult to expose during the operation, and it was blind to separate the tumor during the operation. Moreover, it is prone to bring about important vascular and nerve injury in the PSS, parotid gland leakage in the PSS, and incomplete tumor resection and other adverse consequences [24]. Some scholars believe that the transoral approach is only suitable for benign nonvascular PSS tumors whose diameter in the anterior styloid space is less than 3 cm, especially those with intact capsule protruding to the oropharynx [24]. However, the operative complications and postoperative discomfort induced by transcervical, transparotid and transmandibular approach were significantly more than those induces by transoral approach [25, 26]. In recent years, with the development of endoscopic technology, since the application of endoscopic technology in clinical surgery has the advantages of good lighting, clear surgical field and wide visual angle, it is possible to remove deep tumors through natural spaces such as nasal cavity and oral cavity as much as possible. Endoscope-assisted transoral approach has been widely used, and it has been paid more and more attention by colleagues [27]. We believe that, considering that most PSS tumors are benign, among the many surgical methods of PSS tumor resection, only endoscope-assisted transoral approach can satisfy the complete resection of the tumor. At the same time, the principle of surgical treatment is to retain important vascular and neurological functions and minimize postoperative scars and deformities. Endoscope-assisted transoral approach not only has the advantages of traditional transoral approach but also solves the shortcomings of narrow visual field and difficult exposure [27].

In this study, there was no significant difference in the complete resection rate and residual rate of tumors in the PSS between the two groups ($P > 0.05$), indicating that the two surgical approaches can effectively remove tumors in the PSS, and the curative effect is the same. However, compared with the transcervical approach, the endoscope-assisted transoral approach had shorter operation time and less intraoperative bleeding, the postoperative drainage volume, indwelling time of drainage device, and hospital stay were less than those of transcervical approach, and the difference was statistically significant ($P < 0.05$). In a meanwhile, the VAS score of endoscope-assisted transoral approach was significantly lower than that of transcervical approach at 1st day and 3rd day after operation. In addition, the endoscope-assisted transoral approach can enlarge the surgical area several times, observe the lesions and microvessels that are easy to be ignored by the naked eye, and better remove the diseased tissue. Moreover, the use of low-temperature plasma equipment, cleaning, suction, electrocoagulation, and cutting into one can effectively reduce bleeding and shorten the time of operation. In contrast, the transcervical approach has a wide field of vision. Double-click electrocoagulation and meridian attractor are used during the operation, which need to cut open the skin, subcutaneous tissue and platysma muscle in turn, and suture after operation. Therefore, endoscope-assisted transoral approach surgery requires shorter time and less trauma, so intraoperative bleeding, drainage device indwelling time, hospital stay are less, and the degree of perioperative pain is relatively mild. However, no matter what kind of operation, psychological factors such as treatment operation and worrying about the consequences of treatment will cause patients' fear and anxiety and induce stress reaction to a certain extent. The stress reaction will lead to accelerated breathing, elevated blood pressure, and a series of changes in hormone levels, which is not conducive to the recovery of patients

[27]. The stress response induced by endoscope-assisted transoral approach surgery is mild, which may be due to the shorter operation time and less trauma of the patients. After treatment, some patients in the two groups had complications such as nerve injury (hoarseness, facial paralysis, and weakness of shoulder lifting), dysphagia, mouth opening, massive hemorrhage, and parotid fistula. The total incidence of complications in the observation group was lower than that in the control group, and the difference was statistically significant ($P < 0.05$). Postoperative hoarseness occurred in 3 cases, which was related to cranial nerve injury in the posterior group, but hoarseness recovered within half a year after operation. 1 case with dysphagia had postoperative pain VAS score as high as 7, no pharyngeal edema and dyspnea occurred after operation, and the symptoms were alleviated within 10 days after operation. 3 cases with dysphagia and difficulty in opening mouth considered postoperative infection and cellulitis in parapharyngeal space, and the symptoms improved within a week after treatment with antibiotics and corticosteroids. 2 cases of massive hemorrhage were primary bleeding within 24 hours after operation, and external carotid artery DSA was performed to stop the bleeding. After operation, 2 cases of parotid fistula were caused by parotidectomy of deep lobe of parotid gland. The use of oral prosthetic membrane during the parotid approach surgery can effectively prevent the occurrence of postoperative salivary gland fistula. 4 cases of facial paralysis were considered as injury of marginal mandibular branch of facial nerve, but did not cause fracture and recovered one month after operation, and 1 case of weakness of shoulder lifting was considered as brachial plexus injury after operation. In the comparison of postoperative complications, the total incidence of complications in the transcervical approach group was higher than that in the transoral approach group. It shows that the risk of complications of endoscope-assisted transoral approach is lower than that of transcervical approach. In addition, there is no scar in the neck after the transoral approach, which plays a great role in promoting the physical and mental recovery of the patients. This study has some limitations: the sample size of this study is small, it belongs to a single-center study, and there is a certain deviation. There are patients' own factors and other confounding factors that may interfere with the accuracy of this study. In future research, we will carry out multicenter, large sample prospective studies, or we can draw more valuable conclusions.

To sum up, endoscope-assisted transoral approach is similar to lateral cervical approach in the resection of PSS tumors. But the endoscope-assisted transoral approach has less operation time, less intraoperative blood loss, and postoperative drainage. The indwelling time of the drainage device and the length of hospital stay of the patients treated by endoscope-assisted transoral approach are also shorter. In addition, the degree of perioperative pain and stress reaction of the patients undergoing endoscope-assisted transoral approach are mild, which is beneficial to the physical and mental recovery of the patients with PSS tumors.

Data Availability

The data sets used and analyzed during the current study can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

ESI Special Fund for Medical School of Jiangnan University (Fund Number: 1286010241170330).

References

[1] A. Moffa, V. Fiore, V. Rinaldi, and M. MoffaMagaldiCasaleCassano, "Management of parapharyngeal space tumor using transoral robotic surgery: the tonsillar fossa battlefield," *Journal of Craniofacial Surgery*, vol. 31, no. 6, pp. 1819–1821, 2020.

[2] G. Grilli, V. Suarez, M. G. Muñoz, and J. L. CostalesLlorente, "Parapharyngeal space primary tumours," *Acta Otorrinolaringologica (English Edition)*, vol. 68, no. 3, pp. 138–144, 2017.

[3] M. Polkowski, J. Lenartowicz, J. Zwoliński, and J. PałuckiMrózSokółReguła, "Transoral endoscopic ultrasound-guided fine-needle biopsy of a tumor of the parapharyngeal space," *Endoscopy*, vol. 53, no. 04, pp. E145–E147, 2021.

[4] F. Bootz, S. Greschus, and T. van Bremen, "Therapie parapharyngealer tumoren," *HNO*, vol. 64, no. 11, pp. 815–821, 2016.

[5] K. Chen, C. Luo, and Y. Tan, "CT/MRI imaging analysis of primary tumor in infratemporal parapharyngeal space [J]," *Journal of practical Stomatology*, vol. 34, no. 3, pp. 407–411, 2018.

[6] F. López, C. Suárez, and V. Vander Poorten, "Contemporary management of primary parapharyngeal space tumors," *Head & Neck*, vol. 41, no. 2, pp. 522–535, 2019.

[7] X. Gong, Y. Wei, Z. Lin et al., *Journal of Clinical Otorhinolaryngology, head and neck surgery*, vol. 35, no. 3, pp. 204–208, 2021.

[8] H. Chen, G. Sun, E. Tang, and Q. G Hu, "Clinical analysis and surgical treatment of 23 cases of primary parapharyngeal space tumors [J]," *Chinese Journal of Stomatology*, vol. 54, no. 2, pp. 107–111, 2019.

[9] H. Chen, Z. He, G. Li et al., "Endoscopy-assisted transoral approach to resect parapharyngeal space tumors: a systematic review and meta-analysis," *The Laryngoscope*, vol. 131, no. 10, pp. 2246–2253, 2021.

[10] S. Fan, S. G. Lin, H. Q. Zhang et al., "A comparative study of the endoscopy-assisted transoral approach versus external approaches for the resection of large benign parapharyngeal space tumors," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 123, no. 2, pp. 157–162, 2017.

[11] l. He, J. Xie, and K. Hui, "Clinical analysis of endoscope-assisted resection of tumors in parapharyngeal space via caliber approach [J]," *Journal of Clinical Otorhinolaryngology, head and neck surgery*, vol. 34, no. 9, pp. 824–827, 2020.

[12] K. H. Lien, C. K. Young, S. C. Chin, and S. F. LiaoHuang, "Parapharyngeal space tumors: a serial case study," *Journal of International Medical Research*, vol. 47, no. 8, pp. 4004–4013, 2019.

[13] R. K. Verma, V. Gautam, A. Bahl, and A. Bal, "Malignant peripheral nerve sheath tumor of the parapharyngeal space arising from cervical sympathetic chain: a rare entity," *Journal of Cancer Research and Therapeutics*, vol. 16, no. 3, pp. 630–633, 2020.

[14] W. M. Mendenhall, P. Strojan, J. J. Beitler et al., "Radiotherapy for parapharyngeal space tumors," *American Journal of Otolaryngology*, vol. 40, no. 2, pp. 289–291, 2019.

[15] W. P. Lao, P. S. Han, N. H. Lee, and J. C. GildeInman, "Transoral excision of parapharyngeal tumors," *Ear, Nose, & Throat Journal*, vol. 100, no. 10, pp. NP454–NP458, 2021.

[16] M. L. Kuet, A. V. Kasbekar, L. Masterson, and P. Jani, "Management of tumors arising from the parapharyngeal space: a systematic review of 1,293 cases reported over 25 years," *The Laryngoscope*, vol. 125, no. 6, pp. 1372–1381, 2015.

[17] K. Luna-Ortiz, J. E. Navarrete-Alemán, M. Granados-García, and A. Herrera-Gómez, "Primary parapharyngeal space tumors in a Mexican cancer center," *Otolaryngology-Head and Neck Surgery*, vol. 132, no. 4, pp. 587–591, 2005.

[18] Y. Fang, H. Wu, A. D. Tan, and L. Cheng, *Acta Oto-Laryngologica*, vol. 140, no. 2, pp. 163–169, 2020.

[19] C. Li, Y. Tian, Y. Shen, and Y. WenHe, "18F-FDG PET/CT findings in a rare myoepithelial carcinoma arising in the parapharyngeal space," *Clinical Nuclear Medicine*, vol. 45, no. 11, pp. 894-895, 2020.

[20] C. Hu, Y. Chen, and X. Cao, "A comparative study of MRI findings and pathology of primary parapharyngeal space tumors [J]," *Chinese CT and MRI Magazine*, vol. 16, no. 1, pp. 40–42, 2018.

[21] H. Chen, G. Sun, E. Tang, and Q. Hu, "Surgical treatment of primary parapharyngeal space tumors: a single-institution review of 28 cases," *Journal of Oral and Maxillofacial Surgery*, vol. 77, no. 7, pp. e1–1520, 2019.

[22] L. Tao, X. L. Shi, X. M. Li et al., "Retrospective analysis of 188 cases of parapharyngeal space tumors [J]," *Journal of Clinical Otorhinolaryngology, head and neck surgery*, vol. 32, no. 2, pp. 129–133, 2018.

[23] W. J. Goodwin Jr and J. R. Chandler, "Transoral excision of lateral parapharyngeal space tumors presenting intraorally," *The Laryngoscope*, vol. 98, no. 3, pp. 266–269, 1988.

[24] S. Lin, X. Xu, and C. Yang, "Comparison of endoscope-assisted transoral approach and extraoral approach in the resection of giant benign tumors in parapharyngeal space [J]," *Chinese Journal of Oral and Maxillofacial surgery*, vol. 18, no. 1, pp. 28–32, 2020.

[25] H. Luo, C. Liu, and C. Chen, "Comparative analysis of the effect of tumor resection in parapharyngeal space between two surgical approaches [J]," *Southwest national defense medicine*, vol. 28, no. 11, pp. 1107–1109, 2018.

[26] L. Li, N. R. London Jr, Y. Gao, and X. CarrauChen, "Endoscopic transoral approach for resection of retrostyloid parapharyngeal space tumors: r," *Head & Neck*, vol. 42, no. 12, pp. 3531–3537, 2020.

[27] J. Wang, W. Y. Li, D. H. Yang, and Y. Y. JinNiu, "Endoscope-assisted transoral approach for parapharyngeal space tumor resection," *Chinese Medical Journal*, vol. 130, no. 18, pp. 2267-2268, 2017.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6114061, 16 pages
https://doi.org/10.1155/2022/6114061

Research Article

DA-ActNN-YOLOV5: Hybrid YOLO v5 Model with Data Augmentation and Activation of Compression Mechanism for Potato Disease Identification

Guowei Dai,[1] **Lin Hu**,[1,2] **and Jingchao Fan**[1,2]

[1]*Agricultural Information Institute of CAAS, National Agriculture Science Data Center, Beijing 100081, China*
[2]*National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China*

Correspondence should be addressed to Jingchao Fan; fanjingchao@caas.cn

Received 9 June 2022; Revised 11 July 2022; Accepted 21 July 2022; Published 23 September 2022

Academic Editor: Wenming Cao

To solve the problems of weak generalization of potato early and late blight recognition models in real complex scenarios, susceptibility to interference from crop varieties, colour characteristics, leaf spot shapes, disease cycles and environmental factors, and strong dependence on storage and computational resources, an improved YOLO v5 model (DA-ActNN-YOLOV5) is proposed to study potato diseases of different cycles in multiple regional scenarios. Thirteen data augmentation techniques were used to expand the data to improve model generalization and prevent overfitting; potato leaves were extracted by YOLO v5 image segmentation and labelled with LabelMe for building data samples; the component modules of the YOLO v5 network were replaced using model compression technology (ActNN) for potato disease detection when the device is low on memory. Based on this, the features extracted from all network layers are visualized, and the extraction of features from each network layer can be distinguished, from which an understanding of the feature learning behavior of the deep model can be obtained. The results show that in the scenario of multiple complex factors interacting, the identification accuracy of early and late potato blight in this study reached 99.81%. The introduced data augmentation technique improved the average accuracy by 9.22%. Compared with the uncompressed YOLO v5 model, the integrated ActNN runs more efficiently, the accuracy loss due to compressed parameters is less than 0.65%, and the time consumption does not exceed 30 min, which saves a lot of computational cost and time. In summary, this research method can accurately identify potato early and late blight in various scenarios.

1. Introduction

Potato is the fourth largest food crop globally after wheat, rice, and corn. At present, with potato cultivation area and total production increasing year by year [1], potato disease problems are receiving increasing attention. Among the many leaf diseases of potatoes, early and late blight are common diseases caused by a fungal infection. Potato leaves infected with early blight will wilt prematurely [2], while potato leaves infected with late blight will produce many green spots. Disease damage increases with changes in the external environment; for example, during the rainy season, when humidity is high and favours rapid fungal growth, spots can proliferate within one to two days, causing potato rhizome to become brittle and cracked [3], seriously affecting potato yields. Therefore, it is essential to establish a rapid and effective method for detecting early and late blight of potatoes.

The traditional method of leaf disease diagnosis relies on subjective human experience and is difficult to replicate and extend. With the advancement of agricultural technology, some challenging, expensive, and time-consuming methods of operation have been proposed. However, such methods require specialized equipment and operational skills that make them universally difficult to adopt [4, 5]. With the rapid development of artificial intelligence to promote precision agriculture, artificial intelligence (AI), machine learning (ML), and computer vision (CV) technologies are used for automatic crop leaf disease detection [6–8], which are time-sensitive and efficient and requires less human

intervention, providing a reliable technical means for accurate detection of crop leaf diseases.

In disease identification under natural conditions, traditional computer vision feature extraction is used as a critical technical link to extract colour, texture, and shape features by HSV colour space combined with support vector machine (SVM), random forest (RF), and artificial neural network (ANN) for disease identification. However, the diversity and complexity of leaf spots under actual conditions and the susceptibility of the features to light conditions, especially the poor stability of the colour features, make this method unsatisfactory for identification [9–11]. Compared with traditional methods, convolutional neural networks (CNN) are rapidly developing, and new types of models are emerging with more substantial expressive power in feature extraction [12, 13], and VGG19, AlexNet, SqueezeNet, InceptionV3, Faster R-CNN, and ResNet50 have achieved better results in disease image detection and classification [14–16]. Tiwari et al. [17] proposed to use a pretrained VGG19 model to compare the performance of classifiers such as k-nearest neighbor (KNN), SVM, neural networks and logistic regression on the same dataset and fine-tune the target results using VGG16, VGG19 and InceptionV3 with the help of transfer learning, with a recognition accuracy of 97%, exceeding the test dataset by 8%. D. F. Wang and J. Wang [18] proposed the crop disease classification model TL-SE-ResNeXt-101, improved the deep residual network model SE-ResNeXt-101, and accelerated the convergence of the model using transfer learning techniques and data augmentation with an average accuracy of 97.99%. Yang et al. [19] proposed to train the Faster R-CNN model using a transfer learning approach and to mark out patch regions. The K-means algorithm clustered the established colour and the SIFT features and then passed them into SVM for disease classification with an average accuracy of 90.83%.

Although the above studies have made good progress in disease image classification, the related research is still only at the exploration stage of specific regional datasets, which is challenging to adapt to the current high requirements and standards of precision agriculture [20]. On the one hand, only the accuracy of scenario-specific datasets is considered, ignoring the need for CNN models to consider the impact of the scale of model parameter storage [21], resulting in the difficulty of training CNN models on resource-limited platforms. On the other hand, most methods do not evaluate their performance on unknown images [22] because of the limitations of the dataset sample, where any version of the model can be marked as good when tested on an unknown data sample. In addition, the use of classical neural network models or two-stage models [23] in training on datasets results in greater consumption of storage and computational resources due to their large number of model parameters or the need to compute the region proposal first, brings much less performance gain than one-stage [24] when integrated with other techniques, and is relatively difficult to maintain and extend later.

In this study, by introducing the one-stage YOLO v5 algorithm, balancing the depth of model width, extracting deep-level disease features, and using ActNN adaptation model activation parameter compression to minimize model storage space without compromising detection accuracy, we aim to design a target model that can identify potato early and late blights in a targeted manner while being trainable in multiple equipment environments with multiple model parameters. This work makes the following three contributions: (1) a method combining weather data augmentation with generic data augmentation is proposed for extending the data. (2) Feature visualization is used to understand the extraction of features by each network layer and to obtain global and local information about the objects of interest in each layer's convolution kernel. (3) YOLO v5 is integrated using ActNN to compress the model training memory and improve the training performance of the model in multiple device environments.

2. Materials and Methods

2.1. YOLO v5 Model Structure. The YOLO v5 is the most representative target detection model in the one-stage series, which has high recognition accuracy and fast inference and avoids the recomputation of candidate regions in the two-stage series. The YOLO v5 architecture contains four main model structures: YOLO v5l, YOLO v5x, YOLO v5m, and YOLO v5s, and the complexity of their networks decreases in order. In order to adapt the solution to cell phones, the YOLO v5n model was later proposed, which has the same model depth and half the network width compared to the YOLO v5s, with only 1.9 MB parameters. In this study, the five models included in YOLO v5 have the same architecture (Figure 1), and YOLO v5 refers to the original network structure if not specified.

The YOLO v5 baseline architecture is mainly composed of the Backbone, Neck, and Head. Figures 1(a)–1(d) show the composition of the modules related to the baseline architecture. One of the Backbone structures is a CNN, which combines different fine-grained images and forms image features. The conv module is the basic convolution unit that runs through the architecture and performs 2D convolution, 2D regularization, and SiLU activation operations [25] on the inputs in turn. The c3 module, as the main structure for extracting features, consists of three conv with one bottleneck, and the composing structure is added to the computational graph in turn. It reduces the model size by solving the problem of large-scale repetitive gradient information and integrating gradient changes in the feature graph, reducing the model floating-point operations per second (FLOPs) and parameters and ensuring the accuracy and speed of inference. The concat module connects feature maps of different dimensions; the upsample model is used to sample feature maps. The spatial pyramid pooling (SPP) module, located at layer 9 of Backbone, is designed to increase the perceptual field of the network by transforming feature maps of any size into feature vectors of fixed size.

The Neck structure increases the flow of information through the assembly line. Low-level features can be efficiently propagated by adding feature pyramid network (FPN), a new feature with bottom-up path enhancement property. Similarly, higher-level features can be fused by

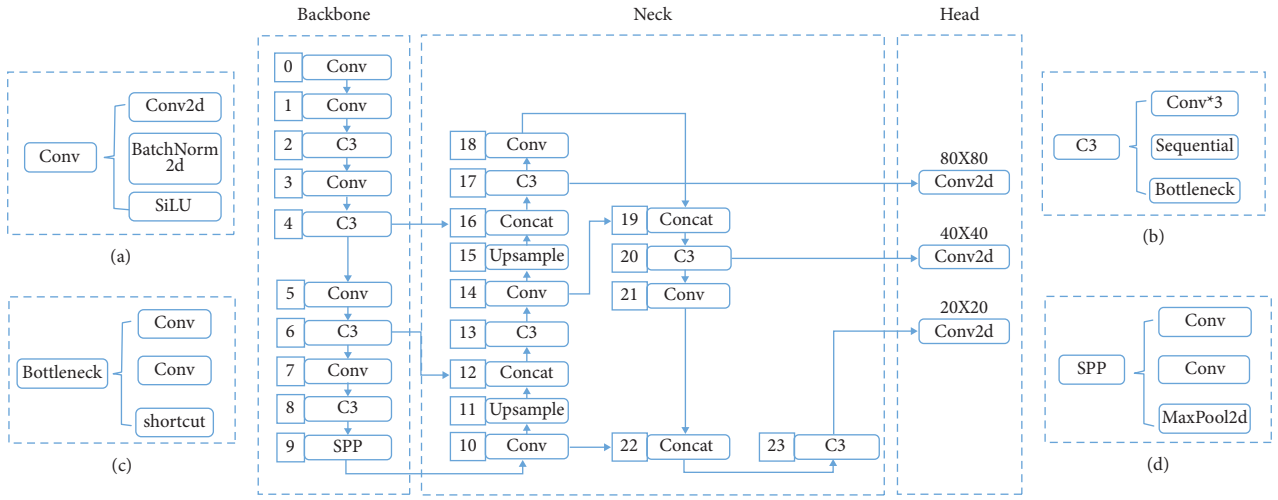


FIGURE 1: Composition and structure of the main modules (a-d) of the YOLO v5 network.

adding path aggregation networks (PAN) to pass features down the line. Feature pyramids, path aggregation networks, and all other features are connected by adaptive pooling to utilise valuable information from each feature layer. The network decides from all layers which features are valid. It improves object localization accuracy by using the correct localization signals from the higher layers with solid semantic features from the lower layers.

The detection network of the Head structure consists of three detection layers, each with an input of a pixel feature map of dimensions 80×80, 40×40, and 20×20, for detecting image objects of various sizes. This study has three detection targets, so each detection layer outputs a 24-channel vector with three categories, one confidence level, four bounding box coordinates, and three anchor boxes $((3+1+4)\times 3)$. The predicted bounding boxes and categories of targets in the original image are generated and labelled to enable the detection of image targets.

2.2. Activation Parameter Compressed ActNN. During training, deep learning models need to store model parameters, intermediate activation results, and optimizer state, which results in an exponential increase in required memory, making it difficult to train large-scale models on a GPU with limited memory. For this reason, Chen et al. [26] proposed ActNN, combining the reduced numerical accuracy BLPA of Chakrabarti and Moseley [27] and the TinyScript and nonuniform quantization strategies proposed by Fu et al. [28]. With ActNN, model parameters can be quickly compressed without negatively impacting prediction accuracy, and common CNN model structures are supported. This ActNN implements a stochastic quantized compressed version of the PyTorch neural network modules commonly used in classification, detection, and segmentation applications.

With ActNN, we implement a dynamic stochastic quantized activation neural network, which minimizes numerical precision by focusing on activation parameter contexts. As a result, training weights, activation parameters, and optimizers can be quantized to reduce numerical precision. The gradient variance convergence is affected by the quantization process, and ActNN contains a mixed accuracy quantization strategy of group quantization and fine-grained quantization, which can approximate the minimization of the gradient variance during the training process with a slight loss in accuracy in the maximum 2-bit quantization case. The model parameters are compressed using the following formula:

$$H^{(l)} = F^{(l)}\left(H^{(l-1)}; \Theta^{(l)}\right), \tag{1}$$

$$\nabla_{H^{(l-1)}}, \nabla_{\Theta^{(l)}} = G^{(l)}\left(\nabla_{H^{(l)}}, C\left(H^{(l-1)}, \Theta^{(l)}\right)\right), \tag{2}$$

$$\begin{cases} \nabla_{H^{(l-1)}} = \nabla_{H^{(l)}} \Theta^{(l)^{\top}}, \nabla_{\Theta^{(l)}} = H^{(l-1)^{\top}} \nabla_{H^{(l)}} \\ C\left(H^{(l-1)}, \Theta^{(l)}\right) = \left(H^{(l-1)}, \Theta^{(l)}\right), \end{cases} \tag{3}$$

where considering each training iteration of the l-layer neural network, let X be the input image and Y be the corresponding label, and one small-batch sample (X, Y) is collected from the dataset, given the input $H^{(0)} = X$ and $H^{(l)}$ as $N \times D^{(l)}$, N is the batch size and $D^{(l)}$ is the number of features; $\Theta^{(l)}$ is the parameter vector, and the forward propagation $F^{(l)}$ contains the N-feature mapping $H^{(l-1)}$ with the model parameters $\Theta^{(l)}$. Given the small-batch loss function $\mathscr{L} = \ell(H^{(L)}, Y)$ for predicting $H^{(L)}$ and label Y, the gradient ∇ is calculated as $\nabla_{\Theta^{(l)}}\mathscr{L}$, $H^{(l-1)^{\top}}$ is the feature mapping transpose, and the model parameters are updated with the optimizer SGD. Since the gradient is always taken together with the loss $\mathscr{L}$, the gradient of the activation parameters is denoted as $\nabla_{\Theta^{(l)}}, \nabla_{H^{(l)}}$. The process of gradient calculation can be expressed as equation (2), where the backpropagation $G^{(l)}$ for the gradient is obtained $\nabla_{H^{(l)}}$; with the context C, C is the information that needs to be saved in the memory for the backpropagation. In essence, this way of keeping $\nabla_{\Theta^{(l)}}$ and $\nabla_{H^{(l)}}$ as full-precision gradient parameters in memory is called full-precision (FP) training. Equation (3) as a particular case, when a linear layer in $H^{(l)} = H^{(l-1)}\Theta^{(l)}$, will carry the exact parameters of the

context. ActNN to achieve 2-bit activation compression, C, $\Theta^{(l)}$, and $\nabla_{H^{(l-1)}}$ of the l-layer neural network are quantized using a random quantization strategy, and the resulting lossy gradient is an unbiased estimate of the original gradient, as shown in the following equation:

$$\% \ \widehat{\nabla}_{H^{(l-1)}}, \widehat{\nabla}_{\Theta^{(l)}} = G^{(l)}\left(\widehat{\nabla}_{H^{(l)}}, \widehat{C}\left(H^{(l-1)}, \Theta^{(l)}\right)\right), \widehat{\nabla}_{H^{(l)}} = \nabla_{H^{(l)}}. \tag{4}$$

A hybrid accuracy quantization strategy is dynamically adjusted by ActNN at runtime to take advantage of the hardware features more effectively. As different network layers have heterogeneous characteristics, the compression algorithm reserves more bits for the most essential activation parameters. By contrast, activation parameters that have a negligible impact on the model's accuracy are compressed using a compression algorithm above the limit level, assigning an average of 2 bits per activation result, maintaining accuracy while reducing activation parameters. Figure 2(b) shows the improvement of ActNN to the actual training process in Figure 2(a). According to Figure 2(b), ActNN defines optional compression parameters for L1 to L5, where L1 and L2 can use 4-bit per-group quantization, but L1 can use 32-bit quantization and only deals with convolutional layers; L3 to L5 use fine-grained mixed precision, swapping, and defragmentation, which acts on the activation parameters of all layers. The effects of the specific processing depend on the proportion of the original model that is processed with ActNN. The processing occurs only during training, and the detection process does not happen. In addition, as shown in (5), the compression algorithm used in L1 to L5 is a superposition of the previous compression levels. When the compression level during the training process is increased, the time it takes to decompress the activation parameters during backpropagation also increases. As a result, the training speed is decreased, even when the hardware conditions remain the same. In terms of the adjustment parameters and the data, increasing the batch size and using high-resolution images both increase the amount of time required for the compression activation (CA) parameter and the decompression activation (DCA) parameter, which in turn slows down the convergence of the model. The process of compressing and decompressing ActNN can be summarized by Algorithm 1.

$$L1 \subsetneq L2 \subsetneq L3 \subsetneq L4 \subsetneq L5. \tag{5}$$

2.3. YOLO v5 with ActNN Integration. In the YOLO v5 network model, the three significant structures are derived directly or indirectly from module under the nn package, an implementation based on PyTorch. The primary operations of these modules include the extraction of features, feature fusion, pooling, batch normalization (BN), and the activation function, which are essential components to measure the number of parameters of each module. With the fine-grained compression of the parameters of constituent modules, optimizations are possible from the model infrastructure without altering the model's functional structure and, consequently, without affecting the model's performance. For most modules, ActNN implements parameter compression. With Qconv and Qconvtranspose, you can create three different sizes of convolutional kernel module versions, while Qbatchnorm performs batch normalization on all three types of modules. In addition, parameter compression is also implemented for the commonly used ReLU, dropout, and maxpool2d operations. As shown in Figures 3(a) to 3(d), ActNN was used to integrate and replace some modules of the original network structure in Figure 1. Qconv, Qbottleneck, Qc3, and Qspp were selected for the replacement design of the corresponding modules. Since the parameter transfer does not affect the size of the model parameters, the upsample and concat of the original network structure are not changed; Conv2d in the Head structure outputs the detection results, during which a large number of feature parameters are passed, and these parameters are compressed using Qconv2d to obtain the improved integrated network structure DA-ActNN-YOLOV5 (Figure 3). It inherits all of the features of YOLO v5 and adds compression parameters and data augmentations. In this study, we have DA-ActNN-YOLOV5s, DA-ActNN-YOLOV5m, DA-ActNN-YOLOV5x, and DA-ActNN-YOLOV5l as the corresponding implementation of the four main YOLO v5 model structures.

2.4. Potato Leaf Dataset. The performance of a deep learning model depends heavily on a suitable and effective dataset. The open dataset of potato foliar diseases in existing studies is based on PlantVillage. This dataset is constructed for a specific region under specific geographical and environmental factors, and potato diseases vary in different parts of the world due to various factors such as shape, variety, and environmental factors. To improve the generalizability of the model, 1,000 early blight leaves (earlyblight), 1,000 late blight leaves (lateblight), and 152 healthy leaves (healthy) images from PlantVillage were selected as experimental subjects in this study. The selected PlantVillage dataset was trained from 0 using DA-ActNN-YOLOV5m with YOLO v5m model and tested on the PBD-IM dataset proposed in this study; as shown in Table 1, the error rate of existing PlantVillage-based models for detecting potato diseases in PBD-IM was generally high. Therefore, there is a need to reconstruct potato leaf disease datasets for the Inner Mongolia region. In this study, real-time data from the potato leaf dataset (PBD-IM) in the southeast Lingnan area of Hulunbeier City, Inner Mongolia Autonomous Region, were collected in videos and pictures. The capture distance of the cell phone camera and the digital camera was 30–40 cm, while the capture distance of the UAV was set to 1.5–3.0 m. Due to the high potato cultivation, potato varieties in the southeast Lingnan area of Hulunbeier City, Inner Mongolia Autonomous Region, were selected for this study: "Jizhang Potato No. 12," "Hysen No. 6," and "Huasong No. 7." Potatoes were planted in rows 7.6 cm apart. Seeds of the crop were cultivated by digging soil pits 15.2 to 20.3 cm deep and 12.7 cm wide. To solve the problem of environmental interference in potato early and late blight identification, the

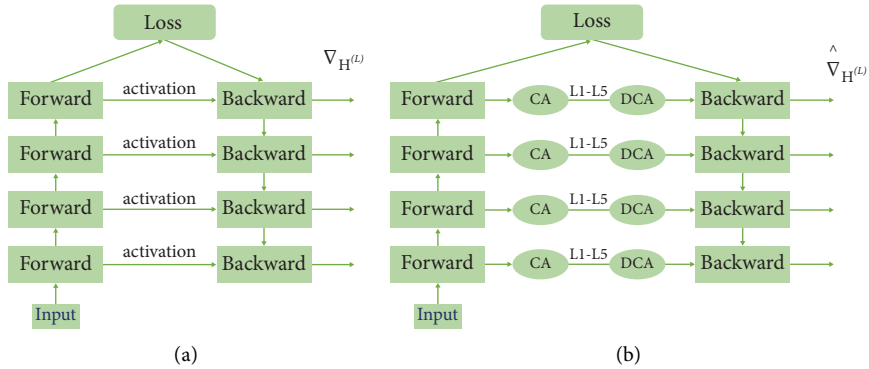


Figure 2: Parameter passing process (a) and ActNN compressed parameter process (b).

Input: training data: **X**
(1) **Convert compressible model:** Replace the **model** module with ActNN
(2) **Initialize Data Loader:** Create mini-batch data loaders based on **X**
(3) **Initialize Model:** Load model parameters or initialize model parameters, Let **W** be the model parameters
(4) **for** number of training iterations **do**
(5) **for** (data, target) of mini-batch data loader **do**
(6) Neural network models complete a computational process based on **data**
Compress the activation (CA) value parameter according to the data in forwarding propagation
Calculate the gradient based on the activation parameter in backpropagation decompression again (DCA)
Calculate the gradient that satisfies (4)
(7) Update the model parameter **W** based on **data**
(8) **end for**
(9) **end for**
Output: Model and model parameters **W**

Algorithm 1: Algorithm description of ActNN compression process. The training process uses minibatch random gradient descent training.

dataset contains leaf samples of different disease cycles, selected for collection in fields with significant differences in leaf shape and plant growth by taking images and videos under different conditions, i.e., morning, evening, noon, cloudy, sunny versus rainy days. Healthy and infected leaves were labelled into PyTorch's YOLO v5 format and XML format using the LabelMe tool, and the labelling process is shown in Figure 4. In the segmentation and leaf extraction stages, the YOLO v5s model, which has a small number of parameters and is fast, was chosen to start training from 0. With the help of YOLO v5s model output and annotations, Python code was used to process the original labelled data and extract potato leaves to form the initial data set. To create the PBD-IM dataset, we selected 4,546 images of healthy and diseased potato leaves from the data. There are 1,828 early blight, 1,514 late blight, and 1,204 normal leaf classes, respectively, as shown in Figure 5.

The growth of potato plants in the same field generally varies with the disease cycle, and to highlight the severity of potato early and late blight in the disease cycle, representative diseased leaves from PBD-IM were selected for this study (Figure 6). There is no specific quantitative definition of the degree of potato disease, and identification is made empirically by plant pathologists. In this study, the number of spots, leaf colour change, and leaf shape distortion on potato leaves were taken as the focus of attention. The cumulative values of three visual characteristics were used to discriminate the degree of disease. The focus was on these characteristics during the actual data collection. Figure 6(a) shows the extent of early potato blight and Figure 6(b) shows the extent of late potato blight. The blue and red gradient lines in the figure represent the severity of the disease, with the left side of the line having a milder degree of disease and the right side having more severe disease. Early detection of the less severe leaves for treatment can have a positive effect on saving the whole plant. Therefore, the data creation method proposed in this study can assist in the early detection of the disease.

2.5. Data Preprocessing General

2.5.1. Data Augmentation Methods. General data augmentation: different data augmentation techniques are applied to the training set using the image data generator of the Keras library in Python to overcome the overfitting problem and enhance the diversity of the dataset. For this reason, this article normalizes the image pixels to 0~1, using smaller pixel parameters and the same range of pixels to reduce the computational cost without changing the properties

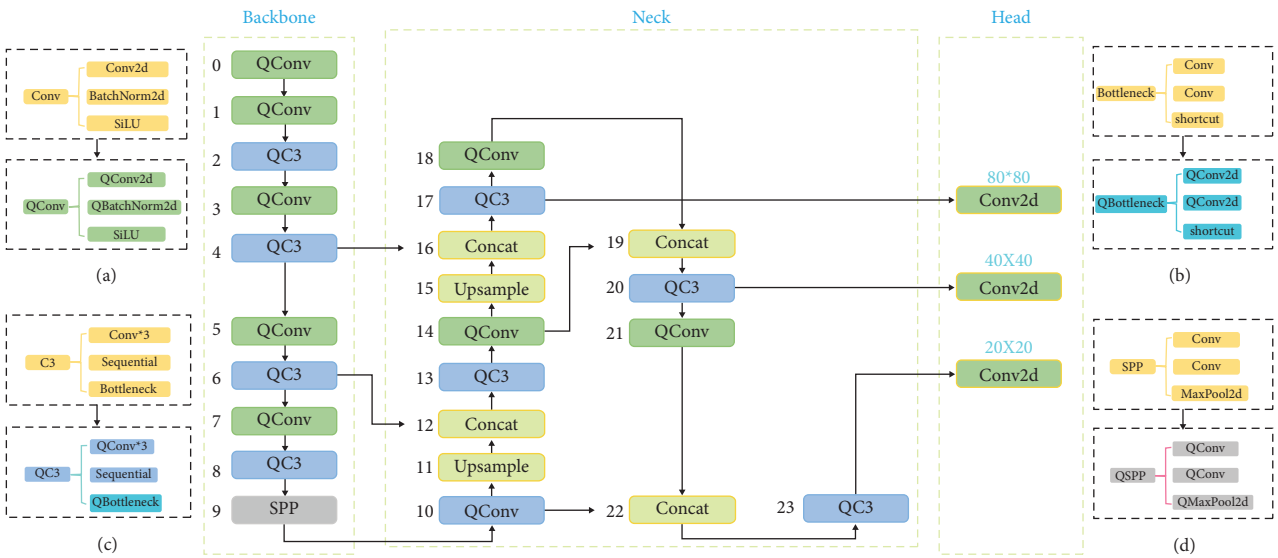


Figure 3: ActNN improves the model structure of YOLO v5 using the modules shown in (a-d).

Table 1: Accuracy of PlantVillage trained models in PBD-IM tests.

Model	Training set	Testing set	Precision (%)			Total images	A_b(%)
			Earlyblight	Healthy	Lateblight		
YOLO v5m	PlantVillage	PBD-IM	94.71	9.82	26.94	900	49.79
DA-ActNN-YOLOV5m	PlantVillage	PBD-IM	96.28	89.39	75.64	900	87.10

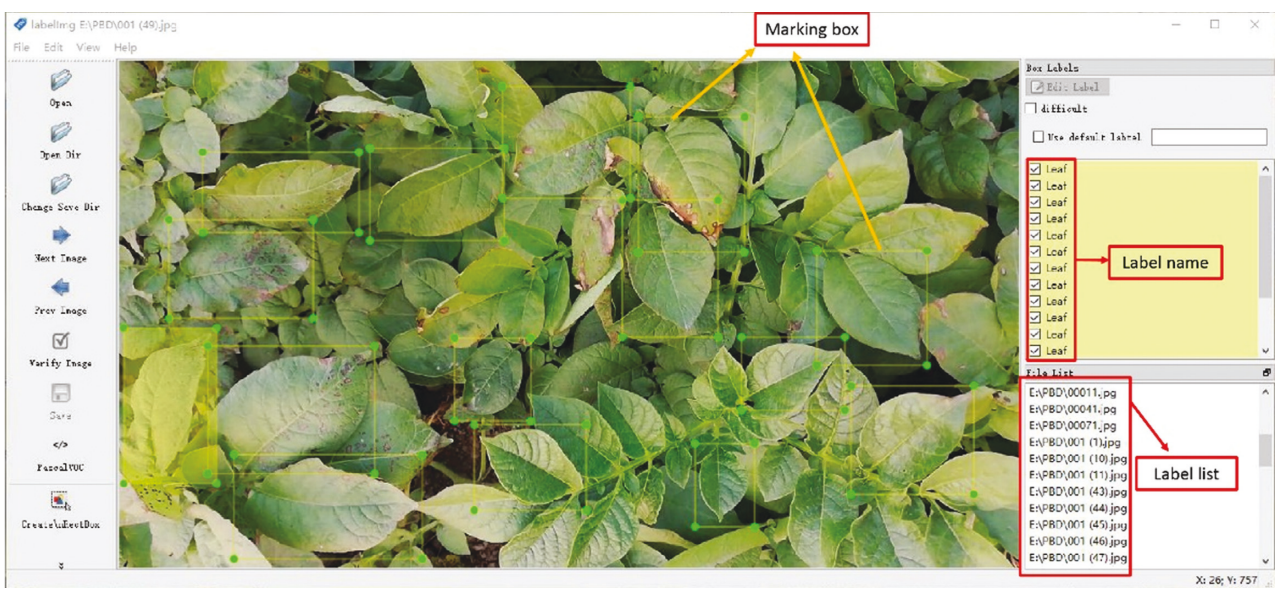


Figure 4: LabelMe labelling potato leaves an example.

possessed by the image itself. This study employs a combination of data augmentation techniques, using a rotation transform (Rotation_transform) to rotate the image to 25°; a width shift transform (Width_transform_range) allows the image to be shifted randomly to the right or left, choosing a width shift parameter value of 0.25. The training image is shifted up or down using a height shift range transform (Height_transform_range) with a range value of 0.25. The shear transform (Shear_transform) fixes one axis of the image and then stretches the other axis to a specific angle, called the shear angle, which is 0.23 applied in this study. The scaling range parameter is applied to perform random scaling transformations (Zoom_transform), >1.0 means expanding the image, and <1.0 is used to scale the image;

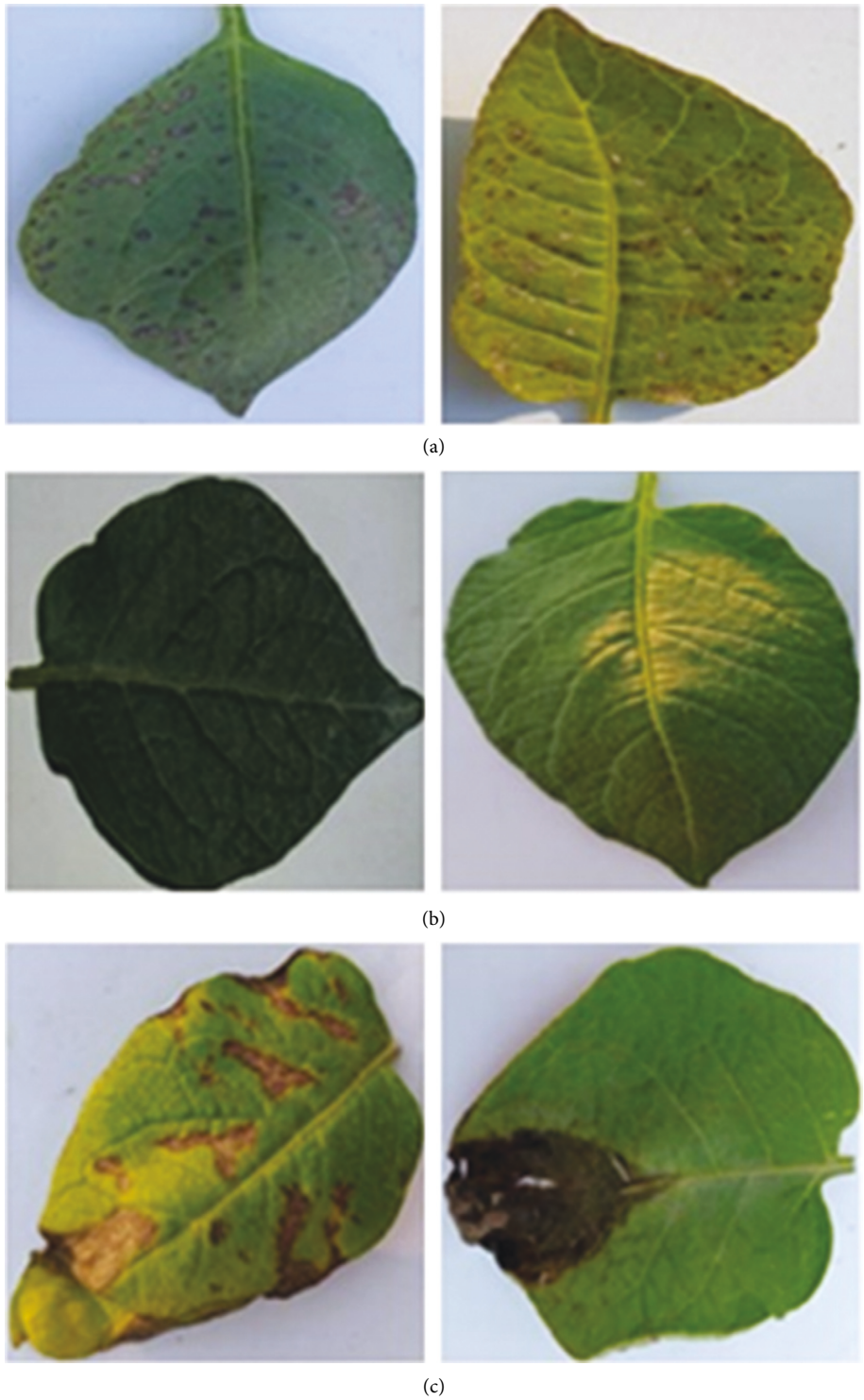


Figure 5: Examples of PBD-IM earlyblight (a), healthy (b), and lateblight (c).

therefore, a 0.2 scaling range is used to transform the image. Flip the image using a random horizontal flip (Horizontal_flip) with a parameter of 0.6. In the applied brightness transformation (Brightness_transform), 0.0 means minimum brightness and 1.0 maximum means brightness; therefore, a zoom range of 0.5~1.0 is used. In the channel shift transformation (Channel_transform_range), a channel shift range of 0.05 is applied, while the fill mode (Fill_-nearest) is used to fill in the proximity pixels.

Weather data augmentation: generic data augmentation methods are used in most studies but with limited performance improvement for object recognition errors caused by weather

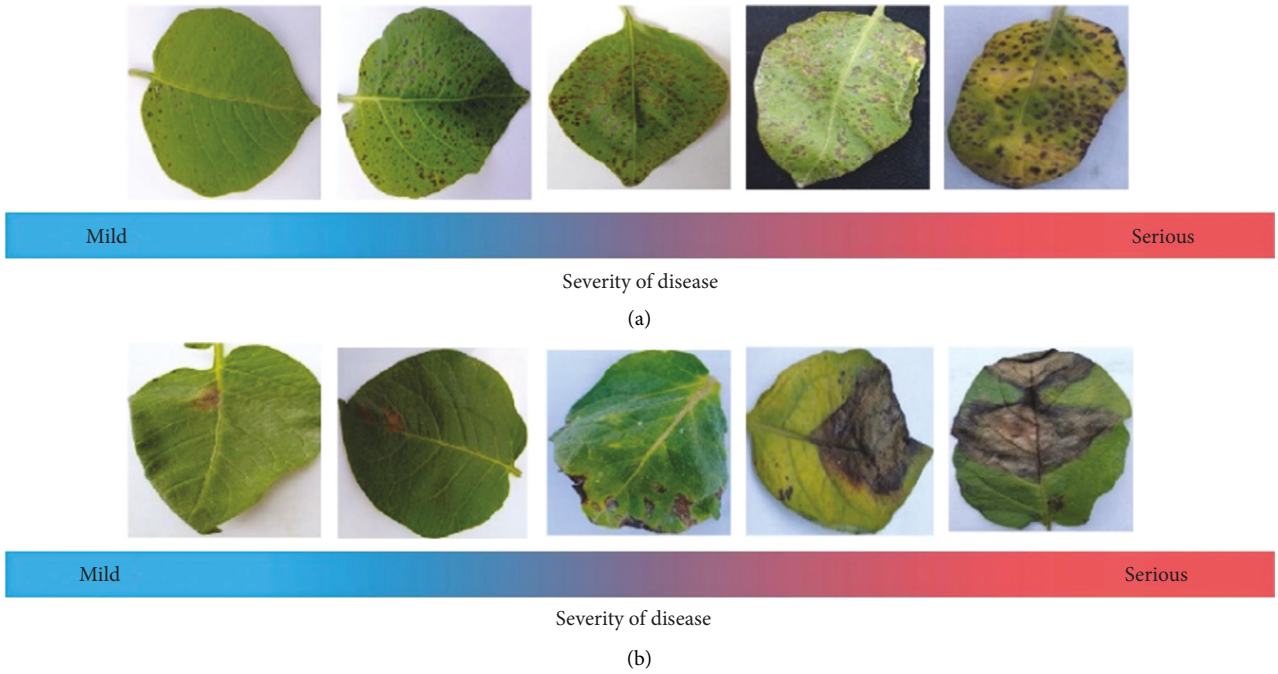


Figure 6: Potato early blight (a) and example of disease cycle image of late blight (b).

changes. For this purpose, we applied different weather data enhancement techniques to the training set using the image data enhancement method of the Albumentations [29] library in Python for simulating the changes in the environment induced by different weather conditions. Add random raindrops to the image via the raindrop transform (Random_rain), select a raindrop size of 1.0, type drizzle, and set the overcast parameter (brightness_coefficient) to 0.6. Due to the large temperature and humidity difference, the natural environment will produce fog. By using the fog transformation (Random_fog), fog can be added randomly to different locations of the image and blur the background; choose a parameter value of 0.1~0.6 for fog intensity (fog_coef) and 0.1 for fog circle transparency (alpha_coef). The life activity of plants is inseparable from light. The solar light transformation (Random_sunflare) is chosen to simulate the image's natural light exposure and light adjustment. The image area where the solar flare appears (flare_roi) is determined by four parameters (x_min, y_min, x_max, y_max), x_min and y_min determine the starting coordinates, x_max and y_max determine the end coordinates; in this article, the upper left corner ((0.0,0.0,1.0,0.5)) and the upper right corner (0.9,0,1,0.5) of the image are chosen as the source of illumination. Potato plants are close to each other, and the neighboring leaves are susceptible to shading to produce the effect of shadow through the shadow transformation (Random_shadow) and to a certain extent can eliminate this effect; this study on the shadow (Shadow_roi) appeared in the region of randomization, the number of shadows in 1 ~ 5 floating, the shadow polygon side parameters set to 4.

2.5.2. Dividing the Training, Validation, and Test Sets. The PBD-IM dataset consists of training, validation, and test sets. The training dataset is used to train DA-ActNN-YOLOV5, while the validation and testing datasets evaluate the performance of the final model. This study divided the training, validation, and test datasets into 80%, 10%, and 10%, respectively. For the PBD-IM dataset, 3,638, 454, and 454 images were used for training, validation, and testing. Different data augmentation techniques were performed on the training set, i.e., rescaling, rotation, width offset, height offset, clipping range, scaling range, horizontal flip, luminance and channel offset, and proximity fill patterns, to increase diversity and enhance the dataset. It will overcome the overfitting problem and thus ensure the model's versatility.

2.6. Experimental Environment and Training Parameters. Experimental environment: the operating platform is Nettrix X640 G30 AI server with Ubuntu 20.04 OS, Intel® Xeon® Gold 6226R CPU, NVIDIA GeForce RTX 3090 GPU, 256 GB RAM, and 7.5 T solid-state drive. The training environment was created by MiniConda3 and configured with Python 3.8.5, PyTorch 1.10.1 with TorchVision 0.11.2 artificial neural network library. CUDA 11.1 deep neural network acceleration library was used.

Training parameter settings: DA-ActNN-YOLOV5 was selected as the PBD-IM dataset to train the network, BCELoss [30] was used for the loss function, SGD was used for the optimizer, the input image size was 640×640 pixels, the learning rate was initialized to 0.0042 and finally 0.15, the momentum parameter was 0.845, the weight decay was set to 0.00056, and a warm-up parameter5 is used to ensure that the model starts training with a certain prior knowledge of the data, and other parameters are kept as default. There is a correlation between the particular dataset and the convergence rate of the model training; when it seems there will no

longer be any change in the model's performance or even a decrease, the training will continue, and the model will not get good convergence; at this time, this problem should be monitored, and timely intervention and the emergence of early stopping mechanism are good solutions to this problem [31]; this study integrated early stopping mechanism in DA-ActNN-YOLOV5 and set the parameter to 50.

3. Experimental Results and Analysis

3.1. Model Evaluation Metrics. To achieve the evaluation of the potato early and late blight detection model, an exhaustive analysis of the performance of the proposed method and other recent methods discussed in this study was carried out, which included various indicators such as Precision (P_b), Recall (R_b), F1-Score ($F1_b$), Accuracy (A_b), and Average Precision (AP). In equations (6)–(11), the formulas are separately written for each measurement indicator.

$$R_b = \frac{\mathrm{TP}_{\mathrm{blight}}}{\mathrm{TP}_{\mathrm{blight}} + \mathrm{FN}_{\mathrm{blight}}}, \tag{6}$$

$$P_b = \frac{\mathrm{TP}_{\mathrm{blight}}}{\mathrm{TP}_{\mathrm{blight}} + \mathrm{FP}_{\mathrm{blight}}}, \tag{7}$$

$$A_b = \frac{\mathrm{TP}_{\mathrm{blight}}}{\mathrm{TP}_{\mathrm{blight}} + \mathrm{FP}_{\mathrm{blight}} + \mathrm{FN}_{\mathrm{blight}}}, \tag{8}$$

$$F1_b = \frac{2 \times P_b \times R_b}{P_b + R_b}, \tag{9}$$

$$AP = \frac{1}{n}\sum_{i=1}^{n} P_{b_i}\left(P_{b_i} = \frac{1}{n}P_{b_1} + \frac{1}{n}P_{b_2} + \cdots + \frac{1}{n}P_{b_n}\right), \tag{10}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{Q} \mathrm{AP}_i}{Q}. \tag{11}$$

In (6)–(8),$\mathrm{TP}_{\mathrm{blight}}$ (true positive), $\mathrm{FP}_{\mathrm{blight}}$ (false positive), and $\mathrm{FN}_{\mathrm{blight}}$ (false negative) refer to the number of positive instances that are properly recognized, the number of negative instances that are wrong recognized, and the number of positive instances mistakenly rejected, respectively. $F1_b$ in (9) is a combined measure of the accuracy and recall. Usually, the accuracy and recall are mutually approximate to give a more balanced response to the model's performance. The calculation of mean average precision (mAP) depends on AP, where AP is defined as the intersection over union (IoU) threshold value of 0.5; for a class with N correctly identified samples, each correctly identified sample will correspond to a P_b value, and the average of the N P_b is taken to obtain the average accuracy of the class; see (10). mAP (IoU > 0.5) is defined as the mean value under all categories of AP, as shown in (11); this study has healthy leaves, early blight, and late blight. As the total number of detection categories, Q is 3, and mAP is the average cumulative value of the average accuracy of multicategory, which can overall demonstrate the comprehensive performance of the model.

3.2. Results and Analysis

3.2.1. Analysis of Data Augmentation Results. To validate the method's superior performance on the PBD-IM dataset using the data augmentation technique applied to the training sets. This study evaluates the model using mAP, precision, and loss values, and Figure 7 depicts the complete training, validation precision, and loss in each epoch. The results show that the method achieves good recognition rates on the PBD-IM dataset using data augmentation techniques on the training set. With the intervention of the early stopping mechanism, the training stops at 350 epochs. The validation and training set's loss value curves depicted in Figures 7(a) and 7(b) drop smoothly. The loss value of the validation set reaches the lowest value at this stage at 275 epochs, and after that, all fluctuate in a very small interval. The curves depicted in Figures 7(c) and 7(d) show an extremely similar consistency, which can be specifically split into two phases. In the first phase, precision and mAP fluctuate in the 0.35–0.40 interval when the epoch is 0 to 100. In the second stage, after 60 epochs of training, the accuracy rate approaches perfect accuracy faster. After that, the accuracy rate is maintained at high accuracy until the end of training.

To evaluate the performance of the proposed data augmentation method in DA-ActNN-YOLOV5, two sets of experiments were conducted. In the first set of experiments, this study applies Table 2 ordinal number 3 data augmentation technique to the training set of the PBD-IM dataset. In Group 2 experiments, to highlight the advantages of all data augmentation techniques, this study trained DA-ActNN-YOLOV5 without using data augmentation techniques. All experimental batch sizes were 32, 100 epochs, and default learning rates. The data augmentation method used in test group 1 produced four sets of samples in each training iteration to compare the three types of data augmentation techniques applied to the PBD-IM dataset. The results are shown in Table 2. The number 1 trial used only three data augmentation technique with 97.56% accuracy; the number 2 trial used five data augmentation techniques and obtained 98.39% accuracy; the number 3 trial used nine data augmentation techniques and obtained 99.75% accuracy; experiment number 4 obtained 99.81% accuracy using 13 data augmentation techniques, with a 0.06% improvement in accuracy compared to experiment number 3, thanks to the contribution of weather data augmentation methods. The data augmentation method proposed in this study achieves high accuracy in the experimental results and generates a significant gain in DA-ActNN-YOLOV5 performance with many data samples.

Table 2 serial number 3 data augmentation techniques compared to the unused results are shown in Table 3. The experimental results showed that the proposed method was 99.38%, 99.58%, and 99.88% accurate for early blight, healthy, and late blight of potatoes. To illustrate the performance on the PlantVillage dataset, DA-ActNN-YOLOV5 was trained on the PBD-IM training set, and then the data was extracted at PlantVillage as a test set for validation, and the test results are shown in Table 4. The PBD-IM dataset

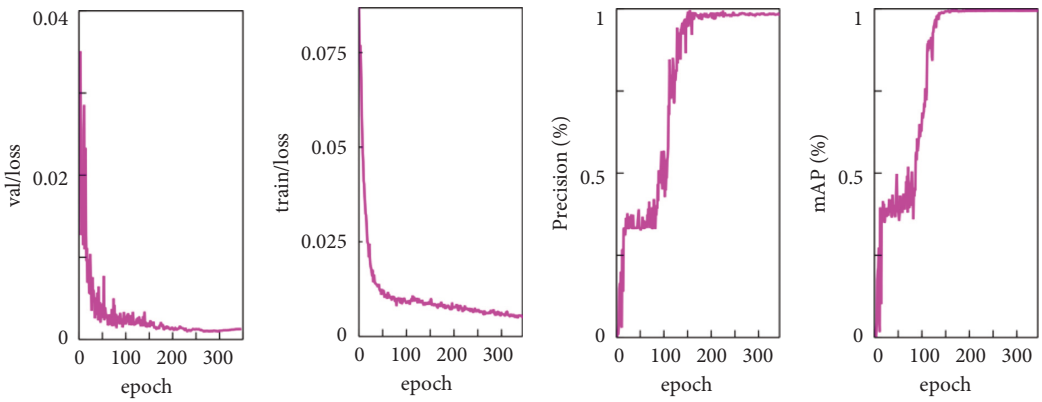


Figure 7: Validation loss (a), training loss, (b) verification accuracy rate, (c) and verification mAP (d).

Table 2: Accuracy of the DA-ActNN-YOLOV5m model using different data augmentation methods in the PBD-IM dataset.

No.	Data augmentation methods	Precision (%)			mAP (%)
		Earlyblight	Healthy	Lateblight	
1	Rotation_transform Width_transform_range Height_transform_range	96.72	97.56	96.43	97.56
2	Rotation_transform Width_transform_range Height_transform_range Shear_transform Zoom_transform	98.87	98.44	97.85	98.39
3	Rotation_transform Width_transform_range Height_transform_range Shear_transform Zoom_transform Horizontal_flip Brightness_transform Channel_transform_range Fill_nearest	99.12	100.00	98.94	99.75
4	Rotation_transform Width_transform_range Height_transform_range Shear_transform Zoom_transform Horizontal_flip Brightness_transform Channel_transform_range Fill_nearest Random_rain Random_fog Random_shadow Shadow_roi	99.98	100.00	99.85	99.81

showed high accuracy in detecting early and late potato diseases for PlantVillage, with a combined performance A_b of 95.76%, which was 8.66% higher than the test results for PlantVillage as a training set, and the performance improvement for the potato health and late blight categories was greater, with at least more than 6.04% A_b improvement. Doing the same test with YOLO v5m as the former, the accuracy rate was reduced by 22.75% compared to the former, proving the superiority of the proposed method over YOLO v5 in this study and also demonstrating the higher accuracy of PBD-IM compared to the PlantVillage dataset for potato disease identification. By comparison, the detection performance of the DA-ActNN-YOLOV5 network model based on the PBD-IM dataset is better.

Table 5 reviews the work related to this study. Most of the image data used to detect potatoes' early and late blight came

Table 3: Test performance of the DA-ActNN-YOLOV5m model on the PBD-IM dataset using data augmentation methods.

Method	Mertics (%)	Earlyblight	Healthy	Lateblight	AP (%)
With data augmentation	A_b	99.38	99.58	99.88	99.61
	P_b	99.01	100.00	100.00	—
	R_b	98.04	100.00	99.03	—
	$F1_b$	98.05	98.06	99.09	—
Without data augmentation	A_b	93.57	85.17	92.45	90.39
	P_b	88.01	93.05	90.00	—
	R_b	91.03	86.04	87.06	—
	$F1_b$	89.06	89.00	88.04	—

Table 4: Accuracy of models trained in PBD-IM tested in PlantVillage.

Model	Training set	Testing set	Precision (%)			Total images	A_b (%)
			Earlyblight	Healthy	Lateblight		
YOLO v5m	PBD-IM	PlantVillage	92.37	56.91	69.76	500	73.01
DA-ActNN-YOLO V5m	PBD-IM	PlantVillage	97.29	94.56	95.43	500	95.76

from the PlantVillage dataset, while Rashid et al. [34] and Afzaal et al. [35] established their datasets. In this study, the shortcomings of the PlantVillage dataset were verified, and the PBD-IM dataset was established by data enhancement. By comparing the methods proposed by other researchers in Table 5, the accuracy rate of the model in this study is 0.06% and 5.81% higher than the models proposed by Rashid et al. [34] and Afzaal et al. [35], respectively. In terms of data enhancement, Chen et al. [37] used a generative adversarial network (GAN) to automatically synthesize diverse images with a 2.08% lower accuracy rate than the generic data enhancement and weather data enhancement methods proposed in this study. While comparing the remaining seven types of methods, the accuracy rate of this study is at least 3% better. Therefore, the comparative results show that the method proposed in this study achieved superior results and outperformed other research methods. In summary, this study used YOLO v5 as the base CNN model. By applying data enhancement methods to PBD-IM data samples, the original problem of high recognition error rate due to specific geographical and environmental factors was solved. Also, it reduced the differences in the degree of leaf deformation of potato early and late blight in different disease cycles in PlantVillage. Early detection of potato leaf diseases could be ensured.

3.2.2. Visualization Feature. Visualization of neural network models can help us explore and understand the black-box learning behavior of deep models. The visualized feature map can be used as a diagnostic tool to observe the extraction of features at each network layer during model training and to diagnose potential problems with the model through a two-dimensional feature representation. By comparing filters layer by layer, we can see the effect of feature visualization of the CNN hierarchy. Each network layer based on DA-ActNN-YOLOV5 has 32 to 512 convolutional kernels, from which representative feature maps are selected as feature representations for each network layer, and Figure 8 shows the 24-layer feature visual representation learned by the model. It can be observed that the convolutional kernels in layers 0 to 1 acquire the ability to extract image edges by learning samples, and the edges of potato leaf and leaf spot images are visible. It is worth noting that different convolutional kernels do not focus on the same object; some focus more on the leaf edges, others on the edges of the disease spots, and a few focus on both leaf and disease spot edges. In the network extracted feature maps of layers 2 to 4, the convolutional kernel focuses on extracting potato leaf texture features, and the overall structure of potato leaf veins is better expressed. In the feature maps extracted by the 5- to 6-layer network, the edge and texture information becomes blurred, and the shape of the generated image changes from fine to coarse, e.g., the image representation obtained using the Gaussian blurring process. From layers 7 to 23, the feature representation becomes more complex with the increase of network layers, from which the information of edge and texture features of the leaf cannot be distinguished. Therefore, the network extracted features can be seen as a mapping process from low-to high-dimensional features, and the increase of network layers and filters makes the features more abstract.

3.2.3. Analysis of ActNN Results. The training will not continue when the training model exceeds the maximum memory capacity supported by the user device. Among the network models provided by YOLO v5, YOLO v5x, YOLO v5m, and YOLO v5s were used most frequently. The three models were tested on the PBD-IM dataset, and except for YOLO v5x, which could not be trained, YOLO v5m and YOLO v5s training were able to proceed normally, but YOLO v5m was close to the device memory threshold in training and could not be trained using a deeper network structure. The usual solution to the problem of training not being possible due to limited device memory is to reduce the batch size or reduce the input image size. However, this approach has limited usefulness, and training may also not be possible when the batch size is 1. Moreover, for the problem that the device memory of the training model is close to the critical value, the effect of reducing the batch size

Table 5: The methodology of this study compared to other research methods.

No.	References	Year	Method description	AP (%)
1	Sholihati et al. [32]	2020	VGGNet16 + VGGNet19	91.31
2	Yang et al. [19]	2020	Faster R-CNN + SIFT + K-means	90.83
3	Barman et al. [33]	2021	Self-build CNN (SBCNN)	96.98
4	D. F. Wang, and J. Wang [18]	2021	SE + ResNet50+DenseNet-121	97.99
5	Rashid et al. [34]	2021	PDDCNN + data augmentation	99.75
6	Afzaal et al. [35]	2021	GoogleNet + VGGNet + EfficientNet	94.00
7	Hou et al. [36]	2022	k-NN + SVM + ANN + RF	97.40
8	Chen et al. [37]	2022	MobileNetV2+GAN + Attention mechanism + octave	97.73
9	Mahum et al. [38]	2022	DenseNet-201 + Efficient DenseNet	97.20
10	Sharma et al. [39]	2022	Deep CNN	97.66
11	Proposed model	2022	YOLO v5+ data augmentation + ActNN (DA-ActNN-YOLOV5)	99.81

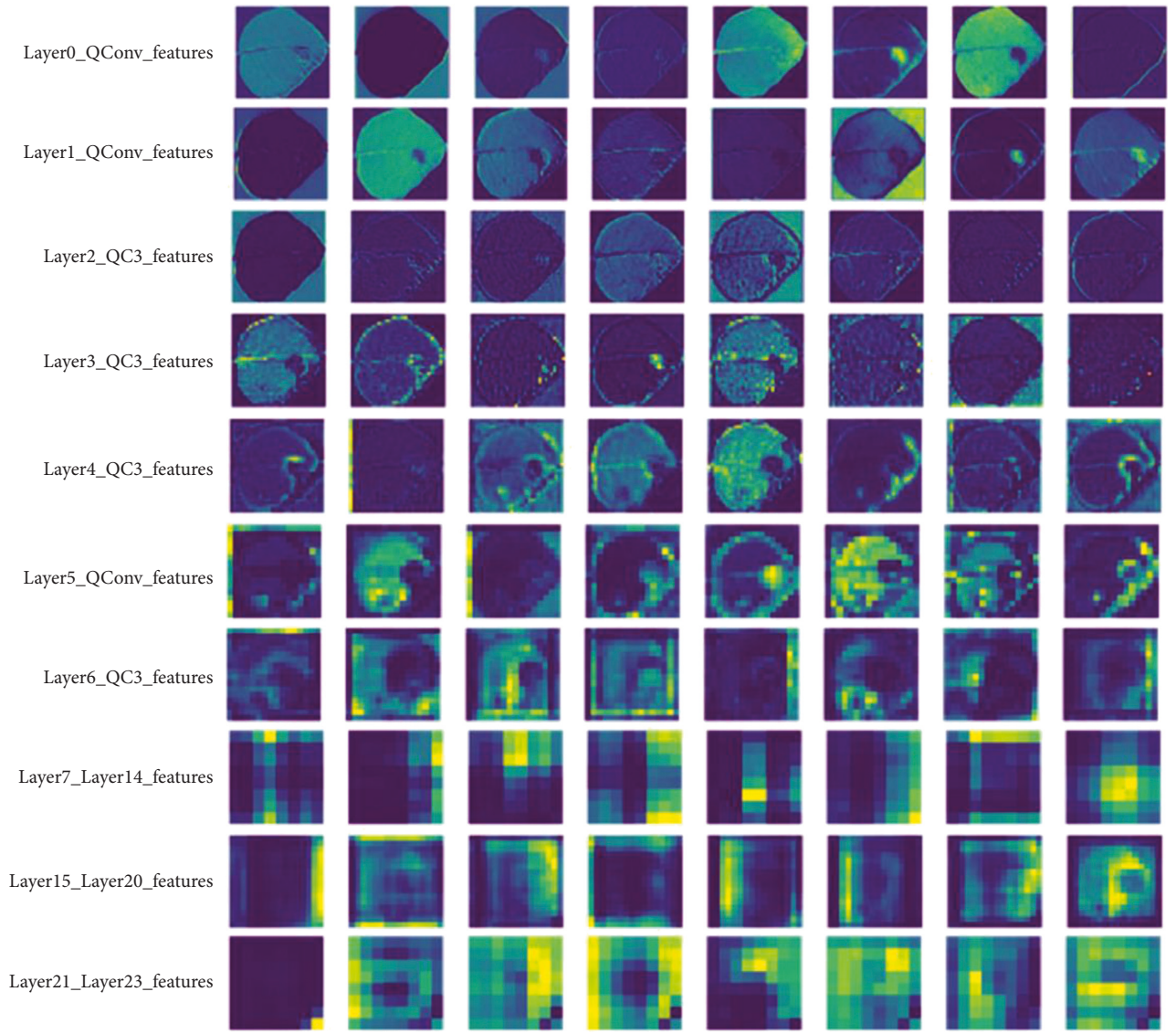


Figure 8: A two-dimensional 24-layer feature map was built using DA-ActNN-YOLOV5.

and the input image size method is more significant. The DA-ActNN-YOLOV5 proposed in this study can use ActNN to compress the memory when out of memory (OOM) occurs in the device.

For this purpose, three sets of experiments were performed to evaluate the performance of DA-ActNN-YOLOV5. Instead of choosing DA-ActNN-YOLOV5s with fewer parameters, experiments were performed using DA-ActNN-YOLOV5l

Table 6: mAP and training elapsed time of DA-ActNN-YOLOV5 on the PBD-IM dataset.

Model	Batch size	Image size	Level	mAP (%)	Time (h)
DA-ActNN-YOLOV5m	32	384	—	98.76	1.852
	64	384	—	99.86	1.579
	128	384	—	98.99	1.286
	64	512	—	99.68	1.685
	128	512	L1	OOM	—
	128	512	L3	98.66	1.663
	128	512	L5	99.12	1.681
DA-ActNN-YOLOV5x	16	384	—	97.23	2.869
	32	384	—	98.68	2.521
	64	384	—	99.78	2.219
	64	512	L1	OOM	—
	64	512	L3	97.99	2.674
	64	512	L5	98.39	2.683
DA-ActNN-YOLOV5l	8	384	—	96.89	3.853
	16	384	—	97.38	3.457
	32	384	—	98.97	3.289
	32	512	—	OOM	—
	32	512	L1	OOM	—
	32	512	L3	98.57	3.565
	32	512	L5	98.19	3.723

with more model parameters. Since the ActNN compression level, batch size, and input image size are different for each group of tests, the model will not get better convergence if a consistent epoch is used, so the tests do not restrict the epoch, and other settings are set concerning the base parameters to ensure that the model converges to a better state for each group of tests. All experimental groups verified 3 compression levels, containing L1, L3, and L5, and L2 and L4 were not selected, mainly because their algorithmic advantages were not obvious. In three groups of experiments, 13 data enhancement techniques were applied to the model trials, and each group was first ensured to train without ActNN. The training process only changed the batch size and image sizes, and ActNN was turned on when the device memory was insufficient. The compression level was tested three times in each group, the optimal values were recorded, and the results are shown in Table 6.

In all three sets of model tests using ActNN, the models achieved a high level of accuracy without any significant degradation in accuracy. They had the best accuracy rate for a batch size of 64. In addition, a smaller batch size corresponds to slightly smaller accuracy due to the low training sample in each iteration, and choosing a larger batch size for the same training can improve the model accuracy and make the model converge quickly. The model complexity of DA-ActNN-YOLOV5m, DA-ActNN-YOLOV5x, and DA-ActNN-YOLOV5l increases sequentially, and when the device memory is insufficient due to changing the batch size and image size, the ActNN compressed model parameters enable the model to be trained normally. In addition, the accuracy loss from compressing model parameters is less than 0.65% for the highest compression level L5 compared to the lowest compression level L1, and the additional time consumed for compressing model parameters is less than 30 min compared to the model without ActNN. In addition, higher compression levels will bring additional computation time consumption, and the compression level of ActNN should be chosen reasonably according to the actual situation.

Table 7: Ablation experiments of the three main structures of ActNN replacement DA-ActNN-YOLOV5l.

No.	Compression network structure			mAP (%)	Time (h)
	Backbone	Neck	Head		
1	✓			97.65	3.375
2		✓		98.16	3.358
3			✓	98.83	3.324
4	✓		✓	97.55	3.412
5		✓	✓	98.07	3.395
6	✓	✓		97.41	3.467
7	✓	✓	✓	97.34	3.474

The proportion of ActNN module replacement selected for the neural network model affects the recognition accuracy and training time to different degrees. In order to analyze the performance of the three network structures of YOLO v5 after replacing the ActNN module, the DA-ActNN-YOLOV5l model is used as an example; the batch size is set to 32, the image size is 384 pixels, and the compression parameter level is selected as L3, and the experimental results are shown in Table 7. When parameter compression is enabled, all three network structures have different degrees of accuracy loss. Among them, since the Head structure is only involved in the output of the feature map and does not involve the intermediate transfer process of the feature map, compressing the parameters of the Head structure has the most negligible impact on the accuracy, and the accuracy loss is less than 0.15%. Ten modules of the Backbone structure were replaced with ActNN modules, accounting for 47.1% of the total replacement modules. Since feature extraction is the main work of the Backbone

structure and the convolutional computation is quite large, compressing the Backbone structure parameters resulted in a 1.32% decrease in accuracy. In addition, the Neck structure accounted for 38% of the total replacement modules and compressed parameters resulted in a 0.81% reduction in accuracy. In a comprehensive analysis, the number of modules involved in compressing parameters is the main reason for the decrease in accuracy and increase in training time. The increase in accuracy loss and training time is positively correlated with the increase in compression level.

4. Conclusion

This study proposes a potato early and late blight detection method based on the deep learning YOLO v5 framework called DA-ActNN-YOLOV5. By replacing the corresponding module of YOLO v5 with ActNN, the fine-grained compression of the activation parameters is achieved. Compared with the model with uncompressed parameters, the highest compression level results in an accuracy loss of less than 0.65%, and the time consumed to compress the model parameters is less than 30 minutes, which is an obvious advantage. When training with models of different complexity, a reasonable choice of different compression levels for the model activation parameters can accelerate the model to converge quickly with approximate original accuracy and compress the model parameters to ensure smooth training when the device memory is insufficient. The accuracy of DA-ActNN-YOLOV5 in identifying early blight, healthy leaves, and late blight of potatoes using 13 data enhancement techniques was 99.38%, 99.58%, and 99.88%, which were 5.81%, 14.41%, and 7.43% higher than without data enhancement; the average accuracy was 9.22% higher. Compared to YOLO v5, performance was improved by at least 6%, with a mAP of 99.81% (IoU $\geq$ 0.5) for the retained test dataset. In addition, the performance of the PBD-IM dataset outperformed PlantVillage, with an 8.66% improvement in accuracy after the DA-ActNN-YOLOV5 test and a more significant performance improvement in the identification of healthy leaves and late blight categories, with an average of 6.04% improvement in accuracy. To obtain the extraction of features by the entire network layer, we visualized the feature maps extracted by each network layer. Each layer of the feature extraction network has its division of labour, with some network layers focusing on object edges and others on object colour textures. The overall feature extraction process can be viewed as a mapping process from low-to high-dimensional features. The results show that the method proposed in this study can effectively discriminate potato early and late blight under different leaf shapes, pathogenesis cycles, and environmental factors; it solves the problem of potato early and late blight staying on specific regional data sets and realizes model training for multiple equipment environments. It can be used as a reference tool for other crop disease detection and building detection models, helping researchers accelerate the application of artificial intelligence in agriculture in other directions and meeting the requirements of precision agriculture falling in the field. In the future, this research will be expanded to include multiple disease detection on individual leaves and enhancements to disease localization and disease severity estimation—development of an IoT-based real-time monitoring system, website development, and mobile application release.

Data Availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

G.D. was responsible for conceptualization, data curation, methodology, software, validation, and writing the original draft. G.D., J.F., and L.H. reviewed and edited the manuscript; J.F. and L.H. were responsible for funding acquisition and obtaining resources and supervised the work. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported by the sub-project of the National Key R&D Program (2019YFF0216202), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2016-AII), and the Major Project of Science and Technology of Inner Mongolia Autonomous Region(2021ZD0043).

References

[1] K. Djaman, K. Koudahe, H. D. Koubodana, A. Saibou, and S. Essah, "Tillage practices in potato (Solanum tuberosum L.) production: a review," *American Journal of Potato Research*, vol. 99, no. 1, pp. 1–12, 2022.

[2] J. Yuen, "Pathogens which threaten food security: phytophthora infestans, the potato late blight pathogen," *Food Security*, vol. 13, no. 2, pp. 247–253, 2021.

[3] G. A. Mora-Romero, R. Félix-Gastélum, R. A. Bomberger, C. Romero-Urías, and K. Tanaka, "Common potato disease symptoms: ambiguity of symptom-based identification of causal pathogens and value of on-site molecular diagnostics," *Journal of General Plant Pathology*, vol. 88, no. 2, pp. 89–104, 2022.

[4] L. Sanchez, A. Ermolenkov, X. T. Tang, C. Tamborindeguy, and D. Kurouski, "Non-invasive diagnostics of Liberibacter disease on tomatoes using a hand-held Raman spectrometer," *Planta*, vol. 251, no. 3, p. 64, 2020.

[5] H. Urbina, T. Smith, C. Jones et al., "Validation of quantitative and digital polymerase chain reaction assays targeting the mating types of phyllosticta citricarpa, the causal agent of citrus black spot," *PhytoFrontiers™*, vol. 1, no. 4, pp. 301–313, 2021.

[6] R. Sharma, "Artificial intelligence in agriculture: a review," in *Proceedings of the 5th International Conference on Intelligent*

Computing and Control Systems (ICICCS), pp. 937–942, Madurai, India, May 2021.

[7] Y. Liu, Y. Hu, W. Cai, G. Zhou, J. Zhan, and L. Li, "DCCAM-MRNet: mixed residual connection network with dilated convolution and coordinate attention mechanism for tomato disease identification," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4848425, 15 pages, 2022.

[8] G. Dai and J. Fan, "An industrial-grade solution for crop disease image detection tasks," *Frontiers of Plant Science*, vol. 13, Article ID 921057, 2022.

[9] P. Patil, N. Yaligar, and S. M. Meena, "Comparision of performance of classifiers - SVM, RF and ANN in potato blight disease detection using leaf images," in *Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–5, Coimbatore, India, December 2017.

[10] M. A. Iqbal and K. H. Talukder, "Detection of potato disease using image segmentation and machine learning," in *Proceedings of the International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pp. 43–47, Chennai, India, August 2020.

[11] Z. H. Kok, A. R. Mohamed Shariff, M. S. M. Alfatni, and S. Khairunniza-Bejo, "Support vector machine in precision agriculture: a review," *Computers and Electronics in Agriculture*, vol. 191, Article ID 106546, 2021.

[12] N. Xu, X. Wang, Y. Xu, T. Zhao, and X. Li, "Deep multi-scale residual connected neural network model for intelligent athlete balance control ability evaluation," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9012709, 11 pages, 2022.

[13] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7445–7455, 2021.

[14] K. K. Chakraborty, R. Mukherjee, C. Chakroborty, and K. Bora, "Automated recognition of optical image based potato leaf blight diseases using deep learning," *Physiological and Molecular Plant Pathology*, vol. 117, Article ID 101781, 2022.

[15] H.-C. Chen, A. M. Widodo, A. Wisnujati et al., "AlexNet convolutional neural network for disease detection and classification of tomato leaf," *Electronics*, vol. 11, no. 6, p. 951, 2022.

[16] S. Verma, A. Chug, and A. P. Singh, "Application of convolutional neural networks for evaluation of disease severity in tomato plant," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 23, no. 1, pp. 273–282, 2020.

[17] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel, and S. Bhardwaj, "Potato leaf diseases detection using deep learning," in *Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 461–466, Madurai, India, May 2020.

[18] D. F. Wang and J. Wang, "Crop disease classification with transfer learning and residual neural networks," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 37, no. 4, pp. 199–207, 2021.

[19] S. Yang, Q. Feng, J. H. Zhang, W. Sun, and G. P. Wang, "Identification method for potato disease based on deep learning and composite dictionary," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 7, 2020.

[20] I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," *Computers and Electronics in Agriculture*, vol. 176, Article ID 105626, 2020.

[21] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—a review," *IEEE Access*, vol. 9, Article ID 56683, 2021.

[22] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 1, 2021.

[23] K. Duan, L. Xie, H. Qi, S. Bai, Q. Huang, and Q. Tian, "Corner proposal network for anchor-free, two-stage object detection," in *Computer Vision – ECCV 2020. ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J. M. Frahm, Eds., Springer, Berlin, Germany, 2020.

[24] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9627–9636, Seoul, Korea (South), October 2019.

[25] V. Singla, S. Singla, S. Feizi, and D. Jacobs, "Low curvature activations reduce overfitting in adversarial training," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16423–16433, Montreal, QC, Canada, October 2021.

[26] J. Chen, L. Zheng, Z. Yao et al., "Actnn: reducing training memory footprint via 2-bit activation compressed training," in *Proceedings of the International Conference on Machine Learning*, pp. 1803–1813, PMLR, July, 2021.

[27] A. Chakrabarti and B. Moseley, "Backprop with approximate activations for memory-efficient network training," in *Advances in Neural Information Processing Systems*Curran Associates, Inc, New York, NY, USA, 2019.

[28] F. Fu, Y. Hu, Y. He et al., "Don't waste your bits! squeeze activations and gradients for deep neural networks via tinyscript," in *Proceedings of the 37th International Conference on Machine Learning*, pp. 3304–3314, PMLR, July 2020.

[29] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.

[30] J. J. Xu, H. Zhang, C. S. Tang, Q. Cheng, B. Liu, and B. Shi, "Automatic soil desiccation crack recognition using deep learning," *Géotechnique*, vol. 72, no. 4, pp. 337–349, 2022.

[31] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, and N. Smith, "Fine-tuning pretrained language models: weight initializations, data orders, and early stopping," 2002, https://arxiv.org/abs/2002.06305.

[32] R. A. Sholihati, I. A. Sulistijono, A. Risnumawan, and E. Kusumawati, "Potato leaf disease classification using deep learning approach," in *Proceedings of the international electronics symposium (IES)*, pp. 392–397, IEEE, Surabaya, Indonesia, September 2020.

[33] U. Barman, D. Sahu, G. G. Barman, and J. Das, "Comparative Assessment of Deep learning to detect the leaf diseases of potato based on data augmentation," in *Proceedings of the International Conference on Computational Performance Evaluation (ComPE)*, pp. 682–687, IEEE, Shillong, India, July 2020.

[34] J. Rashid, I. Khan, G. Ali, S. H. Almotiri, M. A. AlGhamdi, and K. Masood, "Multi-level deep learning model for potato leaf disease recognition," *Electronics*, vol. 10, no. 17, 2021.

[35] H. Afzaal, A. A. Farooque, A. W. Schumann et al., "Detection of a potato disease (early blight) using artificial intelligence," *Remote Sensing*, vol. 13, no. 3, 2021.

[36] C. Hou, J. Zhuang, Y. Tang et al., "Recognition of early blight and late blight diseases on potato leaves based on graph cut segmentation," *Journal of Agriculture and Food Research*, vol. 5, Article ID 100154, 2021.

[37] W. Chen, J. Chen, A. Zeb, S. Yang, and D. Zhang, "Mobile convolution neural network for the recognition of potato leaf disease images," *Multimedia Tools and Applications*, vol. 81, pp. 1–20, 2022.

[38] R. Mahum, H. Munir, Z. U. N. Mughal et al., "A novel framework for potato leaf disease detection using an efficient deep learning model," *Human and Ecological Risk Assessment: An International Journal*, pp. 1–24, 2022.

[39] R. Sharma, A. Singh, K. Z. Jhanjhi, N. M. Masud, E. S. Jaha, and S. Verma, "Plant disease diagnosis and image classification using deep learning," *Computers, Materials & Continua*, vol. 71, no. 2, pp. 2125–2140, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6486876, 9 pages
https://doi.org/10.1155/2022/6486876

Research Article

Short-Term Power Prediction of a Photovoltaic Power Station Based on the SSA-CEEMDAN-FCN Model

Zhaoyang Qu,[1] **Shaohua Qin**,[1,2] **Genxin Xiong**,[3] **Xinpo Zhu**,[3] **Fan Ling**,[3] **Yukun Wang**,[3] **and Juan Kong**[3]

[1]*School of Computer Science and Engineering & School of Software, Guangxi Normal University, Guilin 541004, China*
[2]*Guangxi Key Lab of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin 541004, China*
[3]*Beijing China-power Puhua Co Ltd., Beijing 100085, China*

Correspondence should be addressed to Shaohua Qin; shqin@gxnu.edu.cn

Received 4 June 2022; Accepted 26 August 2022; Published 22 September 2022

Academic Editor: Wenming Cao

Photovoltaic power generation is greatly affected by weather factors. To improve the prediction accuracy of photovoltaic power generation, complete ensemble empirical mode decomposition with an adaptive noise algorithm (CEEMDAN) is proposed to preprocess the power sequence. Then, the full convolutional network (FCN) model optimized based on the sparrow search algorithm (SSA) is used to predict the short-term photovoltaic power. SSA can more reasonably determine the parameters of FCN and improve the prediction performance of FCN. Therefore, the FCN model optimized by the SSA algorithm is used to establish prediction models for subsequences and predict each subsequence, respectively. Finally, the predicted value of each subsequence is superimposed. Taking the actual data of a photovoltaic power station in Jiangsu province of China as an example, by comparing some different common prediction models, it is proved that the proposed method is reasonable and feasible.

1. Introduction

With the increase in power demand, the use of fossil energy has a bad impact on the environment, resulting in huge global climate change. Therefore, a large number of people develop renewable energy. Therefore, accurate short-term photovoltaic power prediction can effectively alleviate the pressure of photovoltaic grid connection on the power system. Therefore, it is urgent to put forward an accurate prediction model, which is of great significance to ensure the stable operation of the power grid and the rational allocation of resources [1, 2].

Solar radiation and various meteorological factors make the prediction of photovoltaic power generation difficult. Therefore, to solve this problem, various prediction methods have been proposed. At present, they can be roughly divided into four categories: physical model, statistical model, machine learning model, and hybrid model:

(1) A mathematical equation [3] is used to describe the physical state of the photovoltaic power generation system, which is classified as a physical model. When the weather is stable, the prediction accuracy of such models meets the demand, but when the weather changes violently, it cannot be predicted accurately [4, 5].

(2) Common statistical models include an autoregressive moving average (ARMA) and its improvement [6–8]. Exponential smoothing and regression are two major categories in statistical models [9–13].

(3) A machine learning model is developed based on a statistical model, but its prediction ability is stronger [14]. At present, the recurrent neural network (RNN) [15], limit learning machine (ELM) [16], support vector machine (SVM) [17], and other related models are commonly used in photovoltaic power generation prediction.

There are certain limitations in the separate prediction of the above three models. Therefore, by combining different prediction technologies, a hybrid model with higher prediction accuracy is designed. For example, the support vector machine (SVM) prediction of particle swarm optimization (PSO) based on wavelet transform (WT) [18] generates a mixture of the countermeasure network (GAN) and the convolutional neural network (CNN) [19].

At present, people usually analyze the relationship between the historical law of photovoltaic power data and external influencing factors, and thus, we establish a prediction model to predict the short-term photovoltaic power. However, at present, hybrid model prediction technology based on modal decomposition is a hot spot, and it cannot improve the prediction effect of photovoltaic power generation. For example, empirical mode decomposition (EMD) [20] decomposes the photovoltaic power sequence to reduce the nonstationary features of the photovoltaic power sequence so that the model can effectively extract subsequence features. EMD overcomes the defect that wavelet decomposition (WT) [21] requires human experience to select the basis function and decomposition levels in advance, but there is mode aliasing in the decomposition process. Ensemble empirical mode decomposition (EEMD) [22], complete ensemble empirical mode decomposition (CEEMD) [23], and CEEMDAN are improved methods of EMD, which eliminate mode aliasing in different ways. CEEMDAN shows good performance and is widely used in many fields [24–26]. In recent years, the application of deep learning in photovoltaic power prediction has achieved very remarkable results. A convolutional neural network (CNN) can learn the key features of local areas in the data sequence. Therefore, some researchers use CNN [27] as a feature extraction module to predict photovoltaic power generation in combination with LSTM. Some people also combined WT with a deep convolutional neural network (DCNN) and proposed a WT DCNN hybrid method for photovoltaic power generation prediction [28]. FCN is more flexible than CNN in processing time series data [29, 30]. Time series data can be input with any sequence length, and more sequence data features can be retained after FCN. For the hyperparametric optimization of neural networks, there are many optimization algorithms, but there is no experiment to prove that a certain algorithm is optimal. A genetic algorithm (GA) is a global optimization algorithm often used in the prediction model. The principle is to screen the population currently studied through the biological action mechanism and gradually select individuals with the highest fitness [31]. The principle of particle swarm optimization (PSO) is to use examples in the population to realize optimization by learning to continuously adjust the position and speed [32]. The sparrow search algorithm (SSA) is inspired by the sparrow's three behaviors of predation, tracking, and reconnaissance. According to the newly proposed prediction model, the sparrow search algorithm has good parameter optimization ability [33].

Therefore, this paper proposes a short-term power prediction model (SSA-CEEMDAN-FCN model) for photovoltaic power plants. The advantages of this model are as follows:

(1) The photovoltaic power generation sequence is preprocessed by CEEMDAN. This improved data preprocessing method can well reduce the complexity of subsequence and improve the prediction performance of the model.

(2) Considering the differences between subsequences, the key super parameters of the FCN model are optimized by SSA to obtain FCN with high prediction accuracy. The combination of SSA-FCN makes the FCN model play a better prediction performance.

(3) Through the combination of CEEMDAN and SSA-FCN, a model with high prediction accuracy for the short-term power of the photovoltaic power station is obtained, which plays an important role in improving the utilization rate of photovoltaic power generation.

2. Methodology

2.1. Complete Ensemble Empirical Mode Decomposition with the Adaptive Noise Algorithm. Based on EMD, CEEMDAN overcomes the phenomenon of mode aliasing by adaptively adding Gaussian white noise and effectively decomposes nonstationary sequences. The specific steps of using CEEMDAN to decompose photovoltaic power sequence are as follows.

Let Y be the historical photovoltaic power sequence, $E_j(*)$ be the j-th order modal component operator generated by EMD, $\omega_n(t)$ be the Gaussian white noise sequence added for the n-th time, IMF be the k-th order eigenmode decomposition sequence obtained by CEEMDAN, and δ_{k-1} be the adaptive coefficient for solving IMF_k.

(1) We add adaptive Gaussian white noise $\delta_0 \omega n(t)$ to the original sequence Y and $n = 1, 2, \ldots$, where N is the number of additions, i.e.,

$$Y_n = Y + \delta_0 \omega_n(t). \tag{1}$$

EMD decomposition is carried out for Y, respectively, so that imf_n^1 is the first-order modal component sequence of Y_n obtained by EMD. Then, the first-order eigenmode component sequence IMF_1 and the first residual component sequence r_1 of CEEMDAN decomposition are, respectively, as follows:

$$\begin{aligned} IMF_1 &= N^{-1} \sum_{n=1}^{N} imf_n^1, n = 1, 2, \ldots, N, \\ r_1 &= Y - IMF_1. \end{aligned} \tag{2}$$

(2) We add adaptive Gaussian white noise $\delta_1 E_1(\omega_n(t))$ and $n = 1, 2, \ldots, N$ to the residual sequence r_1, that is,

$$r_{1n} = r_1 + \delta_1 E_1(\omega_1(t)). \tag{3}$$

EMD decomposition of r_{1n}, respectively, then the second-order eigenmode component sequence IMF_2 of CEEMDAN decomposition is as follows:

$$IMF_2 = N^{-1}\sum_{n=1}^{N} E_1\left(r_1 + \delta_1 E_1\left(\omega_n(t)\right)\right). \tag{4}$$

(3) For $k = 2, 3, \ldots, K$, repeat step (2) to obtain the $k+1$-order eigenmode component sequence and the k-th residual component sequence, namely,

$$\begin{aligned} IMF_2 &= N^{-1}\sum_{n=1}^{N} E_1\left(r_1 + \delta_1 E_1\left(\omega_n(t)\right)\right), \\ r_k &= r_{k-1} + IMF_k. \end{aligned} \tag{5}$$

(4) Until the number of extreme points of the residual component sequence does not exceed 2, the final residual component sequence R is as follows:

$$R = Y - \sum_{k=1}^{K} IMF_k. \tag{6}$$

The historical photovoltaic power sequence Y is decomposed into K eigenmode component sequences IMF_k and a residual component sequence R by CEEMDAN, with a total of $K+1$ subsequences.

2.2. Full Convolutional Neural Network. A convolutional neural network is widely used, but it can only solve the problem of fixed input size. To solve this problem, a full convolutional network is proposed. FCN has no full connection layer, which eliminates its restrictions on the shape of input data and the end-to-end training process of input and output. The processing of time series generally uses a one-dimensional CNN model. Therefore, this paper uses a one-dimensional FCN model to predict the power of a photovoltaic power station.

(1) For the input data, the convolution kernel moves from left to right and from top to bottom, multiplies and sums each site covered by the convolution kernel with the convolution kernel, and finally obtains the output of this layer. The convolution process is shown in Figure 1(a). The convolution layer can extract the features of input data. In a full convolutional neural network, with the deepening of layers and the increase of the receptive field, it can extract the deep features of input data.

(2) The pooling layer of FCN is also known as the lower sampling layer. Using the pooling layer to remove redundant features in FCN can increase the receptive field, further reduce parameters, and prevent the network from overfitting. Figure 1(b) shows these three methods.

(3) Upper layer sampling is the convolution inverse process of the full convolutional network, which is used to enlarge the pooled characteristic map. FCN upper layer sampling uses the double line interpolation method. As shown in Figure 1(c), the values of function F in $G_{11} = (x_1, y_1)$, $G_{12} = (x_1, y_2)$, $G_{21} = (x_2, y_1)$, and $G_{22} = (x_2, y_2)$ are known, and the double line interpolation is used to obtain the value of at point $K = (x, y)$. Therefore, the sampling process is completed by using four adjacent points in FCN.

$$\begin{aligned} F(x, y_1) &= \frac{x_2 - x}{x_2 - x_1} F(G_{11}) + \frac{x - x_1}{x_2 - x_1} F(G_{21}), \\ F(x, y_1) &= \frac{x_2 - x}{x_2 - x_1} F(G_{12}) + \frac{x - x_1}{x_2 - x_1} F(G_{22}), \\ F(x, y) &= \frac{y_2 - y}{y_2 - y_1} F(x, y_1) + \frac{y - y_1}{y_2 - y_1} F(x, y_2). \end{aligned} \tag{7}$$

(4) The traditional FCN network will make the prediction very rough after data processing in the convolution layer and pooling layer. The use of jump structure is to improve the accuracy of prediction.

2.3. Sparrow Search Algorithm. SSA divides sparrows into discoverers and accessors. The identities of discoverers and accessors in sparrows can be exchanged, but the overall proportion remains unchanged. The discoverer is responsible for searching for areas rich in food resources in the overall situation. The participants will monitor the discoverer and search for food or rob the discoverer's food in the area close to the discoverer. When the sparrow population finds danger, sparrows in the marginal area will quickly move closer to the safe area, and sparrows in the central area will move randomly. The steps of SSA are as follows:

(a) We initialize the sparrow population and define its relevant parameters, calculate and sort the fitness of all sparrows, find the sparrow with the best global fitness, and record its fitness value and its global optimal location.

(b) Iterate to update the location of the finder, entrant, and danger sensing sparrow. If the current global optimal fitness value is higher than the previous generation's optimal value, the update operation will be carried out. If not, the update will not be carried out and the iteration will continue.

(c) The fitness function converges or meets the conditions to obtain the global optimal value and the optimal fitness value.

Compared with other traditional optimization algorithms, it is easier to obtain the global optimal solution to the optimization problem. Therefore, SSA is used to optimize FCN, specifically as shown in Figure 2.

2.4. Structure Diagram of the SSA-CEEMDAN-FCN Model. According to the previous basic theory, the overall framework flowchart based on the SSA-CEEMDAN-FCN combination model proposed in this study is shown in Figure 3. First, the power prediction sequence of the photovoltaic power station is preprocessed by CEEMDAN, and then, the

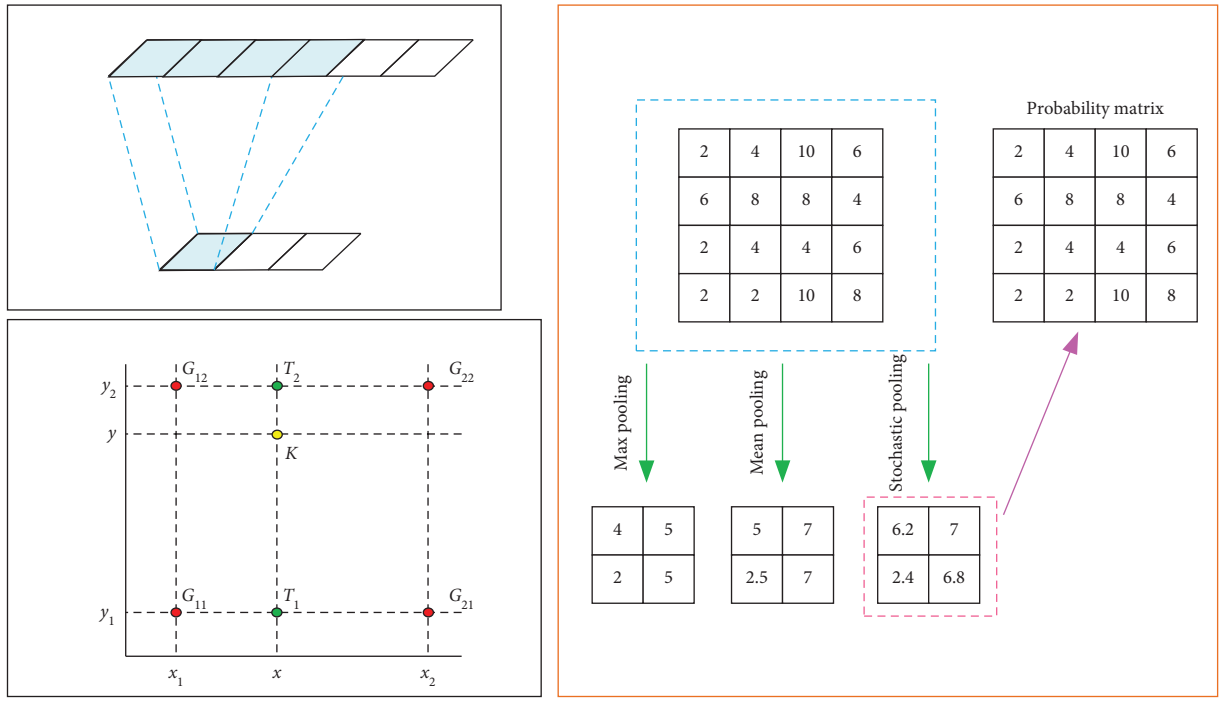


Figure 1: Convolution diagram.

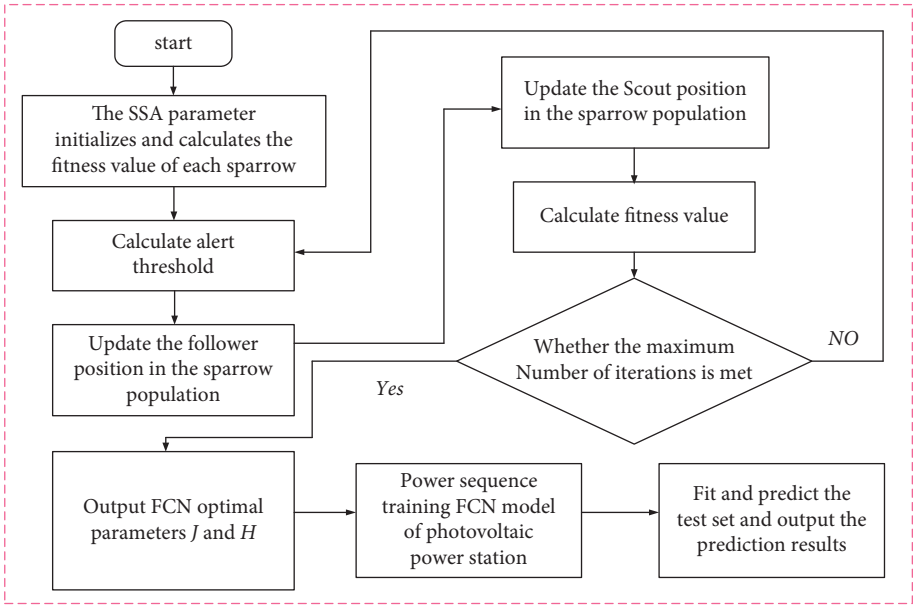


Figure 2: Flowchart of FCN optimized based on SSA.

decomposed subsequences are predicted by the SSA-optimized FCN model. Finally, the predicted values of subsequences are superimposed to obtain the final prediction results.

3. Actual Case Analysis and Verification

3.1. Data Description. Because this study belongs to the ultrashort-term prediction of photovoltaic power generation, from the measured data of photovoltaic power stations in a region of Jiangsu from 2019 to 2029, the data with a time length of 100 days are randomly selected for simulation experiments. Each experiment only needs to take 10 days for training and prediction, and the sampling period is 15 min. Two groups of continuous data with a length of 10 days are selected. The first 8 days of data are used as the training set of the model, the ninth day is used as the verification set, and the last 10 days of data are used as the

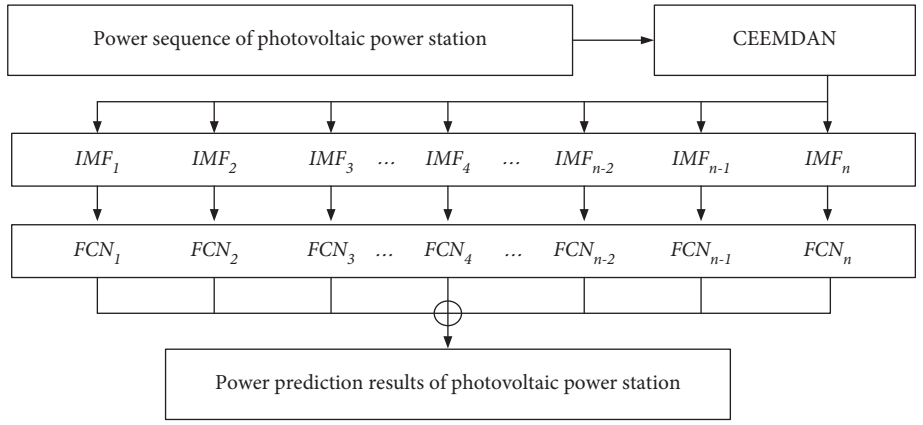


Figure 3: SSA-CEEMDAN-FCN prediction flowchart.

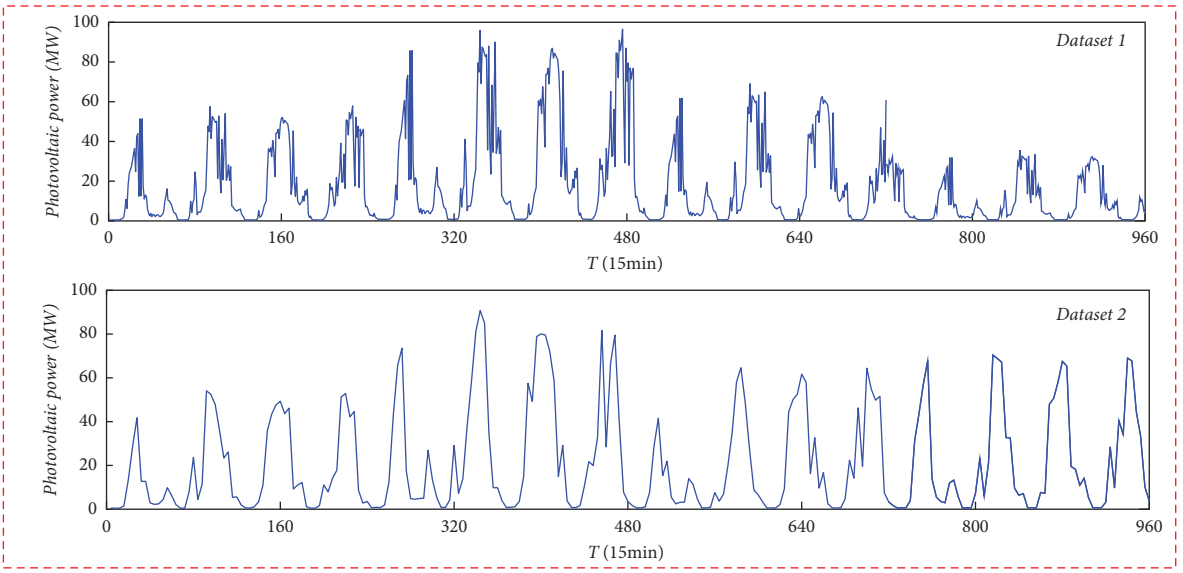


Figure 4: Photovoltaic power sequence diagram.

test. The total installed capacity is 100 MW, and the time is from 06: 00 to 18: 00 every day. The 100-day photovoltaic power sequence and CEEMDAN decomposition sequence are shown in Figures 4 and 5, respectively, where the x-axis of Figure 5 represents the frequency of each subsequence.

It can be seen from Figures 4 and 5 that the subsequence of the CEEMDAN-decomposed sequence has more gentle characteristics. Dataset1 and dataset2 are decomposed into 5 and 6 subsequences, respectively, and then, there is a residual sequence, respectively. CEEMDAN can well reduce the complexity of the sequence.

3.2. Model Prediction and the Evaluation Index. This study uses two classical prediction and evaluation indexes, the mean absolute error (map) and the root mean square error (RMSE). The smaller the value of these two indicators, the better the performance of the model. Their expressions are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|,$$
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}, \quad (8)$$

where n is the predicted total number, Y_i is the actual photovoltaic power value in step i, and Y_i is the predicted value in step i.

3.3. Comparison of Prediction Results between Different Models. To reflect the superior prediction performance of the SSA-CEEMDAN-FCN model proposed in this study, a single model and their combined models will be used for prediction comparison. At present, the commonly used time series models with high prediction accuracy include ARIMA, LSTM, and ELM. The first 80% of the dataset is

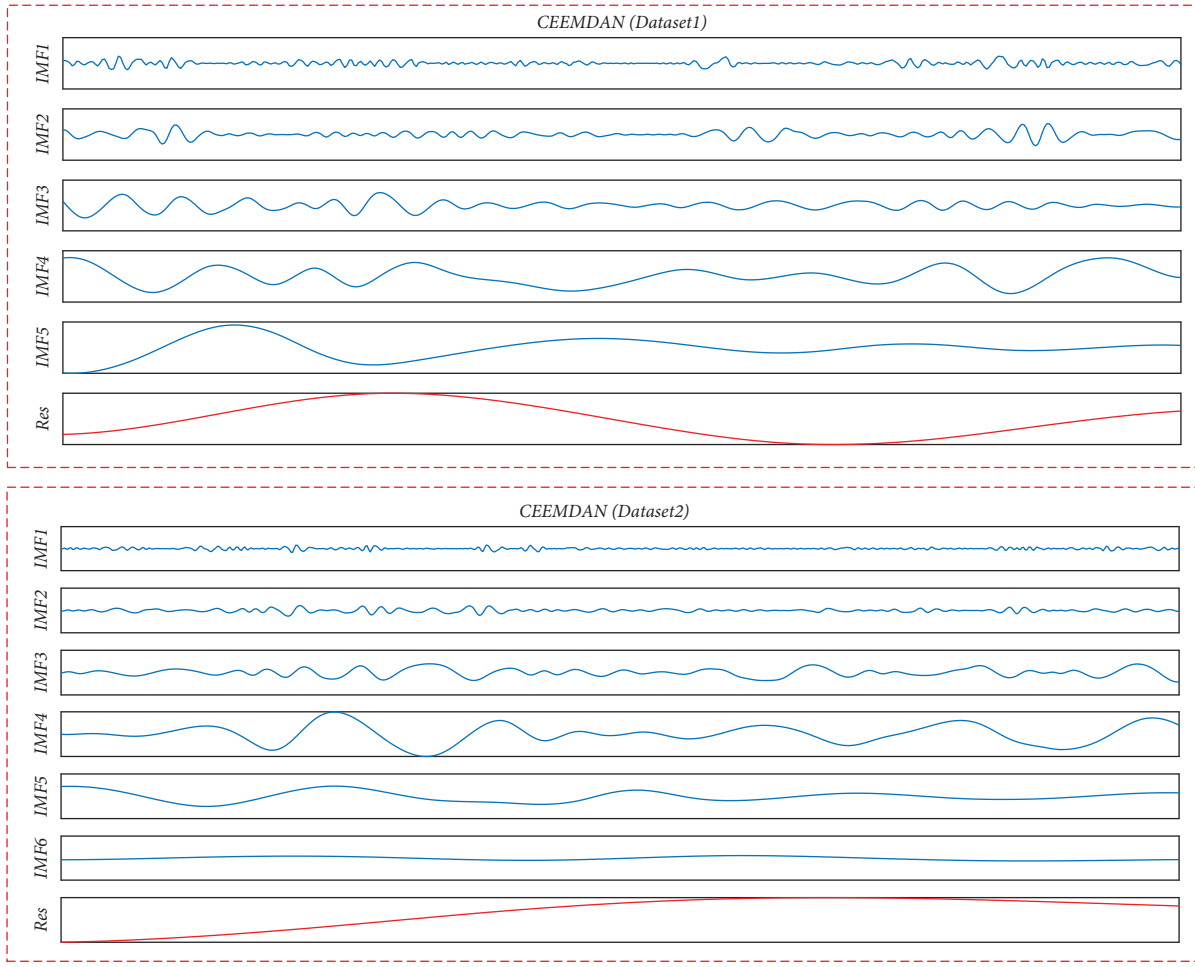


Figure 5: Photovoltaic power decomposition sequence.

used as the training set, 80~90% as the verification set, and the last 10% as the test set.

3.3.1. Prediction of the Single Model in the Dataset. In dataset 1, the four single models ARIMA, LSTM, GRU, and FCN are compared for prediction. Table 1 shows that compared with the other three single models, FCN has the highest prediction accuracy, and the prediction accuracy of the other three models is almost the same, but the prediction error is large.

3.3.2. Comparison of Combined Models with the Decomposition Algorithm. Because the single model is a part of the combined model, it reflects the core prediction part of the combined model to a certain extent. For example, the prediction performance of the FCN model is the highest among several single models, but after the decomposition algorithm is added, the prediction accuracy of different models is improved by different percentages. Therefore, it is also necessary to verify whether the prediction performance of these combined models is improved and whether the prediction accuracy of the SSA-CEEMDAN-FCN model is the highest.

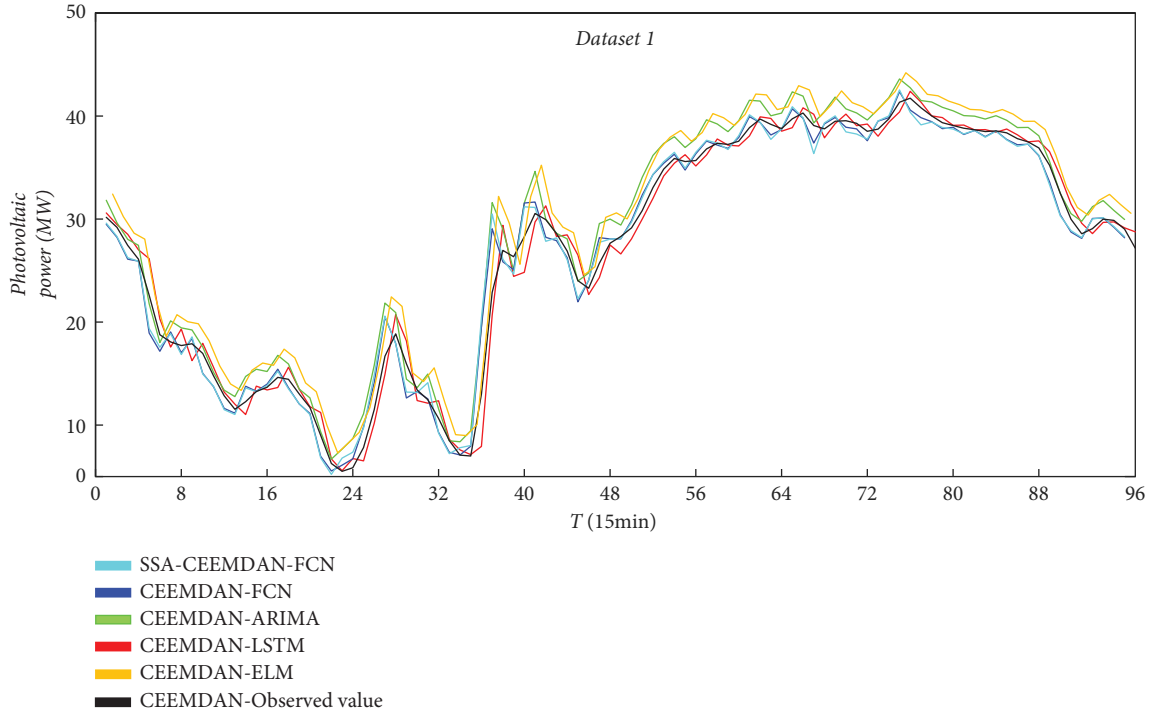
Table 1: Single model prediction performance versus results.

Dataset	Evaluating indicator (MW)	FCN	ARIMA	LSTM	ELM
1	MAE	6.57	9.34	9.21	8.24
	RMSE	7.20	9.21	8.76	8.61
2	MAE	5.47	9.87	8.87	8.17
	RMSE	6.32	10.01	9.76	9.55

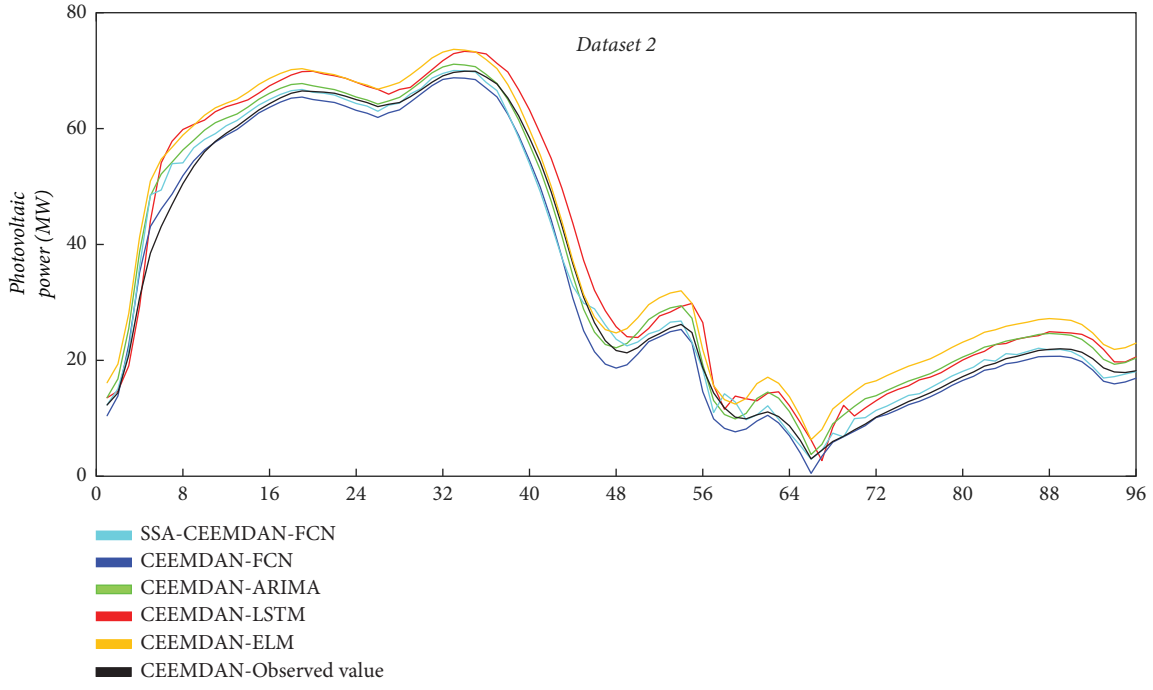
It can be concluded that the prediction curve of the SSA-CEEMDAN-FCN model is the closest to the real observation value, and the prediction index of the SSA-CEEMDAN-FCN model is the smallest in both groups of data, indicating that its prediction accuracy is the highest. At the same time, we can conclude that after adding the CEEMDAN decomposition algorithm, the prediction accuracy of the combined model will be improved compared with that of the single model. Therefore, it also proves the superiority of the combined model. Therefore, the method proposed in this paper can better predict the changing trend of short-term photovoltaic power on the whole, and the prediction error is small. For example, in dataset 2, the prediction indexes RMSE and Mae of SSA-CEEMDAN-FCN are 2.32 and 1.47,

Table 2: Results of combined model prediction performance.

Dataset	Evaluating indicator (MW)	SSA-CEEMDAN-FCN	CEEMDAN-FCN	CEEMDAN-ARIMA	CEEMDAN-LSTM	CEEMDAN-ELM
1	MAE	2.57	3.47	5.34	5.21	4.24
	RMSE	3.20	4.10	5.21	4.76	4.61
2	MAE	1.47	3.27	5.87	4.87	4.17
	RMSE	2.32	4.22	6.01	5.76	5.55



(a)



(b)

Figure 6: Comparison of prediction results of combined models of datasets 1 and 2.

Table 3: Comparison of different optimization algorithms.

Dataset	Evaluating indicator (MW)	FCN (SSA)	FCN (GA)	FCN (BHA)	FCN (GWO)
1	MAE	6.57	7.38	9.19	7.99
	RMSE	7.20	8.11	8.63	8.43
2	MAE	5.47	9.64	7.84	8.11
	RMSE	6.32	9.87	8.33	9.35

respectively, which are far less than those of the other three combined models. The details are shown in Table 2 and Figure 6.

3.3.3. FCN Model Prediction Performance under Different Optimization Algorithms. If only one parameter optimization method is used in the experiment, it will have the disadvantage of insufficient persuasion. Therefore, here, we use several other common classical optimization algorithms to optimize the model and compare their prediction results to prove the rationality of using the SSA method in this paper. The three comparison methods are the genetic optimization algorithm (GA), black hole optimization algorithm (BHA), and gray wolf optimization algorithm (GWO). Their comparison results in the two groups of data are shown in Table 3.

Therefore, from Table 3, we can see that the FCN model optimized by the FCN algorithm has the highest prediction accuracy, so the SSA optimization algorithm is used in the FCN network in the combined model proposed in this paper, which is very reasonable.

4. Conclusion

(1) The photovoltaic power sequence is decomposed into subsequences with different frequencies by CEEMDAN, which can reduce the complexity of photovoltaic power and is conducive to power characteristic analysis, modeling, and prediction.

(2) The FCN model optimized based on the sparrow search algorithm has better prediction performance, because the performance of SSA-CEEMDAN-FCN is better than that of the CEEMDAN-FCN model.

(3) The SSA-CEEMDAN-FCN model is a combination model with excellent performance. At present, the prediction accuracy of LSTM and ELM models and their combined models with high prediction accuracy is lower than that of the SSA-CEEMDAN-FCN model. Therefore, the SSA-CEEMDAN-FCN model combined with the FCN model further improves the prediction accuracy of photovoltaic power.

The decomposition subsequence of photovoltaic power has different characteristic changes, and the prediction method of its higher performance needs to be further studied. Therefore, the photovoltaic power prediction will be further studied from these two aspects in the future.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the State Grid Jiangsu Electric Power 2021 Marketing System Optimization and Business Operation Service Project (B368B121075900ZR000000).

References

[1] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, Article ID 120240, 2021.

[2] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and PV power forecasting: a preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, 2018.

[3] U. K. Das, K. S. Tey, M. Seyedmahmoudian et al., "Forecasting of photovoltaic power generation and model optimization: a review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.

[4] A. Dolara, S. Leva, and G. Manzolini, "Comparison of different physical models for PV power output prediction," *Solar Energy*, vol. 119, pp. 83–99, 2015.

[5] D. Koster, F. Minette, C. Braun, and O. O'Nagy, "Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg," *Renewable Energy*, vol. 132, pp. 455–470, 2019.

[6] J. Boland, M. David, and P. Lauret, "Short term solar radiation forecasting: island versus continental sites," *Energy*, vol. 113, pp. 186–192, 2016.

[7] G. Reikard, "Predicting solar radiation at high resolutions: a comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3, pp. 342–349, 2009.

[8] Y. Li, Y. Su, and L. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 66, pp. 78–89, 2014.

[9] X. Zhang, F. Fang, and J. Liu, "Weather-classification-mars-based photovoltaic power forecasting for energy imbalance market," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 11, pp. 8692–8702, 2019.

[10] G. Wang, Y. Su, and L. Shu, "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renewable Energy*, vol. 96, pp. 469–478, 2016.

[11] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 300–308, 2018.

[12] J. R. Trapero, N. Kourentzes, and A. Martin, "Short-term solar irradiation forecasting based on dynamic harmonic regression," *Energy*, vol. 84, pp. 289–295, 2015.

[13] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for gefcom2014 probabilistic wind power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1074–1080, 2016.

[14] J. MuellerPaul and L. Massaron, *Machine Learning for Dummies*, , 2016.

[15] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," in *Proceedings of the 47th North American Power Symposium*, pp. 1–5, Charlotte, NC, USA, October 2015.

[16] P. Tang, D. Chen, and Y. Hou, "Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting," *Chaos, Solitons & Fractals*, vol. 89, pp. 243–248, 2016.

[17] J. V. Hansen, J. B. McDonald, and R. D. Nelson, "Some evidence on forecasting time-series with support vector machines," *Journal of the Operational Research Society*, vol. 57, no. 9, pp. 1053–1063, 2006.

[18] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and meteorological information," *Renewable Energy*, vol. 118, pp. 357–367, 2018.

[19] F. Wang, Z. Zhang, C. Liu et al., "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy Conversion and Management*, vol. 181, pp. 443–462, 2019.

[20] R. Prasad, M. Ali, P. Kwan, and H. Khan, "Designing a multistage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation," *Applied Energy*, vol. 236, pp. 778–792, 2019.

[21] S. Soltani, "On the use of the wavelet decomposition for time series prediction," *Neurocomputing*, vol. 48, no. 1-4, pp. 267–277, 2002.

[22] Z. Liu, W. Sun, and J. Zeng, "A new short-term load forecasting method of power system based on EEMD and SS-PSO," *Neural Computing & Applications*, vol. 24, no. 3-4, pp. 973–983, 2014.

[23] Y. Ding, Z. Chen, H. Zhang, X. Wang, and Y. Guo, "A short-term wind power prediction model based on CEEMD and WOA-KELM," *Renewable Energy*, vol. 189, pp. 188–198, 2022.

[24] Z. Zhao, H. Nan, Z. Liu, and Y. Yu, "Multi-step interval prediction of ultra-short-term wind power based on CEEMDAN-FIG and CNN-BiLSTM," *Environmental Science and Pollution Research International*, vol. 29, 2022.

[25] Z. Yang, L. Zou, J. Xia, Y. Qiao, and D. Cai, "Inner dynamic detection and prediction of water quality based on CEEMDAN and GA-SVM models," *Remote Sensing*, vol. 14, no. 7, p. 1714, 2022.

[26] F. Zhou, Z. Huang, and C. Zhang, "Carbon price forecasting based on CEEMDAN and LSTM," *Applied Energy*, vol. 311, Article ID 118601, 2022.

[27] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Applied Energy*, vol. 259, Article ID 114216, 2020.

[28] H. Wang, H. Yi, J. Peng et al., "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Conversion and Management*, vol. 153, pp. 409–422, 2017.

[29] T. Tian, Z. Chu, Q. Hu, and L. Ma, "Class-wise fully convolutional network for semantic segmentation of remote sensing images," *Remote Sensing*, vol. 13, no. 16, p. 3211, 2021, https://www.mdpi.com/2072-4292/13/16/3211.

[30] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FCNS for time series classification," *Neural Networks*, vol. 116, pp. 237–245, 2019.

[31] L. Qian, Y. Zheng, L. Li, Y. Ma, C. Zhou, and D. Zhang, "A new method of inland water ship trajectory prediction based on long short-term memory network optimized by genetic algorithm," *Applied Sciences*, vol. 12, no. 8, p. 4073, 2022, https://www.mdpi.com/2076-3417/12/8/4073.

[32] J. Li, Z. Song, X. Wang, Y. Wang, and Y. Jia, "A novel offshore wind farm typhoon wind speed prediction model based on PSO-Bi-LSTM improved by VMD," *Energy*, vol. 251, Article ID 123848, 2022.

[33] J. Jia, S. Yuan, Y. Shi, J. Wen, X. Pang, and J. Zeng, "Improved sparrow search algorithm optimization deep extreme learning machine for lithium-ion battery state-of-health prediction," *iScience*, vol. 25, no. 4, Article ID 103988, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6826573, 10 pages
https://doi.org/10.1155/2022/6826573

Research Article

The Evaluation on the Credit Risk of Enterprises with the CNN-LSTM-ATT Model

Lei Zhang [1,2]

[1]*School of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing 400074, China*
[2]*School of Economics and Management, Chongqing Jiaotong University, Chongqing 400074, China*

Correspondence should be addressed to Lei Zhang; zhangleicqjtu@163.com

Received 26 July 2022; Revised 4 September 2022; Accepted 8 September 2022; Published 22 September 2022

Academic Editor: Man Fai Leung

Credit evaluation is a difficult problem in the process of financing and loan for small and medium-sized enterprises. Due to the high dimension and nonlinearity of enterprise behavior data, traditional logistic regression (LR), random forest (RF), and other methods, when the feature space is very large, it is easy to show low accuracy and lack of robustness. However, recurrent neural network (RNN) will have a serious gradient disappearance problem under long sequence training. This paper proposes a compound neural network model based on the attention mechanism to meet the needs of enterprise credit evaluation. The convolutional neural network (CNN) and the long short-term memory (LSTM) network were used to establish the model, using soft attention, the gradient propagates back to other parts of the model through the attention mechanism module. In the multimodel comparison experiment and three different enterprise data experiments, the CNN-LSTM-ATT model proposed in this paper is superior to the traditional models LR, RF, CNN, LSTM, and CNN-LSTM in most cases. The experimental results under multimodel comparison reflect the higher accuracy of the model, and the group test reflects the higher robustness of the model.

1. Introduction

Small and medium-sized enterprises (SMES) are playing an increasingly important role in supporting the national economy and have made great contributions to innovative output and labor employment. However, due to the difficulties of credit evaluation, small and medium-sized enterprises are currently facing the problem of financing and loans, which has great resistance to the operation and development of enterprises. In this context, it is urgent to build a reliable credit and risk assessment system to solve the problem of the lack of credit and risk assessment in the financing and loan process of small and medium-sized enterprises, and this is of great significance for financial services supervision.

In recent years, most credit risk evaluation models have been paid much attention by many scholars, which are constructed by statistical and machine learning methods [1–9]. The traditional methods such as LR and RF have low accuracy when the feature space is large. Meanwhile, CNN is widely used in computer vision simulation [10, 11] and human speech recognition [12–14]. Due to the explosive development of big data, which drives the development of neural networks, research results in related fields of CNN have achieved considerable progress in a wide range of classification and regression [15, 16]. The problem of serious gradient loss in RNN, LSTM, and neural network has also been developed rapidly in the field of emotion classification under a long sequence [17–20]. With the development of the economy and the development of neural networks, it is necessary to combine neural networks and risk evaluation issues. At the same time, the model combining CNN and LSTM has been applied to text classification [21–24], image recognition [25–28], emotion analysis [21, 29, 30], and other fields. Based on the deep learning method of CNN and LSTM, this paper aims at studying the problem of the enterprise's credit score by using enterprise's behavior data. Based on the enterprise's behavior data, the proposed model used in this paper integrates CNN and LSTM to evaluate the

credit risk of enterprises. LSTM could act in cooperation for the information back and forth in the long sequence, which has the better performance than the general recurrent neural network in the long sequence training [31, 32]. CNN could extract vertical and horizontal features. In this paper, CNN is used to extract features of the behavioral matrix. Moreover, attention mechanism is added to the model to extract important features [33–35]. In the aspect of data processing, the behaviors of the enterprises are represented by the behavior matrix, which could improve the predictive effect and understandability of the model.

Motivated by the above analysis, the CNN-LSTM-ATT composite neural network model is proposed to solve the problem of enterprises credit risk evaluation. The selection experiment for hyperparameter and comparison experiment among different models are conducted on the realistic data set. The accuracy and robustness of the CNN-LSTM-ATT model in solving the problem of credit risk evaluation are comprehensively analyzed. Based on the above discussion, this paper aims at solving the following problems:

(i) First of all, the financial information and behavior data of local enterprises are conducted, and data coding is represented as the two-dimensional matrix. After data preprocessing, the variable data set could finally be used as the input of neural networks, which have better effects on the training and testing of the model. There will be a lot of missing values and unqualified values in the original big data set. The missing values, that are some data of the enterprises may not exist or cannot be filled. For such data, we will fill them by comparing the average value of the whole data of enterprises. Unqualified values are that wrong values and abnormal outliers during the inputting process and are disposed by complementing the default values. After data processing, the data set could improve the accuracy and efficiency of prediction.

(ii) Secondly, the CNN-LSTM model is used for feature extraction of enterprise's behavior data in this paper. CNN model is used for feature extraction of enterprises, which is manifested in enterprise's behavior data. LSTM model is used for feature extraction under the long sequence of enterprises, which could prevent the occurrence that the effect of debt during earlier stage on the normal loan behavior after the later development.

(iii) Thirdly, the parameters of CNN, LSTM, and attention mechanism are adjusted to make sure that the predicted result could reach the expected acceptable result. In the neural network, the influence of parameters on the model results is especially obvious, which has a great influence on the prediction result. The parameters of all models in this paper need to be adjusted to the most reasonable conditions to achieve the best forecast results.

(iv) Finally, the comparison among the traditional single CNN model, single LSTM model, and proposed CNN-LSTM-ATT are carried out by using the control variable method. The prediction effect of the model is verified by calculating the AUC value. And the realistic enterprise's behavior data are substituted into the proposed model to verify the realistic environment. Therefore, it could achieve the correct prediction result.

2. Model Description and Preliminaries

2.1. Long Short Term Memory. Long short-term memory is used for feature extraction under the long sequence of enterprises. LSTM could act in cooperation for the information back and forth in the long sequence, which has the better performance than the general recurrent neural network. LSTM could deal with the explosive gradient and the vanishing gradient in the long sequence training.

LSTM has designed a special structural unit and three unique "gate" structures. This structure can selectively increase or remove the information of passing through the unit. So as to screen the information passing through the unit. The "gate" structure is implemented by Sigmoid function and the Sigmoid value ranges from 0 to 1, which can be regarded as how much information is allowed to pass through. The closer this value is to 0, the more difficult it is in passing through the information, and the more likely it will be abandoned. On the contrary, the closer the value is to 1, the less difficult it is in passing through the information, and the less likely it will be abandoned. The hidden layer of LSTM is formed by the following structure, as shown in Figure 1.

The structure includes the input x_t, the state variable c_t, the temporary state variable $\tilde{c}_t$, the hidden layer state h_t, the forgotten gate f_t, the memorial gate i_t, and the output o_t.

2.2. Convolutional Neural Networks. Convolutional neural networks is a kind of special feedforward neural networks containing convolution computation. It occupies a powerful position in deep learning. In terms of image recognition and human language analysis, CNN has the ability of representation learning, which enables CNN to classify translational invariant on input information according to its own hierarchical structure.

The general neural network structure is composed of the input layer, the hidden layer, and the output layer. The hidden layer of CNN has three different structures from the general NN model: the convolutional layer, the pooling layer, and the fully connected layer. In some more modern algorithms, there may be inception block, residual block, and other complex constructions. In the process of constructing neural networks, the convolutional layer and the pooling layer are unique for the convolutional neural networks.

2.3. Attention Mechanism. The concept of attention mechanism is from the research of human vision and related fields by experts and scholars. Attention mechanism realizes the rapid and effective allocation of resource information processing. In essence, a weight of attention mechanism is

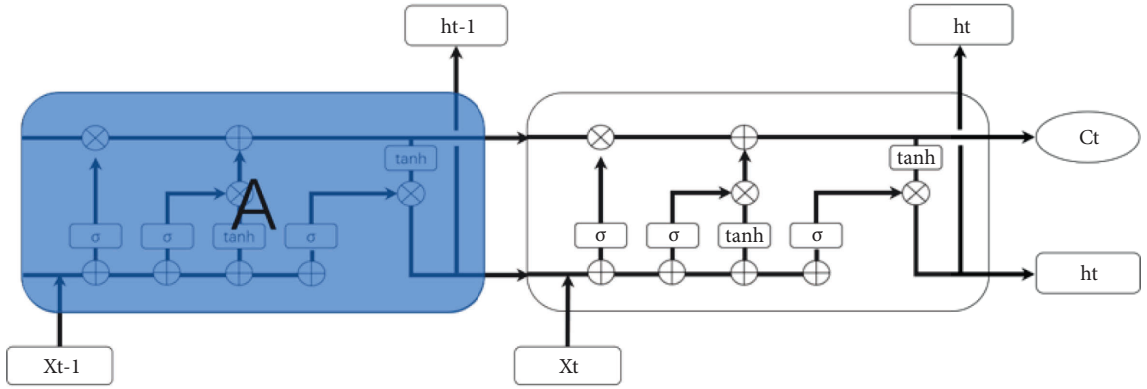


Figure 1: The unit structure of LSTM.

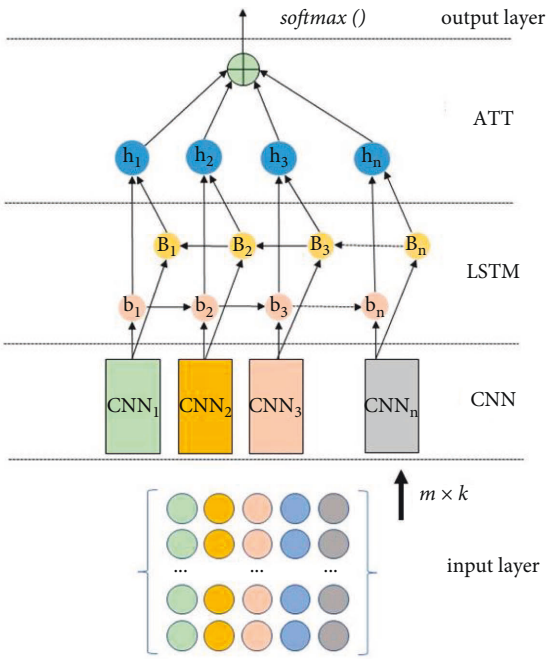


Figure 2: The structure of CNN-LSTM-ATT.

Table 1: The experimental environment configuration.

Experimental environment	Parameter
Operating system	Windows 10 × 64
Processor	CPU intel (*R*) core (TM) i5-9300H@ 2.40 GHz @2.40 GHz
Video card	GTX 1650 4G
Memory	8G
Hard disk	500G
Development language	*Python* (3.7.4)

Table 2: The indexes of calculating ROC.

Indexes	Meaning of indexes
TP	No default and correct prediction
TN	No default and wrong prediction
FP	No default and correct prediction
FN	No default and wrong prediction

added to the hidden layer of the neural network to capture the key features of information. It is found that the importance of feature saliency is positively correlated with the amount of contained information. In other words, the greater the importance of feature saliency, the more information will be contained, which will have a greater influence on the actual demand.

The channel attention mechanism learns the weight of each channel in the attention block to generate channel attention. The attention mechanism contains three parts: the squeeze, excitation, and attention. Firstly, the squeeze function are used as

$$F_{sq}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_c(i, j). \quad (1)$$

Making an average for the global situation, adding all eigenvalues in all channels and taking an average for them. It is essentially a calculation formula of global average pooling. And then the excitation function is

$$F_{ex}(z, w) = \partial(g(z, w)) = \partial(w_2 \delta(w_1 z)), \quad (2)$$

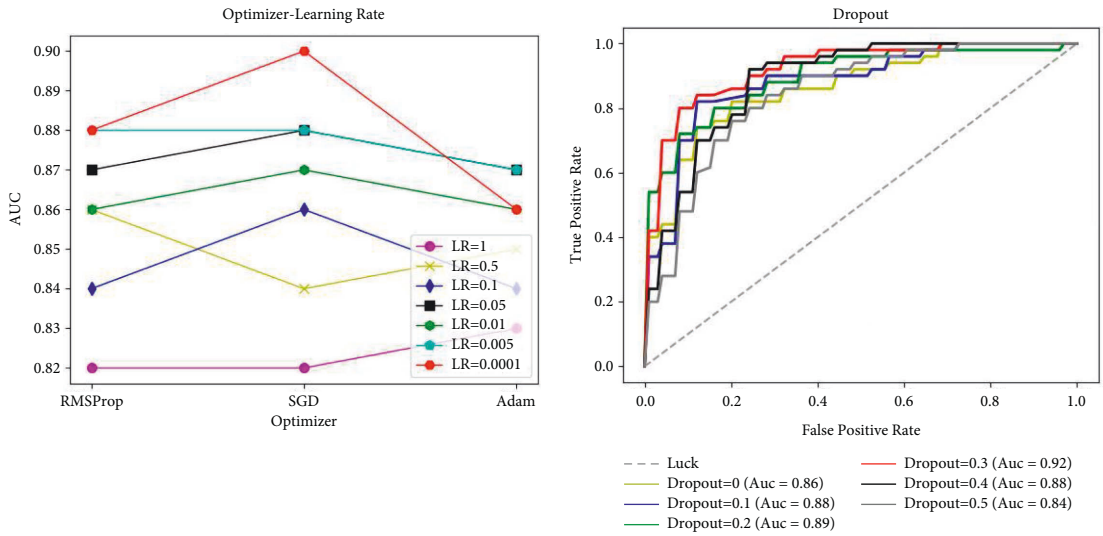
where w_1, w_2 are dimensionality, function ∂ is the Sigmoid function, and function δ is ReLU.

2.4. Model Description. This paper aims at using the neural networks to solve the problem of the enterprise's risk credit evaluation. In recent years, the development of big data has laid a good foundation for us to RNN. The reasonable and effective enterprise's risk evaluation could reduce risks for lending institutions and improve the utilization rate of funds. It helps prevent the frequent occurrence of credit risks which lead to the enterprises bankrupt or running away and other phenomena.

In this paper, the structure of the credit evaluation method based on CNN-LSTM-ATT model is presented as shown in Figure 2. Firstly, the characteristic behavior data of an enterprise is represented in the form of a matrix, and all the behavior characteristics of an enterprise are represented

Table 3: The parameter setting of the CNN-LSTM-ATT model.

Parameter	Meaning of parameter	Value of parameter
time_series	The step length	20
behavior_num	Number of behavioral indexes	14
iB-LSTM_units	Number of neurons for LSTM	64
Kernel_size1	The size of the convolution kernel for convolution layer 1	1×3
Kernel_size1	The size of the convolution kernel for convolution layer 2	3×3
Pool	Pooling method	Max
Stride	Length of pooling or convolutional	1
Dropout	Rate of LSTM dropout	0.3
batch_size	The size of batch	128
Epoch	Number of iterations	20
Optimizer	The optimizer	SGD
Learning rate	The learning rate	0.0001



(a) (b)

Figure 3: Experimental results of super-parameter setting. (a) Optimizer. (b) False positive rate.

in the form of row vector. Then input the data to the CNN. Next, it enters the CNN layer and the LSTM layer. Taking advantages of the CNN and the both way LSTM network to maintain behavior information and extract features. Then, it enters the attention mechanism layer. Using the attention mechanism to identify the most important characteristics of enterprises behavior and measuring the characteristics of LSTM output. Finally, the final classification of the characteristic output could be obtained.

The formulas used in the CNN layer include the convolution formula and the pooling formula:

$$\begin{aligned} c^{(l)} &= f\left(c^{(l-1)} \otimes w + b^{(l-1)}\right), \\ \tilde{c}^{(l)} &= \max_{PZ} c^{(l)}. \end{aligned} \tag{3}$$

The gate formula used in LSTM layer are:

$$\begin{aligned} f_t &= \sigma\left(w_f \cdot \left[h_{t-1}, x_t\right] + b_f\right), \\ i_t &= \sigma\left(w_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \\ o_t &= \sigma\left(w_o \cdot \left[h_{t-1}, x_t\right] + b_o\right). \end{aligned} \tag{4}$$

In the attention mechanism layer, the key special grasping problem in the enterprise is mainly solved and the feature extraction of the previous CNN-LSTM is strengthened more effectively, which improved the accuracy of prediction. The attention mechanism is used to extract the characteristics of enterprise behavior more effectively and to grasp the key points of behavioral data. The output eigenvectors of the LSTM layer are transformed into the matrix H. Denote the length of each vector as n. Use the obtained vector and the following formula to calculate the enterprise behavior:

$$\begin{aligned} M &= \tan h(H), \\ \alpha &= soft\max\left(w^n M\right), \\ r &= H\alpha^n, \end{aligned} \tag{5}$$

where $H \in R^{dn}$, d denotes the dimension of vector; w is the parameter vector, and w^n is the transposition of w; the dimension w, α, and r are corresponding to d, n, and d. Finally, the output of enterprise behavior for classification is as follows:

$$c^* = \tan h(r). \tag{6}$$

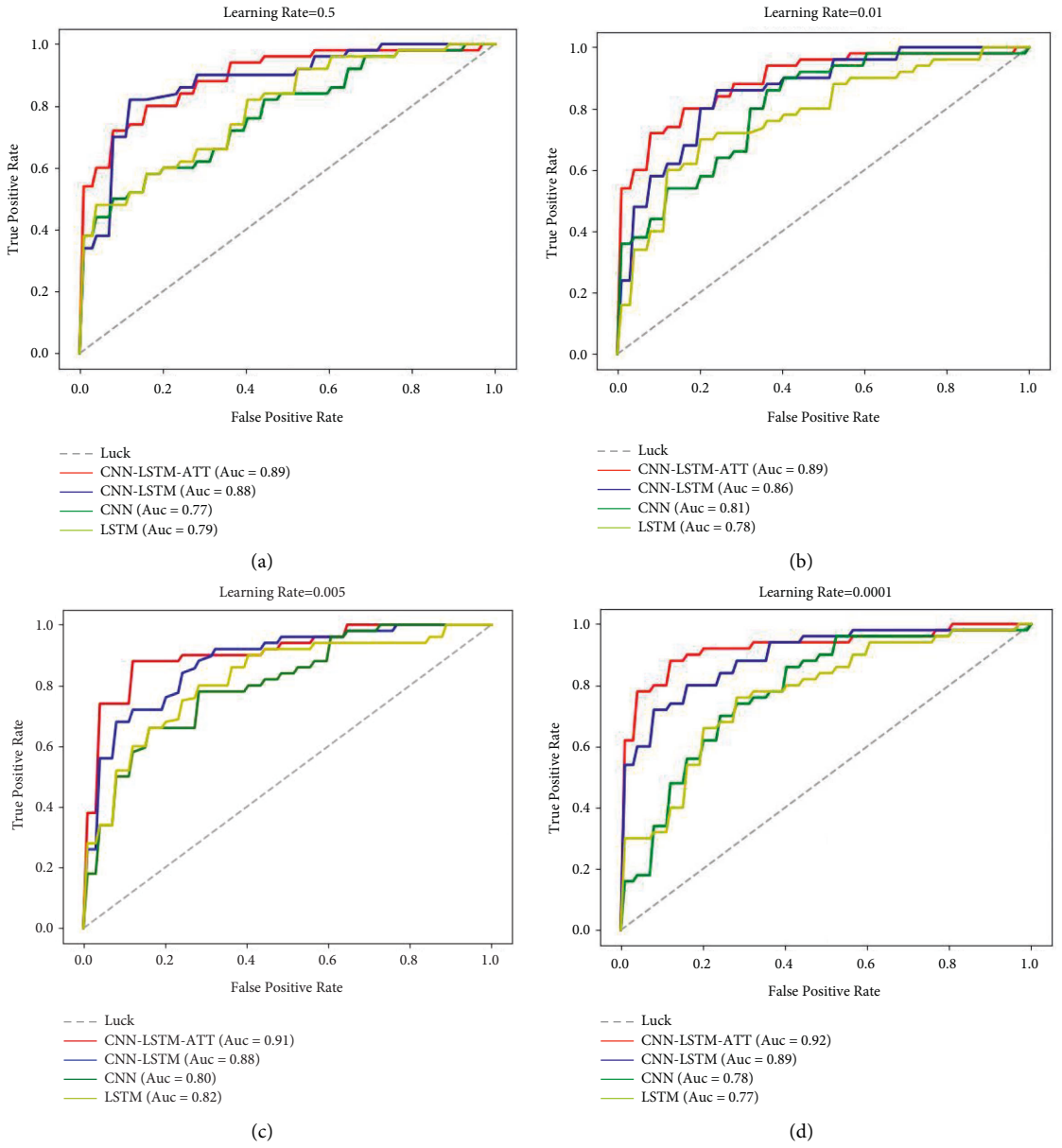


Figure 4: ROC curves of four models under different learning rates.

3. Setting of Experiment

3.1. Experimental Environment. The experimental environment configuration in this paper is shown in the following Table 1:

The development language of experiment in this paper is *Python* (3.7.4) and PyCharm integrated development environment is used. The matrix operation used in the experiment is calculated by NumPy library of *Python*. The Pandas library is used for data processing and data analysis. Scikit-learn library for machine learning and library Matplotlib for drawing tool are used to accomplish the experiment. The Keras library is regarded as a deep learning tool and the TensorFlow is chosen at the back end.

3.2. Parameter Setting of Experiment. In terms of indexes for model evaluation, it is necessary to evaluate the default probability of enterprises. The commonly used index in the field of credit scoring: AUC (area under curve) is selected in this paper. The value of AUC is equal to the area under the receiver operating characteristic (ROC) curve. The AUC value would be larger if the area under the ROC curve is larger. Furthermore, the classification result of the model will be better. ROC curve and AUG are often used to evaluate the strengths and weaknesses of a binary classifier. The horizontal ordinate of the ROC curve is false positive rate (FPR) and the vertical ordinate is True Positive Rate (TPR). Four indexes are needed to calculate FPR and TPR : TP, TN, FP, and FN. The specific meanings are shown in Table 2.

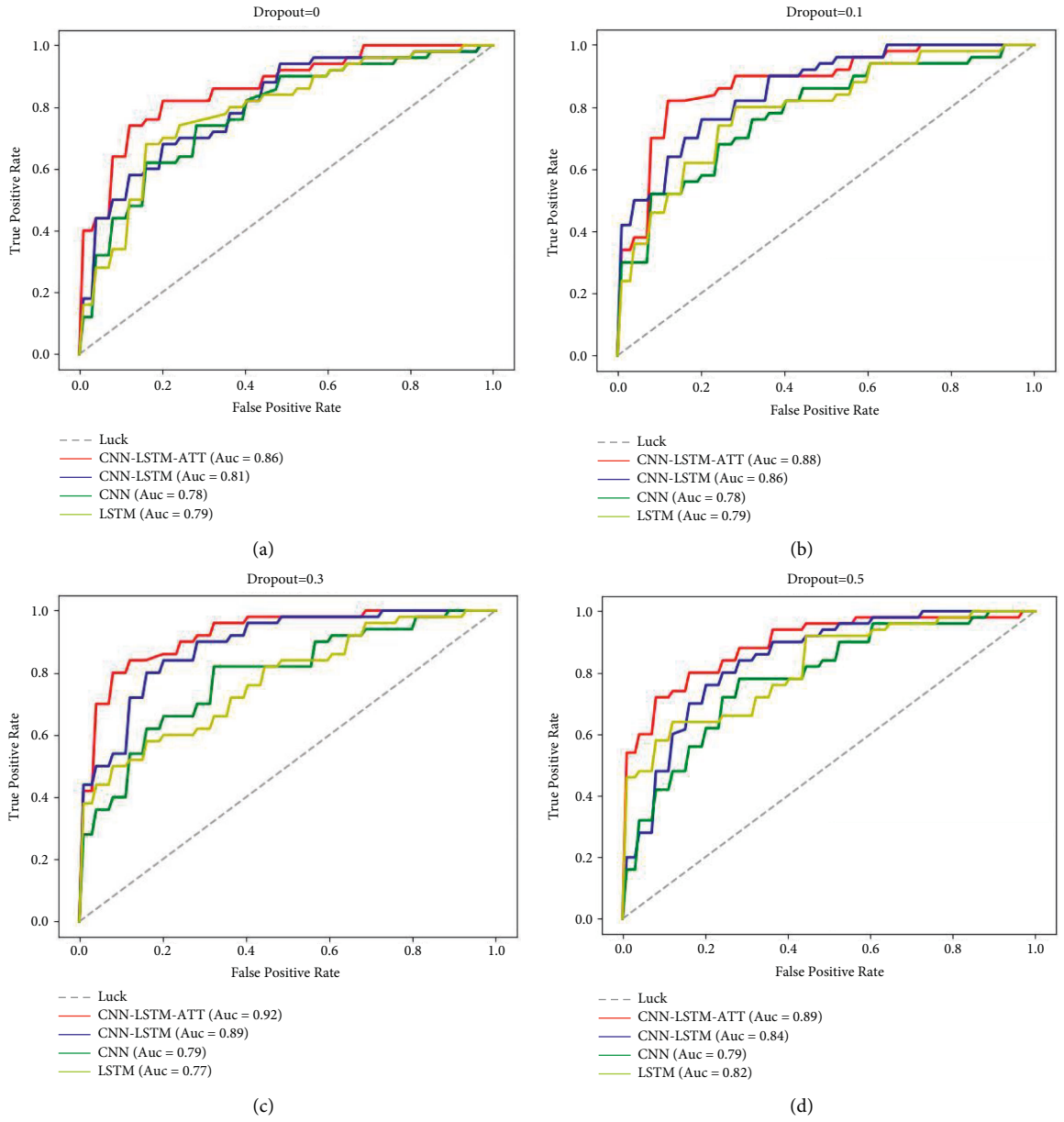


Figure 5: ROC curve of four models under different dropout values.

Then the calculation formula of TPR and FPR is given as follows:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}. \tag{7}$$

Table 3 represents the parameter setting of the CNN-LSTM-ATT model.

In this paper, a super-parameter selection experiment is carried out to determine the parameter values of optimizer, learning rate, and dropout. Figure 3 shows the schematic diagram of the results of the super-parameter selection experiment.

The result of optimization experiment is conducted on the optimizer-learning rate as Figure 3(a). When SGD is selected as the optimizer and the learning rate is 0.0001, the evaluation index achieves the optimal solution. Then make dropout selection experiment with the Optimizer = SGD, Learning Rate = 0.0001, and select Dropout as 0, 0.1, 0.2, 0.3, 0.4, and 0.5 for five times, respectively. The results are shown in Figure 3(b). When the dropout is selected as 0.3, the optimal solution of the evaluation index is obtained. In conclusion, the Optimizer = SGD, Learning Rate = 0.0001, and Dropout = 0.3.

4. The Experimental Comparison among Different Models

In this section, the four models of CNN, LSTM, CNN-LSTM, and CNN-LSTM-ATT are compared and tested.

Table 4: The AUC of different models under three groups data.

Model	Group-0	Group-1	Group-2	Mean
CNN-LSTM-ATT	**0.92**	**0.89**	**0.94**	**0.92**
CNN-LSTM	0.89	0.86	0.94	0.89
CNN	0.82	0.88	0.72	0.80
LSTM	0.83	0.75	0.82	0.80

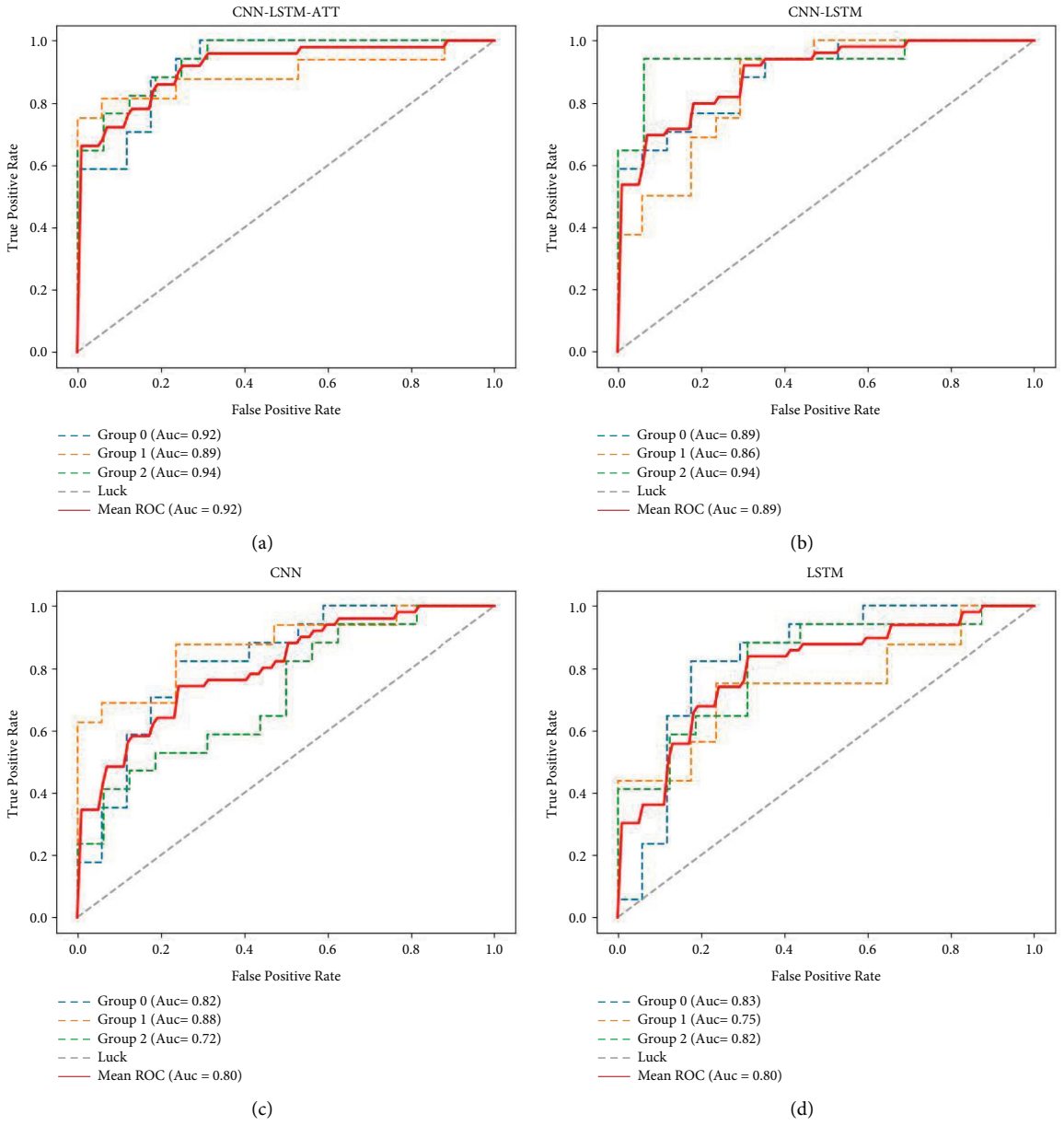


Figure 6: ROC curve of four models under three groups data.

Firstly, the evaluation performance of the four models under different parameters is analyzed and compared, respectively. Then, a set of fixed parameters is selected for the realistic data grouping test and comparison of the four models.

4.1. Comparison of the Four Models under Different Parameters. Similar to the super-parameter setting experiment, select Dropout = 0.3, Optimizer = SGD. The evaluation performance of the four models (CNN, LSTM, CNN-LSTM, and CNN-LSTM-ATT) under different LR parameters is analyzed and compared. The experimental results are shown in Figure 4. It can be seen from Figure 4 that the CNN-LSTM-ATT model used in this paper has the best result among the four models no matter which value the LR takes.

We found that the evaluation results of the four models are more comparative when the Learning Rate = 0.005. The evaluations of CNN and LSTM are near the optimal solution. Therefore, we chose Optimizer = SGD, Learning Rate = 0.005, and then made a comparative analysis experiment on dropout changes of multiple models. The experimental results are shown in Figure 5.

The results of dropout experiments with the Optimizer = SGD and the Learning Rate = 0.005 show that the CNN-LSTM-ATT model used in this paper has a better evaluation result than the other three comparison models in a variety of dropout parameter settings. Thus further illustrates the advantages of the CNN-LSTM-ATT model used in this paper in enterprise credit assessment.

4.2. Data Grouping for Comparison of Four Models. Before grouping test, we choose the Optimizer = SGD, Learning Rate = 0.0001, Dropout = 0.3, Optimizer = SGD, Learning Rate = 0.005, and Dropout = 0.3. Then, we randomly divide the original data sets into three experimental sets Group 0, Group 1, and Group 2, respectively. With three groups of data sets, we carry out AUC evaluation and comparison for four models CNN, LSTM, CNN-LSTM, and CNN-LSTM-ATT.

Table 4 lists the average performance of the four different models under the three experimental data sets. In each data set, the proposed CNN-LSTM-ATT model is superior to the other three models in most cases. The results show that CNN-LSTM-ATT has a good adaptive ability to the uncertainty of enterprise behavior. Compared with the original CNN-LSTM model, the attention-based model has higher accuracy and robustness. The attention module improves the AUC of the CNN-LSTM model. In most cases, the performance of proposed CNN-LSTM-ATT is better than CNN, LSTM, and CNN-LSTM models, which also reflects the advantages of the proposed space-time attention.

AUC predicted by the model in the experiment is shown in Figure 6. In the original data set, the CNN-LSTM-ATT model proposed in this paper has the best accuracy in all three experimental sets. In the data sets of Group-0 and Group-2, the CNN-LSTM-ATT model has the best accuracy in most moments compared with the other three models. At the same time, the performance of the CNN-LSTM model is also better than that of a single CNN or LSTM model. In the Group-1 data set, the performance of the CNN-LSTM model is slightly lower than that of the CNN model, while the CNN-LSTM-ATT model remains optimal, which also indicates that the CNN-LSTM-ATT model in this paper has good robustness and accuracy. The data set has a great impact on the evaluation of the model. The training of the model with data set Group-2 has a better performance, while the average level of the training of the model with Group-1 data set is poor.

5. Conclusion and Future Work

In this paper, the CNN-LSTM-A TT neural network is proposed to solve the problem of enterprise's credit risk evaluation. Super-parameter selection experiment and comparison experiment of different models are conducted on the realistic data set. The validity and robustness of the CNN-LSTM-ATT model in solving the problem of credit risk evaluation are comprehensively analyzed. Super-parameter selection experiments are carried out. It found out that the classification result would be the best and the AUC value turns to 0.9 when the Optimizer = SGD and Learning Rate = 0.0001. After determining the optimizer and the learning rate, dropout selection experiment is put forward. The experiment found that the classification result would be best under Optimizer = SGD and Learning Rate = 0.0001 and the AUC value turns to 0.92 when the Dropout = 0.3. Finally, we determine the value of three super parameters, Optimizer = SGD, Learning Rate = 0.0001, and Dropout = 0.3.

In order to further evaluate the effect of the CNN-LSTM-ATT model on credit risk evaluation, we conduct a comparative analysis experiment of four different models. CNN, LSTM, and CNN-LSTM are chosen as the objects of comparison. First, a comparison experiment is carried out to compare and analyze the evaluation results of the four models by different dropout rate and learning rate. The experiment shows that the CNN-LSTM-A TT model still has the optimal performance effect among the four models under different learning rate and the CNN-LSTM-ATT model also maintains the best result under different dropout. Then, the data grouping for comparison of four models is presented. As a result, four kinds of models are all performance best under the Optimizer = SG, Dropout = 0.3, and CNN and LSTM model are relatively better under Learning Rate = 0.005. In order to guarantee the objectivity, we select two different sets of parameters respectively to ensure that the evaluation of each model is close to the best result. The classification results of four kinds of the model are tested under three separate data sets. The experimental results show that the CNN-LSTM-ATT model has the best accuracy in most moments compared with the other three models on three different data sets.

The proposed LSTM in this paper has solved the problem that RNN appears seriously gradient disappears under the long training sequence. CNN-LSTM-ATT solves the problem that traditional method LR and RF have low accuracy and the shortage of robustness when the feature space is very large. The prediction accuracy is also higher than that of the original CNN and the original LSTM, which also indicates that the CNN-LSTM-ATT model in this paper has better robustness and accuracy.

The future research work will focus on two aspects: one is to consider the impact of abrupt factors on the credit strategy of micro, small, and medium-sized enterprises and build a credit evaluation model; the second is to study the impact of affiliated enterprises on credit risk rating and construct a credit evaluation model.

Data Availability

The data used to support the findings of this study are available from the corresponding author.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grants (62276034), Group Building Scientific Innovation Project for Universities in Chongqing (CXQT21021), and Joint Training Base Construction Project for Graduate Students in Chongqing (JDLHPYJD2021016).

References

[1] J. Nalić, G. Martinović, and D. Žagar, "New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers," *Advanced Engineering Informatics*, vol. 45, Article ID 101130, 2020.

[2] Y. Xia, J. Zhao, L. He, Y. Li, and M. Niu, "A novel tree-based dynamic heterogeneous ensemble method for credit scoring," *Expert Systems with Applications*, vol. 159, Article ID 113615, 2020.

[3] C. Xu, J. Tao, and P. Xu, "Credit decision system of small and medium sized micro enterprises based on big data technology and risk assessment thinking," in *Proceedings of the 2021 6th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 477–480, Kunming, China, May 2021.

[4] L. Zhang and Q. Song, "Multimodel integrated enterprise credit evaluation method based on attention mechanism," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8612759, 12 pages, 2022.

[5] P. Z. Lappas and A. N. Yannacopoulos, "A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment," *Applied Soft Computing*, vol. 107, Article ID 107391, 2021.

[6] M. Li, "Research on the construction of Fintech information sharing system and credit evaluation method based on internet platform," *IEEE*, in *Proceedings of the 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pp. 280–283, Dalian, China, April 2022.

[7] L. Zhang and Q. Song, "Credit evaluation of SMEs based on GBDT-CNN-LR hybrid integrated model," *Wireless Communications and Mobile Computing*, Article ID 5251228, 8 pages, 2022.

[8] Z. Li, L. Liu, L. Zhu, F. Deng, Y. Zhang, and Y. Zhang, "Parallel double-layer prediction model construction and empirical analysis for enterprise credit assessment," *Intelligent Data Analysis*, vol. 26, no. 4, pp. 1007–1022, 2022.

[9] T. Meng, Q. Li, Z. Dong, and F. Zhao, "Research on the risk of social stability of enterprise credit supervision mechanism based on big data," *Journal of Organizational and End User Computing*, vol. 34, no. 3, pp. 1–16, 2022.

[10] F. d L. Weber, V. A. d M. Weber, G. V. Menezes et al., "Recognition of Pantaneira cattle breed using computer vision and convolutional neural networks," *Computers and Electronics in Agriculture*, vol. 175, Article ID 105548, 2020.

[11] A. M. Gafurov and O. P. Yermolayev, "Automatic gully detection: neural networks and computer vision," *Remote Sensing*, vol. 12, no. 11, p. 1743, 2020.

[12] V. Passricha and R. K. Aggarwal, "PSO-based optimized CNN for Hindi ASR," *International Journal of Speech Technology*, vol. 22, no. 4, pp. 1123–1133, 2019.

[13] V. Passricha and R. K. Aggarwal, "A hybrid of deep CNN and bidirectional LSTM for automatic speech recognition," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 1261–1274, 2019.

[14] S. Debnath and P. Roy, "Appearance and shape-based hybrid visual feature extraction: toward audio–visual automatic speech recognition," *Signal, Image and Video Processing*, vol. 15, no. 1, pp. 25–32, 2021.

[15] A. Taherkhani, G. Cosma, and T. M. McGinnity, "AdaBoost-CNN: an adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning," *Neurocomputing*, vol. 404, pp. 351–366, 2020.

[16] K. Uthra Devi and R. Gomathi, "Retracted article: brain tumour classification using saliency driven nonlinear diffusion and deep learning with convolutional neural networks (CNN)," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6263–6273, 2021.

[17] R. K. Pathan, M. Biswas, and M. U. Khandaker, "Time series prediction of COVID-19 by mutation rate analysis using recurrent neural network-based LSTM model," *Chaos, Solitons & Fractals*, vol. 138, Article ID 110018, 2020.

[18] M. Sangiorgio and F. Dercole, "Robustness of LSTM neural networks for multi-step forecasting of chaotic time series," *Chaos, Solitons & Fractals*, vol. 139, Article ID 110045, 2020.

[19] S. Das, D. Das, and A. K. Kolya, "Sentiment classification with GST tweet data on LSTM based on polarity-popularity model," *Sādhanā*, vol. 45, no. 1, pp. 140–217, 2020.

[20] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, Article ID 102656, 2021.

[21] Y. Zhang, J. Zheng, Y. Jiang, G. Huang, and R. Chen, "A text sentiment classification modeling method based on coordinated CNN-LSTM-Attention model," *Chinese Journal of Electronics*, vol. 28, no. 1, pp. 120–126, 2019.

[22] I. Ameer, G. Sidorov, H. Gomez-Adorno, and R. MA. Nawab, "Multi-label emotion classification on code-mixed text: data and methods," *IEEE Access*, vol. 10, pp. 8779–8789, 2022.

[23] N. Cerkez, B. Vrdoljak, and S. Skansi, "A method for MBTI classification based on impact of class components," *IEEE Access*, vol. 9, pp. 146550–146567, 2021.

[24] A. H. Ombabi, W. Ouarda, and A. M. Alimi, "Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks," *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 53–13, 2020.

[25] X. Lili, L. Feng, and C. Xuejian, "The credit analysis of transportation capacity supply chain finance based on core enterprise credit radiation," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 6, pp. 10811–10814, 2021.

[26] H. U. R. Khalid, A. Gorji, A. Bourdoux, S. Pollin, and H. Sahli, "Multi-View CNN-LSTM architecture for radar-based human activity recognition," *IEEE Access*, vol. 10, pp. 24509–24519, 2022.

[27] C. C. Wong, M. Y. Chien, R. J. Chen, H. Aoyama, and K. Y. Wong, "Moving object prediction and grasping system of robot manipulator," *IEEE Access*, vol. 10, pp. 20159–20172, 2022.

[28] A. A. Abdelhamid, E. S. M. El-Kenawy, B. Alotaibi et al., "Robust speech emotion recognition using CNN+LSTM

based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022.

[29] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: sentiment analysis using lexicon integrated two-channel CNN–LSTM family models," *Applied Soft Computing*, vol. 94, Article ID 106435, 2020.

[30] W. Cao, Z. Zhang, C. Liu et al., "Unsupervised discriminative feature learning via finding a clustering-friendly embedding space," *Pattern Recognition*, vol. 129, Article ID 108768, 2022.

[31] L. Zhang, J. He, and Z. Zhao, "Enterprise credit security prediction and evaluation based on multimodel fusion," *Security and Communication Networks*, vol. 2022, Article ID 2754302, 12 pages, 2022.

[32] X. Zhu, L. Li, J. Liu, Z. Li, H. Peng, and X. Niu, "Image captioning with triple-attention and stack parallel LSTM," *Neurocomputing*, vol. 319, pp. 55–65, 2018.

[33] P. Chen, W. Dong, J. Wang, X. Lu, U. Kaymak, and Z. Huang, "Interpretable clinical prediction via attention-based neural network," *BMC Medical Informatics and Decision Making*, vol. 20, no. S3, pp. 131–139, 2020.

[34] R. Liu, X. Yang, X. Dong, and B. Sun, "Credit risk assessment of banks' loan enterprise customer based on state-constraint," *Computing and Informatics*, vol. 40, no. 1, pp. 145–168, 2021.

[35] P. K. Gupta and K. K. Jain, "Credit default prediction for micro-enterprise financing in India using ensemble models," *Global Business and Economics Review*, vol. 26, no. 1, pp. 84–98, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8607760, 7 pages
https://doi.org/10.1155/2022/8607760

Research Article

Effect of Quality Control Circle Activity Nursing Combined with Respiratory Function Exercise Nursing on Patients with Esophageal Cancer

Hairu Yu,[1] Sha Li,[2] and Shasha Shi [1]

[1]*Department of Thoracic Cancer, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China*
[2]*Department of Gastroenterology, Wuhan Third Hospital, No. 241 Shouyi Road, Wuchang, Wuhan, China*

Correspondence should be addressed to Shasha Shi; 631507010122@mails.cqjtu.edu.cn

Received 16 June 2022; Revised 18 July 2022; Accepted 9 August 2022; Published 17 September 2022

Academic Editor: Wenming Cao

Objective. To explore the effect of quality control circle activity nursing combined with respiratory function exercise nursing on esophageal cancer patients' immune function and nutritional status. *Methods*. The clinical case data of 119 esophageal cancer patients admitted to our hospital were selected as the research objects from May 2019 to July 2021. They were divided into the quality control circle activity care group (QCCAC) (9 cases dropped due to incomplete case data, n= 50) and respiratory function exercise care group (10 cases dropped due to incomplete case data, n=50) by the random number table method, the respiratory function exercise care group was treated with respiratory function exercise care, while the QCCAC group was treated with QCCAC. Changes in lung function, immune function, and nutritional status before and after nursing were compared in the two groups of patients. *Results*. Before nursing, there was no significant difference in pulmonary function indexes, immune function indexes, and the level of nutritional status indicators between the two groups ($P > 0.05$). After nursing, the finger pulse oxygen saturation, vital capacity (VC), respiratory rate, forced vital capacity (FVC), deep inspiratory volume (IC), and maximum ventilation (MVV) of the QCCAC improved, and the QCCAC group was significantly higher than the respiratory function exercise care group ($P < 0.05$). After nursing, the QCCAC's CD3+, CD4+, IgG, IgM, and IgA levels increased, and the QCCAC group was significantly higher than the respiratory function exercise care group. The CD8+ level decreased, and the QCCAC was lower than the respiratory function exercise care group ($P < 0.05$). After continuous nursing, the QCCAC's D-lactic acid, DAO, FFMI, Hb, ALB, PA, and other nutritional evaluation indexes all increased, and the QCCAC was significantly higher than the respiratory function exercise care group ($P < 0.05$). *Conclusion*. Quality control circle activity nursing combined with respiratory function exercise nursing can effectively improve the immune function, respiratory function, and nutritional status of esophageal cancer patients.

1. Introduction

Esophageal cancer is one of the most common malignant tumors in China and is the malignant tumor with the fourth highest fatality rate [1]. Esophageal cancer belongs to the body's esophageal epithelial malignant lesions, accounting for 2% of the incidence of all malignant tumors, and the number of patients who die of esophageal cancer worldwide each year reaches about 200,000, and China has a high incidence of esophageal cancer [2]. At present, the main treatment for esophageal cancer is surgery, supplemented by chemotherapy. Cancer itself has a huge psychological impact on patients, coupled with the symptoms caused by tumors, adverse reactions caused by surgery, radiotherapy, and chemotherapy, which bring a huge burden to patients [3]. Patients are often accompanied by negative emotions such as anxiety and depression, which affect the occurrence and development of tumors and adversely affect the quality of life [4]. With the transformation of the medical model to the biological-psychological-social model, paying attention to

the psychological function of cancer patients and adjusting the patients' negative emotions have become an important aspect of cancer treatment and care [5]. In recent years, studies have pointed out that carrying out quality control circle activities in the nursing field can further improve the effectiveness of nursing, improve the quality of nursing services, and encourage patients to cooperate with clinical medical work [6]. Quality control circle activities are used in the rapid rehabilitation of surgical care patients after surgery. The corresponding rehabilitation exercise recovery plan is reasonably formulated based on the patient's condition. The patient is guided and supervised quantitatively and targeted to improve patient compliance and exercise science, thereby improving the prognosis [7]. Cognitive behavioural therapy is a short-term psychotherapy that can eliminate bad emotions and behaviours by changing negative cognition and thoughts or beliefs and behaviours [8]. This therapy is widely used to treat anxiety, depression, insomnia, and other mental and psychological diseases [9]. In recent years, cognitive behavioural therapy has been applied to patients with malignant tumors and has achieved certain results. This study included 119 esophageal cancer patients admitted to our hospital in the past two years as the observation object, aiming to explore the effect of quality control circle activity nursing combined with respiratory function exercise nursing on the respiratory function exercise of esophageal cancer patients to provide esophageal cancer patients new nursing ideas. The report is as follows.

2. Materials and Methods

2.1. General Information. From May 2019 to July 2021, the clinical case data of 119 esophageal cancer patients admitted to our hospital were selected as the research objects. They were divided into the quality control circle activity care group (QCCAC) and respiratory function exercise care group by the random number table method. Nine patients were excluded from the study due to incomplete clinical data, and 50 were finally included. In the respiratory function exercise care group, 10 cases were dropped due to incomplete case data, and 50 cases were finally included. The two groups of patients had no significant difference in general information such as age, gender, body mass index (BMI), type of esophageal cancer, disease course, whether there is a nutritional risk, prognosis, survival time, etc. ($P > 0.05$), and they are comparable, as shown in Table 1. During the 6-month follow-up period, 3 cases died in the QCCAC, 6 cases were lost to follow-up, and 50 cases were finally completed; 3 cases died in the respiratory function exercise care group, 7 cases were lost to follow-up, and 50 cases were finally completed. Inclusion criteria were as follows: (1) all patients underwent chest computed tomography (CT), esophageal barium meal or electronic gastroscopy before treatment, and all had clear pathological diagnoses [10]; (2) newly treated primary esophageal cancer and received radical radiotherapy for the first time; (3) KPS score>70 with the independent reading ability or completed the required questionnaire with the help of the investigator and the patient's age is <75 years old; (4) patients and their families agreed to participate in this study and signed informed consent forms; and (5) patients are eligible for post-follow-up investigation. Exclusion criteria were as follows: (1) patients with a history of other malignant tumors and patients whose surgical plan was changed; (2) patients with mental illness who participated in other similar research during the hospitalization; (3) patients with severe gastrointestinal dysfunction and incomplete clinical data; (4) those who cannot tolerate surgical treatment; (5) patients with cardiac, hepatic, and renal insufficiency; and (6) people with other diseases.

2.2. Method

2.2.1. Respiratory Function Exercise Care Group. The respiratory function exercise care group used respiratory function exercise care and abdominal breathing training: relaxing all auxiliary respiratory muscles, sinking the abdomen when exhaling, and placing a sandbag on the upper abdomen to further increase abdominal pressure. When exhaling, the upper abdomen resists the pressure of the sandbag and slowly bulges the abdomen. When exhaling, the abdomen slowly sinks, and it is repeated many times. Starting from each exercise for 5 minutes, the time is gradually extended until you can breathe steadily for 2 hours with the sand bag placed. The breathing rate changes from fast to slow, eventually reaching about 9 breaths per minute. The sandbag increases from light to heavy, from 2 kg to 5 kg. Generally, patients exercise for 1 or 2 weeks before surgery, while those who are elderly have a poor respiratory function and have a larger scope of surgery exercise for 3 or 4 weeks before surgery and can be extended appropriately. Effective cough training: instruct the patient to take a deep breath and hold their breath, close the glottis, raise the diaphragm to increase the intrapleural pressure, contract the intercostal muscles, then cough, and open the glottis to flush out the gas or sputum. The patient repeats the exercises until they master it.

2.2.2. QCCAC. The implements of QCCAC, (1) establishes a quality control circle activity group. The team members include the head nurse, the head nurse, and the head nurse in charge of the team. The ward nurses are divided into 4 groups, the head nurse manages each group, and the head nurse assigns the detailed work flow. After the head nurse assigns the work, the head nurse assigns the work to the group members and provides professional guidance. The task is arranged from top to bottom, and effective feedback is obtained during the work (recording the problems in work) and then reported from the bottom to the top. Discuss and solve in a unified way. (2) Develop a reasonable solution plan. After the activity time is determined, nurses are arranged in batches to learn the knowledge of the quality control circle. Based on the "Chinese Hospital Quality Control Circle Operation Manual," according to the theme of this activity, patients with esophageal cancer after surgery are guided to perform respiratory function exercises properly and formulate regular and quantitative breathing.

Table 1: Comparison of general information between QCCAC and respiratory function exercise care group (±*s*).

Group	Age (years)	Gender Male/female	BMI	Course of disease (years)
QCCAC ($n = 50$)	58.35 ± 9.43	28/22	23.26 ± 3.88	6.54 ± 1.43
Respiratory function exercise care group ($n = 50$)	59.41 ± 10.65	26/24	22.79 ± 4.12	6.91 ± 1.15
t/χ^2	0.527	0.161	0.587	−1.426
P	0.599	0.688	0.558	0.157

The specific content of the exercise is as follows: 6 to 10 hours after anesthesia: maintain a semirecumbent position on the bed and move your legs twice for 5 minutes each time, after completion of aerosol inhalation, cough and expectorant, take 2 deep breaths for 5 minutes each time. Day 1 after surgery: Move your legs and legs 4 times on the bed for 5 minutes, continue to sit on the bed, cough and cough up sputum after nebulized inhalation, breathe deeply 6 times for 5 minutes each time, stand on the edge of the bed and move 2 times, 100 meters each time. The day after the operation to the discharge: move your legs 6 times on the bed for 5 minutes, sit on the bed all the time, after the aerosol inhalation, cough and sputum production are completed, take 6 deep breaths for 10 minutes each time, stand at the edge of the bed and move 6 times, each time moving 200~-500m. The activity distance gradually increases with time and status. At the same time, improve the nursing staff's professional knowledge of breathing exercise training and guidance ability. After the patient is admitted to the hospital, patiently, calmly, and orderly give the patient breathing exercise guidance to make it clear the importance of respiratory function for rapid postoperative recovery and help them master correct breathing techniques to improve the correct rate of breathing exercises and reduce or avoid pulmonary complications. (3) Nutritional care: patients mainly eat porridge, vegetable soup, and other foods during radiotherapy. They eat sugar-free porridge for breakfast and dinner and vegetable soup (greens, carrots, seaweed, etc.) in the morning, noon, and afternoon. Instruct patients to eat at least 5 times a day, add 1000 ml of water before meals, add 1000 ml between two meals, and drink ≥5000 ml of water daily to prevent vomiting during chemotherapy.

2.3. Observation Indicators. (1) Pulmonary function monitoring: using the Master ScreenDiffusion machine from JAEGER, Germany, professional physicians will perform pulmonary function tests on the first day of hospitalization and 1 day before surgery and observe and record before exercise (the day of admission) and after exercise (1 day before surgery). Changes in lung function indicators: observation items include vital capacity (VC), forced vital capacity (FVC), deep inspiration (IC), and maximum ventilation (MVV). Finger pulse oxygen saturation was measured by clamping the end of the patient's finger with the finger pulse oxygen saturation monitor of NONINMEDICAL. (2) Immune function: before and after 6 months of nursing, peripheral venous blood samples of patients were collected, and use FACS calibur flow cytometer and supporting reagents were to detect and analyze CD3+, CD4+, CD8+ and use flow cytometer to detect and analyze [11]. All operations are carried out in strict accordance with the instrument manual. The immunoglobulin IgG, IgM, and IgA content were detected by enzyme-linked immunosorbent assay, and all operations were performed per the kit instructions [12]. (3) Nutritional status: measure D-lactic acid levels, diamine oxidase (DAO) levels, fat-free body mass index (FFMI), hemoglobin (Hb), albumin (ALB), prealbumin (PA), and other individual nutritional status evaluation indicators [13].

2.4. Statistical Methods. SPSS25.00 statistical software was used to analyze the data. Measurement data were expressed as mean ± standard deviation (±*s*), and a *t*-test was performed between groups; a rank sum test was performed for non-normal distribution, and count data were expressed as a percentage (%). χ^2 test was performed, and $P < 0.05$ indicated that the difference was statistically significant.

3. Results

3.1. Comparison of Lung Function Indicators between the Two Groups. Before nursing, there was no significant difference in pulmonary function indexes between the two groups of patients ($P > 0.05$). After nursing, the QCCAC's finger pulse oxygen saturation, vital capacity (VC), respiratory rate, forced vital capacity (FVC), deep inspiratory volume (IC), and maximum ventilation (MVV) were improved, and the QCCAC was better than the respiratory function exercise care group. The difference was statistically significant ($P < 0.05$) (see Table 2).

3.2. Comparison of the Levels of Immune Function Indicators. Before nursing, there was no statistical difference in the immune function indexes between the two groups ($P > 0.05$). After nursing, the QCCAC's CD3+, CD4+ and IgG, IgM, and IgA levels increased. The QCCAC group was higher than the respiratory function exercise care group. The CD8+ level decreased, and the QCCAC group was higher than the respiratory function exercise care group; the above differences were statistically significant ($P < 0.05$) (see Tables 3 and 4).

3.3. Comparison of Nutritional Status Indicators. Before nursing, there was no statistically significant difference in the level of nutritional status indicators between the two groups ($P > 0.05$). After nursing, the QCCAC's D-lactic acid, DAO, FFMI, Hb, ALB, PA, and other nutritional evaluation

Table 2: Comparison of lung function indexes (±*s*).

Group	Finger pulse oxygen saturation (%)		VC (L)		Respiratory rate (%)	
	Before care	After care	Before care	After care	Before care	After care
QCCAC ($n = 50$)	96.23 ± 10.23	98.77 ± 2.64	3.16 ± 0.81	3.76 ± 0.97	19.82 ± 5.04	14.47 ± 0.43
Respiratory function exercise care group ($n = 50$)	96.46 ± 10.46	96.81 ± 2.19***	3.14 ± 0.31	3.26 ± 0.25***	19.44 ± 4.91	16.09 ± 0.48***
t	0.111	4.040	0.163	3.530	0.382	−17.775
P	0.912	<0.001	0.871	<0.001	0.703	<0.001
Group	*FVC (L)*		*IC (L)*		*MVV (L)*	
	Before care	*After care*	*Before care*	*After care*	*Before care*	*After care*
QCCAC($n = 50$)	2.33 ± 0.23	2.97 ± 0.64	2.26 ± 0.51	2.76 ± 0.77	81.51 ± 3.42	87.23 ± 8.57
Respiratory function exercise care group($n = 50$)	2.36 ± 0.46	2.41 ± 0.19***	2.14 ± 0.41	2.46 ± 0.25**	81.57 ± 3.81	83.06 ± 7.63*
t	0.412	5.931	1.297	2.620	-0.082	2.570
P	0.681	<0.001	0.198	0.010	0.935	0.012

Note. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$ vs. before care.

Table 3: Comparison of CD3+, CD4+, and CD8+ levels (±*s*).

Group	$CD4^+$ (%)		$CD8^+$ (%)		$CD3^+$ (%)	
	Before care	Aftercare	Before care	Aftercare	Before care	Aftercare
QCCAC ($n = 50$)	34.63 ± 7.83	38.87 ± 8.61*	29.34 ± 2.95	27.01 ± 3.42*	61.23 ± 8.57	66.11 ± 8.59*
Respiratory function exercise care group ($n = 50$)	32.76 ± 7.51	35.41 ± 7.31*	30.26 ± 3.64	28.57 ± 3.81*	59.06 ± 7.63	62.53 ± 8.36
t	1.188	2.109	1.355	2.099	1.303	2.048
P	0.238	0.038	0.179	0.039	0.196	0.043

Note. *$P < 0.05$ vs. before care.

Table 4: Comparison of IgG, IgM, and IgA levels (±*s*).

Group	IgG (g/L)		IgM (g/L)		IgA (g/L)	
	Before care	Aftercare	Before care	Aftercare	Before care	Aftercare
QCCAC ($n = 50$)	12.95 ± 4.35	15.82 ± 5.04*	1.17 ± 0.43	1.44 ± 0.46*	2.08 ± 0.12	2.61 ± 0.19
Respiratory function exercise care group ($n = 50$)	13.27 ± 4.46	13.44 ± 4.91	1.09 ± 0.48	1.26 ± 0.41*	2.11 ± 0.14	2.49 ± 0.15***
t	0.354	2.331	1.178	2.012	1.122	3.412
P	0.724	0.022	0.242	0.047	0.265	<0.001

Note. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$ vs. before care.

Table 5: Comparison of D-lactic acid, DAO, and FFMI levels before and after surgery (±*s*).

Group	D-Lactic acid (mg/L)		DAO (U/mL)		FFMI (kg/m^2)	
	Before care	Aftercare	Before care	After care	Before care	Aftercare
QCCAC ($n = 50$)	11.89 ± 6.38	13.62 ± 7.13*	2.76 ± 4.45	4.97 ± 4.8*	15.22 ± 3.21	18.82 ± 3.74**
Respiratory function exercise care group ($n = 50$)	10.21 ± 5.41	16.57 ± 7.22*	2.32 ± 3.96	5.69 ± 4.78*	14.72 ± 2.83	16.71 ± 3.66**
t	1.383	-2.003	0.647	-2.310	0.712	2.779
P	0.170	0.048	0.519	0.023	0.478	0.007

*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$ vs. before care.

Table 6: Comparison of Hb, ALB, and PA levels before and after surgery ($\pm s$).

Group	Hb (g/L)		ALB (g/L)		PA (mg/L)	
	Before care	After care	Before care	After care	Before care	After care
QCCAC ($n = 50$)	116.23 ± 31.23	$148.77 \pm 33.64^{*}$	35.26 ± 7.81	$43.76 \pm 7.97^{**}$	278.15 ± 79.41	$347.46 \pm 83.64^{*}$
Respiratory function exercise care group ($n = 50$)	119.46 ± 29.46	$131.81 \pm 30.19^{*}$	34.94 ± 8.01	$39.26 \pm 8.25^{**}$	277.36 ± 77.63	$309.83 \pm 79.61^{*}$
t	0.518	2.117	0.197	2.704	0.049	2.245
P	0.606	0.012	0.844	0.008	0.961	0.027

Note. $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$ vs. before care.

indexes all increased. The QCCAC groupwas higher than the respiratory function exercise care group; the difference was statistically significant ($P < 0.05$) (see Tables 5 and 6).

4. Discussion

With the transformation of medical models, clinicians have realized that the purpose of modern medicine is not only to treat patients' diseases and improve patients' organ functions but also to pay attention to patients' mental state and improve patients' quality of life [14]. Esophageal cancer is a common clinical malignant tumor. There is currently no specific treatment plan. Only treatment can be used to control the patient's condition, remove the patient's tumor lesions, improve the patient's postoperative survival rate, and improve the quality of life [15]. However, after research and analysis of literature, it is found that esophageal cancer patients' postoperative quality of life is significantly lower than that of ordinary people [16]. Among esophageal cancer patients, the elderly are the majority. The elderly are not well-educated, have weak knowledge structures, and have varying doubts and resistance to clinical treatment. Some patients do not cooperate with clinical treatment due to economic conditions, which makes the overall treatment effect unsatisfactory. The prognosis is poor [17]. Most patients with esophageal cancer have difficulty getting out of bed and coughing after esophageal cancer surgery. Most patients with esophageal cancer still need to return to the hospital for treatment after discharge. The model of nursing care ensures that medical resources are effectively and reasonably used to strengthen nursing management, improve patient compliance with treatment, and encourage them to exercise respiratory function, thereby improving the prognosis [18].

After nursing, the finger pulse oxygen saturation, vital capacity (VC), respiratory rate, forced vital capacity (FVC), deep inspiratory volume (IC), and maximum ventilation (MVV) of the QCCAC improved and the QCCAC was better than the respiratory function exercise care group, indicating that quality control circle activity nursing combined with respiratory function exercise nursing can improve the respiratory function of patients with esophageal cancer. The quality control circle activity was first applied in corporate management. The main purpose of the activity is to encourage all employees to participate in the activity, improve their collective sense of honor, and encourage employees to maintain and improve the products produced by the company, to improve their work efficiency and quality. However, because the quality control circle activity is a very widely adaptable activity, more and more managers realize the high efficiency of its application, so the quality control circle activity covers more and more fields [19]. In recent years, studies have pointed out that carrying out quality control circle activities in the nursing field can further improve the effectiveness of nursing, improve the quality of nursing services, and encourage patients to cooperate with clinical medical work [20]. Quality control circle activities are used in the rapid rehabilitation of surgical care after surgery. The corresponding rehabilitation exercise recovery plan is reasonably formulated based on the patient's condition. The patients are guided and supervised quantitatively and targeted to improve patient compliance and exercise science, thereby improving prognosis [21]. Studies have pointed out that the application of quality control circle activities in the rapid recovery surgical care of patients after esophageal cancer can effectively improve the compliance and the accuracy of breathing exercises, effectively enhance the status of patient's lung function, and significantly shorten the discharge time [22].

After nursing, the CD3+, CD4+ and IgG, IgM, and IgA levels of QCCAC were increased, and the QCCAC group was higher than the respiratory function exercise care group. The CD8+ level decreased, and the QCCAC group was lower than the respiratory function exercise care group. Quality control circle activity nursing combined with respiratory function exercise nursing can improve the immune function of esophageal cancer patients effectively. The immune function of tumor patients is often suppressed. Malnutrition can lead to low immune function. Surgical stress will aggravate the immunosuppressive state, which is manifested as a decrease in the circulating helper lymphocyte subsets CD4+ and a relative increase in the inhibitory lymphocyte subsets CD8+ and CD4+/CD8. The proportion decreases, and the activity of NK cells decreases [23].

After nursing in this study, the QCCAC's D-lactic acid, DAO, FFMI, Hb, ALB, PA, and other nutritional evaluation indexes all increased, and the QCCAC group was higher than the respiratory function exercise care group. Quality control circle activity nursing combined with respiratory function exercise nursing can improve patients' nutritional status. Bacteria produce d-lactic acid in the gastrointestinal tract, and mammals have no fast-metabolizing enzyme system. The detection of blood D-lactic acid levels can reflect changes in intestinal mucosal permeability in time [24]. DAO in the small intestinal mucosa's upper villi can reflect the small intestine's structure and function. DAO activity in

peripheral blood is stable, so its changes in peripheral blood can be measured non-invasively to reflect the state of intestinal mucosa [25].

In summary, the quality control circle activity nursing combined with respiratory function exercise nursing can effectively improve the immune function, respiratory function, and nutritional status of patients with esophageal cancer. Clinical application is recommended. In addition, since this study only observed the results of nursing care without long-term evaluation and the sample size is small, clinical applications should also be considered.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Disclosure

Hairu Yu and Sha Li are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. Y. Dai, Y. C. Yu, Y. S. Leu et al., "Neoadjuvant chemoradiotherapy and larynx-preserving surgery for cervical esophageal cancer," *Journal of Clinical Medicine*, vol. 9, no. 2, p. 387, 2020.

[2] D. J. Uhlenhopp, E. O. Then, T. Sunkara, and V. Gaduputi, "Epidemiology of esophageal cancer: update in global trends, etiology and risk factors," *Clin J Gastroenterol*, vol. 13, no. 6, pp. 1010–1021, 2020.

[3] Q. Zhang, A. Qian, and W. Chen, "Analysis of PICC based on dysfunction module personalized nursing treatment in chemotherapy of advanced esophageal cancer," *J Healthc Eng*, vol. 2021, Article ID 8997915, 10 pages, 2021.

[4] Z. H. Huang, X. W. Ma, J. Zhang, X. Li, N. L. Lai, and S. X. Zhang, "Cetuximab for esophageal cancer: an updated meta-analysis of randomized controlled trials," *BMC Cancer*, vol. 18, no. 1, p. 1170, 2018.

[5] W. K. Chen, C. A. Chen, C. W. Chi et al., "Moscatilin inhibits growth of human esophageal cancer xenograft and sensitizes cancer cells to radiotherapy," *Journal of Clinical Medicine*, vol. 8, no. 2, p. 187, 2019.

[6] Y. Y. Janjigian, S. B. Maron, W. K. Chatila et al., "First-line pembrolizumab and trastuzumab in HER2-positive oesophageal, gastric, or gastro-oesophageal junction cancer: an open-label, single-arm, phase 2 trial," *The Lancet Oncology*, vol. 21, no. 6, pp. 821–831, 2020.

[7] D. Zhang, M. Liao, and T. Liu, "Implementation and promotion of quality control circle: a starter for quality improvement in Chinese hospitals," *Risk Management and Healthcare Policy*, vol. 13, pp. 1215–1224, 2020.

[8] Y. Q. Ju, X. M. Xu, X. Y. Han, J. L. Yao, and H. Y. Tang, "Evaluation of the effect of quality control circle activities on improving the nursing quality of patients with periodontitis," *Shang Hai Kou Qiang Yi Xue*, vol. 28, no. 3, pp. 325–329, 2019.

[9] K. Liu, Z. Xu, X. Wang, Y. Chen, and X. D. Mao, "The application of quality control circle to improve the quality of samples: a SQUIRE-compliant quality-improving study," *Medicine (Baltimore)*, vol. 99, no. 21, Article ID e20333, 2020.

[10] D. Ma, Y. Fan, and Z. Liao, "Expert consensus on screening and endoscopic diagnosis and treatment of early esophageal cancer in China (2014, Beijing)," *Gastroenterology*, vol. 20, no. 4, p. 220, 2015.

[11] J. P. Morgado, C. P. Monteiro, J. Teles et al., "Immune cell changes in response to a swimming training session during a 24 h recovery period," *Applied Physiology Nutrition and Metabolism*, vol. 41, no. 5, pp. 476–483, 2016.

[12] M. A. M. van Delft, M. K. Verheul, L. E. Burgers et al., "The isotype and IgG subclass distribution of anti-carbamylated protein antibodies in rheumatoid arthritis patients," *Arthritis Research and Therapy*, vol. 19, no. 1, pp. 190–202, 2017.

[13] H. Yin, Y. Liu, and W. Zhu, "NRS 2002 screening COPD patients with nutritional risk and its correlation with inspection indicators," *Parenteral and Enteral Nutrition*, vol. 24, no. 5, pp. 301–305, 2017.

[14] Y. Ichijo, Y. Takeda, Y. Oguma, Y. Kitagawa, H. Takeuchi, and A. Z. Doorenbos, "Physical activity among postoperative esophageal cancer patients," *Cancer Nursing*, vol. 42, no. 6, pp. 501–508, 2019.

[15] L. Chen and W. Wang, "Application effect of continuous nursing in patients with advanced esophageal cancer after esophageal stent implantation," *The American Journal of Translational Research*, vol. 13, no. 8, pp. 9743–9749, 2021.

[16] S. Wang, J. Liu, S. Qiu, and J. Yu, "Facile fabrication of Cu9-S5 loaded core-shell nanoparticles for near infrared radiation mediated tumor therapeutic strategy in human esophageal squamous carcinoma cells nursing care of esophageal cancer patients," *Journal of Photochemistry and Photobiology B: Biology*, vol. 199, Article ID 111583, 2019.

[17] G. Cullen, "Esophageal cancer: a case study of a rare metastatic esophageal squamous cell carcinoma," *Clinical Journal of Oncology Nursing*, vol. 24, no. 4, pp. 352–355, 2020.

[18] X. Y. Liu, C. H. Jiao, D. Zhao, Y. Chen, and H. M. Zhang, "Psychological impact of high-quality nursing care on patients with esophageal cancer during perioperative period: a protocol of systematic review," *Medicine (Baltimore)*, vol. 99, no. 43, Article ID e22270, 2020.

[19] D. Zhang, M. Liao, Y. Zhou, and T. Liu, "Quality control circle: a tool for enhancing perceptions of patient safety culture among hospital staff in Chinese hospitals," *International Journal for Quality in Health Care*, vol. 32, no. 1, pp. 64–70, 2020.

[20] Q. Jiang, D. Zhang, J. Majaw et al., "Minimization of the perianal infection rate of hematological malignancies with agranulocytosis by quality control circle activity and patient-hospital-student win-win concept," *Journal of International Medical Research*, vol. 46, no. 6, pp. 2338–2345, 2018.

[21] C. Plüss-Suard, A. Niquille, D. Héquet et al., "Decrease in antibacterial use and facility-level variability after the introduction of guidelines and implementation of physician-pharmacist-nurse quality circles in Swiss long-term care facilities," *Journal of the American Medical Directors Association*, vol. 21, no. 1, pp. 78–83, 2020.

[22] H. Feng, G. Li, C. Xu, C. Ju, and P. Suo, "A quality control circle process to improve implementation effect of prevention measures for high-risk patients," *International Wound Journal*, vol. 14, no. 6, pp. 1094–1099, 2017.

[23] A. C. Tricco, S. M. Thomas, A. A. Veroniki et al., "Quality improvement strategies to prevent falls in older adults: a systematic review and network meta-analysis," *Age and Ageing*, vol. 48, no. 3, pp. 337–346, 2019.

[24] A. Mahlknecht, M. E. Abuzahra, G. Piccoliori, N. Enthaler, A. Engl, and A. Sönnichsen, "Improving quality of care in general practices by self-audit, benchmarking and quality circles," *Wiener Klinische Wochenschrift*, vol. 128, pp. 706–718, 2016.

[25] Y. Luo, Q. Yang, B. Li, and Y. Yao, "Establishment of a quality control circle to reduce biofilm formation in flexible endoscopes by improvement of qualified cleaning rate," *Journal of International Medical Research*, vol. 48, no. 9, Article ID 300060520952983, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4623869, 6 pages
https://doi.org/10.1155/2022/4623869

Research Article

Impact of Mental Health First Aid Training Courses on Patients' Mental Health

Fanli Zeng, Dexia Zhong, Xi Chen, Hongmei Li, and Xiaofei Tian

Neurology Department and Nursing Department, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen 518107, China

Correspondence should be addressed to Xiaofei Tian; tianxf@mail.sysu.edu.cn

Received 24 June 2022; Revised 26 July 2022; Accepted 3 August 2022; Published 12 September 2022

Academic Editor: Wenming Cao

Background. With the prevalence of mental issues worldwide, more and more people are suffering from psychological torture. Mental Health Gap Action Program (mhGAP) has been introduced to improve the life quality of humans. *Objectives*. To explore and synthesize evidence of participants' experience of mental health first aid (MHFA) training course. *Method*. Peer-reviewed qualitative evidence was systematically reviewed and thematically synthesized. Cumulative Index to Nursing and Allied Health Literature (CINAHL), Medical Literature Analysis and Retrieval System Online (MEDLINE), Psychological Information (PsycINFO), PubMed, Psych ARTICLES, Web of Science, Joanna Briggs Institute (JBI), and National Institute for Health and Care Excellence (NICE) databases were searched for the inception of the present study. The study's quality was appraised using the Critical Appraisal Checklist for Qualitative Research of Joanna Briggs Institute (JBI) appraisal tool. All the participants who have attended the MHFA training course (excluding instructors) setting were included. *Results*. Six papers published between 2005 and 2019 were included for thematic synthesis. The review indicated that MHFA had been a positive experience for participants. *Conclusions*. MHFA courses can provide participants with professional knowledge of mental health counseling and improve their knowledge, practice, and attitudes towards their patients. Professional MHFA training courses should therefore be popularized and promoted among other populations.

1. Introduction

With the prevalence of mental issues worldwide, more and more people are suffering from psychological torture. The World Health Organization (WHO) has emphasized the importance of mental health disorders and created a Mental Health Gap Action Program (mhGAP) to improve the life quality of humans. Many people are suffering from suicide and depression [1–3]. This means that governments need to strengthen mental health to promote a better life for people worldwide. Meanwhile, a great number of people in developing countries are living with mental health disorders, which are linked to suicide, alcohol or suicide, and alcohol or drug dependence. This review aimed to review the participants' experience systematically before and after they attended the MHFA training.

There are three parts to the global concern for mental health issues [4]. It is suggested that health care workers should try to help those people with psychological disorders and provide guidance and positive influence/impact to them [5]. Mental health issues are also faced by the general population and related healthcare students [6, 7]. It has been recommended that medical students should seek help from medical schools or mental health systems [8, 9]. The Australian government has designed a Mental Health First Aid (MHFA) course to help those who are developing mental health problems or suffering a mental health crisis. The Mental Health First Aid (MHFA) program is able to help people [10, 11]. People tend to ignore mental health issues partly because they lack insight into mental health and partly because they fear the stigma that comes with mental illness [12, 13]. There were two methods in the educational process

of the MHFA course which included e-learning and face-to-face learning. This review only focused on the effectiveness of the MHFA course. In a specific population, such as nurses and other medical students, they also have heavy stress in hospitals and universities [14]. Hence, it is essential to pay attention to those with mental health disorders to improve their quality of life.

The systematic review was to understand in-depth the experiences of participants who took part in the MHFA training course. Therefore, the review questions are as follows: Firstly, what is the general participants' experience of MHFA training? Secondly, what is the difference between the participants, especially healthcare students, and the other population? The types of participants, exposure, outcome measures, and studies were considered and identified by the following search strategy. Porritt et al. recommended that the PIC (population, phenomena of interest, and context) tool was much more suitable for qualitative studies [15].

2. Methods

This study used systematic review as an approach to answering the specific review questions. More details of participants' experiences are focused on the MHFA training than quantitative data [16]. The review study was identified with different databases, such as Cumulative Index to Nursing and Allied Health Literature (CINAHL), Medical Literature Analysis and Retrieval System Online (MEDLINE), Psychological Information (PsycINFO), PubMed, Psych ARTICLES, and Web of Science. In addition, databases of the Joanna Briggs Institute (JBI) and National Institute for Health and Care Excellence (NICE) were searched to gain more related qualitative articles. This original search terms were shown in Appendix 1. This review will include English-language articles.

2.1. Inclusion and Exclusion Criteria. Nine items were used to make inclusion and exclusion criteria: participants, intervention, intervention setting, study focus, outcomes, study design, time period, publication type, and language [17] (Appendix 2).

2.1.1. Type of Population. This review will use the purposive sampling method to select the final population. This review only focused on the general population, excluding the instructors or facilitators without the limitation on participants' age, gender, religion, and nationality. The phenomenon of interest was participants who have had experience of MHFA training courses after completing the 12-h training or 9-h training. Participants' views of the MHFA, psychology first aid, or physical first aid were also considered.

2.1.2. Type of Context and Study. The MHFA online training courses and Youth MHFA training courses were not considered because of the different teaching methods and the contents of the two-type training [16]. The studies that focused on mental health illness, issues, problems, and disorders were included, as they were relevant to the MHFA training.

The primary qualitative studies, which used grounded theory and phenomenology, were included [18]. The mixed-method literature only focused on the qualitative part to make an analysis. The meta-analysis, case studies, and randomized controlled trial (RCT) were excluded. Therefore, the review focused on qualitative studies, qualitative analysis studies, qualitative research, and mixed-methods papers and did not consider the quantitative studies. All the qualitative studies were written in English to reduce language bias.

2.2. Study Quality Procedures and Assessment. The systematic review needed two reviewers to analyze independently to increase the trustworthiness and decrease the personal variation [18]. The student reviewer would double-check the quality assessment and data extraction every week to ensure that the appraisal and analysis are objective and reliable.

All the included studies were collected and appraised with empirical data procedures, which was an empirical analysis to improve accuracy and estimate errors [19, 20]. This review used the corresponding checklists from the Critical Appraisal Checklist for Qualitative Research of JBI. The details of the comparison were shown in Appendix 3. The JBI checklist contained ten items and made it easier to appraise the qualitative research. Answers to the ten items were categorized as yes/no/not clear/not applicable [21]. The details of quality appraisal are shown in Appendix 4.

2.3. Data Extraction. Data extraction is the process of analyzing and trawling data to retrieve relevant information from data sources of a particular pattern, such as a database [22]. According to Higgins and Green [23], the Cochrane handbook for systematic reviews is a "gold standard" of interventions to guide the review process and implementation. Several data extraction tools include the CASP tool, Evaluation Tool for Qualitative Studies (ETQS), and JBI tool [24].

2.4. Data Synthesis. Data synthesis was an integral part of systematic reviews [22, 25]. According to the Cochrane handbook, a professional synthesis method can help the reviewer gain a deeper analysis of primary studies' findings [26].

3. Results

3.1. Results of the Research. Finally, there were six studies selected, which include five qualitative studies and one mixed-methods paper for inclusion (see Figure 1). It is recommended that PRISMA could make a clear flow chart in systematic reviews and meta-analyses [27–29]. Four were interviews, one was a survey, and another was a questionnaire. These studies were published between 2005 and 2019 and included 238 participants. Three studies were from

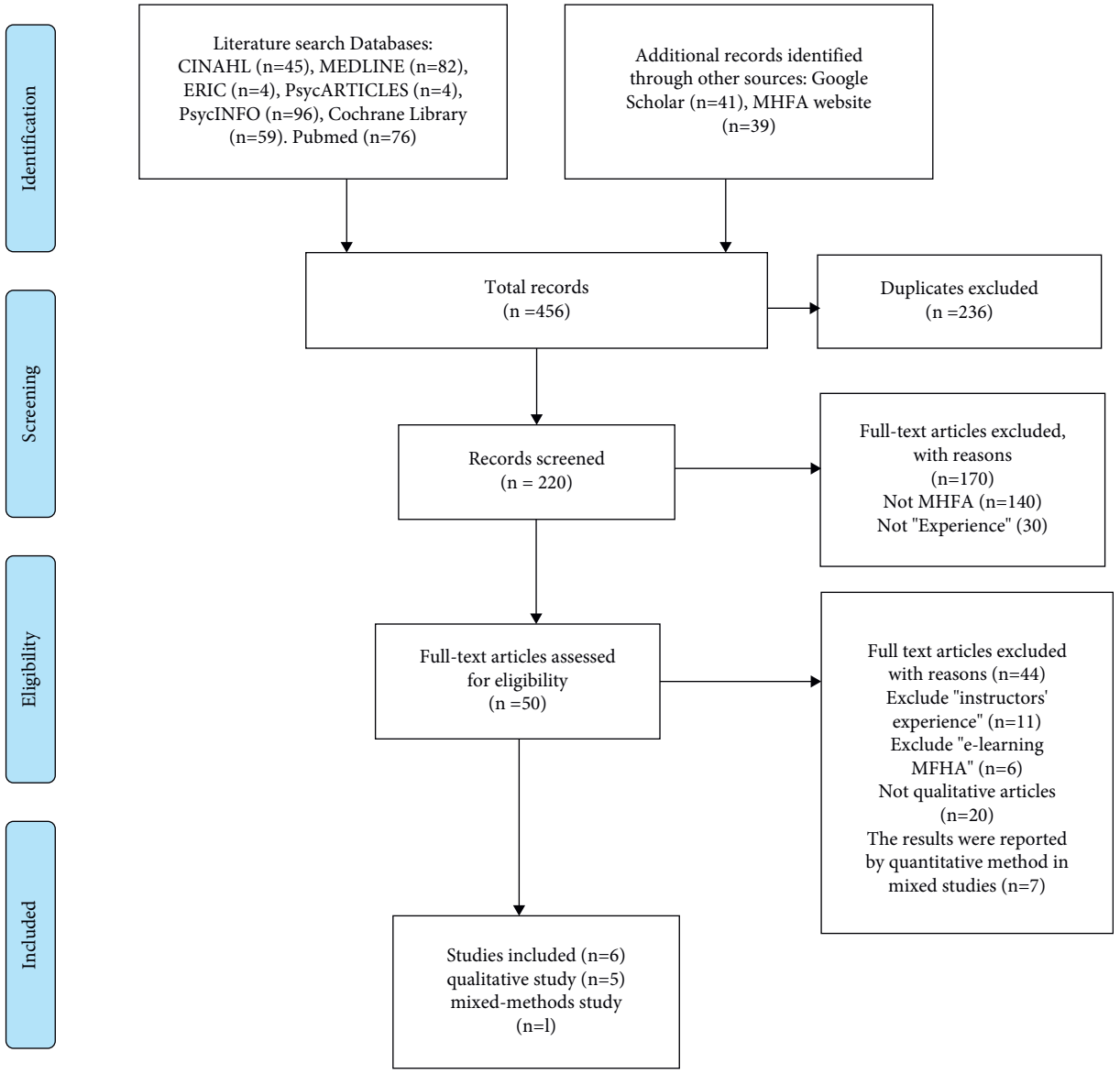


Figure 1: A systematic qualitative review and thematic synthesis is the participants' experiences of the mental health first aid training.

Australia, the USA, Sweden, and Hong Kong (China). Three studies considered the same sample focused on university students, especially nursing and midwifery students [30–32]. One paper had 15 instructors who had finished a 5-day instructor training with the ability to teach the MHFA-USA course to others [11]. The details are shown in Appendix 5.

3.2. Results of Quality Appraisal. Six studies clearly stated their objectives and focused on the different size of studies (see Appendix 6). Three of these papers were high-quality [30, 31, 33], while the other three were medium-quality papers through the quality appraisal [11, 32, 34]. As was shown in Appendix 6, the studies' congruity was high, it was worth analyzing, and these six papers were all included. Two identified analytical themes include to improve individuals' MHFA knowledge and practical ability and to improve individuals' attitude.

3.2.1. Individuals' MHFA Knowledge and Practical Ability Were Improved. All six papers indicated that the MHFA training course could enhance participants' knowledge and understanding of mental health issues [11, 30–34]. Besides, some participants also highlighted that the MHFA course could improve their career confidence and provide more information resources and services in their local contexts [30]. Moreover, one participant described that the MHFA had a positive effect and benefit to "being a good listener and friend and giving helpful advice" [34]. Some participants stated the MHFA training was "very interesting," "very helpful," "beneficial and valuable," and "would encourage everyone to attend it" [34]. On the other hand, the participants noticed that the instructor was "very good," who could teach them professional knowledge of mental health [33].

The papers indicated that the MHFA course could improve participants' mental health essential skills and techniques to support patients [11, 30–32, 34]. Additionally, one

of the trainees reported that the MHFA course was "great" and could teach him/her to interact with people through the ALGEE [11]. It was highlighted that the MHFA course provided trainees with practical skills to people [11].

Four studies indicated that the MHFA course could improve participants' awareness and confidence in dealing with mental health issues [11, 32, 34, 35]. It is suggested that the graduates were more willing and able to help others after joining the MHFA training, as one participant reported that it provided some "tangible tools" [11]. Similarly, Rodgers et al. indicated that some participants replied that the course gave them "more confidence" to ask patient's feelings and thoughts during offering the support [32].

It was interesting that the MHFA course not only improved awareness of other people's mental status but also improved understanding of their own mental health conditions [11, 30, 34]. Likely, some participants promoted the MHFA training to make them "more aware" for themselves including physical and mentally health [11]. The studies indicated that the MHFA training enhanced the sense of achievement and satisfaction [30, 34]. For instance, one participant perceived the MHFA course as a bridge and provided her "more confidence" in dealing with her father's situation and helping him [30].

3.2.2. Improved Individuals' Attitude. The papers indicated that the MFHA course had changed participants' beliefs and values after joining the course [11, 30, 34, 35]. One person expressed that he/she sometimes was "inpatient and annoying" to take care of his/her friend before the MHFA training, and this course made them understand their friend's "negative mood or mindset" [30]. The three studies indicated that the participants became more willing to provide support to people [11, 30, 34]. A nursing student said that she was not "more willing to offer help," as the stress of being a nurse was very high [30]. Lucksted et al. indicated that the MHFA course also increased the intention to help people, especially permission and responsibility [11]. Two studies indicated that MHFA training could reduce participants' stigma and avoidance after joining the course [11, 32]. For instance, many people feared schizophrenia. However, one participant reported that the course made him "more compassionate and a little fearless" to talk with others after the course [11]. Rodgers et al. pointed out that some participants faced challenges in supporting MHFA [32]. One respondent reported they needed to "maintain" and "keep supporting" to help, and another participant stated that he/she lacked the "energy or emotional" ability to deal with the two people at the same time.

4. Discussion

According to this review and thematic synthesis, participants' experiences of the MHFA course after training were categorized into two key themes: improved individuals' MHFA knowledge and practical ability and then improved individuals' attitude. The six subthemes indicated that the MHFA training course improved participators' knowledge and practical support of mental health illness [11, 30, 32, 34, 35]. Five papers indicated that the instructors teach them the first aid actions to offer immediately when someone needs help [11, 30–32, 34]. In terms of attitude outcomes, the attitudes evolved from four different subthemes, which included altered beliefs and values, increased willingness to help, reduced stigma and fear, reduced emotional energy, and increased intelligence [11, 30, 32, 34, 35].

Depending on the duration of the training, the participants improved their negative attitudes and had more willingness to help others in a more positive, patient, and responsible way. When they were eager to help those people who are in need, the more positive outcomes came out at the end of the supporting [11, 32, 34].

This review identified six interviews or questionnaires of MHFA training in a variety of settings. There were generally small to moderate improvements post training and could be seen up to 21 months throughout the results. The findings of the present study indicated that introducing the MHFA training course would benefit the general public, especially nursing students and healthcare workers. The nursing students had more feedback than the general public on the MHFA training course because nursing students had some related knowledge before related to this course.

This section will discuss whether this systematic review can be generalized or not. The papers were a little disparate in terms of their design, populations, exposures, and outcomes, which indicated it was difficult to compare outcomes across the quality studies or perform evidence synthesis. Therefore, the findings of this review should be generalized from the healthcare workers and then spread to the whole population worldwide.

There were a lot of common points in my study. Both Morgan et al. and my study have shared a number of key features, including improving knowledge, confidence, and intentions and altering behaviors after attending the training [36]. My review agreed with these positive experiences of Morgan et al. In terms of methods, this article was a systematic quantitative review, whereas my study was a systematic qualitative review [36]. Furthermore, this article focused on the data from the primary studies, while my review focused on the experiences of the participants [36]. Another related study showed that people who received MHFA improved in mental health literacy, but the changes in adolescents' mental health cannot be detected [37]. However, my study using the PICO tool covered slightly less comprehensively whether MHFA training is needed to be explored further in different countries. Adolescent MHFA is effective in increasing recognition and intention to assist suicidal peers similar to my outcome [38, 39].

The advantages and disadvantages of this systematic review are discussed in this paragraph. The search strategy was comprehensive. The reviewer searched the articles from different databases to ensure all the suitable papers were included. The secondary advantage is that it fully incorporates the subjects' experience of attending an MHFA training course. Another strength is the up-to-date primary studies. However, there was some bias in this review [23].

The articles are from four countries, and their full experience on the MHFA training course is not yet representative of the world's feel about the course [18]. Moreover, the review included a study of which population considered the instructors. This may increase the risk of bias [23]. This systematic qualitative review indicated that the MHFA training course could improve people's knowledge, practical skills, and alter their attitudes. This review can support the evidence who insist that the MHFA training course can reduce their stigma and improve confidence and knowledge.

This review demonstrated the positive effects of participants' experience on the MHFA training course. Although most of the included papers stated positive feedback on MHFA, there were some negative feedbacks from the participators. This may be because they had learned related knowledge before attending this course. What is more, some participants thought that the course should be more flexible and should contain interesting information to improve the efficiency of the course. In future research, the reviewer should analyze more articles such as e-learning MHFA and the YMHFA to undertake this course in-depth and find its meaning in improving people's life quality. For its population, more general participants need to be considered.

In short, different people had different experiences on the MHFA training course, which included both positive and negative outcomes. This was because there were different levels of knowledge of mental health illness; the higher level there was, the more requirements they would acquire. It is a significant outcome that the MHFA training course has different effects on different people, which means that the course still needs some improvements to satisfy a larger population in the future. The MHFA training course can be further developed into a specific course for old people, as their mental health issues differ from those of young people.

5. Conclusion

MHFA courses can provide participants with expertise in mental health counseling, improve their practical skills, and improve their attitude towards their patients.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Fanli Zeng and Dexia Zhong have contributed equally to this work and share first authorship.

Supplementary Materials

Appendix 1: Search strategy for one database used. Appendix 2: Table of include and exclude selection. Appendix 3: The differences among different tools of appraise qualitative research. Appendix 4: Critical appraisal forms. Appendix 5: Table of characteristics of included studies. Appendix 6: Analytical themes forms. (*Supplementary Materials*)

References

[1] V. N. Mutiso, C. W. Musyimi, T. J. Rebello et al., "Perceived impacts as narrated by service users and providers on practice, policy and mental health system following the implementation of the mhGAP-IG using the TEAM model in a rural setting in Makueni County, Kenya: a qualitative approach," *International Journal of Mental Health Systems*, vol. 14, no. 1, pp. 13–56, 2019.

[2] V. N. Mutiso, K. M. Pike, C. W. Musyimi et al., "Feasibility and effectiveness of nurses and clinical officers in implementing the WHO mhGAP intervention guide: pilot study in Makueni County, Kenya," *General Hospital Psychiatry*, vol. 59, pp. 20–29, 2019.

[3] V. N. Mutiso, K. Pike, C. W. Musyimi et al., "Feasibility of WHO mhGAP-intervention guide in reducing experienced discrimination in people with mental disorders: a pilot study in a rural Kenyan setting," *Epidemiology and Psychiatric Sciences*, vol. 28, no. 2, pp. 156–167, 2019.

[4] V. Patel and S. Saxena, "Transforming lives, enhancing communities--innovations in global mental health," *New England Journal of Medicine*, vol. 370, no. 6, pp. 498–501, 2014.

[5] K. Schick-Makaroff and J. L. Storch, "Guidance for ethical leadership in nursing codes of ethics: an integrative review," *Canadian Journal of Nursing Leadership*, vol. 32, no. 1, pp. 60–73, 2019.

[6] R. P. Auerbach, J. Alonso, W. G. Axinn et al., "Mental disorders among college students in the world health organization world mental health surveys," *Psychological Medicine*, vol. 46, no. 14, pp. 2955–2970, 2016.

[7] R. M. Holm-Hadulla and A. Koutsoukou-Argyraki, "Mental health of students in a globalized world: prevalence of complaints and disorders, methods and effectivity of counseling, structure of mental health services for students," *Mental Health & Prevention*, vol. 3, no. 1–2, pp. 1–4, 2015.

[8] R. Puthran, M. W. B. Zhang, W. W. Tam, and R. C. Ho, "Prevalence of depression amongst medical students: a meta-analysis," *Medical Education*, vol. 50, no. 4, pp. 456–468, 2016.

[9] S. K. Lipson, E. G. Lattie, and D. Eisenberg, "Increased rates of mental health service utilization by U.S. College students: 10-year population-level trends (2007-2017)," *Psychiatric Services*, vol. 70, no. 1, pp. 60–63, 2019.

[10] A. P. Yonelinas, J. R. Quamme, K. F. Widaman, N. E. A. Kroll, M. J. Sauvé, and R. T. Knight, "Mild hypoxia disrupts recollection, not familiarity," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 4, no. 3, pp. 393–400, 2004.

[11] A. Lucksted, A. N. Mendenhall, S. I. Frauenholtz, and J. M. Aakre, "Experiences of graduates of the mental health first aid-USA course," *International Journal of Mental Health Promotion*, vol. 17, no. 3, pp. 169–183, 2015.

[12] A. F. Jorm, "Why we need the concept of "mental health literacy"," *Health Communication*, vol. 30, no. 12, pp. 1166–1168, 2015.

[13] J. K. Kern, D. A. Geier, L. K. Sykes, and M. R. Geier, "Relevance of neuroinflammation and encephalitis in autism," *Frontiers in Cellular Neuroscience*, vol. 9, p. 519, 2015.

[14] J. C. Fortney, J. Unützer, G. Wrenn et al., "A tipping point for measurement-based care," *Psychiatric Services*, vol. 68, no. 2, pp. 179–188, 2017.

[15] K. Porritt, J. Gomersall, and C. Lockwood, "JBI's systematic reviews: study selection and critical appraisal, AJN," *AJN, American Journal of Nursing*, vol. 114, no. 6, pp. 47–52, 2014.

[16] J. A. Maxwell, "Designing a qualitative study," *The SAGE handbook of applied social research methods*, vol. 2, pp. 214–253, 2009.

[17] R. Coates, G. Cupples, A. Scamell, and C. McCourt, "Women's experiences of induction of labour: qualitative systematic review and thematic synthesis," *Midwifery*, vol. 69, pp. 17–28, 2019.

[18] E. Aromataris and A. Pearson, "The systematic review: an overview," *AJN, American Journal of Nursing*, vol. 114, no. 3, pp. 53–58, 2014.

[19] G. Tsebelis, C. B. Jensen, A. Kalandrakis, and A. Kreppel, "Legislative procedures in the European Union: an empirical analysis," *British Journal of Political Science*, vol. 31, no. 04, pp. 573–599, 2001.

[20] B. S. Dobronets and O. A. Popova, "Improving reliability of aggregation, numerical simulation and analysis of complex systems by empirical data," *IOP Conference Series: Materials Science and Engineering*, vol. 354, no. 1, Article ID 012006, 2018.

[21] C. Lockwood, Z. Munn, and K. Porritt, "Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation," *International Journal of Evidence-Based Healthcare*, vol. 13, no. 3, pp. 179–187, 2015.

[22] C. M. Castellan, "Quantitative and qualitative research: a view for clarity," *International Journal of Education*, vol. 2, no. 2, p. 1, 2010.

[23] J. P. Higgins and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions*, John Wiley & Sons, no. 4, Hoboken, New Jersey, USA, 2011.

[24] M. B. Elamin, M. Z. Garcia, M. H. Murad, P. J. Erwin, and V. M. Montori, "Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses," *Clinical Endocrinology*, vol. 72, no. 1, pp. 1–10, 2010.

[25] M. Sandelowski, "Reading, writing and systematic review," *Journal of Advanced Nursing*, vol. 64, no. 1, pp. 104–110, 2008.

[26] J. Noyes, A. Booth, K. Flemming et al., "Cochrane Qualitative and Implementation Methods Group guidance series—paper 3: methods for assessing methodological limitations, data extraction and synthesis, and confidence in synthesized qualitative findings," *Journal of Clinical Epidemiology*, vol. 97, pp. 49–58, 2018.

[27] B. Takkouche, G. Norman, A. Liberati, and J. Tetzlaff, "PRISMA statement," *Epidemiology*, vol. 22, no. 1, p. 128, 2011.

[28] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Annals of Internal Medicine*, vol. 151, no. 4, pp. 264–269, 2009.

[29] V. M. Ortiz-Martínez, P. Andreo-Martinez, N. Garcia-Martinez, A. Pérez de los Ríos, F. J. Hernández-Fernández, and J. Quesada-Medina, "Approach to biodiesel production from microalgae under supercritical conditions by the PRISMA method," *Fuel Processing Technology*, vol. 191, pp. 211–222, 2019.

[30] M. S. Hung, S. K. Lam, and M. C. Chow, "Nursing students' experiences of mental health first aid training: a qualitative descriptive study," *Collegian*, vol. 26, no. 5, pp. 534–540, 2019.

[31] J. Kelly and M Birks, "It's the simple things you do first that start the process of help': undergraduate nursing and midwifery students' experiences of the Mental Health First Aid course," *Collegian*, vol. 24, no. 3, pp. 275–280, 2017.

[32] G. Rodgers, S. Burns, and G. Crawford, ""I was able to actually do something useful": evaluating the experiences of university students after completing Mental Health First Aid: a mixed-methods study," *Advances in Mental Health*, vol. 19, pp. 40–62, 2019.

[33] B. Svensson, L. Hansson, and S. Stjernswärd, "Experiences of a mental health first aid training program in Sweden: a descriptive qualitative study," *Community Mental Health Journal*, vol. 51, no. 4, pp. 497–503, 2015.

[34] A. F. Jorm, B. A. Kitchener, and S. K. Mugford, "Experiences in applying skills learned in a Mental Health First Aid training course: a qualitative study of participants' stories," *BMC Psychiatry*, vol. 5, no. 1, pp. 43–52, 2005.

[35] E. Svensson, E. Horváth-Puhó, R. W. Thomsen et al., "Vagotomy and subsequent risk of Parkinson's disease," *Annals of Neurology*, vol. 78, no. 4, pp. 522–529, 2015.

[36] A. J. Morgan, A. Ross, and N. J. Reavley, "Systematic review and meta-analysis of Mental Health First Aid training: effects on knowledge, stigma, and helping behaviour," *PLoS One*, vol. 13, no. 5, 2018.

[37] A. J. Morgan, J. A. A. Fischer, L. M. Hart et al., "Does Mental Health First Aid training improve the mental health of aid recipients? The training for parents of teenagers randomised controlled trial," *BMC Psychiatry*, vol. 19, no. 1, p. 99, 2019.

[38] J. M. Aakre, A. Lucksted, and L. A. Browning-McNee, "Evaluation of Youth Mental Health First Aid USA: a program to assist young people in psychological distress," *Psychological Services*, vol. 13, no. 2, pp. 121–126, 2016.

[39] L. M. Hart, P. Cropper, A. J. Morgan, C. M. Kelly, and A. F. Jorm, "Teen Mental Health First Aid as a school-based intervention for improving peer support of adolescents at risk of suicide: outcomes from a cluster randomised crossover trial," *Australian and New Zealand Journal of Psychiatry*, vol. 54, pp. 382–392, 2019.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8488167, 10 pages
https://doi.org/10.1155/2022/8488167

Research Article

DBN Neural Network Model Combined with Meta-Analysis on the Curative Effect of Acupuncture and Massage

Xiujun Wang

School of Nutrition Clinic, The Third Hospital of Jinan Municipality, Jinan 250132, China

Correspondence should be addressed to Xiujun Wang; 2020212230@mail.chzu.edu.cn

Received 30 June 2022; Revised 1 August 2022; Accepted 14 August 2022; Published 5 September 2022

Academic Editor: Wenming Cao

Acupuncture and massage are among the oldest medical treatments in China. During the acupuncture process, as well as the subsequent needle extraction process, there are differences in the acupuncture intensity, treatment duration, and acupuncture depth. For both medical treatments of acupuncture and massage, this article learns and analyzes a large amount of literature and applies DBN neural network method to build a human skeletal model to simulate and identify medical professional steps such as acupuncture therapy. The research results show that the recognition rate of DBN reaches 92.1% after the training of 1000 samples. After learning all the training samples, the DBN model achieved feature recognition accuracy of 96.4%, 97.68%, 96.66%, and 92.27% for the test samples of mixed needling process, needle insertion operation, needle extraction operation, and rotary needle handling process, respectively. The research in this article can contribute to the modernization of Chinese medicine by maximizing the simulation of the force on the human body when receiving needling and tui-na, as well as the clinical treatment effect.

1. Introduction

Acupuncture and massage, as one of the oldest medical means in China, have a long history of development and play a prominent role in China's medical field [1–3]. In China's medical field, people still habitually choose acupuncture and massage as treatments for some common problems such as acupoint blockage and local pain [4, 5]. Acupuncture is generally used to stimulate specific acupoints in the body to achieve therapeutic goals [6–8]. By stimulating the human body with sanyinjiao, Zusanli acupoints can affect blood circulation and stasis, Tianshu and Liangmen acupoints can promote digestion, etc. [1, 3, 9]. When the needle is inserted into the body, the body reacts to the stimulus and passes it on quickly. Then, the body responds to the action, and the pain-accumulating substances are redistributed, so that the pain is less likely to accumulate locally. After acupuncture enters the body, the immunity of the body can be mobilized, the body's response begins to strengthen, and the body becomes more combative, which can resist the invasion of external evil. Acupuncture can promote the improvement of self-healing ability. To be specific, when the human body is in a mild imbalance state, the body can be restored by self-adjustment. However, when the imbalance state is serious, external stimulation, such as acupuncture, is needed to stimulate the treatment, so as to improve the body's self-healing ability [10, 11].

The same message belongs to the external treatment of Traditional Chinese medicine, massage the specific parts of the body surface to achieve the purpose of regulating the viscera, dredging the meridians, promoting the normal operation of qi and blood, often used for the treatment of various pain, daily massage has the effect of relieving fatigue. Chinese medicine in the human body according to the meridians are acupoints with pushing, taking, lifting, kneading, and other techniques for treatment. Massage is a kind of non-drug natural therapy and physical therapy. The doctor uses his hands to act on the body surface of the patient, the injured part, the uncomfortable place, the specific acupoint, and the painful place, to achieve the curative effect of dredging the meridians, promoting qi and blood, helping the injury and relieving pain, dispelling evil and strengthening, and harmonizing Yin and Yang [12–15]. In addition, because massage is beneficial to circulation and

metabolism, it has a positive therapeutic effect on patients with excessive weakness in general. In the long common field of traditional Chinese medicine, acupuncture and massage are usually combined to provide effective treatment for patients. However, in actual acupuncture and massage therapy for acupuncture needles of processes, stimulation, and subsequent needles pulled up in the process of the healer, stimulation treatment time, the problem such as the depth of the needles, these characteristics as the healer state each are not identical, there are differences, how to efficient, accurate to help doctors to solve these problems put forward the requirement to us.

At present, with the rapid development of science and technology and the improvement of computer networks, the application of virtual reality technology and neural network learning to some difficult training work, in reality, has been increasing [16–20]. In Xu et al. [21], they investigated the uncertainty of the dose-response relationship between acupuncture and major depression, including randomized controlled trials comparing acupuncture with sham acupuncture or antidepressants, data extraction, data quality, and bias risk assessment. Finally, a nonlinear meta-regression method with cubic spline restriction was used to study the dose-effect relationship between acupuncture courses and their effect on Hamilton Depression Scale (HAMD) score. Niu et al. [22] conducted a targeted study on the efficacy of acupuncture and massage, and their results showed that massage and acupuncture had good therapeutic effects on patients, but further regulation of the intensity of massage and acupuncture and other operations as needed. Their studies have shown the positive effect of acupuncture and massage on treatment, but in the process of acupuncture and massage, the strength of acupuncture, the depth of the needle, and the strength of massage need to be further standardized. Combining neural networks to optimize and improve problems is now a commonly used method. Belciug [23] proposed a potential deep learning method, network architecture automation. As the number of network architectures increases exponentially with the number of convolution layers in the network, a differential evolution algorithm is proposed to traverse the search space, generate a group of random individuals, and then mutate, recombine, and select. In each generation, individuals with the worst loss values are eliminated and replaced by more competitive individuals. The results show that the proposed method has a positive effect on medical image recognition. Liu et al. [24] proposed a medical image super-resolution algorithm (SR-DCNN) based on a neural network and using deconvolution, which can be used to recover high-resolution images from low-resolution images, especially for medical images with clinical significance in diagnosis, treatment, and research applications. Interestingly, Anilkumar et al. [25] detected leukemia by analyzing microscopic smear images to detect the presence of leukemia cells using an image processing-based method, and compared the efficacy of three optimization algorithms of stochastic gradient descent with momentum, root means square propagation, and adaptive moment estimation (ADAM) in all the classifications performed. Through the analysis of the above literature, it is shown that the application of deep learning neural networks in the medical field has been more and more extensive, and has achieved good results, which provides a solution to the problems related to acupuncture and massage. In addition, there are differences in some conditions of patients, even in the condition of the same influencing factor, there are differences in different patients and different ages. Therefore, we need to comprehensively compare the heterogeneity of the system, to better serve patients. Silvy et al. [26] used meta-analysis to explore the problem that cumulative meta-analysis can detect the presence or absence of treatment effects early in the treatment process, to avoid unnecessary and expensive new randomized trials. Katsanos et al. [27] analyzed and explained the relationship between statins and cerebral microbleeds (CMB) through meta-analysis, and analyzed and evaluated the association between the existence of CMB and the use of statins. Their study showed that meta-analysis showed excellent and stable performance in comparison and synthesis of research results. Compared with a single study, the integration of all acupuncture and message-related studies can more accurately standardize doctors and is conducive to exploring the consistency of evidence and differences among studies.

Therefore, in our research work, we focused on the acupuncture process of acupuncture, stimulation treatment, as well as the follow-up needle extraction process of the strength of the doctor, the time of stimulation treatment, the depth of acupuncture and other problems, the strength of massage and the corresponding different patients, and different ages of some different performance. We first by choosing DBN neural network for accurate identification of acupuncture and moxibustion in the process of operating characteristics, and through the meta-analysis of heterogeneity and multiple independent research statistic consistency check, to evaluate the effect of acupuncture and massage analysis for us, our research will contribute to the promotion and application of TCM acupuncture and massage.

2. Feasibility of DBN in the Evaluation of the Curative Effect of Acupuncture and Massage

In the context of the booming development of big data [28], due to the robustness of mesh meta-analysis and the accuracy of the analysis model, the computational modeling method driven by big data is more suitable for solving the optimal intervention scheme of acupuncture and massage for the treatment of various TCM bias.

2.1. Design of Multilayer Neural Network. In this article, the gradient descent method is used to train the multilayer neural network shown in the figure. The specific calculation process is as follows:

$$\Delta w = -\eta \frac{\delta E}{\delta w}, \tag{1}$$

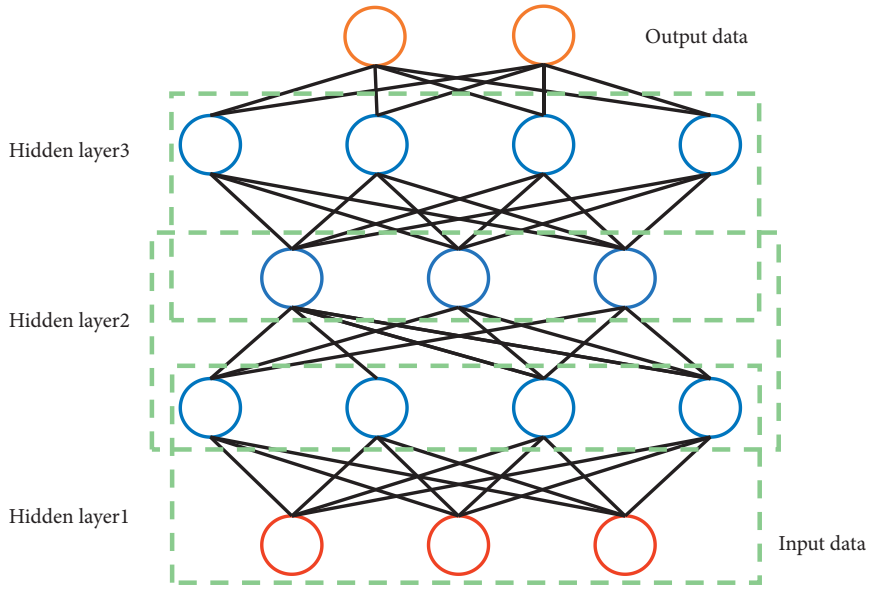


Figure 1: Flow chart of DBN network structure.

where Δw is the adjustment value of the proportion of connection weight w; E is the error between actual results and predicted results; and η is the learning rate.

To update each parameter in the network model of acupuncture and massage efficacy, we need to define the loss function to compare the error between the actual result and the predicted result and update the hyperparameters through the partial derivative of the error function. The smooth $L1$ loss function is used in this article. The specific calculation process is as follows:

$$\text{smooth}_{L1} = \begin{cases} 0.5\,(y - f\,(x))^2, if|x| < 1 \\ |y - f\,(x)| - 0.5, \text{otherwise.} \end{cases} \tag{2}$$

where y is the actual result and $f\,(x)$ is the predicted result.

In common machine learning, the gradient descent direction is the steepest. The smooth $L1$ loss function is minimized by iterating the calculated gradients in the network and updating the neural network model parameters in the reverse gradient direction so that only the minimum matching value of the gradient is used in each calculation. The calculation process is as follows:

$$\begin{aligned} g_t &= \nabla_{\theta t-1} f\,(\theta_{t-1}), \\ \Delta\theta_t &= -\eta g_t, \end{aligned} \tag{3}$$

where g_t is the gradient and θ is the hyperparameter.

In addition, to accelerate the convergence speed of the model, the momentum optimization algorithm is used to update the parameters of the neural network model. The specific calculation process is as follows:

$$\begin{aligned} m_t &= \mu m_{t-1} + g_t, \\ \Delta\theta_t &= -\eta m_t, \end{aligned} \tag{4}$$

where $\mu$is the momentum factor and m_t is the gradient descent loss function after the momentum optimization algorithm is added.

2.2. The Fundamentals of Deep Confidence Networks (DBN). DBN updates and iterates the connection weight ratio in the deep belief network through the learning method of layer-by-layer neurons [29, 30]. Furthermore, the specific data of acupuncture and massage were extracted effectively by using the method of forwarding unsupervised layer-by-layer neuron training, and the therapeutic results of acupuncture and massage were transmitted back using supplementary supervision and fine-tuning to further calculate the optimal intervention scheme for treating various TCM biased constitution. The specific DBN network structure is shown in Figure 1.

In DBN, the neural network was used to generate a random evaluation of acupuncture and massage efficacy to help input acupuncture and massage data to complete probability distribution learning. We build our DBN model through Matlab software, and then we need to deal with basic information such as research objects, intervention measures, measurement indicators, treatment period, number of subjects, gender, physical fitness intervention measures, measurement indicators, and courses of treatment. It is used as the input data of DBN, and then it is continuously trained iteratively. Specifically, the energy function between the units that define the energy function of RBM, the hidden unit that calculates acupuncture and massage efficacy evaluation of each layer, and the units that visually output and input acupuncture and massage efficacy evaluation layer of each layer are described as follows:

$$E(v,h;\theta) = -\sum_{i=1}^{n} b_i v_i - \sum_{j=1}^{n} a_i h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i w_{ij} h_j, \tag{5}$$

where $w = (w_{ij})$, $i\epsilon[1,n] j\epsilon[1,m]$ is the weight ratio matrix between the hidden unit of acupuncture and massage efficacy evaluation and the units of visual output and input acupuncture and massage efficacy evaluation layer; $a = (a_j)$, $j\epsilon[1, m]$ is the bias of the implicit unit in the evaluation of the curative effect of acupuncture and massage; $b = (b_j)$, $i\epsilon[1, n]$ is the bias of units in the evaluation layer of acupuncture and massage efficacy of visual output-input; v represents the input data of acupuncture and massage in DBN neural network; h is the output result of acupuncture and massage in DBN neural network; i is the neural layer; j is the j-th neuron in the i-th neural layer.

$$P(v,h,\theta) = \frac{1}{Z(\theta)} \exp\left(-E(v,h;\theta)\right), \tag{6}$$

where $P(v, h; \theta)$ defines the joint probability between the implicit unit of acupuncture and massage efficacy evaluation and the units of visual output-input acupuncture and massage efficacy evaluation layer based on the theory of thermodynamics.

$$Z(\theta) = \sum_{v}\sum_{h} \exp\left(-E(v,h;\theta)\right), \tag{7}$$

where $Z(\theta)$ is the normalized factor.

Based on the joint probability, the conditional probability and edge probability of the implicit unit of acupuncture and massage efficacy evaluation and the visual output-input of acupuncture and massage efficacy evaluation layer were further obtained. The specific calculation process is as follows:

$$\begin{aligned} P(h;\theta) &= \frac{1}{Z(\theta)} \sum_{v} \exp\left(-E(v,h;\theta)\right), \\ P(v;\theta) &= \frac{1}{Z(\theta)} \sum_{h} \exp\left(-E(v,h;\theta)\right), \end{aligned} \tag{8}$$

where $P(h;\theta)$ is the conditional probability of implicit unit of acupuncture and massage efficacy evaluation and visual output-input layer of acupuncture and massage efficacy evaluation; $P(v;\theta)$ is the edge probability of hidden unit of acupuncture and massage efficacy evaluation and visual output-input of acupuncture and massage efficacy evaluation layer.

In addition, the Sigmoid activation function is used as the activation function of the DBN neural network in this article. The specific calculation process is as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{9}$$

Finally, since there is no connection between the implicit unit of acupuncture and massage efficacy evaluation and the visual output and input acupuncture and massage efficacy evaluation layer, the activation function between neurons in the DBN network is deduced. The specific calculation process is as follows:

$$\begin{aligned} P\left(h_j = 1 \,\middle|\, v;\theta\right) &= \frac{1}{1 + \exp\left(-a - \sum_i v_i w_{ij}\right)}, \\ P\left(v_i = 1 \,\middle|\, h;\theta\right) &= \frac{1}{1 + \exp\left(-b - \sum_i h_i w_{ij}\right)}, \end{aligned} \tag{10}$$

where $P(v_j = 1 \mid h;\theta)$ is the conditional probability of implicit unit of acupuncture and massage efficacy evaluation and visual output-input layer of acupuncture and massage efficacy evaluation obtained through activation function sigmoid processing; $P(h_j = 1 \mid v;\theta)$ is the edge probability of implicit unit of acupuncture and massage efficacy evaluation and visual output-input of acupuncture and massage efficacy evaluation layer obtained by activation function Sigmoid processing. The neural network model applied to evaluate the curative effect of acupuncture and massage; x represents the input data of layer i of acupuncture and massage curative effect; w represents the proportion of connection weight between layer i and layer $i+1$ neurons; b represents bias; z represents the input data of neurons; and a represents the output result of calculation of the curative effect of acupuncture and massage by neurons. The input of neurons in layer $i+1$ is the sum of the output result of acupuncture and massage efficacy of neurons in layer i multiplied by the connection weight proportion w and b bias.

The training process of DBN can be divided into two stages, namely pre-training and reverse fine-tuning. Firstly, the unsupervised greedy learning algorithm is used to train each RBM layer by layer, and the data feature information is transmitted layer by layer. The network parameters are initialized, and the initial connection weights and neuron bias are determined. Then, BP (Back Propagation) algorithm was used for pre-training. The obtained initial weights are fine-tuned from top to bottom, and supervised training is carried out to converge the model to the optimal solution, to determine the structure of the entire DBN network.

Acupuncture and massage input data pretreatment. Classification software was used to classify the initially retrieved acupuncture and massage literature, and the literature that did not meet the standards was screened out. Further screening was conducted by reading the preliminarily obtained acupuncture and massage literature, and the number and reasons for the excluded literature were recorded at the same time. The main process is shown in Figure 2: the databases we searched are mainly divided into Chinese and English databases. The Chinese databases include CNKI Journal Full-text Database (CNKI), CNKI China Doctoral Dissertation Full-text Database, CNKI China Excellent Master's Thesis Full-text Database, Chinese Science and Technology Periodical Database-VIP Information Network (VIP), and Wanfang Database (WANFANG DATA). The English databases include CNKI Journal Full-text Database (CNKI), CNKI China Doctoral Dissertation Full-text Database, CNKI China Excellent Master's Degree Thesis Full-text Database, Chinese Science and

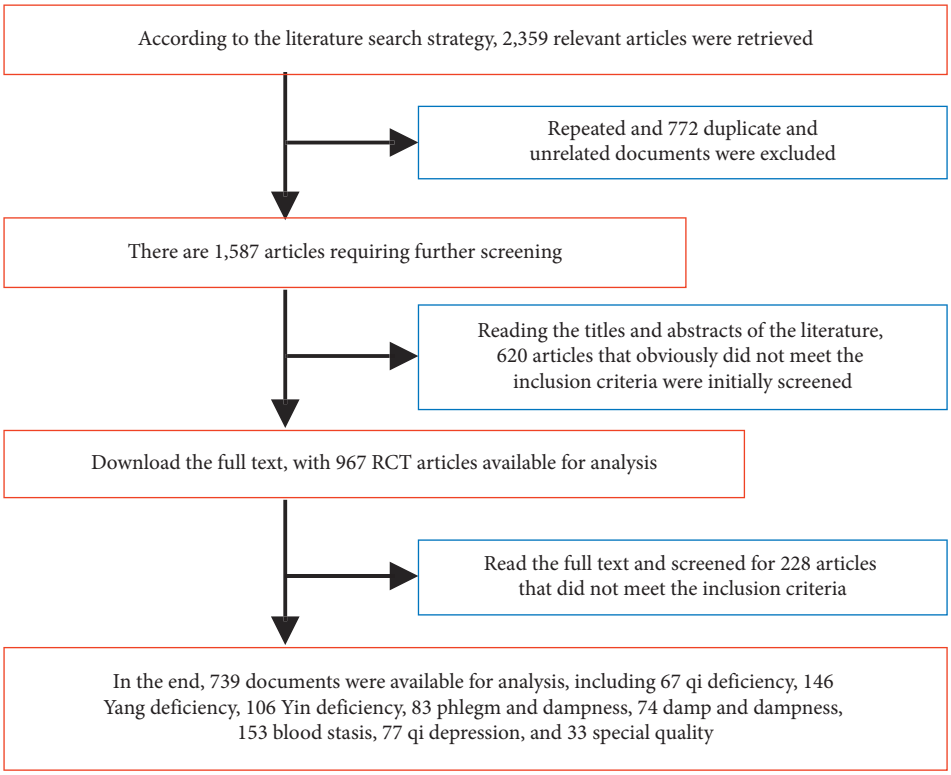


Figure 2: Database literature screening process.

Technology Periodical Database-VIP Information Network (VIP), and Wanfang Database (WANFANG DATA).

In conclusion, in the DBN neural network training process, forward propagation is trained by the input data matrix of acupuncture and massage, and the Gibbs sampling method is adopted to transfer the vector of visual output of RBM into the efficacy evaluation layer of acupuncture and massage to the implicit unit of efficacy evaluation of acupuncture and massage through the RBM network. Then based on the hidden unit of acupuncture and massage efficacy evaluation, the visual output-input acupuncture and massage efficacy evaluation layer was reconstructed and the parameters in DBN network were updated and iterated. Next, the implicit unit vector of acupuncture and massage efficacy evaluation of layer i obtained from training was used as the input vector of layer $i+1$ RBM, and the next calculation was updated. After the loop reaches the maximum number of iterations, the DBN network stops computing and outputs the parameters.

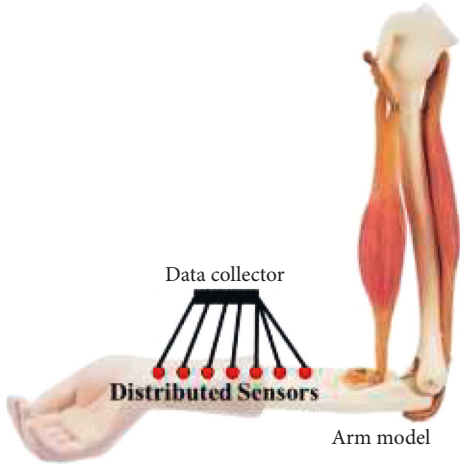


Figure 3: Schematic diagram of arm model.

3. Analysis and Discussion

A deep confidence network (DBN) is a kind of deep neural network based on a traditional constrained Boltzmann machine (RBM). DBN's training scheme is implemented layer by layer, which uses an input vector to calculate and infer the characteristics of the middle layer in the iterative training of each layer, and finally carry out the learning and promotion of the vector of the next layer based on the data and characteristics of the middle layer. In this article, a comprehensive and effective input database was established based on the evaluation of the efficacy of acupuncture and massage by a large number of scholars and researchers collected from the meta-analysis program established earlier. The database includes prior data on various types of patients' physical conditions and previous diagnoses. The analysis data in the database will be input into the established DBN neural network model for feature recognition training. And finally, form a specific program of acupuncture and massage according to the preliminary diagnosis results of patients. During the execution of acupuncture and massage therapy, DBN neural network model will provide precise and comprehensive assistance in strength for physiotherapists by

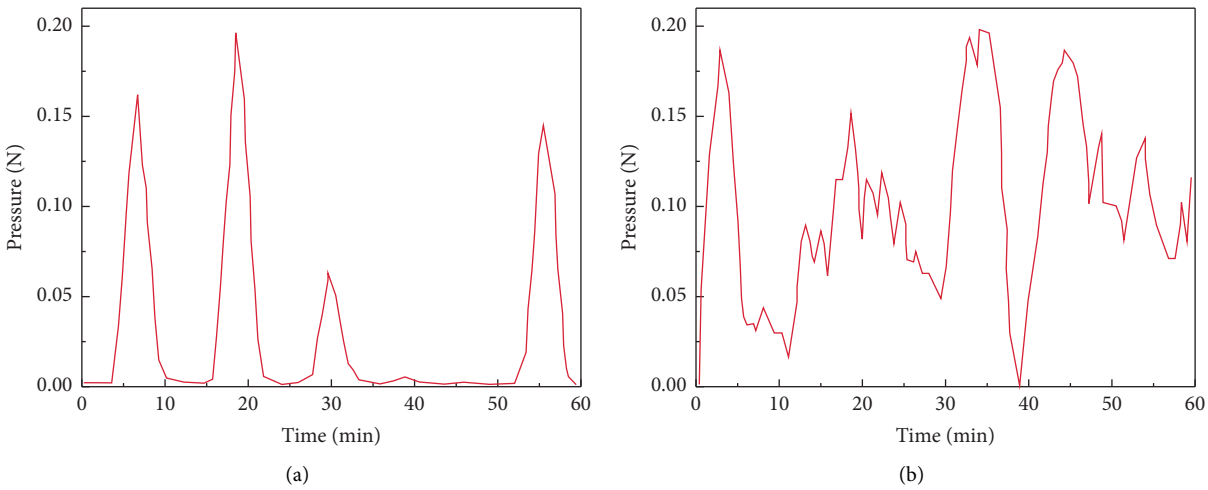


Figure 4: Instantaneous pressure signal diagram of needle: (a) lifting and inserting; (b) rotation.

learning the pressure data of standard acupuncture and massage, and help doctors judge the effectiveness of treatment. Finally, it provides an evaluation of the overall treatment plan and guidance for the follow-up treatment plan.

Training the DBN model has practical guiding significance for factors such as acupuncture and massage strength in the process. In this section, we constructed a human skeleton model to simulate the stress of the actual human body receiving acupuncture and massage and the clinical effect of treatment to the greatest extent. To reduce the amount of calculation and improve accuracy, we focused the actual treatment on the arm part of the human body and carried out targeted acupuncture treatment for the patients. The trained DBN model can be further extended to other parts of patients and other acupoints. The specific composition of the model is shown in Figure 3. The components are silicone imitation of the human arm bone and similar density of flesh tissue model. According to the coordinates and depth of the actual acupoints in the arm, we set up multiple high-precision sensors at the corresponding positions in the arm model. Sensors are evenly distributed on the coordinate line of acupoint depth, and the collected pressure data are read and transmitted to the data collector for recording and feedback. The accuracy of the sensor and data collector has been corrected within a small error range.

Before constructing DBN neural network as a classifier for the specific process of acupuncture and moxibustion treatment for patients, we will first make the DBN model conduct appropriate special diagnosis recognition training for pressure transmission signals in each process of specific acupuncture and moxibustion therapy. For the accuracy and comprehensiveness of DBN model learning, we performed a complete treatment demonstration on the arm model using standard acupuncture treatment courses. The steps of the cadaver include needle insertion, extraction, and needle rotation. Figures 4(a) and 4(b), respectively, show the instantaneous pressure records of the sensor when the needle was inserted and pulled out and rotated in the patient's arm model. Among them, the amplitude of the signal waveform in Figure 4(a) is relatively regular, and the peak represents that we performed a complete needle insertion and extraction for the patient. Specifically, the rise of the crest is the characteristic signal fluctuation of needle insertion, while the decline of the crest in the response corresponds to needle extraction. Figure 4(b) shows the waveform of the pressure signal caused by needle rotation. It is worth noting that the needle rotation is rotated after puncturing the acupoint, so the waveform presented is a discontinuous stimulus signal composed of more pulse signals with smaller amplitude.

We now summarize the input and output of the feature recognition classifier adapted to DBN neural network architecture for the complete arm acupuncture treatment process. For example, in the collection of experimental data, it is found that the main signal recognition feature of needle insertion, extraction, and rotation is different in the waveform diagram in Figure 4. Specifically, we found that the transient pressure value caused by the needle insertion operation was positive, that is, the pressure at the next moment was greater than that at the last moment, which could be explained by the increased tissue penetration resistance encountered by the needle after conducting human skin tissue. On the contrary, the transient pressure value caused by the needle pulling out operation is negative, that is, the pressure at the next moment is less than that at the last moment, which is caused by the small diameter of the tip. For the wave signal of needle rotation, the transient change of pressure is positive and negative, but it is not easy to determine and identify the classification of a single wave peak, so we selected the 60-minute acupuncture treatment process for the overall training and recognition classification in the follow-up. Therefore, for the input data vector of the DBN model, we choose the value of transient change of pressure.

To ensure the reliability, accuracy, and comprehensiveness of the results, we judge the convergence of the trained DBN neural network model. DBN was learned from the signal data of the previous standard acupuncture practices of professional physicians. A total of 2115 sets of needle insertion, extraction, and rotation data were obtained

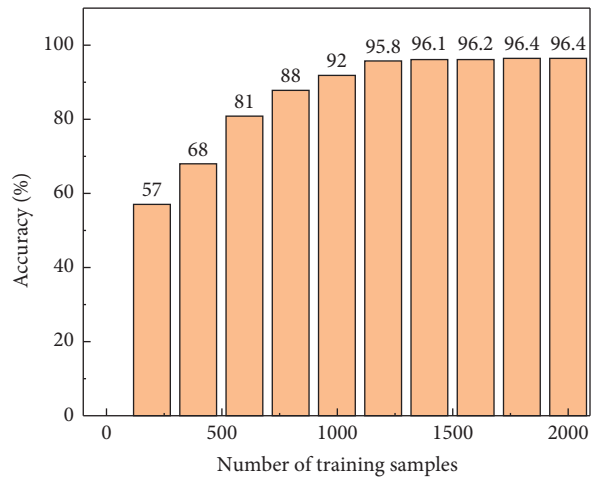


Figure 5: Relationship between identification accuracy and sample number of DBN neural network model.

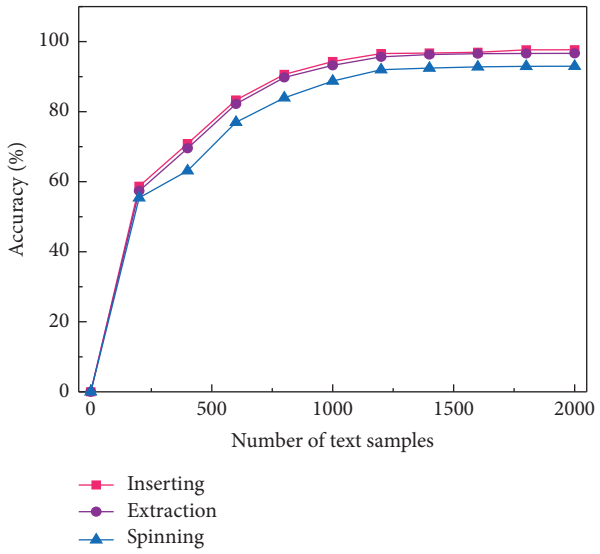


Figure 6: The trend of DBN-based feature recognition rate as a function of the number of tests.

for training. There were 907 mixed tests of needle insertion, extraction, and rotation. Figure 5 shows the trend of feature recognition accuracy of the trained DBN neural network model with the variable of several training samples. It was observed that DBN achieved relatively accurate recognition after learning 1000 training sample species, and its recognition rate reached 92.1%. Then the rising trend of accuracy becomes slow, which indicates that the input of the DBN model and the selection of the input quantity of feature recognition are reasonable. After learning all the training samples, DBN's feature recognition accuracy of the mixed acupuncture process of the test samples reached 96.4%. The reliability of targeted acupuncture therapy has been established to assist physicians.

Further, Figure 6 shows the trend of the feature recognition accuracy of the DBN neural network model for the

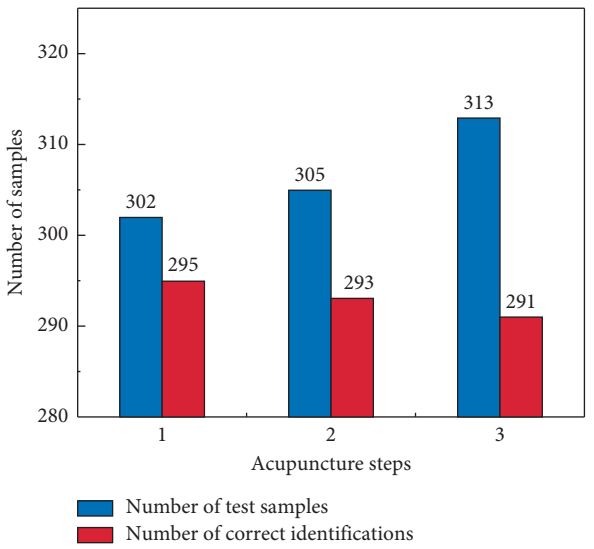


Figure 7: Total sample number and correct sample number of acupuncture steps identified by DBN neural network model.

three steps of insertion, extraction, and rotation as a function of the number of tests. It can be seen that consistent with the accuracy of sample training, the accuracy of test training also increases with the increase of the number of tests, and the overall accuracy is close to that of sample training. When the number of pieces of training reaches 1000, the three-step accuracy rates are all close to 90% accuracy, and then the accuracy rate rises slowly. After reaching 2000 training times, the recognition accuracy of needle insertion and extraction steps reached 97.68% and 96.66%, respectively, and the recognition accuracy of needle rotation therapy was 92.97%.

Subsequently, we made statistics on the accuracy of the prediction of test data of acupuncture steps based on experimental data and the DBN neural network model at the end of the meta-analysis comprehensive training.

4. Results

As shown in Figure 7, it was observed that the recognition accuracy of the trained DBN model for needle insertion and extraction operations reached 97.68% and 96.66%, respectively. The accuracy of DBN in identifying the needle rotation treatment process was 92.97%. The results are similar to the number of training samples and test samples for the three steps of acupuncture. Therefore, we can observe that the trained DBN neural network has excellent recognition accuracy between needle insertion and withdrawal with relatively regular pressure signals and large changes in instantaneous pressure values. For the needle rotation action with more changes in the amplitude of the pressure signal, lower amplitude, and more pulse signals, the recognition accuracy of the DBN model is slightly reduced, which is caused by the complexity of the signal. Therefore, we will conduct further research on the distribution position of the pressure sensor and the noise error generated in the future to reduce the influence of the noise data of the input signal itself on the DBN model.

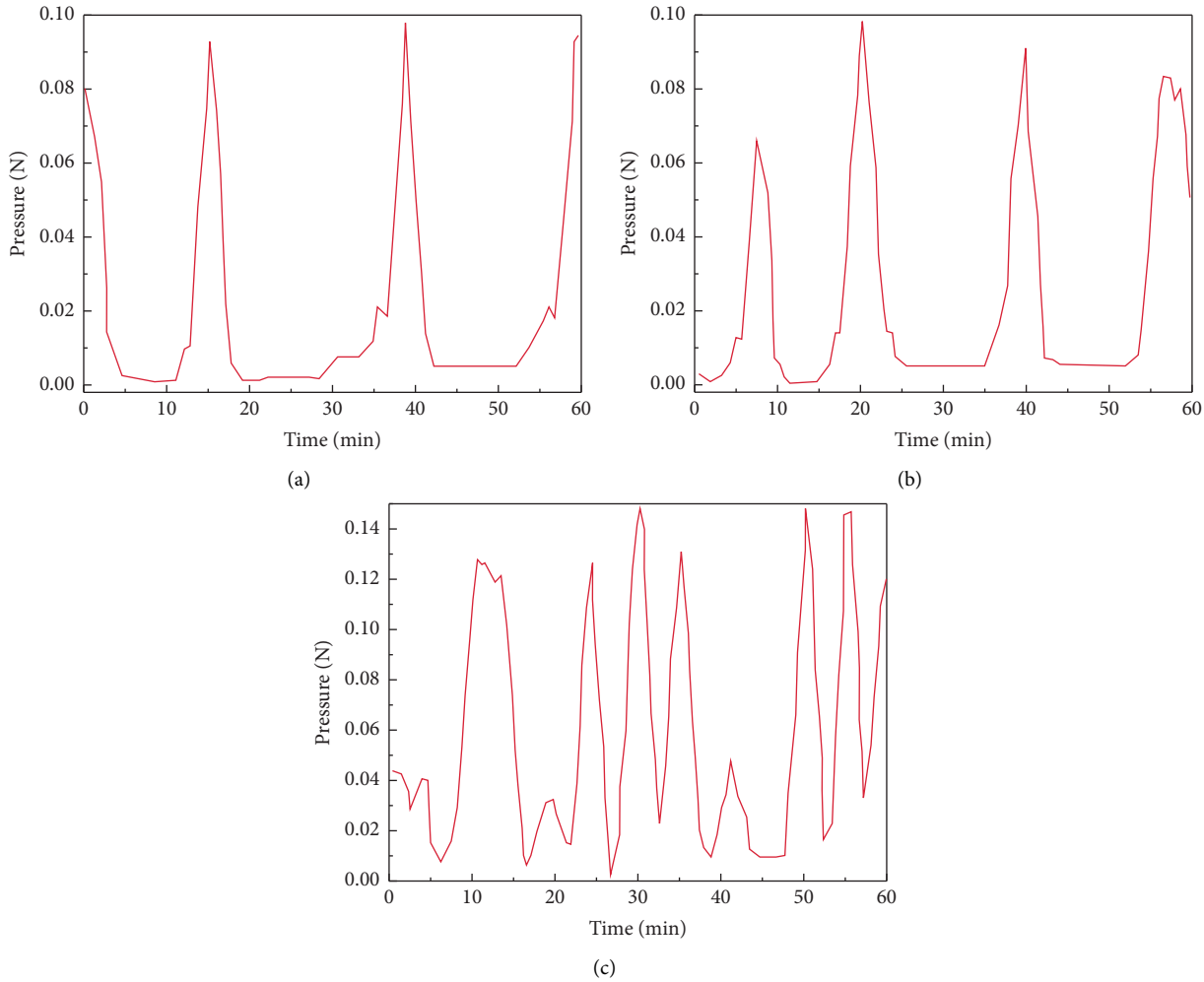


Figure 8: Clinical tests of DBN neural network model to identify acupuncture steps: (a) inserting; (b) extraction; (c) spinning.

Finally, to highlight the help of a trained DBN neural network model for conventional acupuncture treatment, professional physicians randomly performed three typical procedures in clinical practice and identified them using the DBN model. Figure 8 shows the pressure inputs for three typical operations, which DBN identifies as needle insertion, pull out, and rotation operations. In addition, we have built in the appropriate range of the needle insertion and rotation depth, to detect and alarm the optimal needle treatment depth, and to help acupuncturists better accurately targeted treatment for patients.

5. Conclusion

Acupuncture and massage are one of the oldest medical means in China, with a long history of development and an outstanding position in China's medical field. However, in the actual acupuncture and massage treatment, there are differences in the strength, time of stimulation, and depth of acupuncture in the acupuncture process, as well as in the subsequent needle pulling process. Aiming at the basic acupuncture process in acupuncture and massage, DBN neural network model was established based on a large amount of information collected from the meta-analysis database, and the acupuncture techniques and steps of professional physicians were deeply learned. In addition, we have built in the DBN neural network model the appropriate interval range for the depth of needle insertion and rotation, to detect and alarm the optimal needle treatment depth, and to help acupuncturists better accurately target patients for treatment. This is expected to provide powerful help for more efficient clinical diagnosis and treatment of patients by acupuncture and massage physicians in the future.

(1) This article establishes a comprehensive and effective input database based on the evaluation of therapeutic effects of acupuncture and massage by a large number of scholars and researchers collected from the meta-analysis scheme established earlier. At the same time, we constructed a human skeleton model to simulate the actual human body's stress and the clinical effect of acupuncture and massage. The two parts of data are the learning and training input data of the DBN neural network model.

(2) We performed a complete treatment demonstration on the arm model using standard acupuncture

treatment courses. The steps of the cadaver include needle insertion, extraction, and needle rotation. It was observed that the recognition rate of DBN reached 92.1% after learning 1000 training samples. After learning all the training samples, DBN's feature recognition accuracy of the mixed acupuncture process of the test samples reached 96.4%.

(3) The accuracy of the prediction of test data of acupuncture steps was statistically analyzed based on the DBN neural network model after the comprehensive training of experimental data and meta-analysis. It was observed that the recognition accuracy of the trained DBN model for needle insertion and extraction was 97.68% and 96.66%, respectively. The accuracy of DBN in identifying the needle rotation treatment process was 92.97%.

Because of the complexity of the signal, the recognition accuracy of DBN for needle rotation is slightly reduced. Therefore, we will further study the distribution position of the pressure sensor and the noise error generated in the future, to reduce the impact of the noise data of the input signal itself on the DBN model. However, the application of the DBN neural network model in this field is expected to provide powerful help for acupuncture and massage physicians to carry out more efficient clinical diagnoses and treatment for patients in the future.

Data Availability

The dataset can be accessed upon request.

Disclosure

Conflicts of Interest

The authors declared that there are no conflicts of interest.

References

[1] T. Xie, H. Jiang, and C. Zhang, "Clinical observation of pediatric Tuina plus oral Chinese medication for pediatric anorexia due to spleen failing in transportation," *Journal of Acupuncture and Tuina Science*, vol. 20, no. 2, pp. 119–125, 2022.

[2] L. Wang, P. Jin-Lin, Q. Fu-Qiang, and C. Wen-Ming, "Clinical randomized controlled study of acupuncture treatment on children with autism spectrum disorder (ASD): A Systematic Review and Meta-Analysis," *Evidence-based Complementary and Alternative Medicine*, vol. 2021, Article ID 5549849, 2021.

[3] Z. Bo-Wen and C. Ming-Ming, "Nie-pinching the spine, puncturing Sifeng (EX-UE 10) plus Chinese herbs for pediatric anorexia due to dysfunction of spleen in transportation: a randomized controlled trial," *Journal of Acupuncture and Tuina Science*, vol. 19, no. 2, pp. 104–109, 2021.

[4] L. I. Jin-Yan and L. Yang, "Effect of acupunctural therapy combined with acupoint catgut embedding in the treatment of sciatica," *China Modern Medicine*, 2020.

[5] K. Zarabian, A. Wannon, M. Chin, and M. Kogan, "The intersection between integrative medicine and neuropathic pain: a case report," *Explore*, vol. 18, no. 2, pp. 165–169, 2022.

[6] N. Ijaz MSc PhD, S. Welsh PhD, and H. Boon BScPhm PhD, "A survey of acupuncture-providing registered massage therapists in ontario, Canada: motivation, training, and practice characteristics," *International Journal of Therapeutic Massage & Bodywork Research Education & Practice*, vol. 13, no. 3, pp. 18–29, 2020.

[7] Z. Yu and L. Luo, "Clinical effect of acupuncture and massage on cervical spondylosis of vertebral artery type. in 2020 2nd World Congress on Chemistry," *Biotechnology and Medicine*, 2020.

[8] C. Feng, S. Zhou, Y. Qu et al., "Overview of artificial intelligence applications in Chinese medicine therapy," *Evidence-based Complementary and Alternative Medicine*, vol. 2021, pp. 1–6, Article ID 6678958, 2021.

[9] Y. Zhang, Y. Li, and C. Guo, "Acupuncture treating heart diseases based on sixty JiaZi and eight palaces
—mathematical reasoning of treatment principle based on Yin Yang Wu xing theory in traditional Chinese medicine (V)," *Chinese Medicine*, vol. 12, no. 03, pp. 118–184, 2021.

[10] W. W. Tao, H. Jiang, X. M. Tao, P. Jiang, L. Y. Sha, and X. C. Sun, "Effects of acupuncture, tuina, tai chi, qigong, and traditional Chinese medicine five-element music therapy on symptom management and quality of life for cancer patients: a meta-analysis," *Journal of Pain and Symptom Management*, vol. 51, no. 4, pp. 728–747, 2016.

[11] X. Zhan, "Rehabilitation efficacy of acupuncture plus massage on spastic paralysis after stroke," *Clinical Journal of Chinese Medicine*, 2019.

[12] Z. Mao, "Physiotherapy and B.L. Hospital, observation on the clinical curative effect of massage manipulation combined with acupuncture in the treatment of senile periarthritis of shoulder," *Guangming Journal of Chinese Medicine*, 2017.

[13] L. Jia, X. Mei, B. Wang et al., *Observation on the therapeutic effect of acupuncture points of traditional Chinese medicine on vocal diseases*, Chinese Medical Digest(Otorhinolaryngology), China Hong Kong, 2018.

[14] L. Haiming and L. Jianying, "Traditional acupuncture and massage for physical fitness recovery in professional basketball players' high intensity training," *Agro Food Industry Hi-Tech*, vol. 28, no. 1, pp. 2790–2794, 2017.

[15] H. L. Zhao, H. S. Liu, Y. B. Luo, S. S. Xuan, F.-H. Yu, and Y. H. Song, *Traditional Massage Combined with Conventional Acupuncture for Pseudomyopia in Teenagers*, Chinese Manipulation & Rehabilitation Medicine, China Hong Kong, 2018.

[16] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Computers in Industry*, vol. 131, Article ID 103498, 2021.

[17] M. Lngkvist and A. Loutfi, "A Deep Learning Approach with an Attention Mechanism for Automatic Sleep Stage Classification," 2018, https://arxiv.org/abs/1805.05036.

[18] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 566–576, 2017.

[19] B. Lca and B. Hxa, "Deep neural network for semi-automatic classification of term and preterm uterine recordings," *Artificial Intelligence in Medicine*, vol. 105, 2020.

[20] C. Fang, Y. Tao, J. Wang et al., "Mapping relation of leakage currents of polluted insulators and discharge arc area," *Frontiers in Energy Research*, vol. 9, 2021.

[21] G. Xu, H. Lei, L. Huang et al., "The dose-effect association between acupuncture sessions and its effects on major depressive disorder: a meta-regression of randomized controlled trials," *Journal of Affective Disorders*, vol. 310, pp. 318–327, 2022.

[22] J. F. Niu, X. Zhao, H. Hu, J. Wang, Y. Liu, and D. Lu, "Should acupuncture, biofeedback, massage, Qi gong, relaxation therapy, device-guided breathing, yoga and tai chi be used to reduce blood pressure?: recommendations based on high-quality systematic reviews," *Complementary Therapies in Medicine*, vol. 42, pp. 322–331, 2019.

[23] S. Belciug, "Learning deep neural networks' architectures using differential evolution. Case study: medical imaging processing," *Computers in Biology and Medicine*, vol. 146, Article ID 105623, 2022.

[24] H. Liu, J. Xu, Y. Wu, Q. Guo, B. Ibragimov, and L. Xing, "Learning deconvolutional deep neural network for high resolution medical image reconstruction," *Information Sciences*, vol. 468, pp. 142–154, 2018.

[25] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, "Automated detection of leukemia by pretrained deep neural networks and transfer learning: a comparison," *Medical Engineering & Physics*, vol. 98, pp. 8–19, 2021.

[26] L. Silvy, C. Chapelle, J. C. Trone et al., "Early detection of the existence or absence of the treatment effect: a cumulative meta-analysis," *Journal of Clinical Epidemiology*, vol. 124, pp. 24–33, 2020.

[27] A. H. Katsanos, V. A. Lioutas, A. Charidimou et al., "Statin treatment and cerebral microbleeds: a systematic review and meta-analysis," *Journal of the Neurological Sciences*, vol. 420, Article ID 117224, 2021.

[28] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to biconvex optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 8, pp. 2503–2514, 2019.

[29] K. Raza and N. K. Singh, "A tour of unsupervised deep learning for medical image analysis," *Current Medical Imaging Reviews*, vol. 17, no. 9, 2021.

[30] Z. Wang, Y. Liu, D. He, and S. Chan, "Intrusion Detection Methods Based on Integrated Deep Learning Model," *Computers & Security*, vol. 103, Article ID 102177, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3713279, 10 pages
https://doi.org/10.1155/2022/3713279

Research Article

Visual Monitoring Technology for Substation Vulnerable High-Voltage Electrical Equipment Based on ISSA-LSTM Deep Learning Model Coupling Video Overlay Algorithm

Shifeng Wang,[1] **Xueyong Ding**,[1] **and Qingji Tan**[2]

[1]*School of Science and Technology, University of Sanya, Sanya 572099, China*
[2]*School of Mechanical Engineering, Heilongjiang Agricultural Reclamation Vocational College, Harbin 150025, China*

Correspondence should be addressed to Shifeng Wang; shifengwang@sanyau.edu.cn

Received 21 June 2022; Revised 30 July 2022; Accepted 9 August 2022; Published 26 August 2022

Academic Editor: Wenming Cao

To enhance the visualization effect of substation high-voltage electrical equipment vulnerability, this study proposes an ISSA-LSTM coupled video overlay algorithm-based substation high-voltage electrical equipment vulnerability visualization and monitoring model. Using the improved α blending algorithm combined with the inverse sampling of video background color, overlaying visible video as well as infrared video, using the improved adaptive weighted two-dimensional principal component analysis (W2DPCA) to fuse the base layer, selecting the detail layer as the final detail layer, obtaining the final fusion frame, and realizing the visualization and monitoring of substation high-voltage electrical equipment vulnerability, and introducing the improved sparrow search algorithm (ISSA) to establish long and short-term memory network prediction model to reduce the prediction error and improve the monitoring accuracy rate. The experimental results show that the monitoring frames obtained by this method can reflect rich details of substation high-voltage electrical equipment, and the texture color and equipment edge contrast are enhanced to facilitate accurate determination of substation high-voltage electrical equipment vulnerability, and the prediction accuracy of ISSA-LSTM model is as high as 99.85%.

1. Introduction

Substation contains many high-voltage electrical equipment such as transformers, circuit breakers, and voltage transformers, which are important components of power system supply and distribution [1]. When a substation encounters a severe vibration situation such as an earthquake, damage to high-voltage electrical equipment will cause the grid to lose its power supply and distribution functions, and power outage conditions enhance the difficulty of resettling victims after a disaster [2, 3]. Substation high-voltage electrical equipment susceptibility can be applied to equipment earthquake risk assessment, and the equipment status determines the operation of substation transmission and distribution systems [4, 5]. Applying visual monitoring technology to the study of substation high-voltage electrical equipment susceptibility can effectively enhance the overall monitoring needs of electrical systems.

The research methods of seismic capacity and vulnerability of high-voltage electrical equipment in substations mainly include theoretical analysis and simulation calculation methods, shaking table test methods, and statistical analysis of earthquake damage, etc. From the 1980s to the present, many scholars at home and abroad have conducted a lot of research on the seismic performance of high-voltage electrical equipment through the first two methods [6]. Theoretical calculations and numerical simulations, shaking table experiments and other methods focus more on the analysis of seismic response of electrical equipment, simulation of equipment damage process, and damage state, which are powerful methods and means to analyze and study the seismic capacity of equipment, damage mechanism, and develop equipment seismic isolation technology [7, 8]. In the 1980s, the Pacific Earthquake Engineering Center (PEER) and Pacific Gas and Electric jointly established the California Substation Equipment Earthquake Damage Performance

Database, which recorded the damage data of electrical equipment in 60 230 kV to 550 kV substations during 12 earthquakes in California, and thus mapped the vulnerability curves of electrical equipment, and was widely used in the postdisaster assessment of power systems. The American Applied Technology Council (ATC) gave earthquake vulnerability curves for various types of lifeline equipment and facilities based on expert experience [9] for use on the US earthquake emergency (FEMA) seismic risk analysis system HAZAS [10]. Previously, more research has been conducted for substation visualization and monitoring, and some researchers have studied the design of remote video monitoring system for substations based on self-assembling networks [11]. Designing remote video monitoring system for self-assembled substations, which can remotely monitor the operation status of the equipment in the substation; some researchers proposed to study the one-touch programmed control technology based on video integration and intelligent analysis [12], where video integration technology and intelligent analysis technology are applied to programmed control power systems to improve the control performance of many devices in power systems. Some scholars study intelligent analysis of substation video based on image processing [13] and apply image processing technology to intelligent analysis of substation video to improve the performance of substation equipment status analysis, and all the above studies apply visualization technology to substation equipment monitoring, and all achieve certain results [14, 15]. It can be seen that image fusion techniques are mostly used in relevant visualization and monitoring studies, and the existing image fusion techniques include multisensor information fusion techniques, multiple gray-level image fusion techniques, wavelet transform-based image fusion techniques, etc., but jagged and dark-edge situations occur when the abovementioned image fusion techniques are applied to monitoring the vulnerability of high-voltage electrical equipment in substations [16]. In addition, researchers have also found that visual monitoring alone has some uncertainty and false positives, and there is a strong need for an algorithmic optimization approach to exploit the temporal correlation and reduce feature redundancy of visual monitoring data [17].

In response to the above problems, this study proposes a model based on dimension fusion optimization and long short-term memory (LSTM) [18]. In view of the defect that the input layer weights and hidden layer biases of the network model need manual experience tuning, an improved sparrow is introduced. The search algorithm (SSA) optimizes the detection model [19] and visually monitors the vulnerability of high-voltage electrical equipment in substations through a coupled video overlay algorithm [20]. Therefore, the visual monitoring of the vulnerability of high-voltage electrical equipment in substations based on the ISSA-LSTM deep learning model coupled with the video overlay algorithm can realize the visual monitoring of high-voltage electrical equipment in substations, improve the texture color, edge contrast and richness of details of electrical equipment, and realize high-voltage electrical equipment in substations. Real-time detection of the vulnerability of electrical equipment.

2. Construction of Predictive Models

2.1. LSTM Network. The LSTM network is a classical network structure in deep learning. The LSTM contributes to model learning at subsequent moments by passing the weight matrix of the implicit layer at different time steps backward in a coefficient-weighted manner through weight parameter conduction [21]. The accumulation of important information and the abandonment of redundant information are achieved through the collaborative work of input, forgetting and output gates with the help of memory units accumulating the weight states of the implicit layer. Long-term memory is achieved by controlling the gradient transformation range through the synergistic work of the memory unit and the gating structure to effectively avoid the problem of gradients disappearing too quickly [22].

2.2. SSA and Its Improvements. SSA is a class of heuristic optimization algorithms that simulates the behavior of sparrows foraging and avoiding predators. In SSA, the population is divided into discoverers and followers, and the discoverers are responsible for searching for food in the population space, while the followers follow the discoverers to search the let-go space [23].

SSA is prone to fall into local optimum during iteration and to solve this problem, an improved sparrow search algorithm (ISSA) is proposed in this paper, and the main improvement parts of ISSA are as follows [24]:

(1) The logistic chaos algorithm is borrowed to optimize the population initialization, and the characteristics of chaotic pseudorandomness and ergodicity are used to achieve a better initial global search. In this paper, the pseudorandom sequence is generated with the help of logistic mapping, and the strategy is mathematically formulated as follows:

$$Z\colon a_{n+1} = \mu a_n(1 - a_n), \tag{1}$$

where Z is a chaotic variable, and u a is a control parameter. When an initial value a_0 is assigned to the chaotic variable, a set of chaotic initial variables can be obtained by iterating through the logic equation with the help of linear mapping, and the linear mapping scheme is as follows:

$$Z \longrightarrow X\colon X = a + (b - a)Z. \tag{2}$$

(2) In order to better achieve global optimization at the beginning of the iteration and local convergence at the end, this paper proposes an adaptive alert value strategy, which is described as follows:

$$w = w_{\min} + (w_{\max} - w_{\min}) \times \tan\left(\frac{k}{M}\right) \times \frac{\pi}{4}. \tag{3}$$

When the current number k of iterations is small, w is close to $w_{\min}$ to ensure the global let-down capability of the intelligent algorithm, and as the number of iterations increases, w increases in a nonlinear manner to ensure better local let-down convergence in the later iterations, thus allowing the algorithm to flexibly adjust the global let-down and local search capability, that is, M is the total number of iterations.

(3) The adaptive mutation factor is introduced by borrowing the idea of mutation from genetic algorithm. After each iteration is completed, the population has a certain chance to mutate, and the mathematical formulation of the strategy is as follows:

$$p = 0.5 - \frac{k}{2M}, \tag{4}$$

where p is the variation factor, and the probability of population variation decreases as the number of iterative generations increases.

(4) The original movement method is the main reason for standard SSA to fall into local optimum when alerting occurs during reconnaissance. In this paper, we choose a new sparrow change method with the following mathematical formulation:

$$x_{i,d}^{t+1} = \begin{cases} x_{i,d}^{t} + \beta \cdot \left(x_{i,d}^{t} - xb_{i,d}^{t}\right), & f_i \neq f_g, \\ x_{i,d}^{t} + \beta \cdot \left(xw_{i,d}^{t} - xb_{i,d}^{t}\right), & f_i = f_g. \end{cases} \tag{5}$$

β is the correlation coefficient, and f_i and f_g are the parameter values of different forms. When the population produces warning behavior, the optimal sparrow will flee to a random position between the optimal and worst positions, and the remaining sparrows will flee to a random position between themselves and the worst position.

2.3. Video Overlay Algorithm Design. The captured visible substation video as well as the infrared substation video are superimposed using an improved α-blending algorithm combined with background color inverse walk [25]. The so-called "α-blending" algorithm is used to blend the source and target pixels using α-blending vector values to create a sense of transparency in 3D objects. The α-blending algorithm is introduced in this study to give a three-dimensional sense of high-voltage electrical equipment [26]. The α-blending algorithm weighted summation of video near and far views based on a fixed ratio is given by the following equation:

$$I_a = \alpha I_l + (1 - \alpha) I_r. \tag{6}$$

In (6), α and $(1 - \alpha)$ denote the near-field weights and the corresponding far-field weights in the video, respectively; I_l and I_r denote the near-field pixel points and the corresponding far-field pixel colors, respectively, I_a denotes the output value after superimposing the video using the α mixing algorithm.

Based on the grayscale, coordinates and color of the pixel points obtained by the pixel desampling algorithm, the desampling equation is implemented for the corresponding locations in the video far field as follows:

$$I_b = I_r + (I_e - I_r) J_g. \tag{7}$$

Equation (7), I_b that the body of the video screen and the distant field on the antisample pixel color; I_e and J_g that the distant field pixels are located in the same position in the near field corresponding pixel color and based on the corresponding antisample to obtain the near field pixel gray level.

The new near-field screen is obtained through (7), and the new screen is based on the far-field basis to implement antisampling to avoid black edges and jagged boundaries between the far-field and near-field. The acquired near-field and far-field can be well integrated, but the background color in the original screen and the far-field fusion leads to the alpha channel cannot be mixed using the background color detection, the transformation formula (7) is as follows:

$$I_b = \begin{cases} I_l, I_l = I_k, \\ I_r + (I_e - I_r) J_g, I_l \neq I_k. \end{cases} \tag{8}$$

Formula (8), I_k is the original video screen background color.

Using the above process to obtain a mix of channel α, the inverse walk-through on the local video screen of the telepresence to achieve the mix formula is as follows:

$$I_{out} = \begin{cases} I_r, & I_b = I_k, \\ \alpha I_b + (1 - \alpha) I_r, & I_b \neq I_k. \end{cases} \tag{9}$$

In (9), indicates the final video overlay output value.

Based on the background color to determine the channel mix within the channel, the implementation of inverse sampling of the distant screen contains the original background color and the new pixel points, the new background color is also presented in the video screen.

The new background color within the channel discriminant may not be discriminated to be located within the image element, which can be transformed into (9) as follows:

$$I_{\text{out}} = \begin{cases} I_r, & I_l = I_k, \\ \alpha I_l + \left(1 - \alpha J_g\right) I_r, & I_l \neq I_k. \end{cases} \tag{10}$$

In (10), I_{out} indicates the final video overlay result of the channel mix.

Analysis of the above process shows that when the algorithm is used to superimpose the video, the blending weights are transformed from fixed values to variable blending parameters, and the corresponding far-field weights of the variable blending parameters can be adjusted according to the gray level of the near-field pixel points at each position, and the far-field within the blending weights are smaller when the near-field gray level is higher, and the far-field within the blending weights are larger when the

near-field gray level is lower, resulting in the following formula:

$$I_{\text{out}} = \begin{cases} I_r, & I_l = I_k, \\ I_a + \left(1 + J_g\right)\alpha I_r, & I_l \neq I_k, \end{cases}$$
$$\eta = \left(1 - J_g\right)\alpha, \tag{11}$$
$$I_{\text{out}} = \begin{cases} I_r, & I_l = I_k, \\ I_a + \eta I_r, & I_l \neq I_k. \end{cases}$$

Equation (11) shows that the algorithm adds the correction function to the traditional n-blending algorithm and sets the blending weights to compensate for the correction value based on the near-field gray level so that the blended video does not have jaggedness and dark edges after superposition.

When the near-field blending weight is large and the gray level of the pixel is fixed, and the gray level of the blended pixels involved in the superposition is low, the far-field of the superposition will be covered by the near-field, and the larger the near-field blending weight is, the more obvious the jaggedness and dark edges will be, so it is necessary to increase the correction amount to compensate for this. When the gray level of the pixel points in the near-field superposition is too low and the blending weight is fixed, the jaggedness and dark edges exist at the boundary of the superposition element when the far-field blending weight is not the maximum value, and the gray level of the near-field pixel points is negatively correlated with the correction amount.

The video overlay algorithm is applied to the established substation model. This is when the video realizes the visual monitoring of the vulnerability of high-voltage electrical equipment in the substation, and the pixel information of the close-view screen existing in RGB space does not have gray-level information. RGB space data are displayed using YUV space gray-level information with the following conversion formula:

$$Y = 0.299R + 0.587G + 0.114B, \tag{12}$$

where R and B are both parameters of the equation.

2.4. Fusion Algorithm Based on Bootstrap Filter and W2DPCA

2.4.1. Visual Surveillance Video Frame Decomposition with Bootstrap Filters. The fusion of visible video and infrared video after video overlay to obtain higher-quality visual surveillance results so that the visual surveillance video is not affected by the surrounding environment, the adaptive weighted two-dimensional principal component analysis (hereinafter referred to as adaptive W2DPCA algorithm) is combined with the bootstrap filter to fuse visible video as well as infrared video, and this fusion method has the advantage of low-computational complexity and is able to obtain the visual surveillance results by the frame-by-frame fusion. This fusion method has the advantage of low-computational complexity and is able to obtain visual surveillance results by fusing visible and infrared video on a frame-by-frame basis [27]. The fusion method mainly consists of dividing the visual monitoring layers by using the guidance filter, obtaining the base and detail layers of the visible and infrared frames monitoring results, selecting different planning methods to fuse the base and detail layers obtained by video layering, and finally combining the fused base and detail layers to obtain the final substation high-voltage electrical equipment vulnerability visual monitoring results:

$$I_B^i = G_{r,\xi}\left(I^i, I^i\right),$$
$$I_B^v = G_{r,\xi}\left(I^v, I^v\right). \tag{13}$$

In the above equations, G and r denote the bootstrap filter function and the filter radius, respectively; I^i and I^v denote the source infrared frame and the source visible frame, respectively; ξ and I_B^v denote the regularization parameter and the visible frame base layer, respectively; I_B^i denotes the infrared frame base layer.

The visible frame and the infrared frame detail layer are obtained using the difference between the visible base layer and the infrared base layer obtained from the source visible frame and the source infrared frame [14]. The obtained infrared frames and visible detail layers are given by

$$I_D^i = I^i - I_B^i,$$
$$I_D^v = I^v - I_B^v. \tag{14}$$

The base and the detail layers obtained by the above process retain the large variance region and texture information within the source video, respectively, and the target within the infrared image exists within the video base layer.

2.4.2. W2DPCA Fusion Base Layer. The quality of video frames directly affects the final result of image fusion [28], and only infrared thermal images need to be added to high-quality visible frames to obtain the best fusion results. The process of fusing visible frames with IR frames using the adaptive W2DPCA algorithm is as follows:

The W2DPCA algorithm is first implemented on the infrared frame and visible frame base layer, and the set of infrared frames and visible frames is represented by matrix P. The matrix P covariance matrix G is obtained as follows:

$$G = \frac{1}{2}\sum_{i=1}^{2}\left(P^i - \mu\right)^T\left(P^i - \mu\right). \tag{15}$$

In Equation (19), (P^1) denotes the infrared frame, P^2 and u denote the visible frame as well as the average of visible and infrared frames, respectively.

The eigenvectors of the covariance matrix G are arranged in descending order according to the size of the eigenvalue λ. The eigenvectors corresponding to the largest eigenvalue are placed on the leftmost side of the matrix, using U to represent the extracted Eigenmatrix, and the Eigenmatrix of size $n \times d$ is composed of the first d eigenvectors, and the matrix expression is as follows:

$$U = \left(U_1, U_2, \ldots, U_d\right). \tag{16}$$

The equations for obtaining the feature image Q_B^i and Q_B^v are as follows:

$$\begin{aligned} Q_B^i &= I_B^i \times U = \left[Q_{B_1}^i, Q_{B_2}^i, \ldots, Q_{B_d}^i\right], \\ Q_B^v &= I_B^v \times U = \left[Q_{B_1}^v, Q_{B_2}^v, \ldots, Q_{B_d}^v\right]. \end{aligned} \tag{17}$$

The feature image consists of the infrared frame and the first d principal components of the photon frame. The different frame weights change as the pixel size changes. The visible light frame and infrared frame weights are calculated by using the feature image with smaller size, which can effectively reduce the calculation amount.

The fusion process needs to detect whether there is a difference in contrast between the edge regions of the visible light frame and other regions [29]. The quality of visible light is reduced in poor environments, but still contains useful texture information. Fully considering the frame quality to obtain adaptive visible light weights in the fusion process, the visible light frame weight formula based on regional variance is as follows:

$$W^v = \delta^2 = \frac{1}{n \times d} \sum_{i=1}^{n} \sum_{j=1}^{d} \left(Q_B^v(i, j) - \overline{Q_B^v}\right)^2. \tag{18}$$

In formula (18), δ represents the standard deviation of the source image. Higher and lower pixel values in the infrared frame indicate hotter and cooler regions, respectively. The grayscale values of all pixels in the infrared frame are significantly lower in the low-temperature case than in the high-temperature case, and the grayscale values of hot objects in the image are significantly higher than those of other pixels. The fusion process should focus on the infrared thermal objects, using the zero-mean operation to obtain the infrared weights Wi equation as follows:

$$W^i = \delta^2 = \frac{1}{n \times d} \sum_{i=1}^{n} \sum_{j=1}^{d} \left(Q_B^i(i, j) - \overline{Q_B^i}\right)^2. \tag{19}$$

It can be seen by (19) that the larger weights will be assigned to the higher pixel fraction.

Using the weighted average method to obtain the fused feature image Q_B of infrared frames as well as visible frames, the fused feature map can retain all the detail information of visible frames by this method, and the fused feature map equation is as follows:

$$Q_B = \frac{W^v Q_B^v + W^i Q_B^i + Q_B^v}{W^v + W^i + 1}. \tag{20}$$

The following equation is used to approximate the reconstruction of the fusion base layer equation as follows:

$$I_B = Q_B \times U^T. \tag{21}$$

2.4.3. Detail Fusion Layer. The real texture information within the image may destroy the large amount of useless information contained in the IR frame detail layer, and the IR frame image detail layer may be discarded during the fusion process, and the detail layer fusion equation is as follows:

$$I_D = I_D^v. \tag{22}$$

2.4.4. Final Fused Video Frames. The final fusion formula for the fused base layer and the fused detail layer is as follows:

$$I_F = I_B + I_D. \tag{23}$$

The final fusion frame obtained by (23) can effectively represent the thermal target and texture information contained in the infrared frame as well as the visible frame.

3. Modeling Method

3.1. Data Preprocessing. To prevent the loss of low order of magnitude features, this paper uses a normalization strategy to preprocess the original features, and the original set of features is normalized to the data segment of [0, 1], and the normalization mathematics is expressed as follows:

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{24}$$

where $x_{\max}$ is the maximum value in the sample features and $x_{\min}$ is the minimum difference in the sample features.

3.2. Model Training. A total of 70% of the data were selected as training model and 30% as validation model. PCA is used to process the input data to achieve data dimension reduction, and the input format of data is processed by sliding window to train the LSTM detection model.

3.3. Intelligent Algorithm Optimization and Its Improvement. Compared with other swarm intelligence optimization algorithms, SSA has high search accuracy, fast convergence, good stability, and strong robustness. However, when SSA search approaches global optimum, the population diversity will decrease and fall into local optimum. In this paper, SSA is improved and ISSA algorithm is proposed. According to the fan data modeling requirements, the search dimension was set as 3D, the population size was 20, and the number of iterations was 1000. Different test functions were selected for testing, and the results were shown in Table 1.

3.4. Prediction Accuracy. PrecisionRecall curve derived from the evaluation of correlation in information retrieval was used to reflect the accuracy of fault decision of icing model, that is, recall rate and accuracy:

$$\begin{aligned} \text{recall} &= \frac{TP}{TP + FN}, \\ \text{precision} &= \frac{TP}{TP + FP}, \\ F_1 &= 2 \times \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \end{aligned} \tag{25}$$

In the formula, T/F represents whether the predicted result is consistent with the actual situation: that is, if the real situation is a positive sample (P) and the prediction is a positive sample (P), it is T; if the real situation is a negative sample (N), the prediction is Negative samples (N), it is T; if the true case is P, the prediction is N, it is F; if the true case is N, the prediction is P, it is F. That is, TP is a positive sample that is actually a positive sample; FN is a negative sample that is actually a positive sample; FP is a positive sample that is actually a negative sample. Recall represents recall rate, precision represents precision, and fraction represents harmonic mean of precision rate and recall rate.

4. Results' Analysis

4.1. Video Overlay Algorithm Analysis. In order to verify the effectiveness of the visualization and monitoring of substation high-voltage electrical equipment susceptibility based on video overlay algorithm for the study of substation high-voltage electrical equipment susceptibility, a power system composed of a national grid is selected as an example analysis object, which is characterized by a large amount of industrial electricity consumption and high population density. A total of 29 substations above 110 kV in the region were selected as the sample for the statistical analysis of high-voltage electrical equipment vulnerability study: 2 500 kV substations, 8 220 kV substations, and 19 110 kV substations. Design a set of substation high-voltage electrical equipment vulnerability visualization monitoring system, the system uses infrared sensors and visible sensors of online monitoring devices to obtain substation high-voltage electrical equipment image source, using the upper computer to run the system using the method of this paper to process the collected images, through the C language using VS compiler platform programming this paper software, the system has infrared image and visible image fusion, infrared and visible imaging information complementary, substation high-voltage electrical equipment vulnerability analysis, manual operation, and other functions. This method uses the video overlay method to superimpose the video so that the presented video has no jaggedness; the visual monitoring video has the equipment perspective function, which can set the transparency of each high-voltage electrical equipment in the substation, so as to avoid the inability to accurately analyze the vulnerability of the equipment due to some equipment being obscured. A pair of infrared frames and visible frames are selected from the video of high-voltage electrical equipment in the source substation, and the objective indexes of the image quality before and after the fusion process are shown in Table 1, using this method to superimpose the above two original images.

The index data in Table 2 show that this method can fuse outdoor substation scenes well, and the fused images can show the texture of high-voltage electrical equipment in the visible video and the equipment information in the infrared video well, and the edge performance is more natural, and the visualization effect of the fused video is significantly improved compared with that before the fusion, and the shadow part of high-voltage electrical equipment is enhanced, and the contrast is improved, and the details of the equipment can be strengthened effectively in the fusion result, and the color of the equipment texture and the contrast of the equipment edge are enhanced, and the overall details of the high-voltage electrical equipment in the monitoring video of the substation are richer.

Based on the three-dimensional model of the substation established in this paper, Matlab simulation software is used to simulate a real earthquake disaster, given that the nodes sampled by the substation information are 200, the maximum daily load is 120 kW, the active power of the substation is 42 kW, and the simulated earthquake disaster level is 6.7. Due to the difference between different regions and the distance from the earthquake source, the damage of high-voltage electrical equipment in different regions is different, and the distance between the substation and the earthquake source is used as the standard to divide the studied region into six regions from A to F. Based on the above settings, the damage of high-voltage electrical equipment in each region is shown in Table 3.

The results in Table 3 show that the number of damaged devices decreases as the distance from the source increases, which are consistent with reality and fully verifies the effectiveness of the simulation of high-voltage electrical equipment in the substation. Based on the visualization and monitoring interface, it can be seen that the main damage states of the porcelain column-type equipment in the substation, except for transformers, are fracture damage and crack damage. Equipment fracture damage affects the use of equipment, equipment cracks do not affect the continued use of equipment cracks set to undamaged, the equipment fracture state is classified as complete damage. Transformer damage state in the earthquake is mainly manifested as wheel rail fixing device damage, oil pillow damage, and other component damage, it can be seen that the transformer has a low vulnerability. Transformer damage can be divided into two states: damaged parts and undamaged parts, and transformer parts are undamaged when the damage to the parts is less than one-fifth of the transformer itself and does not affect the normal operation of the transformer; otherwise, the transformer parts are damaged.

4.2. SSA-LSTM Prediction Optimization Analysis. The sliding window width is set to 10, the learning rate of the optimizer is 0.001, the number of neurons in the hidden layer is 50, the step size parameter is 64, the model is trained 200 times, and the Adam optimizer is selected to optimize the model. When the training samples are 20,000, the LSTM model training loss is shown in Figure 1. In Figure 1, the dashed line represents the loss variation of the model validation set, and the solid line represents the loss variation of the training set. Analysis of the curves in the figure shows that the training loss of the LSTM model has converged to around 0.13 at about 105 iterations. The samples after feature fusion processing were selected as training data to construct ISSA-LSTM model and compared with the optimized models of other swarm intelligence optimization algorithms to verify the optimization of ISSA, SSA, PSO, and GWO on

Table 1: Benchmark function results.

Benchmark function	Dimension	Search space	Optimal value	SSA	ISSA
$F_1(x) = \sum_{i=1}^{n} x_i^2$	3	$[-100, 100]^n$	0	4.006×10^{-3}	4.006×10^{-89}
$F_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	3	$[-100, 100]^n$	0	1.049×10^{-1}	6.023×10^{-17}
$F_3(x) = \sum_{i=1}^{n} (\sum_{i=1}^{i} x_j)^2$	3	$[-100, 100]^n$	0	2.891×10^{-1}	1.343×10^{-89}
$F_4(x) = \max_i \{\lvert x_i \rvert, 1 \le i \le n\}$	3	$[-100, 100]^n$	0	1.779×10^{-1}	1.624×10^{-46}
$F_5(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	3	$[-30, 30]^n$	0	3.262×10^{-2}	1.0667×10^{-3}

Table 2: Objective index comparison of image quality before and after fusion.

Objective indicators	Prefusion results	Postfusion result
Entropy	7.55	7.57
Standard deviation	55.39	56.48
Root mean square error	6.63	2.84
Peak signal-to-noise ratio	70.36	87.33
Spatial frequency	11.026	15.51
Average gradient	5.192	6.09

Table 3: Degree of damage to high-voltage electrical equipment.

Divide area	Substation	Total number of high-voltage electrical equipment	Quantity of damaged equipment
A	6	243	197
B	4	185	115
C	6	285	97
D	5	219	76
E	4	197	58
F	4	193	39

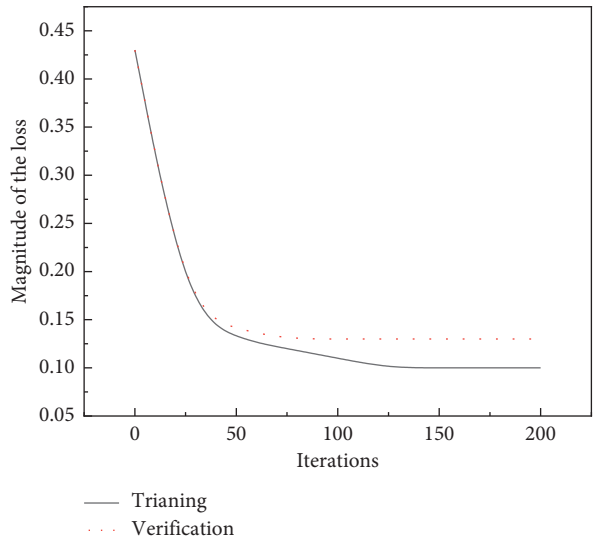


Figure 1: LSTM model training loss.

the detection model, and the model validation set accuracy rates are shown in Table 4.

From Figure 2 and Table 4, it can be seen that the LSTM and its coupled models ISSA-LSTM, SSA-LSTM, PSO-LSTM, and GWO-LSTM are in a stable state after reaching a certain number of training sessions and slowly improve, increasing with the number of training sessions. However, it is found that the prediction effect of LSTM model is

Table 4: Accuracy of swarm intelligence algorithm model.

Training set	LSTM	ISSA-LSTM	SSA-LSTM	PSO-LSTM	GWO-LSTM
1000	93.25	95.24	93.96	93.37	94.31
2500	95.45	97.59	97.11	96.17	96.44
5000	96.64	98.33	97.23	96.93	97.14
10,000	97.50	99.36	98.69	98.33	98.48
15,000	97.74	99.67	99.26	99.07	98.90
20,000	97.77	99.85	99.46	99.16	99.49

significantly lower than its coupled model, and the prediction accuracy of LSTM model is 97.77% after 20,000 training times, while the prediction accuracy of PSO-LSTM model, which is the worst prediction effect among the coupled models, is 99.07, which is significantly higher than that of LSTM model. In addition, the improved SSA model, whose improved model ISSA coupled with LSTM model has the highest prediction accuracy of 99.85% after 20,000 training cycles, which indicates that the LSTM model alone has a certain prediction deficiency, probably due to its own model factors, but coupling it with other model algorithms can significantly improve its prediction accuracy and prediction effect.

In addition, this study takes 20,000 training samples, and the rest parameters are kept consistent for the experiments, and the parameters of the group intelligence algorithm optimization model are shown in Table 5. Compared with SSA, PSO, GWO, and other group intelligence optimization

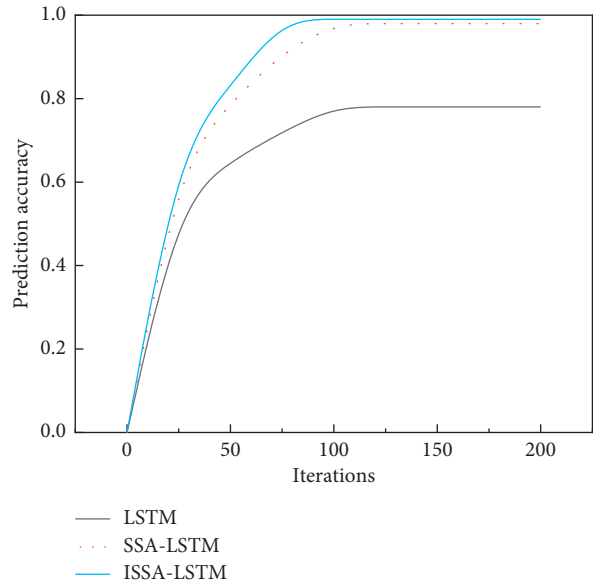


Figure 2: The relationship between the number of iterations and model accuracy.

Table 5: Comparison of optimized parameter results.

Swarm intelligence optimization algorithm	Positive sample judgment accuracy rate (%)	Negative sample judgment accuracy rate (%)	Positive sample recall rate (%)	Negative sample recall rate (%)	Positive sample $F1$ score	Negative class sample $F1$ score
ISSA	99.89	99.81	99.38	99.96	0.996	0.998
SSA	99.58	99.40	98.55	99.84	0.990	0.996
PSO	99.47	99.03	97.63	99.80	0.985	0.994
GWO	99.89	99.34	98.50	99.96	0.991	0.996

algorithms, the parameters of ISSA are better than those of other group intelligence algorithms, and the judgment accuracy of 99.85% can be achieved when the training sample data is 20,000, which fully proves the effectiveness and reliability of ISSA-LSTM model. Meanwhile, the $F1$ scores of both positive and negative samples show a decreasing trend. However, the accuracy and recall rates of ISSA and GWO are extremely close, which indicates that GWO may be able to be used as a potential coupled analysis model to enhance the prediction effect of the LSTM model.

5. Conclusion

The ISSA-LSTM coupled video overlay algorithm is applied to the visual monitoring of high-voltage electrical equipment vulnerability in substations. The video overlay algorithm can avoid the jaggedness and dark edges of the borders of the close view of the overlaid video images, which meet the practical requirements of high-voltage electrical equipment vulnerability monitoring in substations. The researched method has high 3D visualization monitoring performance, strengthens the details of electrical equipment, enhances its texture color and edge contrast, improves the richness of electrical equipment details in the monitoring video, accurately analyzes the vulnerability of high-voltage electrical equipment using substation visualization monitoring, and improves the safe operation of substations. The following conclusions were obtained: (1) the images after the video overlay algorithm can show the texture of high-voltage electrical equipment in the substation in the visible video and the information of the equipment in the infrared video well, and the edge performance is more natural. (2) The prediction effect of the LSTM model is significantly lower than that of its coupled model, and the prediction accuracy of the LSTM model is 97.77% after 20,000 times of training, while as the coupled prediction accuracy of PSO-LSTM model, which has the worst prediction effect, is 99.07, which is significantly higher than that of LSTM model. (3) Compared with SSA, PSO, GWO, and other group intelligence optimization algorithms, the parameters of ISSA are better than other group intelligence algorithms, and the judgment accuracy of 99.85% can be achieved when the training sample data are 20,000.

Data Availability

The dataset can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declared that there are no conflicts of interest.

Acknowledgments

This work was supported by the Hainan Natural Science Foundation of 2023, project name: Research on Optimal Configuration of Intelligent Cooperative Charging Station Based on Competitive Algorithm (SQ2023MSXM0113).

References

[1] B. Yang, J. Tang, Z. Luo et al., "Optimal collaborative expansion planning of integrated cooling and power system for low-latitude distribution networks," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.

[2] Y. Wang, M. Zheng, and Y. Wang, "Analysis of the Design of the Power Supply and Distribution System in the Medical Technology District of Lei Shenshan Hospital," *Electrical Technology of Intelligent Buildings*, 2020.

[3] Z. Hou, X. Sun, Z. Ju, Z. Li, Q. Niu, and S. Liu, "Design of the monitoring system for power distribution area by clean energy intelligent integration terminals," *IOP Conference Series: Earth and Environmental Science*, vol. 804, no. 3, Article ID 032051, 2021.

[4] G. Xuan, W. Wang, Y. Liu, and B. Lang, "The reliability evaluation of power supply distribution system based on distribution automation," *Journal of Physics: Conference Series*, vol. 1585, no. 1, Article ID 012034, 2020.

[5] X. Kong, C. Yong, C. Wang, P. Li, L. Yu, and Y. Chen, "Multi-objective power supply capacity evaluation method for active distribution network in power market environment," *International Journal of Electrical Power & Energy Systems*, vol. 115, Article ID 105467, 2020.

[6] T. Yang, S. Wang, and W. Liu, "Experimental Study on Seismic Performance of High Voltage Electrical Equipment GIS," *World Earthquake Engineering*, vol. 1, pp. 146–155, 2016.

[7] X. Guo, Y. Liu, and Y. Fu, "Analysis of influence of sliding fittings on seismic performance of interconnected high voltage electrical equipment," *Gaoya Dianqi/High Voltage Apparatus*, vol. 54, no. 10, pp. 221–226, 2018.

[8] Z. Liu, Y. Sun, and P. Gao, "Simulation analysis of the seismic performance of electrical equipment by redundant length of flexible conductors in interconnected circuits," *Journal of Physics: Conference Series*, vol. 1639, no. 1, Article ID 012010, 2020.

[9] M. Lian, B. Guan, Q. Cheng, H. Zhang, and M. Su, "Experimental and numerical study of seismic performance of high-strength steel fabricated framed-tube structures with replaceable shear links," *Structures*, vol. 28, pp. 2714–2732, 2020.

[10] M. Gan, Y. Yu, and H. Zhang, "An experimental study on the seismic performance of high-strength composite shear walls," *Frontiers in Materials*, vol. 8, 2021.

[11] H. Jiangtao, "Discussion on the construction of substation security video surveillance system," *IOP Conference Series: Materials Science and Engineering*, vol. 563, no. 3, Article ID 032004, 2019.

[12] Y. Liu and C. Zhang, "The auxiliary system of video surveillance in smart substation," *Journal of Physics: Conference Series*, vol. 2195, no. 1, Article ID 012030, 2022.

[13] Y. Wang, J. Zhang, L. Zhu, S. Zhongwei, and L. Jun, "A Moving Object Detection Scheme Based on Video Surveillance for Smart Substation," in *Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP)*, IEEE, Beijing, China, August 2019.

[14] F. Sun, L. Wei, and J. Fan, "Application of Video Image Recognition Technology in Substation Equipments Monitoring," in *Proceedings of the International Conference on Electrical & Control Engineering*, IEEE, Yichang, China, September 2011.

[15] C. Yang, J. Wu, B. Tao, X. Zhang, L. Li, and H. Qu, "Study of internet of things technology applied in intelligent substation auxiliary system," *IOP Conference Series: Materials Science and Engineering*, vol. 719, no. 1, Article ID 012009, 2020.

[16] R. Zhao, W. Shang, and T. Wang, "State Evaluation of Isolating Switch in Transformer Substation Based on Image Processing," *Computer Measurement & Control*, 2016.

[17] J. Zhao, "Application of image processing technology in substation patrol inspection," *IOP Conference Series: Materials Science and Engineering*, vol. 677, no. 4, Article ID 042060, 2019.

[18] J. Ma, X. Zeng, X. Xue, and R. Deng, "Metro emergency passenger flow prediction on transfer learning and LSTM model," *Applied Sciences*, vol. 12, no. 3, p. 1644, 2022.

[19] R. Dube, C. D. Rais, K. Y. Wang, and S. Tripathi, "Signal stability-based adaptive routing (SSA) for ad hoc mobile networks," *IEEE Personal Communications*, vol. 4, no. 1, pp. 36–45, 1997.

[20] S. K. Boral, D. Agarwal, A. Das, D. Chakraborty, and T. K. Sinha, "A novel video overlay guided enlargement of area of ILM peeled versus inverted flap technique: a long-term study in large macular holes," *European Journal of Ophthalmology*, vol. 31, no. 6, pp. 3277–3283, 2021.

[21] M. Obrubov and S. Kirillova, "Using Lstm Network for Solving the Multidimentional Time Series Forecasting Problem," *National Association of Scientists*, vol. 2, no. 68, pp. 43–48, 2021.

[22] C. Tao, J. Lu, J. Lang, X. Peng, K. Cheng, and S. Duan, "Short-term forecasting of photovoltaic power generation based on feature selection and bias compensation–LSTM network," *Energies*, vol. 14, no. 11, p. 3086, 2021.

[23] P. Kathiroli and S. Kanmani, "Multi-objective sparrow search algorithm for energy efficient clustering and routing in wireless sensor networks," in *Proceedings of the International Conference on Intelligent Cyber Physical Systems*, Jaipur, India, January 2021.

[24] C. Ouyang, D. Zhu, and F. Wang, "Application of improved sparrow search algorithm in SVM optimization," *Journal of Physics: Conference Series*, vol. 1966, no. 1, Article ID 012008, 2021.

[25] T. Darbyshire, *AI3SD Video: Practical Ethics for Data Science and Algorithm Design, AI3SD, PSDS & Patterns Failed it to Nailed it: Getting Data Sharing Right Seminar Series 2020*, 2020.

[26] Q. Sun, A. Taherin, Y. Siatitse, and Z. Yuhao, "Energy-Efficient 360-degree video rendering on FPGA via algorithm-architecture Co-design, FPGA'20," in *Proceedings of the 2020 ACM/SIGDA International Symposium on Field-*

Programmable Gate Arrays, Association for Computing Machinery, New York, NY, USA, February 2020.

[27] Z. M. Zhuang, X. F. Zhong, and W. Xiao, "Face recognition based on WT-W2D2DPCA algorithm," *Journal of Shantou University (Natural Science Edition*, 2012.

[28] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to mixed-integer optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 36–48, 2021.

[29] C. Fang, Y. Tao, J. Wang et al., "Mapping relation of leakage currents of polluted insulators and discharge arc area," *Frontiers in Energy Research*, vol. 9, p. 694, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5953322, 15 pages
https://doi.org/10.1155/2022/5953322

Research Article

Study on Thermodynamic Characteristics and Heat Transfer Method of Uncontrolled Fire in Coal Mine Gangue Mountain Spontaneous Combustion Based on System Dynamics

Guangxing Bai

College of Safety Science and Engineering, Xi'an University of Science and Technology, Shaanxi, Xi'an 710054, China

Correspondence should be addressed to Guangxing Bai; baiguangxing@xust.edu.cn

Received 15 June 2022; Revised 10 July 2022; Accepted 19 July 2022; Published 26 August 2022

Academic Editor: Wenming Cao

Coal fire disaster caused by the spontaneous combustion of coal has always been one of the serious problems that threaten the safety of coal mining. In our study, we first solved the dynamic parameters and mechanical functions of coal gangue using the Achar differential method and Coats-Redfern integral method. Then, based on the flow and heat transfer mechanism of hot rod vapor-liquid two-phase flow, combined with coal spontaneous ignition conditions, influencing factors, coal pile spontaneous combustion temperature field structure distribution, etc., the heat transfer process of the hot rod in the coal pile (coal gangue mountain) was analyzed. Results show the average activation energies of the second stage of different types of coal mine gangue. Upon comparing the characteristic parameters of coal gangue in different regions and ages, it is found that 8# coal gangue has better combustion and burn-out characteristics, and its comprehensive combustion characteristics are second only to those of the coal gangue in Datong and higher than those in Panzhihua, Pingdingshan, and Hancheng. The higher the content of volatile matter and fixed carbon in coal gangue, the lower the ash content, and the better the comprehensive combustion performance of the coal gangue. Under the condition of sufficient oxygen supply combustion, the larger the fuel ratio, the better the burnout performance of the coal gangue. The test of the influence of the hot rod on the temperature field distribution inside the coal pile shows that the maximum cooling rate of a single hot rod to the coal pile during the test period is 33.4°C, and the maximum cooling rate reaches 39.6%. The calculated heat dissipation of the 80 h hot rod is 1.0865, 2.1680, and 3.3649 MJ, respectively.

1. Introduction

Coal fire disasters caused by the spontaneous combustion of coal have always been one of the serious problems threatening the safe mining of coal mines. Coal fire disasters spread all over the world, such as in China, Australia, Poland, Portugal, the United States, India, etc [1]. Coal fires caused by the spontaneous combustion of coal burn a lot of resources every year and emit greenhouse gases and toxic and harmful gases, producing CO, CO_2, CH_4, SO_2, NO_X, etc [2, 3]. The CO_2 produced by coal fires in China accounts for about 0.1%–0.22% of global carbon emissions from fossil fuels. The emissions from these coal fires pollute the atmosphere, destroy the ozone layer, cause the greenhouse effect and global warming, and seriously threaten ecological security. Therefore, controlling the coal fires caused by the spontaneous combustion of coal has become a huge issue facing the world today [4]. The spontaneous combustion of coal gangue is a relatively special combustion system, which has the characteristics of large heat storage and easy reignition. Coal and coal gangue mountains stacked in the open often spontaneously combust, resulting in the wastage of resources, environmental pollution, and serious accidents. Aiming at the key scientific problems in the prevention and control of spontaneous combustion of coal gangue mountain, based on the research idea of "heat transfer and cooling," the prevention and control methods are carried out focusing on the kinetic characteristics of uncontrolled spontaneous combustion and heat transfer process of coal gangue mountain. The scientific management of fire has important theoretical guiding significance.

The research on the spontaneous combustion mechanism of coal began in the 17th century, however, the research on the spontaneous combustion mechanism of the coal gangue only started in the late 19th century [5]. As humans find that the hazard of the spontaneous combustion of coal gangue is becoming more serious and the world has begun to pay attention to the concept of environmental protection, scientists in various professional fields from all over the world have carried out in-depth research on the spontaneous combustion mechanism of the coal gangue since the 21st century [6]. The spontaneous combustion of coal gangue is a complex physical and chemical reaction process, which has been an important subject of human exploration and research for centuries. In the 17th century, people believed that coal gangue was formed because of the reaction of pyrite (FeS2) in the air to generate $FeSO_4$, $Fe_2(SO_4)_3$, $Fe(OH)_3$, SO_2, CO_2, CO, H_2S, and other products. At the same time, heat gradually warmed the coal. Eventually, it led to the self-heating and spontaneous combustion of coal gangue. Hence, the theory of spontaneous combustion caused by pyrite is proposed [7]. In 1870, Rachtan discovered through experiments that coal had oxygen-absorbing properties. Jones (Jones E·R) further confirmed the oxygen absorption of coal in 1945, pointing out that bituminous coal can absorb 0.4 ml/g oxygen when exposed to air at room temperature. In the 1960s, the Fushun Coal Research Institute of China proposed a chromatographic oxygen absorption identification method for the spontaneous combustion tendency of coal. In the 1990s, Australian Chinese scholars Humphreys and Ren et al. proposed to use the average self-heating rate index R70 (self-heating rate index) of the adiabatic heating process of coal between 40°C and 70°C to characterize the spontaneous combustion tendency of coal. After that, the R70 indicator to measure the spontaneous combustion tendency of coal has been widely studied and applied in many countries, such as Australia, New Zealand, and China [8].

In recent years, with the development of science and technology and the continuous enrichment of technical means, many advanced experimental techniques and methods, such as adiabatic heating, temperature programming, the spontaneous combustion of large coal, thermal analysis, in situ spectroscopy and thermal analysis, and quantum chemistry theory have been successively adopted. It is used to study the mechanism of coal spontaneous combustion and reveal the micromechanism and macrodynamic characteristics of the coal spontaneous combustion process [9]. For example, Ren et al. studied the relationship between the size of coal spontaneous combustion and the number of active groups [10], and Beamish et al., based on the quantum chemical theory method, simulated and calculated the active order of coal active groups to reveal the nature of coal spontaneous combustion [11].

The current research contents on the coal gangue mountain control at home and abroad mainly include the following points: research on the stability of coal gangue mountain slopes, research on the prevention and control technology of coal gangue mountain spontaneous combustion, research on mountain antipollution treatment and transformation technology, research on restoration and treatment of coal gangue mountain greening environment, etc. [12]. The fire-fighting methods of the gangue mountain mainly include excavating the fire source method, hydraulic fire-fighting method, wrapping method, covering fire-fighting method, and grouting fire-fighting method [13]. The two conditions of sulfur combustibles, ventilation, and oxygen supply are as follows: make full use of flame retardant and water in filling the slurry to reduce the temperature of the coal gangue in the deep area, wrap sulfur-containing combustibles with mud and lime slurry, fill the gaps, and isolate oxygen to reduce the temperature of coal gangue in deep areas. To achieve the purpose of blocking combustion, cooling, and extinguishing, the coal gangue no longer has the combustion conditions [14].

The waste heat resource is the residual heat after the combustion of primary energy and combustibles and is regarded as the fifth largest energy source after coal, oil, natural gas, and water power. China's energy utilization rate is about 10% lower than that of developed countries, and at least more than half of the industrial energy consumption is directly discarded in the form of waste heat [15]. Waste heat resources can be divided into three types according to temperature grade: low-temperature waste heat (<300°C), medium-temperature waste heat (300–600°C), and high-temperature waste heat (>600°C). The residual heat energy in spontaneous combustion coal gangue hills and coalfield fire areas is the heat stored after coal combustion and has a wide temperature distribution range, a wide fire area, and a large amount of heat storage, which has potential utilization. At present, the methods for cooling the spontaneous coal gangue mountain fire areas include water injection, blasting, dry ice, liquid nitrogen injection, liquid CO_2 injection, grouting, glue injection, etc., which belong to the phase change absorption heat of the water or inert gas to reduce the high-temperature fire zone temperature [16]. Liquid nitrogen cooling and fire extinguishing technology have been used earlier in the field of mine fire-fighting. With its excellent low-temperature cooling and safety characteristics, colorless, odorless, noncorrosive, and low temperature liquid nitrogen has become an important fire-fighting technical means and has now formed a mature technical system, equipment, and process [17].

In 1942, Gaugler proposed the heat pipe (HP) by envisioning the manufacture of refrigerators. The heat pipe is a high-efficiency heat transfer element with gas-liquid two-phase flow circulation [18]. It was invented in the 1960s to solve the special requirements for heat transfer in spaceflight. Because of its huge role in energy conservation, its theoretical research and engineering applications have made great achievements in temperature uniformity, temperature control, waste heat recovery, new energy development, etc., and they have been widely used in chemical, energy, power, and mining fields [19]. In terms of the engineering application research of hot rods, the Institute of Cold Regions of the Chinese Academy of Sciences, Lanzhou University, and other units have successively carried out research on the thermal stability of hot rods to protect frozen soil subgrades in the Qingshui River section of the Qinghai-

Tibet Railway and found that the effective heat transfer radius of the hot rods is 1.5 m, the maximum heat transfer influence range is 2.16 m, the numerical simulation shows that the installation distance between the heat pipes can be designed to be 3.3–3.8 m, and the upper limit of the frozen soil of the subgrade can be raised by 0.8–1.2 m [20].

The research on the heat transfer and cooling of hot rods in the coal gangue mountains or coal piles began in the 1990s. In 1991, P. Muthukumar and M. Groll experimentally tested the cooling effect of gravity hot rods on the coal pile in the Hepingmen coal yard of Nanjing Fuel Company. Compared with the dense-hole cooling method, the hot rod is more effective in preventing coal spontaneous combustion [21]. A. Adamus established a heat transfer calculation model for hot rods from the perspective of engineering design and analyzed the technical and economic feasibility of using hot rods for heat transfer in deep coal gangue mountains. The first stage is inserted into the area where the temperature does not exceed 250°C. The second stage is inserted into the self-ignition center area by staged cooling measures [22]. In 2014, M. Mochizuki et al. discussed the advantages of using hot rods in the prevention and control of coal gangue mountains, analyzed the technical difficulties of using hot rods in the spontaneous combustion of coal gangue mountains to remove heat and cool down, and studied the heat transfer and cooling of coal spontaneous combustion prevention and fire-extinguishing heat rods. The performance test method shows that the cooling ability of the hot rod to the coal pile decreases with the increase of the distance, the cumulative heat transfer and cooling ability of the hot rod to the coal pile increase with time, and the radius of the distance to the hot rod is 20 mm. The cooling effect is the most significant, with a maximum temperature reduction of 13.3°C, a decrease of up to 29% [23].

A large number of pieces of research have been carried out on the spontaneous combustion characteristics, process, and prevention and control methods of coal gangue mountains at home and abroad, however, the research on the spontaneous combustion mechanism of the coal gangue and its heat transfer prevention and control methods mainly have the following shortcomings: (1) influenced by the thermal storage environment, oxygen supply conditions, and self-oxidation characteristics, the kinetic process of coal gangue oxidation and spontaneous combustion is complex, and there are many influencing factors, which are typically noncontrollable. It is difficult for existing research to accurately determine the high temperature area in the coal gangue mountain and its development trend. (2) The unsteady characteristics of heat generation and accumulation during the spontaneous combustion of coal gangue mountains cannot be accurately grasped, and it is difficult to reveal the kinetic mechanism of heat transfer during the spontaneous combustion of coal gangue mountains, and the kinetic mechanism of nonheat-controlled combustion of coal gangue is unclear.

To this end, this study plans to use a combination of theoretical analysis, experimental testing, and numerical simulation to build a coal gangue multifield-coupled uncontrolled combustion kinetic model based on the characteristics of the thermal migration behavior of coal gangue mountain spontaneous combustion. The research provides theoretical guidance for grasping the spontaneous combustion mechanism of coal gangue hills and is of great significance for preventing and controlling spontaneous combustion and the reignition of coal gangue hills.

At present, a large number of studies have been carried out on the spontaneous combustion characteristics, processes, and prevention and control methods of coal gangue mountains at home and abroad, however, there are several deficiencies in the research on the spontaneous combustion mechanism of coal gangue and its heat transfer prevention and control methods, which are as follows: (1) there are few studies on the spontaneous combustion characteristic parameters, such as gas index, critical temperature, and oxygen consumption rate, in the spontaneous combustion process of coal gangue. (2) Existing research is difficult to accurately determine the high temperature area and its development trend inside the coal gangue mountain. (3) The unsteady characteristics of heat generation and accumulation during the spontaneous combustion of coal gangue mountains cannot be accurately grasped, and it is difficult to reveal the kinetic mechanism of heat transfer during the spontaneous combustion of coal gangue mountains. The kinetic mechanism of the non-heat-controlled combustion of the coal gangue is unclear. (4) There are few pieces of research on key parameters and methods, such as heat transfer technical characteristics, heat transfer radius, and heat transfer effect evaluation of hot rods, implanted in coal gangue mountains, and the basic parameters for the directional heat transfer control of hot rods in the coal gangue mountain fire areas have not been determined.

Therefore, in our study, we first introduced the research methods and thermodynamic theory including similarity theory, fluid mechanics, thermodynamics, heat transfer, numerical heat transfer, and other theories in Section 2. Then, we conducted the coal gangue reaction kinetics solution, the analysis of the kinetic parameters of coal gangue oxidative combustion, the analysis of comprehensive combustion characteristics of spontaneous combustion coal gangue, and a study on the heat transfer method of the hot rod in the gangue mountain in the results and discussion section.

2. Research Methods and Thermodynamic Theory

2.1. Research Methods. Using the similarity theory, fluid mechanics, thermodynamics, heat transfer, numerical heat transfer, and other theories, by analyzing the law of the natural ignition of the coal gangue mountains, and according to the flow and heat transfer law of vapor-liquid two-phase flow inside the hot rod, the heat transfer theory and the methods of the hot rod are organically combined with the spontaneous combustion mechanism and temperature field distribution of the gangue mountain to master the temperature field distribution law, influencing factors, and the ignition law of the gangue mountain oxidation and spontaneous combustion process, determine the conditions

and influencing factors of gangue spontaneous combustion, and determine the dangerous area of the gangue mountain.

Using a variety of advanced experimental methods, such as industry analysis, chemical composition analysis, element analysis, and thermal physical parameter analysis, experimental research on coal gangue was carried out, and the correlation between the internal parameters of the coal gangue and the natural ignition characteristics was obtained. The trend of temperature change, combined with the measured parameters of the coal gangue, and the results of the experimental analysis are used to study the spontaneous ignition mechanism of the coal gangue.

2.2. Thermodynamic Theory. The process of the spontaneous combustion of the coal gangue starts from the accumulation of heat. The moisture and adsorbed free gas on the surface and pores of the gangue particles evaporate and desorb, and then, they oxidize and separate into a volatile matter. When the outside temperature is high enough and there is sufficient oxygen supply, the volatile matter burns, and the fixed carbon is cracked and ignited under high-temperature burning. The oxidative combustion of the coal gangue is controlled by chemical kinetics, in which the precipitation of volatiles controls the combustion process. In the process of the thermal analysis of coal, the relationship between physical parameters, such as coal mass and heat, and the reaction rate can be expressed in the following two forms:

integral form:

$$M(\alpha) = kt \tag{1}$$

differential form:

$$\frac{d\alpha}{\mathrm{d}t} = kf(\alpha), \tag{2}$$

where α is the conversion rate of coal at time t, k is the reaction rate constant, $M(\alpha)$ is the integral form of the reaction mechanism function, and (α) is the differential form of the reaction mechanism function. The relationship between $f(\alpha)$ and $M(\alpha)$ can be expressed as follows:

$$f(\alpha) = \frac{1}{M'(\alpha)} = \frac{1}{d[M(\alpha)]/d(\alpha)}. \tag{3}$$

The relationship between k and the reaction temperature T (absolute temperature) can be expressed according to the Arrhenius equation as follows:

$$k = A \times \exp\left(\frac{-E}{RT}\right), \tag{4}$$

where E is the apparent activation energy of the reaction, kJ/mol. A is the apparent pre-exponential factor of the reaction, R is the universal gas constant, i.e., 8.314 J/(mol·K), and T is the absolute temperature.

Since the experimental sample is under the temperature-programmed experimental conditions of a certain heating rate, the relationship between the thermodynamic temperature and the time of the experimental sample is as follows:

$$T = T_0 + at, \tag{5}$$

where T_0 is start temperature, K, and a is temperature rate, K·min^{-1}. Therefore, the common kinetic equations for heterogeneous systems under nonisothermal conditions can be obtained.

The differential form is as follows:

$$\frac{d\alpha}{dT} = \frac{A}{a} f(\alpha)\exp\left(\frac{-E}{RT}\right). \tag{6}$$

The integral form is as follows:

$$\begin{aligned} M(\alpha) &= \int_{T_0}^{T} \frac{A}{a}\exp\left(-\frac{E}{RT}\right)dT \\ &\approx \int_{0}^{T} \frac{A}{a}\exp\left(-\frac{E}{RT}\right)dT = \frac{AE}{aR}H(u). \end{aligned} \tag{7}$$

In the calculation process of the nonisothermal kinetic equation used, the calculation of $H(u)$ does not converge. Hence, it does not have an accurate analytical formula. In this study, the general kinetic analysis method of the heating rate under nonisothermal conditions is adopted, and the fitted correlation coefficients are consistent. The dynamic mode function and parameters of the coal gangue at different stages can be determined by the characteristics.

Achar integral method. Upon transforming equation (6), we can obtain the following another form:

$$\frac{a}{f(\alpha)}\frac{d\alpha}{dT} = A\exp\left(-\frac{E}{RT}\right). \tag{8}$$

Upon taking the logarithm of both sides directly, we obtain the following:

$$\ln\frac{d\alpha}{f(\alpha)dT} = \ln\frac{A}{a} - \frac{E}{RT}. \tag{9}$$

Equation (9) is called as the Achar–Brindley–Sharp–Wendworth equation.

$$\frac{d\alpha/\mathrm{d}t}{f(\alpha)} = \ln A - \frac{E}{RT}. \tag{10}$$

Based on (10), the relationship between $\mathrm{d}\alpha/\mathrm{d}t/f(\alpha)$ and $1/T$ is linear. By substituting different solid reaction mechanism functions, a linear fitting analysis is performed to solve the corresponding reaction apparent activation energy E, pre-exponential factor A, and mechanism function $f(\alpha)$.

The Coats-Redfern integration method for activation energy is as follows:

For equation (1), a common kinetic equation in heterogeneous systems is as follows:

$$\frac{d\alpha}{dT} = \frac{A}{a}\exp(-E/RT)f(\alpha). \tag{11}$$

According to the Coats-Redfern method,

$$\ln\left[\frac{M(\alpha)}{T^2}\right] = \ln\left(\frac{AR}{aE}\left(1-\frac{2RT}{E}\right)\right) - \frac{E}{RT}. \quad (12)$$

where $\alpha = m_0 - m/m_0 - m_\infty$ is reaction conversion, and $R = 8.314\ \mathrm{J\cdot mol^{-1}\cdot K^{-1}}$. Because of the general reaction temperature range and most E values, $E/RT \geq 1$, $1 - 2RT/E \approx 1$, and (12) can be changed to the following:

$$\ln\left[\frac{M(\alpha)}{T^2}\right] = \ln\left(\frac{AR}{aE}\right) - \frac{E}{RT}. \quad (13)$$

According to equation (13), integral form of the reaction mechanism function can be obtained.

Coal gangue is a complex mixture, and its spontaneous combustion oxidation reaction is not controlled by a single reaction mechanism. The whole process involves moisture and gas evaporation and desorption (I), oxidative cracking (II), intense combustion stage (III), and activation phase transition (IV). For exploring the mechanism of coal spontaneous combustion, only the second and third stages need to be considered, namely the oxidative cracking stage and the combustion process. For different reaction mechanisms, the selection of the integral function $M(\alpha)$ corresponding to the kinetic model function $f(\alpha)$ has different forms [24].

Taking the maximum constant rate point as the limit, the spontaneous combustion of the gangue is divided into two stages fitting equations: low-temperature oxidation II and high-temperature combustion III. Compare and analyze the consistency of the calculation results of the two kinetic calculation methods under the condition of nonisothermal heating rate. If the selected reaction model is reasonable, the values of E and A (or ln A) obtained by the two calculation methods should be similar. The values of parameters E and A should be in the normal value range of the kinetic parameters of energetic materials, and the coefficient R^2 of the fitting degree should be above 0.98. It is considered that the oxidation process of the coal gangue conforms to the kinetic model function.

3. Results and Discussion

3.1. Coal Gangue Reaction Kinetics Solution. Activation energy refers to the lowest energy that a coal gangue can undergo oxidation reaction at. At a certain temperature, the energy of the activated molecule is higher than the average energy of other molecules, and it is these energies that exceed the average energy to hit the original molecule, weaken or destroy its internal chemical bonds, and rearrange the atoms in the molecule to form new products. If the energy of the impacting molecule is less than this energy, the reaction cannot take place. For different substances or the types of reactions, the energy levels required for their reactions are not the same as activation energy. In other words, the larger the activation energy, the smaller the tendency of the coal gangue to spontaneously ignite. The smaller the activation energy, the easier it is to spontaneously ignite.

According to the analysis results of coal gangue thermogravimetric experiments, the Achar differential method and the Coats-Redfern integral method were used to analyze the kinetics of the second stage (116.2–353.0°C) of the spontaneous combustion oxidation of the coal gangue, and the oxidation kinetic parameters of this stage under the experimental conditions were calculated. Substitute the relevant data into the commonly used solid reaction mechanism functions to obtain the activation energy and the pre-exponential factor of the differential and integral methods, respectively. The results are shown in Table 1.

According to the inference of the Bagchi method of the reaction mechanism function, the linear correlation coefficients R^2 obtained by the Avrami–Erofeev function is relatively high, which are 0.987 and 0.997, respectively, and the obtained E and ln$A$ values are relatively close. After a comprehensive comparison, the Avrami–Erofeev function was selected for the most probable mechanism function of the second stage of the coal gangue. Substitute the experimental data of the second stage of the coal gangue into $f(\alpha)$ and $G(\alpha)$ of the Avrami–Erofeev function, plot, and fit, respectively, and for results, see Figure 1.

In the same way, the kinetic parameters of the third stage of the coal gangue samples were solved, and the kinetic analysis of the third stage of the coal gangue combustion was carried out. For results, see Table 2. According to the Bagchi method of the reaction mechanism function, the values of E and ln A obtained by the Avrami-Erofeev function are relatively close, and their linear correlation coefficient, R2,is 0.997 and 0.998, respectively., After comprehensive comparison, we choose the Avrami-Erofeev function as the most probable mechanism function of coal gangue stage III. Substitute the experimental data of coal gangue stage III into $f(\alpha)$ and $G(\alpha)$ of Avrami-Erofeev and plot and fit them, respectively, as shown in Figure 2.

3.2. Analysis of Kinetic Parameters of Coal Gangue Oxidative Combustion. According to the method and procedure of the thermodynamic analysis of 2# coal gangue sample, the Achar differential method and Coats-Redfern integral method are used to carry out the thermodynamic calculation and analysis of the coal gangue stage II and III, and the most probable reaction mechanism is deduced according to the Bagchi method. The thermodynamic parameters of 2#, 4#, 8#, and 12# coal gangue are obtained as shown in Table 3.

For the 4# coal gangue sample, according to the Bagchi method of the reaction mechanism function, after comprehensive comparison, the reaction order function is selected as the most probable mechanism function of the second stage of the coal gangue, and the Avrami-Erofeev function is selected as the most probable mechanism function of the third stage of the coal gangue. Substitute the experimental data of the second and third stages of the coal gangue into $f(\alpha)$ and $G(\alpha)$ of the reaction order and plot and fit them, respectively, as shown in Figures 3 and 4, respectively.

Because of the characteristics of low volatile content and high ash content of Gongwusu coal gangue, according to the kinetic parameters of the second and third stages of the coal gangue, in the initial stage of oxidative spontaneous combustion, the temperature is low, the decomposition and

Table 1: Phase II kinetic parameters of coal gangue at a heating rate of 20°C/min.

Function form	Differential			Integral		
	E	ln$A$	Correlation R	E	ln$A$	Correlation R
Avrami–erofeev equation	89.996	25.321	0.987	96.211	20.997	0.997

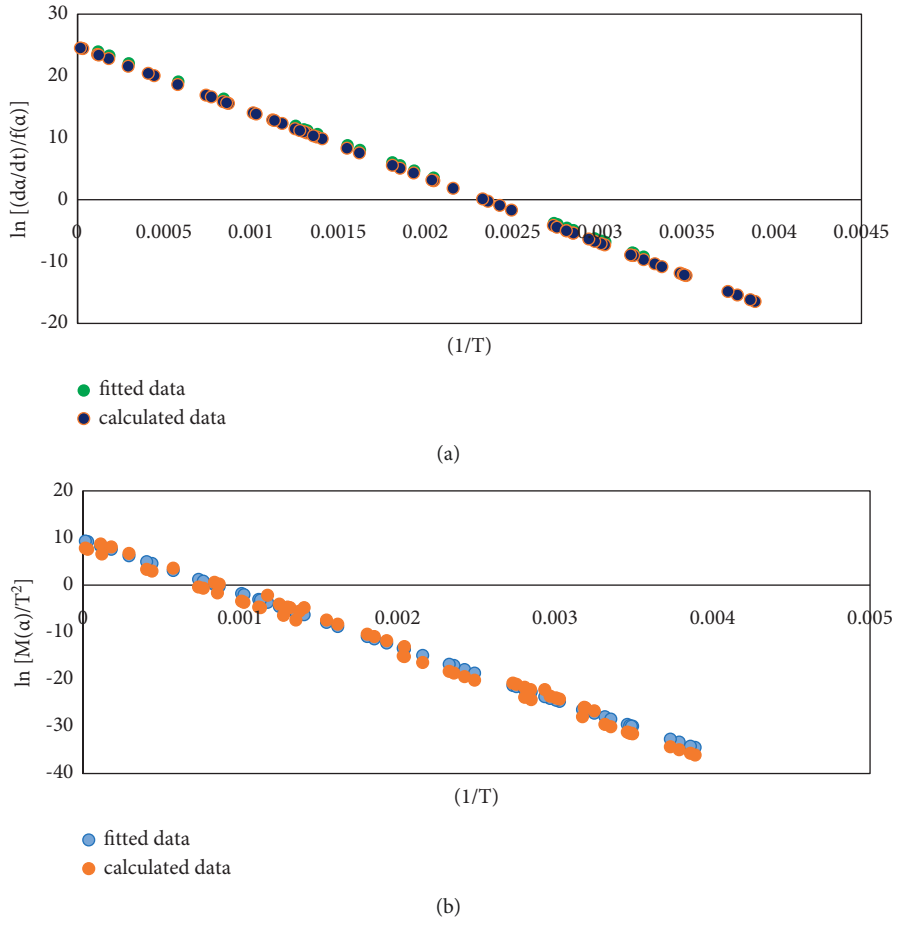


Figure 1: Fitted data of Avrami–Erofeev function for the second stage of coal gangue. Differential FOR (a) and integral FOR (b).

Table 2: Phase III kinetic parameters of coal gangue at a heating rate of 20°C/min.

Function form	Differential			Integral		
	E	ln$A$	Correlation R	E	ln$A$	Correlation R
Avrami- erofeev function	250.996	36.896	0.997	261.125	36.559	0.998

precipitation rate of volatile matter is low, and the ash content is high. To a certain extent, it hinders the precipitation of volatile matter in the coal gangue. Hence, the smooth oxidation reaction needs to consume a lot of energy, the activation energy and frequency factor are relatively high, the fixed carbon in the coal gangue starts to burn, the activation energy is relatively low at this time, and the oxidation reaction is easy to proceed.

3.3. Analysis of Comprehensive Combustion Characteristics of Spontaneous Combustion Coal Gangue. The ignition temperature of the coal gangue is much lower than the ignition temperature of coal with poor combustion performance in the mixture. The higher the volatile content in the coal gangue, the lower the characteristic temperature of pyrolysis, the faster the rate of volatile analysis, and the greater the total weight loss rate after pyrolysis. The fuel ratio of the coal gangue, i.e., the ratio of the fixed carbon to volatile matter (FC_{ad}/V_{ad}), is used to qualitatively judge the combustion performance of the coal gangue. The smaller the fuel ratio, the better the combustion performance of the coal gangue. NIE Qi–hong defined the comprehensive combustion characteristic index Sn to evaluate the combustion performance of coal [25]. For the slow-burning process of coal, the burning rate is expressed approximately with Arrhenius's law, and (14) is its definition.

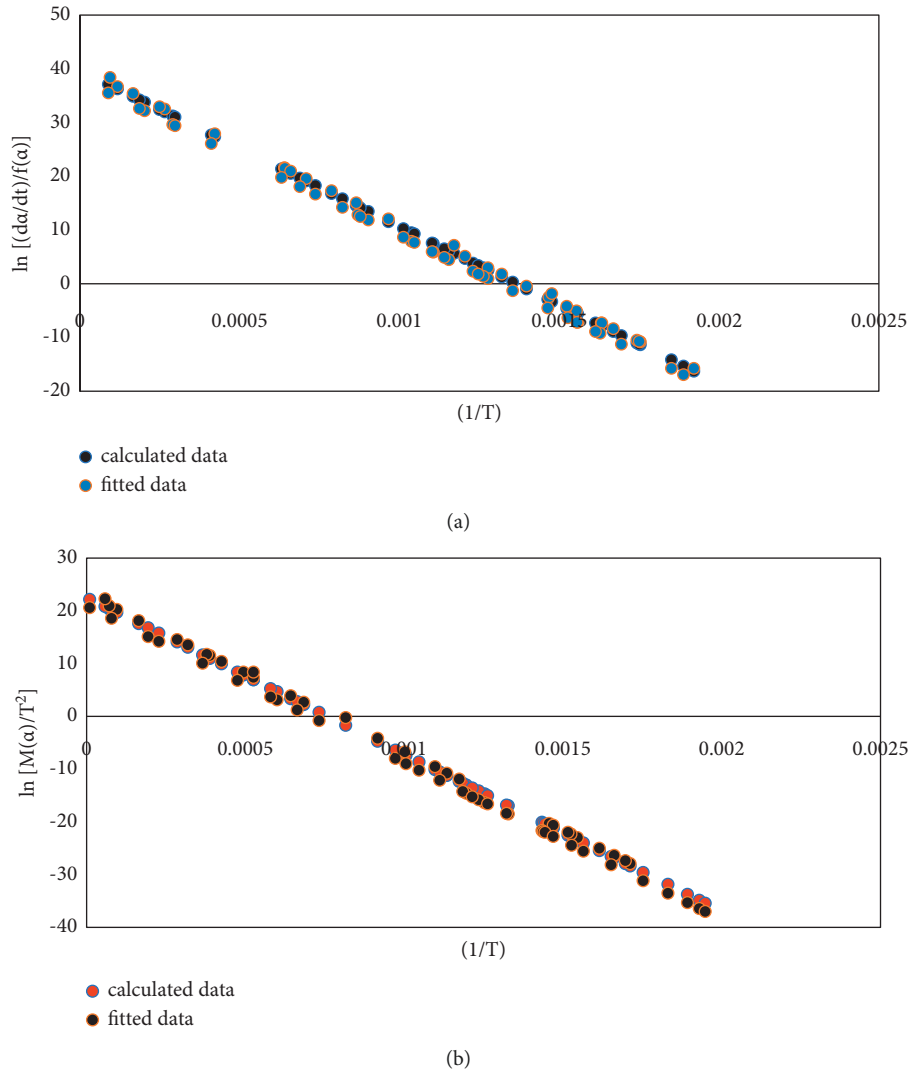


Figure 2: Fitted data of Avrami–Erofeev function for the third stage of coal gangue. Differential FOR (a) and integral FOR (b).

Table 3: Kinetic parameters of the second and third stages of coal gangue at a heating rate of 20°C/min.

Sample	Oxidation stage	Mechanism function	Activation energy (kJ·mol^{-1})	Pre-exponential factor (log A/s^{-1})	Relationship
2	II	Avrami–Erofeev	88.664	28.662	0.993
	III	Avrami–Erofeev	253.557	45.325	0.994
4	II	Reaction order	48.365	20.162	0.991
	III	Avrami–Erofeev	115.526	19.632	0.986
8	II	Avrami–Erofeev	105.345	36.23	0.988
	III	Avrami–Erofeev	354.594	66.335	0.998
12	II	Avrami–Erofeev	189.63	55.632	0.996
	III	Avrami–Erofeev	398.454	67.562	0.995

$$\frac{dv}{d\tau} = A\ \exp\left(\frac{-E}{RT}\right), \tag{14}$$

where $dv/d\tau$ is the combustion rate, %/min. A is the pre-exponential factor, and E is the activation energy, kJ/mol. T is the particle temperature, and R is atmospheric constant, 8.31 J (mol K).

Derivate (14) as follows:

$$S_n = \frac{dv/d\tau_{\max} dv/d\tau_{ave}}{T_i^2 T_h}, \tag{15}$$

where $dv/d\tau_{\max}$ is the maximum combustion rate, %/min. $dv/d\tau_{\text{ave}}$ is the average burning rate of the combustibles, %/min. T_i is fire temperature, K. T_h is the burnout temperature. Ignition index (D_i) and extinction index (D_h) are calculated by equations (16) and (17), respectively. They can

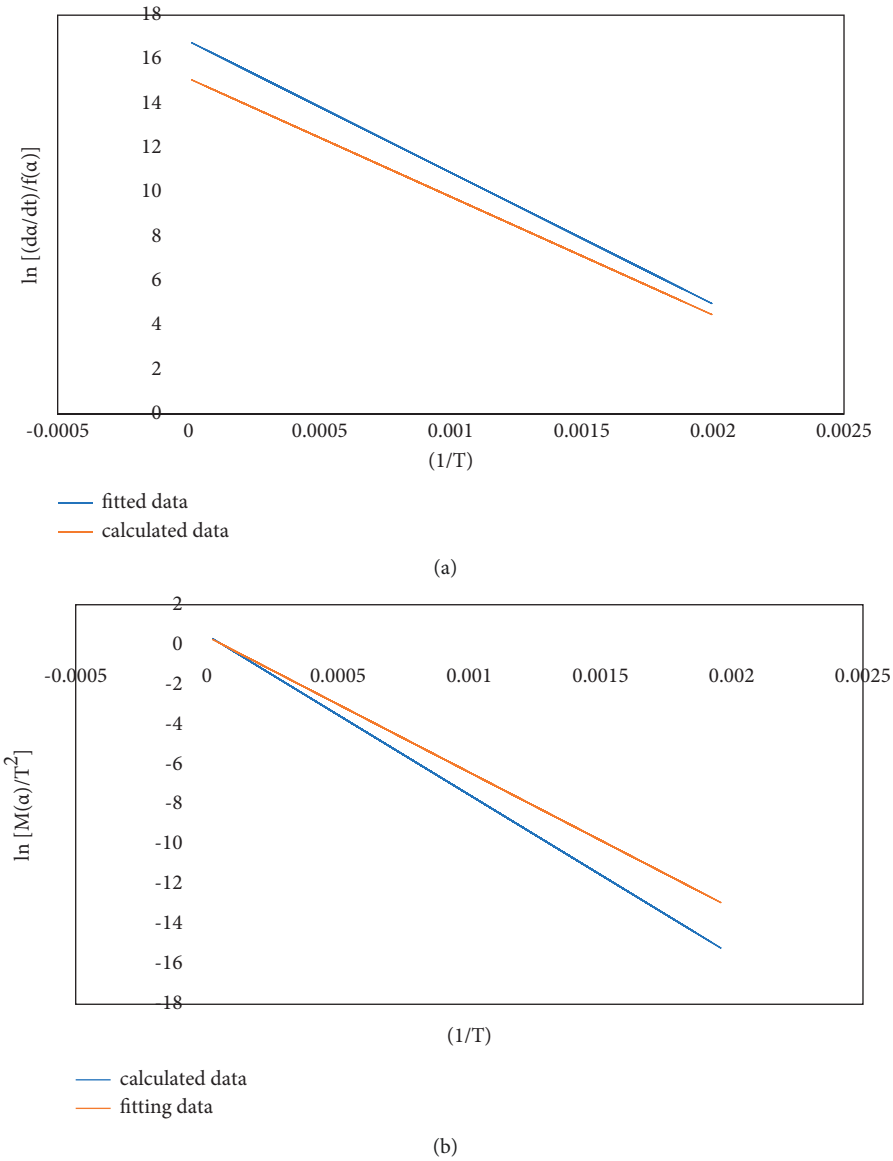


Figure 3: $\ln[(\mathrm{d}a/\mathrm{d}t)/f(a)]$, $\ln[M(a)/T^2]$, and $1/T$ curves of coal gangue 4# Phase II. Reaction order function with differential form FOR (a), $R^2 = 0.965$, with integral form FOR (b), $R^2 = 0.978$.

evaluate the ignition and extinguishing characteristics of the gangue.

$$D_i = \frac{DT\ G_{\max}}{t_p t_i}. \tag{16}$$

$$D_h = \frac{DT\ G_{\max}}{\Delta t_{1/2} t_p t_h}, \tag{17}$$

where $DT\ G_{\max}$ is the maximum burning rate, t_p is the corresponding time at $DT\ G_{\max}$, ignition time of $t_1 = t_3$, $\Delta t_{1/2}$ is the time when $DT\ G/DT\ G_{\max} = 0.5$, $t_h = t_5$, $T_i = T_3$, °C, and T_4 is the maximum combustion rate of the combustible substances that corresponds to the temperature. Combined with the results of the thermogravimetric test of the coal gangue and coal, according to formulas 15–17, the characteristic parameters of the oxidative combustion process of the coal gangue and coal are calculated as shown in Table 4.

We can conclude from Table 4 that the comprehensive combustion index $S_4 > S_8 > S_{12} > S_2 > S_{13}$. Among the 5 coal gangues, 4# coal gangue has the largest comprehensive combustion index, high volatile content, low pyrolysis characteristic temperature, faster volatile analysis rate, and the largest total weight loss rate after pyrolysis. It is corroborated by the experimental results. The corresponding 2# coal gangue has high lime content, low volatile and fixed carbon content, and its comprehensive combustion index is the smallest, which is not easy to burn. (2) Ignition index $D_{i\,8} > D_{i\,12} > D_{i\,4} > D_{i\,2}$. Among the 5 coal gangues, the ignition index of 8# is the largest, which is better than that of 12#, 4#, and 2# coal gangue. Industrial analysis shows that

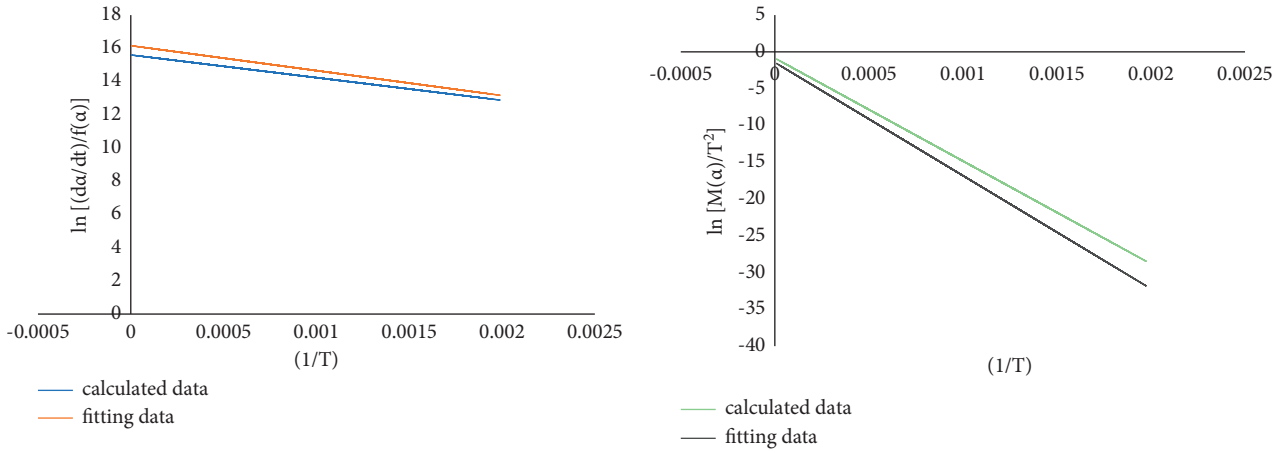


Figure 4: $\ln[(\mathrm{d}a/\mathrm{d}t)/f(a)]$, $\ln[M(a)/T^2]$, and $1/T$ curves of coal gangue 4# Phase III. Avrami–Erofeev function with differential form FOR (a), $R^2 = 0.989$, with integral form FOR (b), $R^2 = 0.991$.

Table 4: The results of the characteristic parameters of the oxidative combustion process of coal gangue and coal.

Number	$dv/d\tau_{\max}$ (%min^{-1})	$dv/d\tau_{\text{ave}}$ (%)	$D_i \times 10^4$	$D_h \times 10^4$	S_n (10^{11}%^{2})
2	−0.862	−0.421	14.325	0.896	4.985
4	−1.455	−1.201	17.987	0.621	29.221
8	−2.181	−0.721	40.220	2.067	29.125
12	−1.775	−0.521	32.667	2.981	16.541
13	−6.895	−1.584	200.215	6.785	211.952

Table 5: Characteristic parameters and combustion index of coal gangue in different regions.

Types of coal gangue	Total weight loss (%)	T_5 (%°C)	$V_{a\,d}$ (%)	$FC_{a\,d}/V_{a\,d}$	$S_n \times 10^{11}$ (%^{2}/min · K^3)
Experimental coal gangue 8#	32.5	410.2	16.35	0.93	31.02
Pingdingshan coal gangue	16	420	9.1	1.3	9.1
Datong coal gangue	25	400	18.2	1.1	39.2
Hancheng coal gangue	23	460	6.88	4.1	9.1
Panzhihua coal gangue	22	450	10.26	1.1	13.5

the 8# coal gangue has a high fixed carbon content, and its ignition index may be related to its fixed carbon content. (3) Burnout index $D_{h\,12} > D_{h\,8} > D_{h\,2} > D_{h\,4}$. The larger the burnout index, the worse the burnout performance. Among the 5 coal gangue samples, 12# has the worst burnout performance, 8# is the second, and 4# has the best burnout performance. It shows that in the later stage of coal gangue combustion, the ash content hinders the diffusion of oxygen molecules in the direction of combustible substances to a certain extent. Therefore, the higher the content of fixed carbon in the coal gangue, the worse the burnout performance.

The characteristic parameters and comprehensive combustion characteristics of coal gangue in different regions and ages are compared, as shown in Table 5.

We can demonstrate from Table 5 that the ignition temperature T_5 of Gongwusu coal gangue used in the test is 410.2°C, the lowest ignition temperature, and the highest ignition temperature of Hancheng coal gangue indicating that the coal gangue contained in Gongwusu coal gangue has a low ignition point, and the coal gangue is prone to ignition. Under the condition that the particle size of the coal gangue is not much different, comparing the comprehensive combustion index S of the coal gangue in different regions, it can be seen that the comprehensive combustion index of the Gongwusu coal gangue used in the test is higher, which is $31.02 \times 10 - 31.02 * 11\%\text{min-2K-3}$. It is second only to the coal gangue in the Datong area and higher than the coal gangue in the Panzhihua, Pingdingshan, and Hancheng areas, indicating that the Gongwusu coal gangue has better characteristics of burning and burning out. According to the industrial analysis data of the coal gangue, it is found that the level of volatile matter content and the comprehensive combustion characteristic index S are consistent. The ratio of fixed carbon to volatile matter in the coal gangue used in the test is 0.93, which is the lowest among the five coal gangues, and its weight loss rate of 32.5% is the highest among the five coal gangues, which shows good performance of burnout and the high total weight loss rate of oxidative combustion.

3.4. Study on Heat Transfer Method of Hot Rod in Gangue Mountain

3.4.1. Thermal Behavior Characteristics of Coal Pile Temperature Field under the Influence of a Single Hot Rod. Coal gangue mountain spontaneous combustion is a relatively special combustion system, and its ignition, combustion maintenance, fire zone transfer, and combustion extinguishing are very different from ordinary fires. According to the unique heat conduction characteristics of the hot rod to eliminate the heat storage environment of the coal gangue mountain, it can destroy the spontaneous combustion conditions and prevent the spontaneous combustion from occurring.

Where there is a temperature difference, there must be a natural phenomenon of heat transfer from high temperature to low temperature. Of the three ways of heat transfer—heat radiation, convection, conduction,and the rate of heat transfer is the fastest. The hot rod is a heat exchange element with extremely high heat transfer efficiency. It relies on the internal working fluid phase change and continuous working fluid convective heat transfer and circulation to achieve efficient heat transfer. It is the device with the highest heat transfer efficiency in passive cooling systems. Hot rods are also called heat pipes because they are generally larger in mining engineering. Hence, they are called hot rods or hot piles.

In our study, we tested the influence of a single hot rod on the thermal behavior of the coal pile, and the temporal and spatial variation of the coal pile temperature field with and without the hot rod was compared and analyzed. The hot rod is 100 mm away from the heating surface in the X direction. The initial temperature of the experiment is set to 100°C, and the data is read and observed every 10 hours. We conducted the temperature response inside the coal pile, and its influence on the temperature field were tested under the action of two hot rods. Observations determine the behavioral effect of the distance between the two hot rods on the temperature field of the coal pile. The buried depth of the evaporation section of the hot rod is 490 mm. The distance between the hot rod and the heating surface in the X direction is 100 mm, and the distance between the two hot rods is 100 mm, 200 mm, and 300 mm. We read and observe data every 10 hours. Finally, the enhanced cooling effect of the hot rod on the coal pile was studied under different heating temperature conditions. The buried depth of the evaporation section of the hot rod is 490 mm, and the distance between the hot rod and the heating surface in the X direction is 100 mm. Three groups of initial heating temperatures of 100, 150, and 200°C are selected for testing, and the data is read and observed every 10 hours. The data processing method of this experiment is mainly to compare the slope of each group of curves (the rate of temperature change) and the cooling range. This method can intuitively see the size of the heat exchange and the heat exchange efficiency of the heat pipe under different conditions.

To study the effect of hot rods on the distribution of the temperature field inside the coal pile, the changes in the temperature field inside the coal pile before and after the effect of the hot rods were observed experimentally under the condition that there is an ignition source inside the coal pile, as shown in Figure 5.

The conclusion from Figure 5 shows that the flat plate heater is set to a constant temperature of 150°C to simulate the temperature of the fire source to heat the coal pile, and the temperature of the coal pile gradually rises without the action of the hot rod. The faster the rate, the higher the temperature at No. 1 measuring point, reaching about 84°C. After heating for 300 hours, the temperature of each measuring point in the coal pile is stable. At this time, the heat input of the heater to the test coal pile and the heat dissipation of the coal pile to the environment are in a dynamic balance, and the temperature field inside the coal pile is in a relatively stable state. After the temperature of the coal pile is stable, insert the hot rod into the predetermined position of the coal pile. It can be seen from Figure 5 that the monitoring temperature in the coal pile drops rapidly after the hot rod is inserted. After inserting the hot rod for about 50 h (the test time is about 350 h), the temperature of the coal pile gradually stabilized again. Among them, the temperature drop of No. 1, 2, 3, and 4 points was obvious because of the enhanced heat dissipation effect of the hot rod, and the temperature of No. 5 test point decreased significantly. The temperature change at the No. 6 measuring point is not obvious. It is generally believed that the critical temperature range for the coal spontaneous combustion is 60–80°C. When the coal temperature is lower than the critical temperature of spontaneous combustion, the oxygen consumption rate and oxidation heating rate of coal increase slowly. When the temperature of coal is higher than the critical temperature of spontaneous combustion, it is regarded as the spontaneous combustion of coal.

After the test was carried out for 460 hours (the working time of the hot rod was about 150 h), the temperature in the coal pile became stable again, and the temperature data was processed to analyze the enhanced cooling effect of the hot rod on the coal pile under the test conditions. Figure 6 shows the cooling range and cooling rate of each measuring point in the coal pile under the action of the hot rod. It can be seen that the hot rod has the most obvious effect on the temperature of the No. 1 measuring point, where the cooling range of the coal pile is 33.4°C, and the cooling rate reaches 39.6%. The cooling effect here is weak. The results show that the cooling effect of the hot rod on the coal pile is inversely proportional to the distance between the hot rod and the coal body. The closer the hot rod to the coal body, the greater the cooling range of the coal body. As the distance from the edge of the hot rod increases, the influence of the hot rod on the distribution of the temperature field inside the coal pile gradually decreases.

3.4.2. Thermal Behavior Characteristics of Coal Pile Temperature Field under the Influence of Double Hot Rods. Figure 7 is the temperature change curve of the coal pile with two hot rods arranged at 10 cm parallel to the heating plate. With the increase in heating time, the overall temperature of the coal pile is on the rise, and the temperature rise rate is the

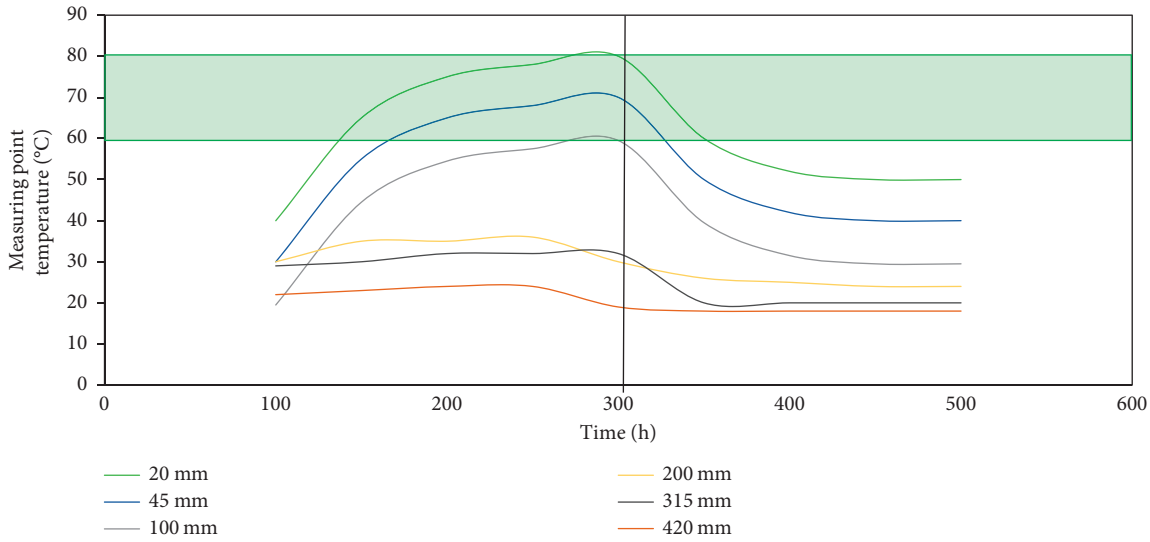


Figure 5: Variation of coal pile temperature curve with time. The shaded part represents the critical temperature range of coal spontaneous combustion. 0~300 h is without the hot rod stage, and 300~600 h is with the hot rod stage.

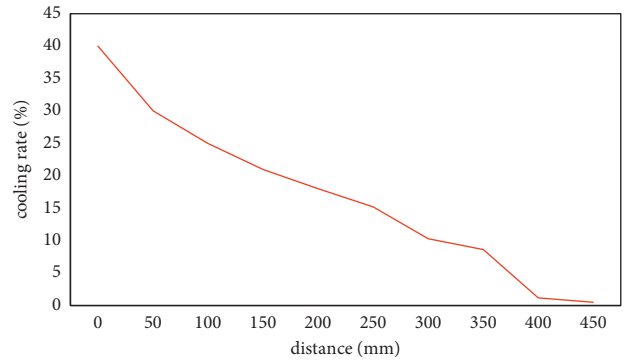


Figure 6: The cooling effect of each measuring point under the action of the hot rod.

fastest in the first 10 hours and the amplitude is the largest. After that, the heating rate gradually decreases with the increase of time, which is similar to that of the coal pile under the action of a single hot rod. The law of temperature field changes is consistent. In addition, the curve in the figure presents a "saddle" shape, because at the center point ($x = 30$ cm) of the line connecting the two hot rods, the temperature of the coal pile is up to 105°C. However, under the action of the external heat source for 132 hours at the central part of the installation of the hot rod (see Figure 7), the temperature of the coal body is still lower than 50°C. The cooling effect of the hot rod weakens with the increase of the radius, and the coal pile at the center point of the connection line is difficult to reduce rapidly in a short time. Therefore, the temperature field in the coal pile has a "saddle" shape because of the superposition of heating.

Once the coal temperature exceeds 200°C, the oxidation reaction rate of the coal is accelerated, and the spontaneous combustion stage quickly enters the combustion stage. To test the effect of the hot rod on the internal temperature field of the coal pile under different fire source temperature conditions, the initial temperature of the plate heater was set to 100, 150, and 200°C to simulate the heat source temperature. The internal temperature of the coal pile is changed with time when the coal pile is heated for 80 hours with hot rods (control group) and without hot rods (normal group). Based on the analysis of the heat transfer performance of the hot rod, it can be seen that the closer the coal pile is to the hot rod, the more obvious the effect of the hot rod on the temperature of the coal pile. Hence, the No. 1 measuring point is selected for analysis. The variation curves of coal pile temperature with time obtained from the three groups of tests are shown in Figure 8.

It can be seen from Figure 8 that under the different conditions of the initial heating temperature of the three groups, with the increase of time, the temperature of the coal pile is on the rise as a whole. During the heating process within 80 hours, the temperature of the measuring point in the normal group (without the hot rod) rose to a maximum of around 46°C, and the temperature of the *b* normal group and *c* normal group reached around 78°C and 104°C, respectively. The reason for the analysis is that the temperature of the heat source in the normal group *a* is low, and the temperature at the measuring point after heating is still lower than the critical temperature, while the normal groups *b* and *c* are heated without a hot rod. The temperature of the measuring point rises rapidly and exceeds the critical temperature. It can be considered that spontaneous combustion of coal has occurred there. Under the action of the hot rod, the temperature of the coal pile was kept in a low range in the three groups of tests, and the maximum coal temperature did not exceed 60°C. It shows that the hot rod can control the temperature of the coal body below the critical temperature, which can remove the heat inside the coal pile, delay the spontaneous combustion process of the coal, and prevent the spontaneous combustion of the coal body.

Figure 9 is a graph of the cooling range of the coal pile under different heating temperatures. The greater the

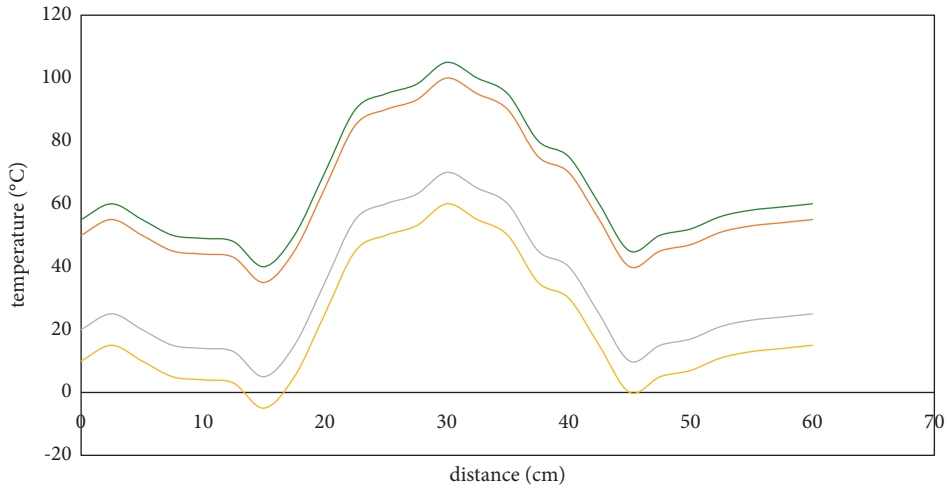


Figure 7: .Temperature changes of two hot rods at different times.

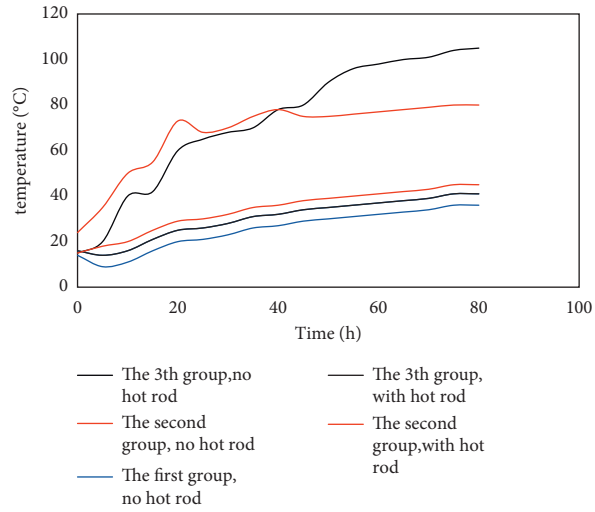


Figure 8: Temperature curve of No. 1 measuring point with time under different heat source temperature conditions. The bold line is the experimental group without a hot rod, and the thick line is the control group with a hot rod.

intensity of the heat source, the greater the cooling rate of the hot rod to the coal pile. In group *a* (heat source temperature of 110°C), the cooling range of the coal pile was maintained at 6–7°C. In group *b* (heat source temperature of 150°C), the cooling range was maintained at 25–28°C. In the heat source temperature of 200°C, the cooling amplitude value increases the fastest, and the cooling amplitude can reach up to 57.5°C. With the increase of time, the cooling amplitude has a further increasing trend. Analyzing the reasons, the heating temperature of group *a* is low, the heat flow inside the coal pile is small, and the heat dissipation power of the hot rod is also small. When the temperature is 200°C, the power of the hot rod is the largest among the three groups of tests, and its heat dissipation capacity to the coal pile is also the strongest. It shows that the cooling ability of the hot rod to the coal pile increases with the increase of the internal temperature of the spontaneous combustion coal pile.

As an index to quantitatively evaluate the cooling effect of the hot rod on the coal pile, the heat dissipation is of great significance. The amount of heat dissipation is related to the structure of the hot rod, and the cooling effect of the hot rod on the coal pile is also related to the particle size, porosity, thermophysical properties, air temperature, and internal temperature of the coal pile.

The change of heat dissipation trend can objectively evaluate the performance of the hot rod [26]. The heat dissipation of the hot rod is proportional to the temperature difference between the coal pile and the environment and the working time of the hot rod, and it is inversely proportional to the thermal resistance of the hot rod. During the test, the average indoor temperature was 11°C, and the test time was 80 h. The average temperatures of the coal piles under the three groups of test conditions were 39.82, 42.11, and 52.18°C, respectively, and the total thermal resistance of the coal pile-hot rod system was 5.0263°C/W. According to the

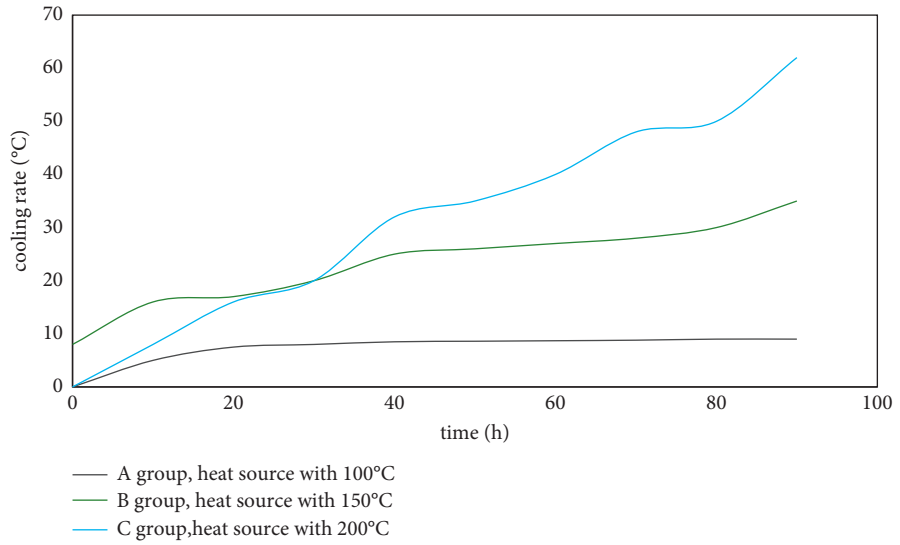


Figure 9: Variation of the cooling amplitude value of no. 1 measuring point with time at different heating temperatures.

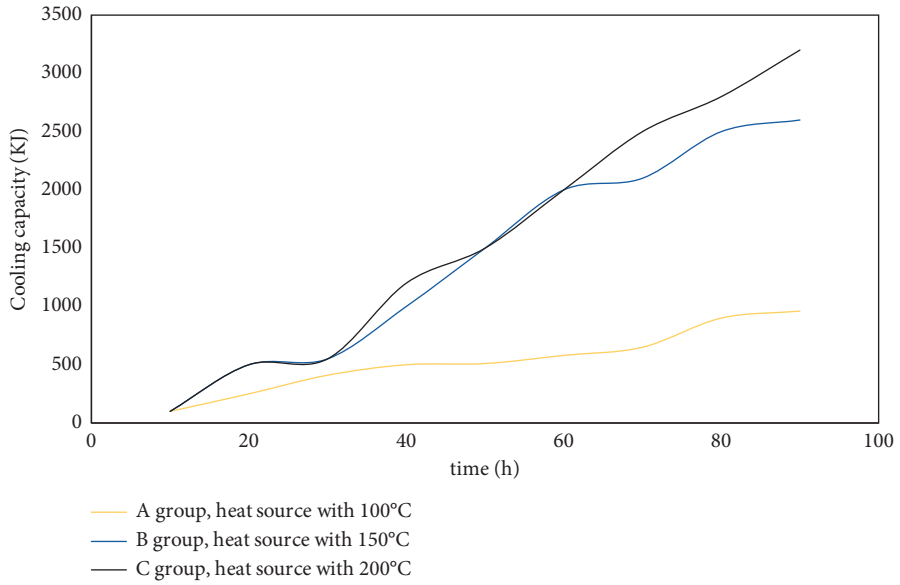


Figure 10: Variation curve of heat dissipation of hot rod with time under different heat source temperature conditions.

formula (15), the cooling capacity of the hot rod under the test conditions is obtained as shown in Figure 10.

4. Conclusions

In our study, we first solved the dynamic parameters and mechanical functions of the coal gangue using the Achar differential method and Coats-Redfern integral method, analyzed the comprehensive combustion characteristics of the coal gangue, and then, based on the flow and heat transfer mechanism of hot rod vapor-liquid two-phase flow, combined with coal spontaneous ignition conditions, influencing factors, coal pile spontaneous combustion temperature field structure distribution, etc., the heat transfer process of hot rod in coal pile (coal gangue mountain) was analyzed. The excellent characteristics and working characteristics of the hot rod are studied, the thermal turnover process of the hot rod in the coal pile is studied, the thermal resistance network of the hot rod and the mathematical expression of the heat transfer power are analyzed, and the unsteady mathematical model of the heat transfer of the hot rod is constructed. The quantitative standard for evaluating and judging the thermal turnover effect of the hot rod is studied. The main conclusions are as follows.

(1) The average activation energies of the second stage of 2#, 4#, 8#, and 12# coal gangue are 88.664, 48.365, 105.345, and 189.63 kJ/mol, respectively. According to the Bagchi method, the most probable mechanism functions of the second stage are inferred as Avrami–Erofeev equation, reaction order equation, and

Avrami–Erofeev equation with different stages; the average activation energies of the third stage are 253.557, 115.526, 354.594, and 398.454 kJ/mol, respectively, and all of the machine functions are Avrami–Erofeev.

(2) Upon comparing the characteristic parameters of the coal gangue in different regions and ages, it is found that 8# Gongwusu coal gangue has better combustion and burn-out characteristics, and its comprehensive combustion characteristics are second only to those of the coal gangue in Datong, and they are higher than those in Panzhihua, Pingdingshan, and Hancheng. The ratio of the fixed carbon to the volatile matter of 8# Gongwusu coal gangue is 0.88, which is the lowest among the five different coal gangues compared. Its weight loss rate is 30.30%, which is the highest among the five coal gangues, indicating that the volatile analysis of 8# coal gangue has fast speed, easy to ignite and burn, good burnout performance, the high total weight loss rate of oxidative combustion, and low ignition point of coal contained, which is prone to spontaneous combustion.

(3) The correlation between S_n and $(V_{a\,d} + FC_{a\,d})/A_{a\,d}$ is positive. It represents that the higher the content of volatile matter and fixed carbon in coal gangue, the lower the ash content, and the better the comprehensive combustion performance of the coal gangue. The comprehensive combustion index, ignition index, and burnout index were positively correlated with $FC_{\mathrm{ad}}/V_{\mathrm{ad}}$. The burnout index and the linear correlation of the fuel ratio FC_{ad}/V_{ad} is the highest, followed by the ignition index, and the comprehensive combustion index is the lowest. Under the condition of sufficient oxygen supply combustion, the larger the fuel ratio, the better the burnout performance of the coal gangue.

(4) The test of the influence of the hot rod on the temperature field distribution inside the coal pile shows that the maximum cooling rate of a single hot rod to the coal pile during the test period is 33.4°C, and the maximum cooling rate reaches 39.6%. However, under the action of the external heat source for 132 hours, the temperature of the coal body was still lower than 50°C. In the three sets of tests with different heating powers, as the temperature of the heat source increases, the heat flux density in the coal pile increases, and the heat dissipation power of the hot rod also increases.

(5) The calculated heat dissipation of the 80 h hot rod is 1.0865, 2.1680, and 3.3649 MJ, respectively. The increase in the heat dissipation of the hot rod can strengthen the cooling efficiency of the coal pile, effectively reduce the internal temperature of the coal pile, and thus inhibit the spontaneous combustion of the coal pile. The cooling effect of the hot rod on the coal pile is inversely proportional to the distance between the hot rod and the coal body. The closer the hot rod is to the coal body, the greater the cooling range of the coal body. The cooling ability of the hot rod to the coal pile increases with the increase of the internal temperature of the spontaneous combustion coal pile. The heat dissipation of the hot rod has been increasing with the increase of time, indicating that the cooling capacity of the hot rod to the coal pile will continue to increase with time.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest to report regarding the present study.

Acknowledgments

This work was supported by National Nature Fund Project Study on thermodynamic characteristics and regional heat transferring methods in the process of spontaneous combustion of coal gangue, No. 52004211.

References

[1] C. Kuenzer, J. Zhang, A. Tetzlaff et al., "Uncontrolled coal fires and their environmental impacts: investigating two arid mining regions in north-central China," *Applied Geography*, vol. 27, no. 1, pp. 42–62, 2007.

[2] J. Ciesielczuk, "Coal mining and combustion in the coal waste dumps of Poland," *Coal and Peat Fires: Global Perspectives*, pp. 464–473, 2012.

[3] E. Heffern and D Coates, "Geologic history of natural coal-bed fires, Powder River basin, USA," *International Journal of Coal Geology*, vol. 59, no. 1-2, pp. 25–47, 2004.

[4] G. B. Stracher and T. P. Taylor, "Coal fires burning out of control around the world: thermodynamic recipe for environmental catastrophe," *International Journal of Coal Geology*, vol. 59, no. 1–2, pp. 7–17, 2004.

[5] Z. Song and C. Kuenzer, "Coal fires in China over the last decade: a comprehensive review," *International Journal of Coal Geology*, vol. 133, pp. 72–99, 2014.

[6] L. Rosso, N. Koneva, and V. Fernicola, "Development of a heat-pipe-based hot plate for surface-temperature measurements," *International Journal of Thermophysics*, vol. 30, no. 1, pp. 257–264, 2009.

[7] R. Y. C. Lu, "Investigation of dry machining with embedded heat pipe cooling by finite element analysis and experiments [J]," *International Journal of Advanced Manufacturing Technology*, vol. 31, pp. 905–914, 2007.

[8] T. Kaya and J. Goldak, "Three-dimensional numerical analysis of heat and mass transfer in heat pipes," *Heat and Mass Transfer*, vol. 43, no. 8, pp. 775–785, 2007.

[9] G. A. Carlson, "Modeling of coal structure by using computer–aided molecular design," *ACS Symposium Series*, vol. 1, pp. 461–465, 1991.

[10] T. X. Ren, J. S. Edwards, and D. Clarke, "Adiabatic oxidation study on the propensity of pulverised coals to spontaneous combustion," *Fuel*, vol. 14, pp. 1611–1620, 1999.

[11] B. B. Beamish, M. A. Barakat, and J. D. S. George, "Adiabatic testing procedures for determining the self-heating propensity of coal and sample ageing effects," *Thermochimica Acta*, vol. 12, no. 1-2, pp. 79–87, 2000.

[12] B. B. Beamish, M. A. Barakat, and J. D. St George, "Spontaneous-combustion propensity of New Zealand coals under adiabatic conditions," *International Journal of Coal Geology*, vol. 45, no. 2-3, pp. 217–224, 2001.

[13] X. Qi, H. Xin, D. Wang, and G. Qi, "A rapid method for determining the R70 self-heating rate of coal," *Thermochimica Acta*, vol. 571, pp. 21–27, 2013.

[14] D. Wang, H.–hui Xin, Xu–yao Qi, G. L. Dou, G. S. Qi, and L. Y. Ma, "Reaction path way of coal oxidation at low temperatures: a model of cyclic chain reactions and kinetic characteristics [J]," *Combustion and Flame*, vol. 163, p. 447, 2016.

[15] A. K. Singh, R. V. K. Singh, M. P. Singh, H. Chandra, and N. Shukla, "Mine fire gas indices and their application to Indian underground coal mine fires," *International Journal of Coal Geology*, vol. 69, no. 3, pp. 192–204, 2007.

[16] J. Xie, S. Xue, W. Cheng, and G. Wang, "Early detection of spontaneous combustion of coal in underground coal mines with development of an ethylene enriching system," *International Journal of Coal Geology*, vol. 1, pp. 123–127, 2011.

[17] C. Wu and Z. Li, "A simple method for predicting the spontaneous combustion potential of sulphide ores at ambient temperature," *Mining Technology*, vol. 114, no. 2, pp. 125–128, 2005.

[18] J. Deng, Y. Xiao, Q. Li, J. Lu, and H Wen, "Experimental studies of spontaneous combustion and anaerobic cooling of coal," *Fuel*, no. 157, pp. 261–269, 2015.

[19] J. Deng, J. Zhao, Y. Zhang, and G. Ruilin, "Study on coal spontaneous combustion characteristic temperature of growth rate analysis," in *Proceedings of the2014 ISSST", 2014 International Symposium on Safety Science and Technology*, pp. 796–806, Beijing, China, July 2014.

[20] G. Desrayaud and A. Fichera, "Laminar natural convection in a vertical isothermal channel with symmetric surface-mounted rectangular ribs-mounted rectangular ribs[J]," *International Journal of Heat and Fluid Flow*, vol. 23, no. 4, pp. 519–529, 2002.

[21] P. Muthukumar and M. Groll, "Metal hydride based heating and cooling systems: a review," *International Journal of Hydrogen Energy*, vol. 35, no. 8, pp. 3817–3831, 2010.

[22] A. Adamus, "Review of the use of nitrogen in mine fires," *Mining Technology*, vol. 111, no. 2, pp. 89–98, 2002.

[23] M. Mochizuki, R. Singh, T. Nguyen, and T. Nguyen, "Heat pipe based passive emergency core cooling system for safe shutdown of nuclear power reactor," *Applied Thermal Engineering*, vol. 1, pp. 699–706, 2014.

[24] R. Y. Chiou, L. Lu, J. S. J. Chen, and M. T. North, "Investigation of dry machining with embedded heat pipe cooling by finite element analysis and experiments," *International Journal of Advanced Manufacturing Technology*, vol. 9, pp. 905–914, 2007.

[25] Y. Quan and Q. Mai, "Investigation of the cooling effect of heat pipe-embedded cutter in dry machining with different thermal conductivities of cutter/workpiece materials and different cutting parameters," *International Journal of Advanced Manufacturing Technology*, vol. 5, pp. 1161–1169, 2015.

[26] Y. Xiao, KQ. Zhong, JY. Tian, L. Yin, Y. Tian, and C. M. Shu, "Thermal extraction from a low-temperature stage of coal pile spontaneous combustion by two-phase closed thermosyphon," *Journal of Thermal Analysis and Calorimetry*, vol. 144, no. 2, pp. 587–597, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9835236, 1 page
https://doi.org/10.1155/2023/9835236

Retraction

Retracted: Analysis of the Application Effect of Multidisciplinary Team Cooperation Model in Chronic Heart Failure under WeChat Platform

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Huang, Y. Su, and X. Mao, "Analysis of the Application Effect of Multidisciplinary Team Cooperation Model in Chronic Heart Failure under WeChat Platform," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4051955, 7 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4051955, 7 pages
https://doi.org/10.1155/2022/4051955

Research Article

Analysis of the Application Effect of Multidisciplinary Team Cooperation Model in Chronic Heart Failure under WeChat Platform

Jieyu Huang,[1] Yu Su,[2] and Xiucai Mao [3]

[1]*Department of Cardiovascular Medicine, Hezhou People's Hospital, Second Ward, Hezhou, China*
[2]*Department of Nephrology, Hezhou People's Hospital, Hezhou, China*
[3]*Hezhou People's Hospital Nursing Department, Hezhou, China*

Correspondence should be addressed to Xiucai Mao; 18402437@masu.edu.cn

Received 21 June 2022; Revised 15 July 2022; Accepted 27 July 2022; Published 25 August 2022

Academic Editor: Wenming Cao

Background and Purpose. Chronic heart failure (CHF) is a complex clinical syndrome with a gradually increasing incidence that has become a public health problem of global concern. With the improvement of treatment modalities, great progress has been made in its treatment, but patients' recovery outcomes are poor. The aim of this study was to apply the multidisciplinary teamwork model under the WeChat platform to the treatment of patients with chronic heart failure. *Methods*. From April 2020 to May 2021, 56 patients with CHF who were discharged from the cardiology department of our hospital after treatment were randomly divided into two groups: experimental group ($n = 28$) and control group ($n = 28$). The control group was given conventional nursing measures and health education and discharge instructions, while the experimental group received collaborative multidisciplinary team nursing care based on the WeChat platform on the basis of the control group, all for 3 months. All enrolled patients underwent the Self-Care of Heart Failure Index Version 6.2 (SCHFI v6.2), the Minnesota Living with Heart Failure Questionnaire (MLHFQ), and the 6-minute walking test (6MWT test). The SCHFI v6.2 and MLHFQ scores, 6 MWT test results, and readmission rates within 3 months were observed and compared between the two groups. *Results*. There was no significant difference between the SCHFIv6.2 and MLHFQ scores of the two patients at admission and at discharge, and the scores of the experimental group were significantly higher than the scores of the control group at the end of 3 months after discharge. On the other hand, the SCHFIv6.2 and MLHFQ scores of the two groups were significantly higher at discharge compared to admission; the 6-minute walking distance of the experimental group was significantly higher than that of the control group at the end of 3 months. The readmission rate in the experimental group was significantly lower than that in the control group. *Conclusion*. The multidisciplinary teamwork model based on the WeChat platform can significantly improve the self-care ability and quality of life of CHF patients and reduce the readmission rate.

1. Introduction

Chronic heart failure (CHF) has become a public health problem of global concern, affecting approximately 1% to 2% of the world's population. It poses a serious threat to patients' health, reduces their ability to work and quality of life, and consequently consumes enormous medical resources, and places a heavy burden on families and society [1, 2]. CHF refers to a persistent state of heart failure. The main symptoms are dyspnoea, ankle swelling, fatigue, signs of elevated jugular pressure, lung rale, and peripheral oedema. The current treatment of heart failure includes general, pharmacological, and nonpharmacological treatments, which have reduced the rate of sudden death and improved clinical symptoms in patients with CHF [3, 4], whereas, although clinical treatment can control symptoms quickly, there are many adverse drug reactions such as electrolyte disorders and drug resistance. Furthermore, the goal of CHF treatment is not only to improve symptoms and hospital outcomes, but also to reduce mortality and

readmission rates [5]. Thus, patients need to follow not only standardised medication, but also self-care behaviours that focus on various aspects of diet, activity, physical status monitoring, behavioural and emotional management, proper recognition of symptoms, and assessment of treatment effectiveness.

Domestic and international studies have shown that a patient-centred, multidisciplinary team-based service model that provides comprehensive and individualised treatment, care, and rehabilitation for patients can improve patient self-management and quality of life and can reduce the number of patient readmissions [6, 7]. However, such multidisciplinary collaborative teams involve more people, are more affected by time and space, and are difficult to organise and implement at a fixed time, in a fixed hospital, or in the community [8]. At present, WeChat, as a rising and rapidly developing mobile new media force, has the advantages of low cost, strong interaction, featured consultation, and intelligent customization of public platform functions, which can penetrate scientific healthcare concepts into people's social circles and social groups, and provide personalized health management services [9–11]. Given the above research base, we believe that the WeChat platform can provide an ideal virtual space for collaborative multidisciplinary interventions to overcome the difficulties in implementing real-life home interventions for patients with CHF, and is important for improving patients' self-care behaviours, improving care outcomes, and reducing the number of patient readmissions. From April 2020 to May 2021, a 3-month multidisciplinary collaborative team-directed intervention study was implemented on 56 patients with stable CHF via the WeChat platform to observe the clinical value of its evaluation on patients, which is reported below.

1.1. Study Subjects. A convenience sampling method was used to select 56 patients who were admitted to the cardiology department of our hospital from April 2020 to May 2021 and were diagnosed with CHF as the study population. Diagnostic criteria: meeting the criteria of the guidelines for the diagnosis and treatment of CHF [12]. Diagnostic criteria for cardiac function classification refer to the 1994 New York Heart Association (NY-HA) classification criteria [13]. Inclusion criteria: those who meet the western diagnostic criteria for CHF with class II-III cardiac function; age 18–75 years; systolic insufficiency with left ventricular ejection fraction (LVEF) ≤ 50% on echocardiography; those who sign the informed consent form; the clinical data of the patients was complete. Exclusion criteria: heart failure due to failure of vital organs such as the liver and kidneys; women who are pregnant or lactating, have severe skin infections or are particularly sensitive to certain drugs; patients with severe primary diseases of the liver, kidneys, endocrine system, and haematopoietic system; patients with psychiatric abnormalities and unwillingness to cooperate; patients with factors that increase mortality, such as cardiogenic shock, severe ventricular arrhythmias, complete atrioventricular block, obstructive cardiomyopathy, unrepaired valve disease, constrictive pericarditis, pericardial tamponade, pulmonary embolism, and uncontrolled hypertension (hypertensive crisis, etc.). Fifty-six patients who met the relevant diagnostic and inclusion criteria were divided into experimental and control groups using a random number table. During the study period, there was 1 case of shedding in the control group and no shedding in the experimental group. 28 patients in the experimental group, 15 males and 13 females, age 58–76 years, mean (68.56 ± 5.28) years; cardiac function status: 14 cases of class II, 10 cases of class III, and 4 cases of class IV. In the control group, there were 27 cases, 13 males and 14 females, age 60–78 years, mean (68.45 ± 5.33) years; cardiac function status: class II 15 cases, class III 9 cases, and class IV 3 cases.

1.2. Case Shedding Criteria. Those who failed to complete the whole study process due to nonchanges in the disease or other reasons; those who received other treatments that may affect the results and affected the later evaluation; and those who experienced deterioration of the disease or other serious complications during the study period that required urgent treatment.

2. Study Methods

2.1. Intervention Methods. Six experienced cardiology nurses with solid professional knowledge were responsible for case management and all had worked in the specialty for more than one year, with a college degree or above and nurse practitioner or above titles. The research team, based on an extensive literature review, combined with the practical work experience of the group members in CHF diagnosis and management nursing, prepared a CHF nursing instruction manual, including knowledge related to CHF disease, knowledge of medication management, home oxygen therapy, breathing techniques, exercise training, and management of acute exacerbation period.

2.2. Control Group. Patients were given systematic CHF treatment in accordance with their medical characteristics and the CHF Treatment Guidelines. In the control group, the nurse in charge of the patient was given an admission education after admission, and during the hospital stay, the patient was given routine nursing measures and health education, including guidance on diet, medication, lifestyle adjustment, regular follow-up, and other routine knowledge, and at the time of discharge, the patient was given routine discharge instructions, and a personalized discharge plan was formulated according to the patient's specific situation. Patients received discharge health education, the content of which was mainly individualised according to the patient's specific mastery of self-management content and practical problems, and patients were encouraged to give feedback after discharge. Patients were followed up by telephone within 1 week after discharge. To reinforce and consolidate the relevant knowledge and skills in the CHF Nursing Manual, to help patients to answer their questions and encourage them to self-manage.

2.3. Experimental Group. On the basis of the control group, a multidisciplinary teamwork based on the WeChat platform was implemented to guide the patients to self-manage at home. The intervention period was 3 months. Briefly, a CHF multidisciplinary collaborative guided intervention WeChat group was established with the head nurse of the cardiology department as the group leader and members including the patient, 2 cardiovascular physicians, 6 cardiology nurses, 1 rehabilitation physiotherapist, 1 nutritionist, and 1 psychological counsellor. Responsibilities and division of labour: cardiologists were responsible for comprehensive assessment of patients' conditions, formulation of disease diagnosis and treatment, and management plans; nurses were case managers, each responsible for 5 patients, and established personal WeChat contact with patients. In addition to sending WeChat educational content to patients, they were also responsible for reminding patients to take their medication on time through WeChat, guiding patients to adhere to medical advice, writing the name, method of administration and effect of medication on the medication box, and not to stop taking medication on their own or missing medication. For some patients with poor memory and missed doses, patients were advised to use reminders such as alarm clocks, reminder cards or links to important events or objects in their daily lives to urge them to take their medication on time. In addition, patients were educated to recognise early symptoms such as occult oedema, reduced activity endurance, increased exertional dyspnoea, increased peripheral oedema, or a resting heart rate that was 15–20 beats/min faster than usual and to consult a doctor promptly in order to control the progression of the disease. The rehabilitation physiotherapist was responsible for setting rehabilitation goals and rehabilitation programmes, formulating individualised exercise programmes according to the classification of cardiac function, and encouraging patients to exercise appropriately. For patients in cardiac function class IV, the nurse instructed bedside physical activities and bedside sitting up exercises; for patients in cardiac function class III, they were given bedside chair training and bedside walking assistance; for patients with class II cardiac function, they were instructed to train ward aisle walking, walking, and stair walking, to gradually increase the amount of exercise, all without exertion, and to control their heart rate to 110 beats/min or less after activity, after discharge from hospital, they were instructed to take appropriate exercises such as jogging, cycling, and soft gymnastics, and gradually transition to general physical activities. In addition, we should avoid inducement, prevent upper respiratory tract infection caused by catching cold to aggravate heart failure, maintain emotional stability, avoid fatigue, defecated regularly, keep defecation unobstructed, do not hold breath and exert too much force during defecation, so as not to lead to heart failure, aggravation, or even sudden death. The dietitian assessed the patient's nutritional status, formulated a nutritional support plan, and teached the patient and family to use a beer cap or coke cap and measuring cup for salt and water and to use <5 g/d of salt for cardiac function class II, <2.5 g/d for cardiac function class III, and <1 g/d for cardiac function class IV, or to avoid salt. If taking diuretics, excessive salt restriction in the diet was not necessary. The dietitian also provided a list of common foods that contained water so that patients could calculate their daily intake, provided a list of common foods to facilitate the calculation of daily water intake, and instructed the patient to eat a light and easily digestible diet rich in high quality protein and vitamins. The psychological counsellor provided regular psychological counselling to patients, guided them in self-regulation and emotional control, actively assisted them to participate in the appropriate medical insurance so that they had sufficient expenses to ensure smooth treatment, conducted psychological assessment of patients, provided targeted psychological care for patients, and at the same time encouraged family members to communicate more with patients. Reduced the patient's anxiety and helped him/her to build up the confidence to overcome the disease. On the other hand, the nurse, as a case manager, sent a piece of knowledge and linked about CHF disease to the patients under her charge through WeChat every week. The content was more detailed and comprehensive than the CHF nursing instruction manual, which was discussed and formulated by the team. The case management nurses collected the information separately, and after the team consultant physician had reviewed and agreed, the case management nurses operated and sent it, and the nurse manager implemented supervision of the pushing process. The push was sent in the form of text, pictures, videos, or voice messages, and was sent at a fixed time, once a week for 3 months. Patients were followed up through phone twice a week to solve their problems and difficulties.

2.4. Observation Indicators

2.4.1. Patient's Ability to Care for Themselves. Self-Care of Heart Failure Index Version 6.2 (SCHFI v6.2). This scale was developed by Riegel et al. [14], based on the Self-Management Scale. The scale consists of 22 items in 3 dimensions: self-care maintenance (10 items), self-care management (6 items), and self-care confidence (6 items), using a 4-point Likert scale with a total score of 22 to 88, with higher scores indicating better self-care status of the patient.

2.4.2. Patient Quality of Life. The Minnesota Heart Failure Quality of Life Questionnaire (MLHFQ) [15]was used to assess patients' quality of life using this scale. The scale includes physical, fatigue, emotional, and economic aspects, with 21 entries, using a Likert 6-point scale with a total score of 105. The assessment time was between 5 and 10 min. Higher scores indicated a poorer quality of life for the patient.

2.4.3. 6MWT. The main objective was to measure the distance of a 6-minute walk, which was an important indicator of objective assessment of cardiac function in patients with CHF. Each patient was picked up by a case management nurse at discharge and at the end of 3 months to the unit where the 6MWT test was performed under medical

supervision. The evaluation was performed according to the 2002 American Thoracic Society (ATS) guidelines for the application of the 6-minute walk test [16]. Evaluation criteria: in 1993 Bittner classified the results of the 6MWT into 4 levels according to the condition: level 1 < 300 m, level 2 300–375 m, level 3 375–450 m, and level 4 > 450 m, with lower levels indicating poorer cardiopulmonary function. Termination criteria: the experiment was terminated by the presence of any of the following: chest pain, dyspnoea, lower limb cramps, pallor, etc. [17].

2.4.4. Readmission Rate. This was the incidence of rehospitalisation for exacerbation of preexisting disease in the same patient after discharge from hospital. Patients who were hospitalised within 3 months of discharge for exacerbation of chronic heart failure for any reason were considered to have been readmitted. Excluding hospital admissions for surgical conditions, a rehospitalisation within 7 days of discharge was considered a single hospitalisation for the same patient.

2.5. Data Collection Methods. Prior to the survey, the first author of this study provided uniform training to the investigators. The SCHFIv6.2 and MLHFQ scales were administered one-to-one with patients' consent at admission, at discharge and 3 months after discharge, respectively; patients' readmission rates were counted 3 months after discharge. Patients were assisted to complete the questionnaire without the use of suggestive words.

2.6. Statistical Methods. All data were analysed using SPSS 19.0 statistical software. Measures were expressed by t-test and statistical analysis by χ^2 test.

3. Results

3.1. General Information Comparison. When comparing the baseline data of patients in the experimental and control groups, the differences in age, gender, 6MWT, NY-HA cardiac function classification, and BNP levels were not statistically significant ($P > 0.05$) and were comparable; see Table 1.

3.2. Comparison of Self-Care Capabilities. There was no significant difference between the SCHFIv6.2 scores of the two patients at admission and at discharge, and at the end of 3 months after discharge, the scores in the experimental group were significantly higher than those in the control group. At the end of 3 months after discharge, the experimental group scores were significantly higher than the scores at discharge; however, the control group scores were lower than at discharge (Table 2).

3.3. Quality of Life Comparison. There was no significant difference between the MLHFQ scores of the two patients at admission and discharge, and the scores of the experimental group were significantly higher than the scores of the control group at the end of 3 months after discharge. At the end of 3 months postdischarge, the scores in the experimental group were significantly higher than the scores at discharge; however, the scores in the control group were lower than at discharge (Table 3).

3.4. 6MWT Distance Comparison. The difference in 6MWT between the two groups at admission and discharge was not statistically significant. At the end of 3 months, 6MWT was (471.32 ± 32.43) m in the experimental group and (215.41 ± 22.56) m in the control group, excluding one case of termination of the experiment due to angina pectoris. The results showed that at the end of 3 months after discharge, the difference in 6MWT between the two groups was statistically significant, with the experimental group significantly higher than the control group. See Table 4.

3.5. Comparison of Readmission Rates. The readmission rate within 3 months was 14.28% (4 cases) in the experimental group and 44.45% (12 cases) in the control group. The results showed that the readmission rate in the experimental group was significantly lower than that in the control group, and the difference was statistically significant. Among them, one case in the experimental group was lower than two cases in the control group for rehospitalisation caused by improper medication; three cases in the experimental group were lower than nine cases in the control group for rehospitalisation caused by improper life management (infection, exertion, and diet). For details, see Table 5.

4. Discussion

CHF has become a serious public health problem worldwide, and epidemiological data in recent years has reported that the prevalence of heart failure in China is 0.9%, with 4.5 million patients suffering from heart failure and about 500,000 new cases of heart failure each year [18–20]. Various end-stage cardiac diseases can lead to heart failure to varying degrees, including coronary heart disease, hypertension, arrhythmias, heart valve disease, and cardiomyopathy. Of these, coronary artery disease and hypertension account for the highest proportion [21]. The multidisciplinary collaborative model has become the more respected model of CHF management in national and international academia [1, 22, 23]. The current multidisciplinary model of CHF is mainly implemented during the patient's hospitalisation, and it is difficult to organise a multidisciplinary team to provide comprehensive and individualised care and rehabilitation after the patient is discharged from the hospital [24–26]. With the popularity of smartphones, the WeChat platform as a new mobile media has its own advantages in the prevention and treatment of chronic diseases, such as large amount of information, wide coverage; wide range of users, rapid dissemination; peer-to-peer communication, personalized mutual assistance; and multimedia communication, low cost [10, 27, 28]. The WeChat platform is a good solution to the problem of inconsistency in time and place, and team members can communicate with each other

Table 1: Comparison of baseline information between the two groups of patients [$\overline{\chi} \pm S$, (n, %)].

Groups	n	Age	Gender (n, χ^2)		6MWT	Classification of cardiac function			BNP
		(year $\overline{\chi} \pm S$)	Male	Female	(m, $\overline{\chi} \pm S$)	II	III	IV	(pg/ml, $\overline{\chi} \pm S$)
Experimental group	28	68.56 ± 5.28	15	13	195.46 ± 20.13	14	10	4	342.13 ± 282.14
Control group	27	68.45 ± 5.33	13	14	196.52 ± 21.36	15	9	3	350.46 ± 267.25
Statistical values	—	0.078[a]	0.162[b]		0.191[a]		0.212[b]		0.113[a]
p-value	—	0.938	0.687		0.849		0.899		0.910

Note. [a] is the value of t and [b] is the value of χ^2. $P < 0.05$ means the difference is statistically significant.

Table 2: Comparison of SCHFIv6.2 scores between the two groups of patients ($\overline{\chi} \pm S$).

Groups	n	On admission	On discharge from hospital	At the end of 3 months after discharge from hospital
Experimental group	28	33.45 ± 4.65	54.23 ± 5.87	60.14 ± 5.84
Control group	27	33.56 ± 3.98	54.12 ± 5.23	50.47 ± 6.23
t-value	—	0.095	0.074	5.992
p-value	—	0.924	0.941	0.000

Note. $P < 0.05$ means the difference is statistically significant.

Table 3: Comparison of MLHFQ scores between the two groups ($\overline{\chi} \pm S$).

Groups	n	On admission	On discharge from hospital	At the end of 3 months after discharge from hospital
Experimental group	28	75.65 ± 6.58	55.42 ± 4.73	45.23 ± 3.85
Control group	27	74.81 ± 6.84	54.65 ± 5.21	55.24 ± 4.74
t-value	—	0.089	0.579	8.674
p-value	—	0.929	0.565	0.000

Note. $P < 0.05$ means the difference is statistically significant.

Table 4: Comparison of 6MWT distances between the two groups of patients ($\overline{\chi} \pm S$).

Groups	n	On admission	On discharge from hospital	At the end of 3 months after discharge from hospital
Experimental group	28	195.46 ± 20.13	250.45 ± 22.47	471.32 ± 32.43
Control group	27	196.52 ± 21.36	251.32 ± 23.14	215.41 ± 22.56
t-value	—	0.191[a]	0.143	34.278
p-value	—	0.849	0.887	0.000

Note. $P < 0.05$ means the difference is statistically significant.

Table 5: Comparison of readmission rates between the two groups within 3 months of discharge (n, %).

Groups	N	Rehospitalisation	Reason for rehospitalisation	
			Improper administration of medication	Inappropriate diet and lifestyle habits
Experimental group	28	4 (14.28)	1 (3.57)	3 (10.71)
Control group	27	12 (44.45)	2 (7.14)	9 (32.14)
χ^2- value	—	0.014	—	—
p-value	—	0.030	—	—

Note. $P < 0.05$ means the difference is statistically significant.

and collaborate in multiple disciplines in the virtual network of WeChat [29]. China's chronic disease self-management WeChat intervention projects focus on diabetes, coronary heart disease, hemodialysis, chronic hepatitis, chronic obstructive pulmonary disease, etc. [30, 31], there is less research on self-management WeChat intervention for CHF patients. This study attempted to realize the multidisciplinary cooperation of CHF through WeChat platform and guided patients to carry out family self-management through point-to-many and point-to-point methods. The research results showed that patients had high recognition for the multidisciplinary collaborative team intervention mode based on WeChat platform.

There is a widely accepted concept in academia that "self-management is the best option for chronic disease management" [32]. However, it has been reported that ≤7% of people with chronic diseases are actually self-managing worldwide. The vast majority of patients have difficulty with the question of how to self-manage their chronic disease [33]. This study addressed this substantial issue by conducting a self-management intervention guided by a multidisciplinary facilitated team via a WeChat platform in

patients with stable CHF. It was found that the multidisciplinary collaborative team intervention based on the WeChat platform all significantly improved the self-management ability of stable CHF patients and patients' quality of life, which may be related to the fact that the multidisciplinary collaborative team intervention increased patients' awareness of the disease to enhance their self-management compliance and helped them to develop good rehabilitation behaviours. In addition, after 3 months of the WeChat-based multidisciplinary team intervention, the patients' knowledge of the specialty and the 6-minute walking distance markedly improved.

According to statistics, the readmission rate of CHF patients is 24.8% within 30 d after discharge and 67.4% within 1 year [34, 35]. Reducing hospital readmissions for CHF patients has become an urgent issue for countries around the world [36–38]. A research has proved that multidisciplinary collaboration is effective in reducing readmission rates for CHF patients [39]. In the current study, we found that the patients in the experimental group, under the dynamic observation and continuous supervision of the case management nurses over a period of 3 months, mastered the adverse effects of the drugs they were taking and the importance of adhering to them, their own dietary structure, and specific intake; learned to arrange their work and rest schedules and exercise reasonably; and gradually improved their ability to manage themselves out of the hospital and their quality of life, resulting in a significant reduction in readmission rates.

Patients with severe heart, lung and kidney disease, cardiogenic shock, severe ventricular arrhythmias, complete atrioventricular block, obstructive cardiomyopathy, unrepaired valvular disease, constrictive pericarditis, pericardial tamponade, pulmonary embolism, and uncontrolled hypertension (hypertensive crisis, etc.) were excluded from this study. Therefore, the sample size was small and a larger sample would have been included in the study.

5. Conclusion

In summary, despite the improving treatment modalities for CHF, the readmission rate for patients with CHF remains high. A collaborative multidisciplinary team model based on the WeChat platform has been used in a study to reduce readmission of CHF patients, with the intervention providers being predominantly nurse practitioners and multidisciplinary teams. A discharge plan is developed for patients with CHF who are predischarged from the hospital. Postdischarge patients are given multidisciplinary guidance on medication, diet, health, and psychological guidance, and enhanced management through a microsoft platform with real-time follow-up and feedback. Strengthening medication comparison and carrying out remote monitoring are conducive to facilitating timely and continuous care services for CHF patients during their transition from hospital to home, improving patients' self-management and quality of life, and reducing readmission rates.

Data Availability

The datasets used and analysed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] E. J. Brennan, "Chronic heart failure nursing: integrated multidisciplinary care," *British Journal of Nursing*, vol. 27, no. 12, pp. 681–688, 2018.

[2] J. Spinar, L. Spinarova, and J. Vitovec, "Pathophysiology, causes and epidemiology of chronic heart failure," *Vnitrní Lékarství*, vol. 64, no. 9, pp. 834–838, 2018.

[3] C. de Gregorio, "Physical training and cardiac rehabilitation in heart failure patients," *Advances in Experimental Medicine & Biology*, vol. 1067, pp. 161–181, 2018.

[4] C. M. Campbell, R. Kahwash, and W. T. Abraham, "Optimizer Smart in the treatment of moderate-to-severe chronic heart failure," *Future Cardiology*, vol. 16, no. 1, pp. 13–25, 2020.

[5] X. Q. Li, "Effects of high-quality nursing care on psychological outcomes in patients with chronic heart failure," *Medicine (Baltimore)*, vol. 98, no. 41, Article ID e17351, 2019.

[6] M. Berggren, A. Karlsson, N. Lindelof et al., "Effects of geriatric interdisciplinary home rehabilitation on complications and readmissions after hip fracture: a randomized controlled trial," *Clinical Rehabilitation*, vol. 33, no. 1, pp. 64–73, 2019.

[7] N. M. L. Rasmussen, K. Belqaid, K. Lugnet, A. L. Nielsen, H. H. Rasmussen, and A. M. Beck, "Effectiveness of multidisciplinary nutritional support in older hospitalised patients: a systematic review and meta-analyses," *Clinical Nutrition ESPEN*, vol. 27, pp. 44–52, 2018.

[8] W. M. Vaughn, P. K. Bunde, K. Remick-Erickson, S. Rebeck, and D. Denny, "Forging multidisciplinary collaboration to improve mental/behavioral health," *NASN School Nurse*, vol. 32, no. 5, pp. 298–301, 2017.

[9] K. Zhou, W. Wang, W. Zhao et al., "Benefits of a WeChat-based multimodal nursing program on early rehabilitation in postoperative women with breast cancer: a clinical randomized controlled trial," *International Journal of Nursing Studies*, vol. 106, Article ID 103565, 2020.

[10] J. He and J. Xia, "Effect of a WeChat-based perioperative nursing intervention on risk events and self-management efficacy in patients with thyroid cancer," *Am J Transl Res*, vol. 13, no. 7, pp. 8270–8277, 2021.

[11] S. H. Zhou, S. T. Huang, N. Xu, L. W. Chen, and Q. Chen, "Application of the wechat platform to implement continuous nursing for patients after percutaneous coronary intervention," *Medical Science Monitor*, vol. 26, Article ID e925444, 2020.

[12] P. Ponikowski, A. A. Voors, S. D. Anker et al., "ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC," *European Journal of Heart Failure*, vol. 18, no. 8, pp. 891–975, 2016.

[13] C. Bredy, M. Ministeri, A. Kempny et al., "New York Heart Association (NYHA) classification in adults with congenital

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 7906135, 6 pages
https://doi.org/10.1155/2022/7906135

Research Article

Analysis and Prediction of Cross-Border e-Commerce Scale of China Based on the Machine Learning Model

Qiaoping Chen

School of Foreign Studies, Yiwu Industrial and Commercial College, Yiwu, Zhejiang 322000, China

Correspondence should be addressed to Qiaoping Chen; chenqiaoping@ywicc.edu.cn

Received 1 June 2022; Revised 9 July 2022; Accepted 18 July 2022; Published 23 August 2022

Academic Editor: Man Fai Leung

In the context of the rapid development of Internet technology, the integration of the world economy has been strengthened, and the continuous innovation of technology and foreign trade business forms has promoted the rapid development of cross-border e-commerce. Due to the lack of relevant data on cross-border logistics empirical research, this paper conducts a prediction study on the scale of China's cross-border e-commerce market based on machine learning models and combines the relevant financial reports of listed companies to determine the proportion of performance costs to turnover. Forecast of the scale of cross-border e-commerce in China. Combined with the total economic volume, industrial structure, domestic and foreign trade, online shopping development, people's life, and express development, the index system is established, and 11 indicators are initially established with reference to the selection principle of indicators. Combined with the research object, multiple regression and gray prediction methods are established. The relevant prediction model is tested, and the established model is tested to ensure the prediction accuracy. The forecast results show that by 2027, the size of China's cross-border e-commerce market will reach 30.8133 trillion yuan.

1. Introduction

Since the establishment of the Shanghai Free Trade Zone in 2013, China has established 11 free trade zones including Guangdong and Tianjin. In 2015, China proposed the Belt and Road development strategy. In the context of the national macro development strategy of the Belt and Road Initiative and the Pilot Free Trade Zone, China's cross-border e-commerce has become a dark horse in foreign trade. According to customs statistics, China's total import and export of goods in 2016 was 24.33 trillion yuan, a year-on-year decrease of 0.9 percentage points. While China's import and export trade declined, the scale of China's cross-border e-commerce transactions reached 6.3 trillion yuan in 2016, an increase of 23.5%. In the face of the rapid development of cross-border e-commerce, the State Council has also issued relevant guidelines to make an overall plan for the development of cross-border e-commerce, and the development of cross-border e-commerce in China has also entered a new stage of development.

Premier Li Keqiang emphasized a series of achievements in opening up in recent years and reviewed the progress of the Belt and Road Initiative. The inclusion of RMB into the SDR marks a new level of internationalization. Shenzhen-Hong Kong Stock Connect was launched import and export levels stabilized. The advanced experience of the Shanghai Free Trade Zone is also being comprehensively promoted in an orderly manner. Thanks to the promotion of policies, China's foreign trade development level has developed in a favorable direction as a whole. In the face of the complicated foreign trade situation, cross-border e-commerce, as a special form of foreign trade, has drawn much attention to its operation. Coupled with the active guidance and support of policies, the research on cross-border e-commerce is very valuable. Cross-border e-commerce logistics is an important part of cross-border e-commerce transactions, and the two depend on each other. At present, there are many obstacles and problems in cross-border logistics, which hinder the improvement of consumer experience and thus limit its development. Therefore, it is urgent to carry out research on cross-border logistics.

Foreign research on forecasting started earlier. In 1983, when Makridakis was studying problems related to time series, he conducted an empirical study on combined forecasting for the first time, and finally came to the conclusion that combined forecasting can effectively improve forecasting accuracy, and the more methods, the more accurate the forecast. The more accurate the prediction results [1]. In terms of the application of prediction methods, Terasvirta et al. [2] in 2004 used the monthly macroeconomic data of 7 countries for 47 months, using the smooth transition autoregressive model (STAR) and neural network (NN) model. In 2007 [2], Aburto and Weber [3] took the demand of supermarket commodities in Chile as the research object, using the ARIMA model and neural network. The model makes combined predictions, and the prediction results help supermarkets to effectively reduce the inventory level [3], and the practical application of prediction has achieved corresponding economic benefits. In 2017, facing the dynamic logistics demand, Sheu and Kundu [4] proposed a three-tier supply chain framework based on the Marr model. The logistics interaction model of the Kov chain and the validity of the established model were tested with the help of the logistics data of PetroChina, and the results proved the reliability of the model [4]. In terms of cross-border e-commerce, Gomez-Herrera et al. [5] in 2014 discussed the driving force and obstacle factors in the development of EU cross-border e-commerce. Studies have shown the convenience of cross-border e-commerce transactions, but also increase other related costs [5]. In 2015, Asosheh et al. [6] proposed a localization process for cross-border B2B e-commerce. The paper divides the process of cross-border B2B into three layers: message layer, business process layer, and content layer. Each level must have corresponding standard processes and corresponding solutions. Process optimization helps cross-border B2B e-commerce operate efficiently [6]. In terms of cross-border logistics, in 2013, Davis and Friske [7] took the United States and Canada as examples to study the impact of public-private partnerships on cross-border logistics using grounded theory. The quality of the operation of the environment logistics [7]. In 2014, Wong et al. [8] took the Pearl River Delta region of China as the research object and proposed the importance of bonded warehouses for cross-border logistics. On this basis, a PMS performance evaluation framework is proposed. The PMS framework includes four dimensions: time, cost, quality, and flexibility. Research shows that the PMS framework can help enterprises to improve competitiveness and resource allocation efficiency [8]. In 2016, Ai et al. [9] studied for the performance evaluation of cross-border logistics, a corresponding performance evaluation model is constructed using the two-sided market theory. On the other hand, it points out that cross-border payment is the bottleneck of cross-border e-commerce development and puts forward corresponding policy suggestions on cross-border payment [9].

The development speed of China's cross-border e-commerce is eye-catching, especially in academia. This field is a new research direction in recent years. Noticing the importance of cross-border logistics, more and more scholars are focusing on research-related issues of cross-border logistics [10–15]. The current research mostly focuses on the analysis of the status quo, solutions (overseas warehouses, bonded warehouses), and policy recommendations, and there is little empirical analysis [16, 17]. Therefore, the research purposes of this paper are two: one is to establish a corresponding market scale forecasting system, and with the help of the corresponding index system, to do a current research and short-term forecast on the scale of China's cross-border e-commerce market. The second is to make corresponding predictions on the scale of China's cross-border e-commerce logistics market. The theoretical significance of the research is to enrich the application scope of multiple regression forecasting and gray forecasting. Both prediction models are classic models, and when faced with new problems, the classic prediction models are adopted to enrich their applications. Another aspect of theoretical significance lies in enriching the empirical research on China's cross-border logistics. The empirical research based on data has reliable data sources, and the research conclusions are objective and have good guiding significance. The forecast of China's cross-border e-commerce market size and cross-border e-commerce logistics has important reference value for the government to formulate development plans for cross-border logistics and integrate logistics resources. With the deepening of China's opening to the outside world, the economic pillar effect of foreign trade has become more and more obvious. As a form of trade with the fastest growth in recent years, cross-border e-commerce has a good impact on the development of China's foreign trade guiding value.

2. Theoretical Background

2.1. Cross-Border e-Commerce. Cross-border e-commerce refers to a business form in which buyers and sellers in different countries use cross-border e-commerce platforms to reach purchase agreements, buyers pay for goods, and sellers are responsible for delivery and distribution. In a broad sense, it can be considered that foreign trade e-commerce means that buyers and sellers in different customs will digitize the traditional foreign trade process, complete the distribution with the help of international logistics, and finally achieve the purpose of the transaction. In a narrow sense, we can understand cross-border e-commerce as international retail. That is, the transaction activities of consumers in one country purchasing goods from other countries through cross-border e-commerce platforms. Cross-border e-commerce can be divided into cross-border export and cross-border import according to the logistics direction. In the cross-border export process, foreign goods are finally delivered to consumers or enterprises through e-commerce platforms, international payment companies, domestic and foreign customs, and domestic and foreign logistics companies. The cross-border import process is similar to the export process, but in the opposite direction. There is currently no mature theoretical system to support cross-border e-commerce, but the essence of cross-border e-commerce is international trade, and its

economic essence lies in the international relationship between production factors circulation. The economic connotation of the development of cross-border e-commerce lies in reducing the mismatch between supply and demand markets, reducing the transaction costs of producers and consumers in the global market, and optimizing the allocation of resources.

2.2. Heckscher–Ohlin Model (H-O Model). Heckscher–Ohlin (H-O) Model is the basic theory of international trade, it points out that resource difference is the fundamental reason for trade. In the process of international trade, a country should export its relatively rich products and import scarce products. The economic principle of the H-O model is shown in Figure 1. Suppose two countries realize commodity circulation through international trade, one country exports commodity X and imports Y, and the other country exports commodity Y and imports X. As the import and export proceed, the prices of the two commodities change accordingly, thus affecting the consumption. Equilibrium is achieved when the relative prices of the same commodity in the two countries are the same. The result of the equilibrium is that both sides of the trade increase the level of domestic consumption and gain benefits from international trade. For China, cross-border e-commerce, as a new form of international trade, can give full play to the advantages of China's resources and population and provide assistance for China's transformation and development.

2.3. Cross-Border Logistics. Cross-border logistics refers to logistics services carried out between different countries. From the definition of cross-border logistics, it can be seen that the two sides of the cross-border logistics trade cannot come from countries, and the goods traded need to go out of the country where the sender is located, and then complete the entry and exit procedures after arriving in the country where the consignee is located, and finally delivered to consumers through logistics and distribution. The entire process of cross-border logistics can be divided into three parts, domestic logistics in the country where the goods are sent, international logistics, and logistics and distribution in the country where the consumer is located. The difference between cross-border logistics and domestic logistics is that it needs to go through the domestic and destination country customs and will be subject to customs supervision. Therefore, the operation process is more complicated, and the operation requirements of enterprises are higher. At present, few enterprises can complete all cross-border logistics alone operate.

Compared with domestic logistics, the characteristics of cross-border logistics are as follows: (1) Cross-border logistics has a long timeliness. The most significant difference between cross-border logistics and domestic logistics is the geographical distance between the origin and the destination. In general, the transportation mileage of cross-border logistics is significantly longer than that of domestic logistics, which determines that the timeliness of logistics is difficult to guarantee. (2) The cost of cross-border logistics is high. Due to the increase in transportation mileage, the cost of logistics and transportation has increased significantly, eroding the profits of cross-border e-commerce, and thus increasing commodity prices and affecting consumer demand. (3) The cross-border logistics process is complex. Cross-border logistics involves cross-border trade, which involves a series of links such as domestic first-way delivery, customs clearance, and destination delivery. A flaw in one of these links will directly lead to the failure of the transaction.

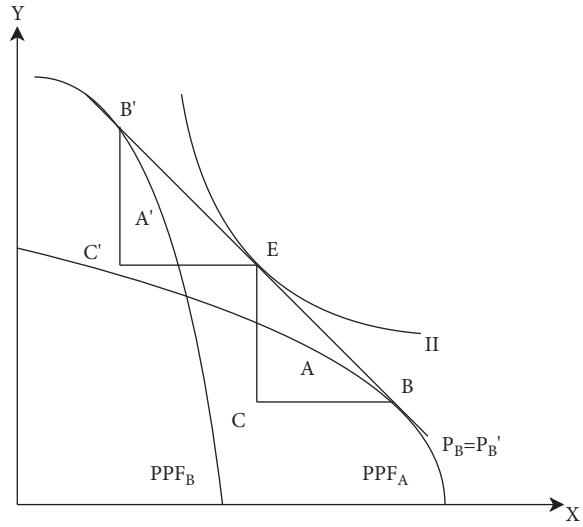


Figure 1: Heckscher-ohlin model.

3. Analysis and Prediction of Cross-Border e-Commerce Scale of China Based on Machine Learning Model

3.1. Establishment of a Forecasting Indicator System for the Scale of China's Cross-Border e-Commerce Market. For the scale of China's cross-border e-commerce market, we can use the corresponding statistical indicators to describe the scale of China's cross-border e-commerce transactions. The gross domestic product (GDP) is used to describe the economic aggregate index in the previous literature, and this index is also used in this paper. As far as industrial structure indicators are concerned, we usually use the added value of the tertiary industry or the ratio of the tertiary industry to represent the industrial structure of a specific region. For domestic and foreign trade, we use two indicators: total retail sales of consumer goods and total imports and exports. The Internet development indicators closely related to the development of China's cross-border e-commerce are described by the number of Internet users and the scale of online shopping users. We use the express volume and express business income to study the indicators of China's express delivery development. In addition, we use the per capita disposable income and consumption level of urban residents to describe the economic level of consumers. Therefore, we initially obtained the relevant indicators for the prediction of the scale of China's cross-border e-commerce market, as shown in Table 1.

Table 1: Cross-border e-commerce market scale forecast indicator system.

The target layer	The primary indicators	The secondary indicators
China cross-border e-commerce market size forecast	Economic aggregate	Gross domestic product
	Industrial structure	The added value of the primary industry
		The added value of the secondary industry
		The added value of the tertiary industry
	Domestic and foreign trade	The total retail sales of social consumer goods
		Total import and export
	Online shopping development	Scale of online shopping users
	people's life	Per capita disposable income of urban residents
		Resident consumption level
	Express development	Express volume
		Express business income

Table 2: Gray correlation degree.

Index	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Correlation	0.6638	0.6455	0.6423	0.6904	0.6973	0.6126	0.8512	0.6521	0.6688	0.8014	0.9447

3.2. Gray Relational Model. The gray correlation degree is a measurement tool to judge the degree of correlation between indicators. The principle is to judge the geometric similarity between the research sequence and the target sequence, so as to obtain the degree of association between the target systems. The precondition for calculating the gray correlation degree is to clearly describe the characteristic behavior variables and related factor variables of the research system and then determine the sequence of its corresponding variables according to the actual sample data. And build a matrix, and finally analyze the results according to the expertise in the target field.

The correlation coefficient refers to the degree of correlation between each target indicator and the reference indicator in each year, so there will be a value for each year. This results in too much information, which cannot well reflect the overall correlation between the influencing factor and the predictor at all times. Therefore, it is necessary to integrate the correlation coefficients of each year into a single value, so as to facilitate the comparison between the various factors. In this paper, the average value of each index over the years is selected to represent the degree of correlation, and the specific formula is:

$$R_i = \frac{1}{k}\sum_{i}^{n} p_i(k), \tag{1}$$

where R is the degree of association, k is the number of indicators, and p is the impact value. Here, we take the average of the correlation coefficients of each indicator over the years as the final indicator correlation degree. As can be seen from Table 2, the priority order of strength and weakness affecting the scale of China's cross-border e-commerce is: express delivery business revenue X_{11}, online shopping user scale X_7, express delivery volume X_{10}, total retail sales of consumer goods X_5, tertiary industry added value X_4, household consumption level X_9, GDP X_1, and per capita disposable income of urban residents X_8. The correlation between these 8 factors and the scale of cross-border e-commerce in China is more than 0.65. It can be seen from Table 2 that these 8 factors have a high correlation with the market scale. In the following prediction model modeling of market size, in order to improve the prediction accuracy, according to the strong correlation principle of index selection, this paper selects 8 factors whose gray correlation degree is greater than 0.65 among the above 11 indicators as independent variables for modeling analyze.

3.3. Multiple Linear Regression Model for Predicting the Scale of China's Cross-Border e-Commerce Market. The first step in regression analysis is to identify independent and dependent variables. Usually, the object to be studied is the dependent variable, and the relevant factors are the independent variables. In this paper, the dependent variable is the size of China's cross-border e-commerce market, and the independent variables are the eight indicators selected in the previous section. Therefore, we construct a multiple regression model. Due to the multivariate brings computational difficulties. In order to solve this problem, this paper uses the data analysis software SPSS to complete the calculation. Let the random variable Y and the general variables $X_1, X_2, \ldots, X_k$ have the following linear relationship:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \mu, \tag{2}$$

where $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the regression coefficients, Y is the explained variable, X is the explanatory variable, and u is the random error. Use the historical data and time T of each indicator to establish a univariate regression model, and finally obtain the forecast results from 2023 to 2027 with the help of formula (2). Combining with Table 3, it can be seen that by 2027, the size of China's cross-border e-commerce market will exceed 20 trillion mark, reaching 30.8133 trillion yuan. It shows that under the current macroeconomic situation and the slow recovery of the world economy, the scale of China's cross-border e-commerce market will achieve steady growth, and it also verifies that the forecast results are in line with objective facts.

Table 3: Forecast result.

Time	2023	2024	2025	2026	2027
Value	20.61024	22.71897	25.22368	27.57899	30.81331

4. The Impact of Intelligent Tourism Loop on Traffic between Scenic Spotsevelopment Environment of China's Cross-Border e-Commerce

4.1. Political Environment. With the development of e-commerce in China, the development of domestic e-commerce has entered a period of stable development. At this stage, the demographic dividend is no longer obvious. In order to pursue profit growth, domestic e-commerce companies will focus more on the shopping experience of consumers. In this process, the government's regulatory policy has played a crucial leading role from beginning to end. Therefore, in the process of studying cross-border e-commerce, we must pay attention to policy guidelines and policy orientations. The development of cross-border e-commerce is also inseparable from policy support to a certain extent.

As far as China is concerned, large-scale cross-border e-commerce emerged relatively late. Compared with domestic e-commerce, the development of cross-border e-commerce lags behind domestic e-commerce due to its special transaction model. However, in recent years, due to the attention of the capital market., the momentum of development is fierce, injecting new vitality into China's cross-border trade. In order to regulate the development of the industry, China has also issued corresponding policies and regulations in recent years. The promulgation of these policies and regulations not only points out the direction for the development of cross-border e-commerce, but also improves the legal environment, so that there are laws and rules to follow when dealing with relevant conflicts and disputes.

After interpreting the relevant policies, we found that the national macro level has a positive attitude toward the development of cross-border e-commerce in China and has created many convenient conditions. It can be seen that the development of cross-border e-commerce in China in recent years has attracted the attention of the state, and its impact on China's foreign trade has also been affirmed by the state. At the same time, the country has also seen discordant aspects in the development process and has issued relevant management rules and regulations. Under the guidance and supervision of policies, China's entry into the e-commerce market will inevitably make great strides in the direction of healthy development. After a transitional period, the new cross-border e-commerce policy will inevitably lead the industry to gradually become standardized. Under the guidance of policies and regulations, some backward small enterprises will inevitably be eliminated, corresponding market-leading enterprises will be cultivated, and industry innovation will be led to lead industry norms. In terms of delivery mode, bonded warehouses have natural logistics advantages, and centralized procurement of explosives is conducive to large-scale operations and lower operating costs, which is conducive to the healthy development of the cross-border e-commerce industry.

4.2. Economic Environment. The internationalization of the RMB aims to promote the international status of the RMB as a globally recognized currency for pricing, settlement, and reserves. In 2016, the RMB was officially included in the SDR. There are currently five currencies in the SDR currency basket, of which the US dollar accounts for the highest proportion, reaching 41.73%, followed by the euro, accounting for 30.93%, and the third is the RMB, accounting for 10.92%, the last Japanese yen (8.33%), and the British pound (8.09%).

The inclusion of RMB into the SDR is of strategic significance, and at the same time, it further accelerates the process of RMB marketization. The addition of the RMB to the SDR not only enhances the international currency status of the RMB, but also enhances the pricing power of the RMB in commodities. The impact of RMB internationalization on China's cross-border e-commerce is reflected in the payment and settlement links, and the corresponding currency advantages will gradually emerge. The internationalization of the RMB has provided very convenient basic conditions for the development of cross-border e-commerce in China.

Thanks to the continuous development of the global economy and the deepening of the degree of integration, cross-border e-commerce, as a brand-new trading method, benefits from the simplicity of its transaction links and reduces the operating costs of foreign trade price advantage. In particular, the Belt and Road national strategy has created a new situation for China's foreign trade. Taking this opportunity, China can allow more enterprises to go to the world through cross-border e-commerce and promote enterprises development to accelerate transformation and upgrading in the new foreign trade situation.

4.3. Social Environment. There are many reasons for the upgrading of people's consumption concept, one is the progress of society, the other is the rise of living standards, and the third is the improvement of the level of education for all. When domestic consumer goods cannot meet people's consumption needs, people naturally turn their attention to the world and look for satisfactory products around the world. Cross-border e-commerce provides convenient conditions for meeting people's needs, allowing people to understand commodity information around the world without leaving home, and meeting people's diversified consumption needs. The development of information technology and the change of consumption concepts have made Online shopping an extremely common thing and have been recognized by more and more consumers. Therefore, the upgrading of the consumption concept provides convenient conditions for the development of the industry.

The improvement of networking infrastructure and the popularization of mobile Internet have brought great network convenience to people. China's network infrastructure has developed from the earliest narrowband to the later low-

speed broadband to the current high-speed optical fiber communication. China's Internet broadband level has developed by leaps and bounds, and the network scale and the scale of netizens are also expanding year by year. In terms of mobile Internet, with The big investment in 4G networks of the three major operators and the continuous expansion of the coverage of wifi hotspots in cities and operators have greatly improved the surfing environment of mobile terminals. The improvement of Internet infrastructure is an essential prerequisite for the entire e-commerce industry.

5. Conclusion

The research on China's cross-border e-commerce and cross-border logistics is still in its infancy, and the relevant research literature is still relatively small. This paper refers to the selection of indicators and method applications for regional logistics forecasting in the past, selects appropriate indicators to analyze China's cross-border e-commerce market, and indirectly predicts the market size of cross-border e-commerce logistics by predicting the scale of China's cross-border e-commerce market. The work done in this paper is as follows: (1) Indicator selection. In the process of index selection, we mainly refer to the relevant indicators of previous logistics forecasting and adjust them according to the characteristics of cross-border logistics. The corresponding indicators are initially selected from five aspects: total volume, industrial structure, online shopping development, people's life, and express delivery development. (2) Index test. Due to the strong subjectivity in the process of index selection, this paper selects gray correlation analysis to test the degree of correlation between each index and the size of the cross-border e-commerce market. Through the test, it is found that 8 indicators have passed the correlation analysis. In the forecasting process, we use these 8 indicators for analysis and forecasting. (3) Choose a forecasting method. There are also many commonly used market size forecasting methods. This paper refers to the previous research literature and combines its own academic ability to select multiple regression forecasting models and gray forecasting models for forecasting research. Based on the regression forecasting results and gray forecasting results, the final cross-border forecast is completed through comprehensive integrated forecasting. E-commerce market size forecast.

Data Availability

The raw data supporting the conclusions of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] W. Makridakis, "Forecasting freight transportation demand with the space-time multinomial pro ait model," *Transportation Research Part A*, vol. 34, pp. 403–418, 1983.

[2] T. Teräsvirta, D. van Dijk, and M. C. Medeiros, "Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination," *International Journal of Forecasting*, vol. 21, no. 4, pp. 755–774, 2005.

[3] L. Aburto and R. Weber, "Improved supply chain management based on hybrid demand forecasts," *Applied Soft Computing*, vol. 7, no. 1, pp. 136–144, 2007.

[4] J. B. Sheu and T. Kundu, "Forecasting time-varying logistics distribution flows in the One Belt-One Road strategic context [J]," *Transportation Research Part E Logistics & Transportation Review*, vol. 117, pp. 5–22, 2017.

[5] E. Gomez-Herrera, B. Martens, and G. Turlea, "The drivers and impediments for cross-border e-commerce in the EU," *Information Economics and Policy*, vol. 28, no. 1, pp. 83–96, 2014.

[6] A. Asosheh, H. Shahidi-Nejad, and H. Khodkari, "A model of a localized cross-border E-commerce," *iBusiness*, vol. 04, no. 02, pp. 136–145, 2012.

[7] D. F. Davis and W. Friske, "The role of public-private partnerships in facilitating cross-border logistics: a case study at the U.S./Canadian border," *Journal of Business Logistics*, vol. 34, no. 4, pp. 347–359, 2013.

[8] D. W. C. Wong, K. L. Choy, H. K. H. Chow, and C. Lin, "Assessing a cross-border logistics policy using a performance measurement system framework: the case of Hong Kong and the Pearl River Delta region," *International Journal of Systems Science*, vol. 45, no. 6, pp. 1306–1320, 2014.

[9] W. Ai, J. Yang, and L. Wang, "Revelation of cross-border logistics performance for the manufacturing industry development," *International Journal of Mobile Communications*, vol. 14, no. 6, p. 593, 2016.

[10] R. H. Coase, "The nature of the firm," *Economica*, vol. 4, no. 16, pp. 386–405, 1937.

[11] Ş. Erengüç, N. C. Simpson, and A. J. Vakharia, "Integrated production/distribution planning in supply chains: an invited review," *European Journal of Operational Research*, vol. 115, no. 2, pp. 219–236, 1999.

[12] L. Wang, J. Yang, and S. Yin, "Electronic commerce international logistics performance influence factor analysis," *International Journal of Mobile Communications*, vol. 13, no. 5, p. 498, 2015.

[13] S. Moodley, "Steering in uncertain territory: E-Business, globalisation and the South African automotive components industry," *South African Journal of Economic and Management Sciences*, vol. 5, no. 1, pp. 123–153, 2002.

[14] S. Moodley and M. Morris, "Does e-commerce fulfil its promise for developing country (South African) garment export producers?" *Oxford Development Studies*, vol. 32, no. 2, pp. 155–178, 2004.

[15] J. Sutanonpaiboon and A. M. Pearson, "E-commerce adoption: perceptions of managers/owners of small- and medium-sized enterprises (SMEs) in Thailand," *Journal of Internet Commerce*, vol. 5, no. 3, pp. 53–82, 2006.

[16] P. Chatzoglou, D. Chatzoudes, and N. Kipraios, "The impact of ISO 9000 certification on firms' financial performance," *International Journal of Operations & Production Management*, vol. 35, no. 1, pp. 145–174, 2015.

[17] C. Nan and Y. Yang, "The impact of customer experience on consumer purchase intention in cross-border E-commerce——taking network structural embeddedness as mediator variable [J]," *Journal of Retailing and Consumer Services*, vol. 59, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1405139, 11 pages
https://doi.org/10.1155/2022/1405139

Research Article

Research on a Machine Learning-Based Method for Assessing the Safety State of Historic Buildings

Xiao-Hong Peng[1] **and Zi-Hao Zhang**[2]

[1]*School of Architecture, Anhui Science and Technology University, Bengbu, Anhui 233000, China*
[2]*School of Architecture, South China University of Technology, Guangzhou, Guangdong 510000, China*

Correspondence should be addressed to Xiao-Hong Peng; pengxh@ahstu.edu.cn

Received 23 June 2022; Revised 18 July 2022; Accepted 25 July 2022; Published 23 August 2022

Academic Editor: Wenming Cao

Historic and protected buildings are increasingly valued due to their valuable historical and cultural value. The assessment of the safety state of historic buildings has received more attention. Emerging machine learning algorithms, with their excellent computational performance, provide new ideas and new means to solve practical problems in various fields. Therefore, this paper proposes a method for assessing the safety state of historic buildings based on machine learning techniques. Firstly, based on the analysis of the characteristics of historical buildings and common security problems, the application of wireless sensor networks to the security monitoring of historical buildings is proposed in order to improve the automation of monitoring. Then, in order to improve the accuracy of the assessment, a combination of kernel canonical correlation analysis (KCCA) and support vector machine (SVM) is used to establish the security monitoring model. The experimental results show that by choosing a suitable KCCA function, the redundant features of the data can be reduced while the comprehensiveness of the building structure identification features can be retained, thus effectively improving the prediction accuracy of the SVM. The KCCA-SVM model can accurately predict the physical quantities such as relative structural displacement of historical buildings with good reliability.

1. Introduction

Outstanding historical buildings are either the former residences of great men and celebrities or traditional buildings with unique architectural styles and cultural connotations. These buildings are a distillation of the history of a city or region and document the architectural culture of the area. Historic and protected buildings are a proud urban landscape and a rare and valuable cultural heritage [1–4]. Therefore, we need to strengthen the protection of historic buildings through relevant laws and regulations, and at the same time adopt better technical means to protect the safety of historic buildings.

The first step should be to ensure the safety of the structure from the point of view of structural safety. On the one hand, due to the great age of these historic buildings, the performance of the building materials has deteriorated severely. There are some historical buildings that have undergone many alterations, and their use has changed. In addition, most of these historical buildings designed and built decades or centuries ago have not been considered for earthquake resistance. On the other hand, the rapid development of modern cities, the emergence of high-rise buildings, and urban metros have caused varying degrees of impact on these old buildings in the vicinity, all of which are potential safety hazards. Structural safety monitoring technology has a prominent role to play in monitoring and maintaining the safety of buildings [5–9]. Therefore, installing structural monitoring systems on historic buildings to predict their safety state is an effective technical tool.

Safety monitoring techniques differ from traditional nondestructive evaluation (NDE) techniques, which usually measure the physical state of a building structure directly [10–13]. The results of NDE evaluations depend heavily on the resolution and accuracy of the measurement equipment. Monitoring techniques, on the other hand, predict the state

of a structure based on changes in measurements at different times at the same location. Historical data are therefore crucial, and the accuracy of the predictions is dependent on the sensors and interpretation algorithms. Advanced structural safety monitoring technology is a real-time automated system that requires no human intervention and is capable of automatically assessing the safety state of a building via a local area network or remote center. It is generally accepted that a structural safety monitoring system should consist of 2 main components [14–16]: (1) a sensor system, including the selection of sensing elements and the arrangement scheme of the sensor network in the structure, and (2) a data acquisition and analysis system. The working principle of the safety monitoring system is shown in Figure 1.

It is more important to take preventive measures to protect historic buildings than to restore them in a state of imminent destruction. Existing security monitoring systems are less automated and less real time, which makes it difficult to meet the actual needs. The use of wireless sensor networks [17–19] for security monitoring of historic buildings is a more advanced technology than current methods of building security monitoring. As shown in Figure 1, one of the key steps in a security monitoring system is the security state assessment [20–22]. However, it is often difficult to accurately describe the nonlinear relationships between complex data in the security state assessment of historic buildings. With the emergence of various machine learning algorithms in recent years, machine learning algorithms are used to solve this problem and good computational results can be achieved. Based on monitoring data from wireless sensor networks, machine learning algorithms are combined with traditional security monitoring theory to fully exploit the information inherent in the monitoring data, thereby improving the accuracy of the security posture assessment of historic buildings.

The rapid development of machine learning algorithms has led to computers being able to better mimic human learning behavior. Machine learning algorithms continuously acquire new knowledge through autonomous learning and achieve self-renewal for solving new problems [23–26]. Currently, many machine learning algorithms are widely used with their good search capability and fast computing speed, providing new means to solve various problems in multiple fields. Similarly, machine learning algorithms have been widely used in traditional construction engineering, such as the application of Bayesian learning, genetic algorithms, and neural networks. Goodfellow et al. [27] used fuzzy mathematical methods to identify horizontal and vertical displacements and displacement distributions of buildings. Hejazi et al. [28] investigated the fuzzy relationship between various influencing factors and the displacement of offshore buildings. Di Napoli et al. [29] proposed the use of CNNs in the fitting and prediction analysis of building landslide monitoring data.

Compared to neural network-based algorithms, support vector machine (SVM) has obvious advantages in solving small samples, nonlinear, and high-dimensional data processing [30]. Duarte and Wainer [31] used least squares support vector machines for building deformation prediction, and the designed model has good feasibility, validity, and high prediction accuracy. Jain et al. [32] used support vector machines for building safety early warning models with high model accuracy. Tamilarasi and Prabu [33] used particle swarm algorithms to optimize support vector machines in order to perform inverse analysis of building safety model parameters. However, the SVM-based monitoring model will fully extract a large number of nonlinear features and noise, which inevitably increases the complexity of the model operations and affects the accuracy of the prediction.

Kernel canonical correlation analysis (KCCA) is an important method for multidimensional feature correlation analysis [34], in which variable features of different dimensions are correlated in order to remove redundant features. KCCA can reduce the dimensionality of variables while reducing the interference of noise, which helps to reduce the computational complexity of the monitoring model and improve the accuracy of the final prediction.

Therefore, this paper attempts to combine the two approaches to build a KCCA-SVM-based security monitoring model for historic buildings. Firstly, wireless sensor networks are applied to the security monitoring of historic buildings in order to improve the automation of monitoring. Secondly, KCCA technique is used for feature correlation analysis to reduce the dimensionality of a large amount of nonlinear data. Then, SVM's advantages in handling nonlinear and high-dimensional data are fully utilized to predict a wide range of physical quantities of historic buildings, improving the accuracy of security state assessment. The aim of this study is to automatically assess the security state of historic buildings using a KCCA-SVM-based security monitoring model. The proposed method helps to achieve automated monitoring of historic buildings over time to ensure the safety of these buildings.

The main innovations and contributions of this paper include the following.

(1) Application of wireless sensor networks to the security monitoring of historic buildings in order to improve the automation of monitoring.

(2) A historical building safety monitoring model based on KCCA-SVM is developed to address the problems of variable dimensionality and noise interference in the traditional monitoring model based on SVM. KCCA-SVM can improve the final prediction accuracy while reducing the operational complexity of the monitoring model.

The rest of the paper is organized as follows: In Section 2, the characteristics of historic buildings and common safety issues are studied in detail, while Section 3 provides the principles associated with the KCCA-SVM model. Section 4 provides the security monitoring model for historic buildings based on KCCA-SVM. Section 5 provides the project examples. Finally, the paper is concluded in Section 6.

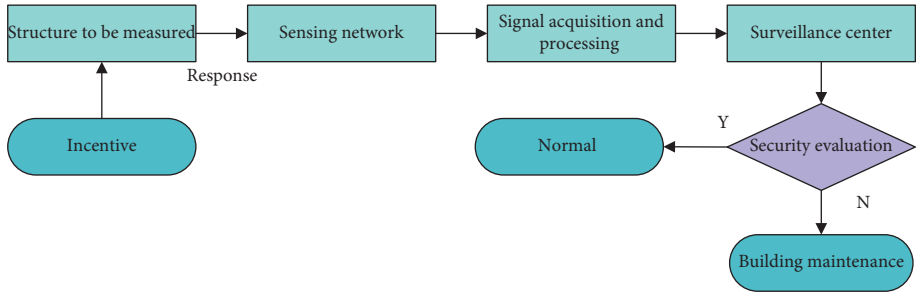


Figure 1: Working principle of the security monitoring system.

2. Characteristics of Historic Buildings and Common Safety Issues

2.1. Characteristics of Historic Buildings. As history has continued to develop, a rich and diverse range of architectural types has developed in each region and is also a visual representation of the extent of economic and cultural development. To this day, the surviving historic buildings show distinctive characteristics in terms of three aspects: architecture, architectural style, and supporting facilities.

2.1.1. Building Structure. Due to the limitations of the technical means, the historical buildings formed a structure mainly made of wood. At the same time, modern architectural structures have been added in modern times to form a complete system. Due to the direct exposure of wooden structures to the air, historic buildings are highly susceptible to spontaneous combustion in the event of prolonged heat or lightning strikes. Historic buildings are also susceptible to ignition when other open flames are present.

2.1.2. Architectural Style. According to the different architectural styles, historical buildings can be roughly divided into traditional ancient buildings and recent historical buildings. Traditional ancient buildings are mostly residential buildings, which are characterized by their layout according to the axis. Modern historical buildings have consciously retained the appearance of traditional buildings and absorbed some Western architectural styles. However, both types of historic buildings perform poorly in terms of fire resistance and seismic performance. According to the Third National Cultural Relics Census in 2011, there are 766,722 immovable cultural relics in China, of which 34.42% are in the category of ancient buildings and 18.45% are in the category of modern historical buildings, as shown in Figure 2.

2.1.3. Supporting Facilities. Firstly, due to the lack of funds for renovation and the difficulty of retrofitting, the electrical installations of many historic buildings are seriously deteriorating. Secondly, many historic buildings cannot be equipped with natural gas pipelines and therefore have to use liquefied petroleum as a domestic energy source, which

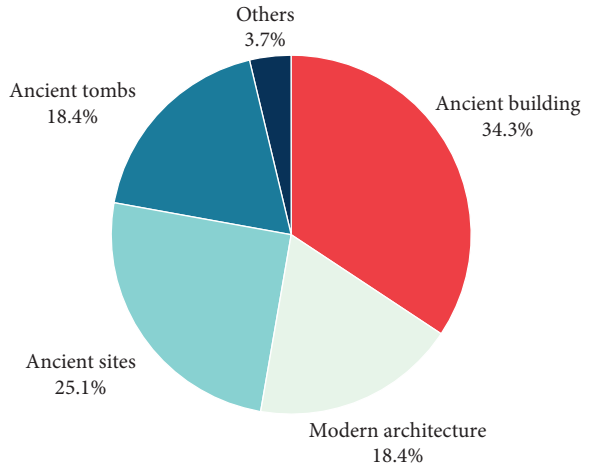


Figure 2: Percentage of types of immovable cultural objects.

undoubtedly poses a huge safety hazard. Finally, due to the age of the buildings, some of them lack effective structural support and are less able to withstand earthquakes.

2.2. Common Safety Issues. Geological and meteorological hazards do not occur as frequently as fires, but when they do, the damage to historic buildings is often very significant. Geological hazards are mainly classified into transient and slow-onset geological hazards. Earthquakes, landslides, and ground subsidence are classified as transient geological hazards. These "natural" disasters are devastating to historic buildings once they have been affected. Soil erosion and ground subsidence, on the other hand, are slow-onset disasters. These hazards are characterized by the cumulative damage that they cause to historic buildings. Once they reach a critical point, they can damage the building itself.

2.3. Wireless Sensor Network-Based Security Monitoring of Historic Buildings. Currently, traditional bus-based monitoring systems are used for the safety monitoring of historic buildings. Bus-based monitoring systems can control a variety of disaster detectors and fire-fighting equipment. However, the biggest drawback of the bus system is the wired connection, which is an inherent problem of traditional technology. In wireless monitoring systems, data are

transmitted wirelessly. No wires are required to connect the sensors to the collection units, which greatly reduces the amount of labor required for on-site installation and minimizes damage to historic buildings. The wireless sensing unit in the wireless monitoring system realizes the acquisition and wireless transmission of signals, is small in size, consumes little power, and can be battery powered.

A wireless sensor network is a multi-hop, self-organizing system that uses wireless communication. A wireless sensor network consists of multiple discrete sensor nodes randomly deployed in a monitoring area. As a new environmental monitoring technology, wireless sensor network has the advantages of real time, large range, automation, and all-weather. The use of wireless sensor networks for historical building monitoring can improve the automation of historical building safety monitoring and enhance the real-time nature of monitoring. The three common topologies of wireless sensor networks are shown in Figure 3.

For historical building safety monitoring, the environmental data to be monitored include physical quantities such as the settlement of the house and the tilt of the walls. Therefore, inclination sensors, displacement sensors, and pressure sensors need to be deployed at relevant locations. In this paper, a wireless sensor network with a tree topology is used to implement a historical building safety monitoring system. The nodes of the wireless sensor network transmit the collected physical quantities to the routing node in a multi-hop manner, and the routing node sends the data to the monitoring computer. The structure of the wireless sensor network-based historical building safety monitoring system is shown in Figure 4.

The main components of a wireless sensor network node include a microprocessor module, a wireless transceiver module, a power supply module, a debugging interface module, and a sensor module. The first four modules are common to the node, while the node has different sensor modules for different functions. The microprocessor module is the core of the sensor node, which is mainly responsible for collecting and processing local data. The sensor node controls the wireless transceiver module to complete tasks such as data transmission. The microcontroller of the node is an STM32F103 chip from STMicroelectronics. The wireless RF chip is an Atmel AT86RF231 chip. The wireless RF chip supports the IEEE802.15.4 standard, works in the 2.4 GHz band, and also supports communication protocols such as RF4CE, Zigbee, and 6LoWPAN. The wireless sensor network node is shown in Figure 5.

3. Principles Associated with the KCCA-SVM Model

3.1. Typical Correlation Analysis. Typical correlation analysis (CCA) is commonly used to quantify the correlation between two multidimensional data [35], and its main structure is illustrated in Figure 6.

Let a set of mean-zero treated samples be $X = (x_1, x_2, \ldots, x_M)$, $Y = (y_1, y_2, \ldots, y_M)$. The objective of the CCA method is to combine equations (1) and (2) to find the maximum correlation between them.

$$x^* = \boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} x, \tag{1}$$

$$y^* = \boldsymbol{\varphi}_{\mathbf{y}}^{\mathrm{T}} y. \tag{2}$$

The coefficients corresponding to the maximum value of the correlation are $\varphi(\mathbf{x})$ and $\varphi(\mathbf{y})$.

Let the covariance matrices of X and Y be C_{xx} and C_{yy}, respectively, and the mutual covariance matrix be C_{xy}.

$$\begin{aligned}
\rho &= \frac{\boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} C_{xy} \boldsymbol{\varphi}_{\mathbf{y}}}{\sqrt{\boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} C_{xx} \boldsymbol{\varphi}_{\mathbf{x}} \boldsymbol{\varphi}_{\mathbf{y}}^{\mathrm{T}} C_{xy} \boldsymbol{\varphi}_{\mathbf{y}}}}, \\
C_{xx} &= \operatorname{cov}(x, x) = E\left(x^2\right) - (E(x))^2, \\
C_{yy} &= \operatorname{cov}(y, y) = E\left(y^2\right) - (E(y))^2, \\
C_{xy} &= \operatorname{cov}(x, y) = E(xy) - E(x)E(y),
\end{aligned} \tag{3}$$

where $E()$ denotes the expected solution.

$$\begin{cases} \max\limits_{\varphi_x, \varphi_y} & \boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} C_{xy} \boldsymbol{\varphi}_{\mathbf{y}}, \\ \text{s.t.} & \boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} C_{xx} \boldsymbol{\varphi}_{\mathbf{x}} = 1, \quad \boldsymbol{\varphi}_{\mathbf{y}}^{\mathrm{T}} C_{yy} \boldsymbol{\varphi}_{\mathbf{y}} = 1. \end{cases} \tag{4}$$

The best $\varphi(\mathbf{x})$ and $\varphi(\mathbf{y})$ can be obtained by solving the above equation using the Lagrangian function method.

3.2. Nuclear Typical Correlation Analysis. The kernel function is introduced on the basis of CCA to construct the kernel typical correlation analysis (KCCA) method, which better solves the correlation analysis between two different dimensional features [36], and the main structure is shown in Figure 7.

Let the mapping function $\phi(x)$ satisfy $K(x, y) = \langle \phi(x), \phi(y) \rangle$, then $K(x)$ is said to be the kernel function. The normalized samples ($X = (x_1, x_2, \cdots, x_N)$, $Y = (y_1, y_2, \cdots, y_N)$) are mapped to the ϕ function, and then the correlation coefficients of the samples X and Y are solved according to equation (1).

$$\rho = \frac{\boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} K(x, y) \boldsymbol{\varphi}_{\mathbf{y}}}{\sqrt{\boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} K(x, x) \boldsymbol{\varphi}_{\mathbf{x}} \boldsymbol{\varphi}_{\mathbf{y}}^{\mathrm{T}} K(y, y) \boldsymbol{\varphi}_{\mathbf{y}}}}. \tag{5}$$

Then, we can get to the required constraint.

$$\begin{cases} \max\limits_{\varphi_x, \varphi_y} & \boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} K(x, y) \boldsymbol{\varphi}_{\mathbf{y}}, \\ \text{s.t.} & \boldsymbol{\varphi}_{\mathbf{x}}^{\mathrm{T}} K(x, x) \boldsymbol{\varphi}_{\mathbf{x}} = 1, \quad \boldsymbol{\varphi}_{\mathbf{y}}^{\mathrm{T}} K(x, y) \boldsymbol{\varphi}_{\mathbf{y}} = 1. \end{cases} \tag{6}$$

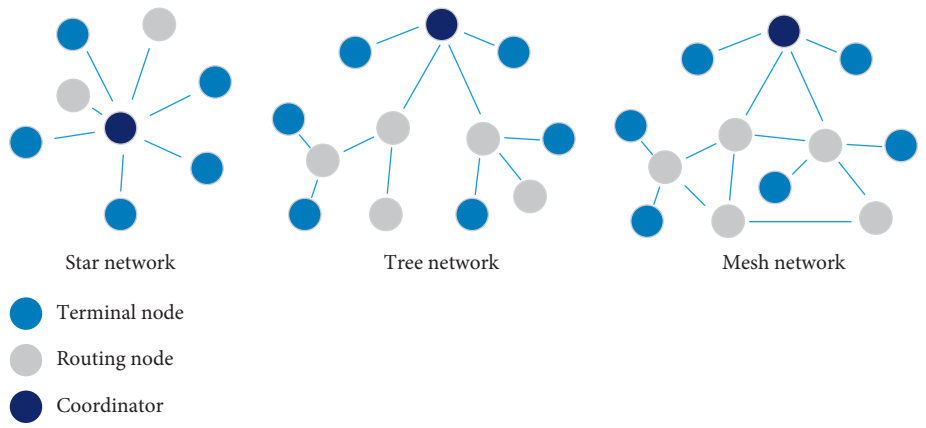


Figure 3: Three common topologies for wireless sensor networks.

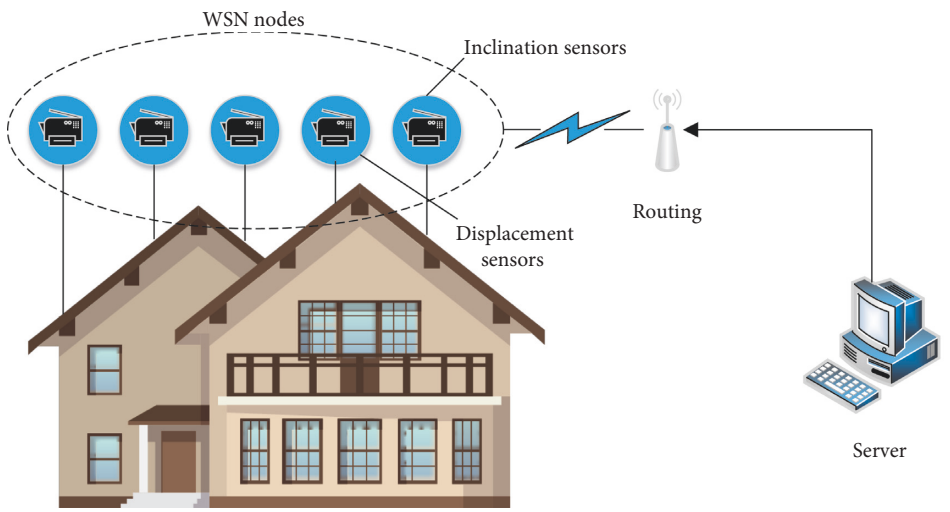


Figure 4: Wireless sensor network-based safety monitoring system for historic buildings.

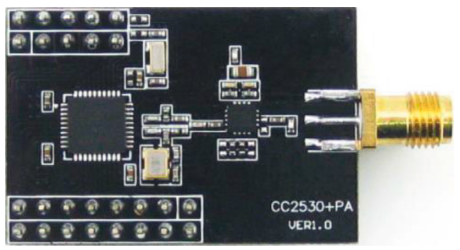


Figure 5: Wireless sensor network nodes.

3.3. Support Vector Machines. Let the sample set (x_i, y_i) can be mapped by a nonlinear support vector machine to obtain the linear equation, $i = 1, 2, \ldots, n$.

$$y_i\left(\mathbf{w}^T x_i + b\right) - 1 \geq 0, \quad i = 1, 2, \ldots, n, \tag{7}$$

where $\mathbf{w}^T$ is the weight matrix and b represents the bias.

The solution to equation (7) is converted to solving for the minimum of $\phi(\mathbf{w}) = 1/2\|\mathbf{W}\|^2 = 1/2(\mathbf{w}^T\mathbf{w})$. A Lagrangian transformation is performed to obtain the new solution equation.

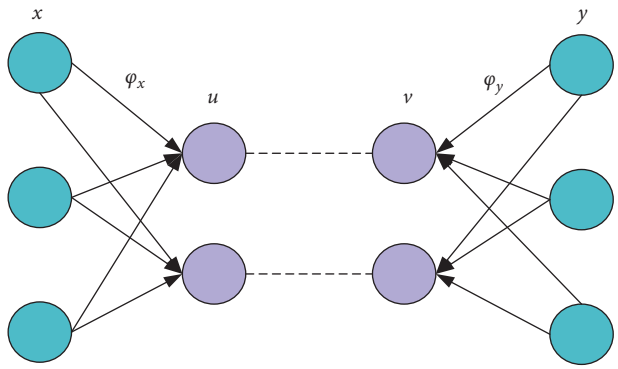


Figure 6: CCA schematic diagram.

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2}\left(\mathbf{w}^T\mathbf{w}\right) - \sum_{i=1}^{n} a_i\left[y_i\left(\mathbf{w}^T x_i + b\right) - 1\right], \tag{8}$$

$$\text{s..t.} \sum_{i=1}^{n} \alpha_i y_i = 0,$$

where a_i is the Lagrangian coefficient. Carry out bias derivative of equation (8) for $\mathbf{w}$ and b, respectively.

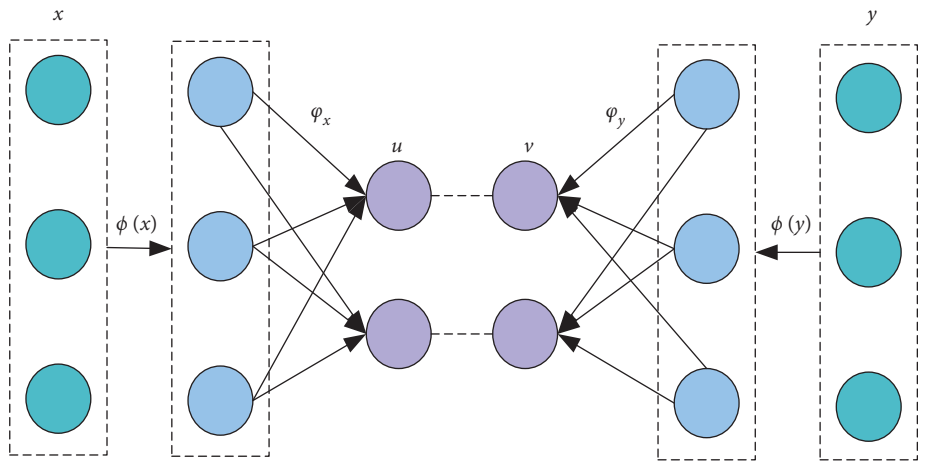


Figure 7: KCCA schematic diagram.

$$Q(\mathbf{a}) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum a_i a_j y_i y_j \left(x_i x_j\right). \tag{9}$$

Solve $Q(\mathbf{a})$ to obtain the maximum value corresponding to $\mathbf{a}^*$.

$$\mathbf{w}^* = \sum_{i=1}^{n} y_i a_i^* x_i. \tag{10}$$

Finally, the optimal SVM can be calculated as follows.

$$f(x) = \operatorname{sgn}\left(\left(\mathbf{w}^*\right)^T + b \right). \tag{11}$$

4. KCCA-SVM-Based Security Monitoring Model for Historic Buildings

4.1. Monitoring Model Implementation Process. In this paper, a KCCA-SVM-based safety monitoring model for historical buildings is proposed. Firstly, KCCA is used to preprocess the independent variables of the original data and extract the principal components (principal components represent the information synthesized by the independent variables according to different weights), so as to reduce the dimensionality of the data and eliminate the noise. During the training process, the kernel parameters can be adjusted to improve the fitting ability of the SVM. The best-fitting combination of parameters is selected as the model parameters. Figure 8 shows the implementation process of the KCCA-SVM-based historical building safety monitoring model.

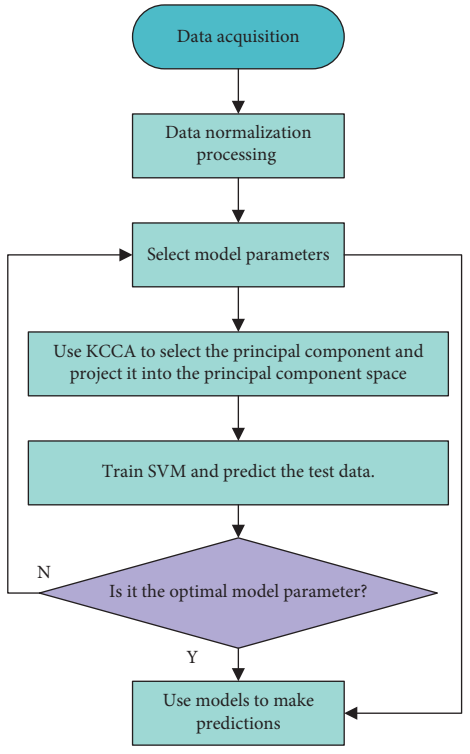


Figure 8: The flowchart of safety monitoring model based on KCCA-SVM.

4.2. Validation of the Machine Learning Dataset. To verify the classification performance of KCCA-SVM, simulation tests were conducted using the commonly used UCI machine learning dataset, which is shown in Table 1.

Table 1: UCI machine learning dataset.

Sample set	Number of samples	Number of attributes
ORL	400	30
Yale	400	32
AR	400	28
PIE	400	30

4.2.1. Influence of Different Kernel Functions. The selection of a suitable kernel function has a large impact on the feature extraction effect of KCCA, which directly affects the classification performance of SVM. Therefore, in this paper, different kernel functions are selected for KCCA analysis and then SVM classification. The KCCA-SVM recognition accuracies of different kernel functions are shown in Table 2.

Table 2: KCCA-SVM recognition accuracy of different kernel functions.

Datasets	Kernel functions	RMSE		
		Minimum value	Average value	Maximum value
ORL	Gaussian kernel	$8.322e-2$	$8.581e-2$	$8.676e-2$
	Sigmoid nucleus	$8.693e-2$	$8.825e-2$	$8.991e-2$
	Polynomial kernel	$9.592e-2$	$9.674e-2$	$1.201e-1$
Yale	Gaussian kernel	$1.044e-1$	$1.124e-1$	$1.265e-1$
	Sigmoid nucleus	$9.817e-2$	$9.924e-2$	$1.063e-1$
	Polynomial kernel	$1.249e-1$	$1.535e-1$	$1.922e-1$
AR	Gaussian kernel	$8.465e-2$	$8.573e-2$	$8.834e-2$
	Sigmoid nucleus	$7.835e-2$	$7.902e-2$	$8.064e-2$
	Polynomial kernel	$9.437e-2$	$9.611e-2$	$9.947e-2$
PIE	Gaussian kernel	$7.916e-2$	$8.111e-2$	$8.259e-2$
	Sigmoid nucleus	$8.146e-2$	$8.297e-2$	$8.481e-2$
	Polynomial kernel	$1.098e-1$	$1.244e-1$	$1.367e-1$

From Table 2, KCCA has lower RMSE values for the ORL and PIE sets using Gaussian kernels, while KCCA has lower RMSE values for the Yale and AR sets using sigmoid kernels. Therefore, the Gaussian kernel is more accurate in the ORL and PIE sets, and the sigmoid kernel performs better in the Yale and AR sets.

4.2.2. Influence of Different Variable Dimensions. To further validate the performance of KCCA-SVM, different variable dimensions were selected for KCCA analysis, followed by SVM identification. The recognition accuracies of the different dimensions are shown in Table 3.

From Table 3, the variable dimensionality has a significant impact on the accuracy of KCCA-SVM. At dimension 10, the KCCA-SVM recognition accuracy is the lowest. The accuracy of KCCA-SVM was higher when the number of dimensions was 20 and 25, and the two values were very close to each other. When the number of dimensions is small, the variables cannot contain important feature information, resulting in a low recognition accuracy. And when the number of dimensions was increased to 20, the accuracy did not appear to improve significantly when the number of dimensions continued to increase. This is mainly because after the dimensionality reaches 20, the selected variable features can already contain the sample attributes in a more comprehensive way. Therefore, even if the number of variable features is increased further, the accuracy does not increase significantly.

The simulation of recognition stability for different dimensions continues below, and the statistical results are shown in Table 4.

From Table 4, the RMSE values of the KCCA-SVM are decreasing as the number of dimensions increases, which indicates that an increase in the number of dimensions has a significant positive effect on stability. The RMSE values are still decreasing when the dimensionality is increased from 20 to 25, indicating that the full extraction of variable features is more beneficial to stability improvement. The dimensionality of variables can improve the stability of recognition, but it also brings a greater amount of recognition operations, which affects the recognition efficiency, so the dimensionality of image recognition should be selected according to the actual situation.

Table 3: Recognition accuracy of KCCA-SVM with different dimensions.

Datasets	Dimension	Accuracy		
		Minimum value	Average value	Maximum value
ORL	10	0.728	0.735	0.751
	15	0.841	0.857	0.866
	20	0.921	0.922	0.923
	25	0.922	0.922	0.923
Yale	10	0.769	0.782	0.791
	15	0.838	0.849	0.858
	20	0.902	0.903	0.906
	25	0.906	0.906	0.906
AR	10	0.842	0.855	0.861
	15	0.882	0.897	0.908
	20	0.943	0.952	0.958
	25	0.950	0.956	0.958
PIE	10	0.812	0.821	0.833
	15	0.863	0.872	0.884
	20	0.887	0.896	0.901
	25	0.907	0.913	0.916

5. Project Examples

5.1. Historical Architectural Context. In this paper, the Shanghai Great World, which was rebuilt in 1924 on South Xizang Road in Huangpu District, is a reinforced concrete frame structure with an L-shaped plan. Shanghai World has a site area of 6537 m^2 and a building area of 13580 m^2. Its architectural style is mixed, including Western classical and Chinese traditional forms. Shanghai World is one of the representative buildings of modern entertainment architecture. Due to factors such as the excavation of the underground in the vicinity of Great World, safety issues have arisen in the structure of the building, such as tilting and cracking of the walls. In order to monitor the structural safety of Great World, a wireless sensor network was used to monitor the overall characteristics of the building structure.

Table 4: Recognition accuracy of RMSE of KCCA-SVM with different dimensions.

Datasets	Dimension	RMSE		
		Minimum value	Average value	Maximum value
ORL	10	$8.715e-2$	$8.927e-2$	$9.142e-2$
	15	$8.646e-2$	$8.893e-2$	$9.013e-2$
	20	$8.321e-2$	$8.581e-2$	$8.676e-2$
	25	$8.136e-2$	$8.327e-2$	$8.435e-2$
Yale	10	$1.253e-1$	$1.277e-1$	$1.294e-1$
	15	$1.175e-1$	$1.263e-1$	$1.381e-1$
	20	$9.817e-2$	$9.924e-2$	$1.063e-1$
	25	$9.656e-2$	$9.721e-2$	$9.826e-2$
AR	10	$8.486e-2$	$8.579e-2$	$8.721e-2$
	15	$8.101e-2$	$8.243e-2$	$8.375e-2$
	20	$7.836e-2$	$7.903e-2$	$8.064e-2$
	25	$7.811e-2$	$7.885e-2$	$7.975e-2$
PIE	10	$8.721e-2$	$8.835e-2$	$8.907e-2$
	15	$8.367e-2$	$8.472e-2$	$8.563e-2$
	20	$7.916e-2$	$8.112e-2$	$8.259e-2$
	25	$7.875e-2$	$8.002e-2$	$8.157e-2$

In this paper, the manual monitoring data of vertical displacement from 8 July 2010 to 8 July 2019 were selected as the research samples to build a safety monitoring model based on KCCA-SVM. A total of 205 groups were sampled, with the first 190 groups used as training samples and the last 15 as testing samples. The main factors of settlement affecting building safety include 3 aspects, namely, temperature, pressure, and time duration, and each factor consists of a number of vectors. Therefore, a total of 14 factors were selected as the initial input vectors, including 4 temperature factors, 8 pressure factors, and 2 aging factors.

5.2. Results of the Safety State Assessment. In the security monitoring model, the kernel matrix of KCCA (190×190) is obtained from the input sample matrix (190×14). The number of principal components extracted by KCCA may be greater than the number of independent variables in the initial sample (14). The highest number of samples was 190. The exact number of principal components to be extracted should be determined through analytical studies. When the kernel parameter is $g = 25.6$ and $g' = 5.76$, the SVM outperforms the SVM with other kernel parameters, regardless of the number of principal components extracted. Therefore, the kernel parameter is fixed to $g = 25.6$ and $g' = 5.76$. The relationship between the number of principal components extracted by KCCA and the computational results of the security monitoring model is investigated, as shown in Table 5 and Figure 9.

When the SVM kernel parameters are fixed, the mean absolute error (MAE) of the KCCA-SVM-based security monitoring model first tends to decrease as the number of principal components extracted by KCCA increases. When the number of principal components is 7, the MAE reaches a minimum and then increases slightly when the number of principal components increases to 8. Subsequently, as the number of extracted principal components increases, the MAE fluctuates slightly but gradually plateaus. Therefore, the number of principal components in the security monitoring model is not as large as possible. When the number of extracted principal components is 7, the prediction accuracy of SVM has reached a high level and there is no need to extract more principal components as input independent variables. Further increase in the number of principal components will introduce more noise and affect the prediction accuracy of the model.

Table 5: The relationship between the number of principal components extracted of KCCA and the monitoring model.

Number of principal components	Cumulative contribution (%)	Average absolute error (mm)
1	25.9	8.14
2	36.78	6.68
3	45.27	2.62
4	52.87	2.31
5	60.1	1.81
7	65.32	0.94
10	69.27	0.59
11	72.57	0.96
12	75.52	0.88
13	77.68	1.11
14	79.74	0.83

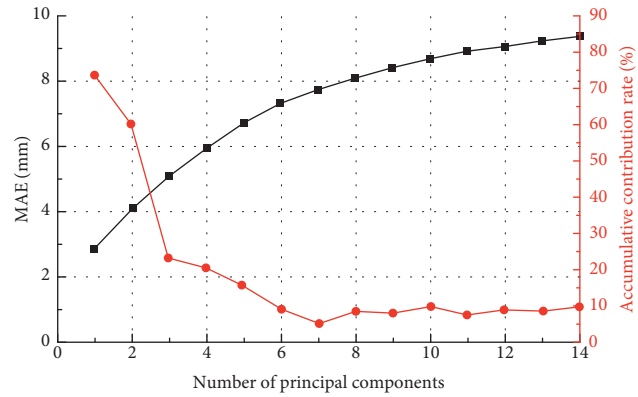


Figure 9: The relationship between the number of principal components extracted of KCCA and the monitoring model.

In the safety monitoring model, a reasonable number of principal components are extracted using KCCA, which can achieve the purpose of eliminating data noise, reducing data dimensionality and improving model prediction accuracy. The number of principal components extracted for the safety monitoring model is determined to be 7, and the corresponding cumulative contribution rate is 69.3%. Compared with the 14 input vectors of the original data, the data dimensionality reduction is very obvious. Figure 10 shows a comparison of the fitted values of the training data for the conventional statistical regression model (HST) and the KCCA-SVM model.

It can be seen that the fitting effect of the conventional HST model deviates significantly with the measured settlement data as a benchmark. The KCCA-SVM-based safety monitoring model, on the other hand, has a significantly

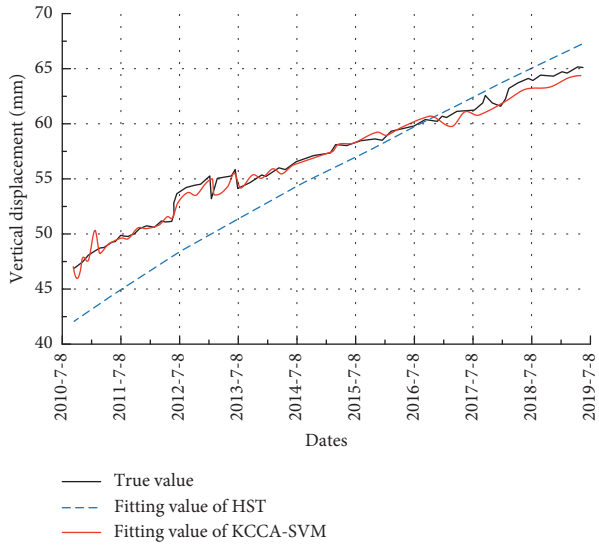


Figure 10: The settlement fitting value of HST and KCCA-SVM model.

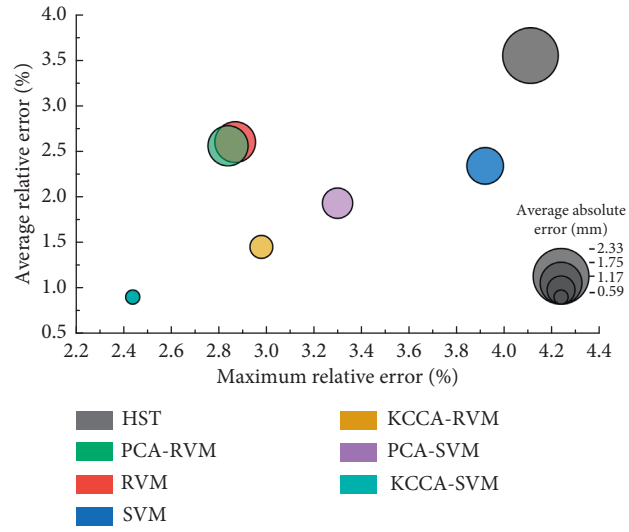


Figure 11: Comparison of errors of different settlement prediction models.

Table 6: Error comparison of different settlement prediction models based on training data.

Models	Maximum relative error (%)	Average relative error (%)	Average absolute error (mm)
HST	4.11	3.55	2.3256
RVM	2.87	2.6	1.7025
SVM	3.92	2.34	1.5357
PCA-RVM	2.84	2.56	1.6789
PCA-SVM	3.3	1.93	1.2641
KCCA-RVM	2.98	1.45	0.953
KCCA-SVM	2.44	0.9	0.5915

better fit. The good fit of the model provides a basis for the subsequent accurate prediction of settlement of historic buildings.

In order to better represent the fitting ability and generalization ability of the KCCA-SVM-based security monitoring model, this paper has built HST, SVM, RVM, PCA-SVM, PCA-RVM, and KCCA-SVM models simultaneously. These models use the same training samples and make predictions on the same test data. The prediction effectiveness was evaluated using the maximum relative error, mean relative error, and mean absolute error metrics. A comparison of the prediction accuracy is shown in Table 6.

We can see that the traditional HST model has a significant error. The prediction accuracy of machine learning algorithms SVM and RVM is significantly higher than that of the HST model. Compared to not using the preprocessing algorithm, the prediction accuracy was slightly improved by using the PCA model to extract the principal components from the input data and then using the SVM and RVM models to make predictions. This indicates that PCA has some noise removal effect. The prediction accuracy of both the KCCA-SVM and KCCA-RVM models improved substantially after using the KCCA model for nonlinear principal component extraction of the input data. The main reason for this is that KCCA has a better handling of the nonlinear features present in the original subsidence data. Among the various compared algorithms, SVM is significantly faster than RVM, due to the sparsity of the results of SVM. For a more visual display, a bubble diagram was used to represent Table 6, as shown in Figure 11.

In summary, the KCCA-RVM-based historical building safety state prediction model has the advantages of reduced data dimensionality, noise elimination, fast calculation speed, and high prediction accuracy.

6. Conclusion

This paper attempts to combine two approaches to build a KCCA-SVM-based security monitoring model for historic buildings. Firstly, wireless sensor networks are applied to the security monitoring of historic buildings in order to improve the automation of monitoring. Secondly, KCCA technique is used for feature correlation analysis to reduce the dimensionality of a large amount of nonlinear data. SVM is then used to take full advantage of its strengths in handling nonlinear and high-dimensional data to predict multiple physical quantities of historic buildings, improving the accuracy of security state assessment. Test results on commonly used machine learning datasets and engineering examples show that the KCCA-SVM model can accurately predict physical quantities such as relative structural displacements of historic buildings.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. Biagini, P. Capone, V. Donato, and N. Facchini, "Towards the BIM implementation for historical building restoration sites," *Automation in Construction*, vol. 71, pp. 74–86, 2016.

[2] J. Xu, L. Ding, and P. E. Love, "Digital reproduction of historical building ornamental components: from 3D scanning to 3D printing," *Automation in Construction*, vol. 76, no. 4, pp. 85–96, 2017.

[3] M. A. Hadianfard, R. Rabiee, and A. Sarshad, "Assessment of vulnerability and dynamic characteristics of a historical building using microtremor measurements," *International Journal of Civil Engineering*, vol. 15, no. 2, pp. 175–183, 2017.

[4] G. Stella, L. Almeida, L. Basilio et al., "Historical building dating: a multidisciplinary study of the Convento de São Francisco (Coimbra, Portugal)," *Geochronometria*, vol. 45, no. 1, pp. 119–129, 2018.

[5] A. Barontini and P. B. Lourenço, "Seismic safety assessment of mixed timber-masonry historical building: an example in Lima, Peru," *Journal of Earthquake Engineering*, vol. 25, no. 5, pp. 872–891, 2018.

[6] C. Santagati and M. Lo Turco, "From structure from motion to historical building information modeling: populating a semantic-aware library of architectural elements," *Journal of Electronic Imaging*, vol. 26, no. 1, Article ID 011007, 2016.

[7] A. Galatioto, G. Ciulla, and R. Ricciu, "An overview of energy retrofit actions feasibility on Italian historical buildings," *Energy*, vol. 137, no. 10, pp. 991–1000, 2017.

[8] S. Fais, G. Casula, F. Cuccuru, P. Ligas, and M. G. Bianchi, "An innovative methodology for the non-destructive diagnosis of architectural elements of ancient historical buildings," *Scientific Reports*, vol. 8, no. 1, p. 4334, 2018.

[9] T. J. Mason, "Ultrasonic cleaning: an historical perspective," *Ultrasonics Sonochemistry*, vol. 29, pp. 519–523, 2016.

[10] E. Jones, J. Sciard, and J. Corcoran, "The influence of creep induced grain boundary separation on electrical non-destructive evaluation measurements," *Journal of Nondestructive Evaluation*, vol. 41, no. 1, pp. 24–15, 2022.

[11] Z. Zhuang, J. Zhang, G. Lian, and B. W. Drinkwater, "Comparison of time domain and frequency-wavenumber domain ultrasonic array imaging algorithms for non-destructive evaluation," *Sensors*, vol. 20, no. 17, p. 4951, 2020.

[12] M. R. Bhat and V. D. Ragupathy, "Characterisation of friction stir weld discontinuities by non-destructive evaluation," *Transactions of the Indian Institute of Metals*, vol. 72, no. 11, pp. 2971–2979, 2019.

[13] T. Ma, T. Inagaki, and S. Tsuchikawa, "Non-destructive evaluation of wood stiffness and fiber coarseness, derived from SilviScan data, via near infrared hyperspectral imaging," *Journal of Near Infrared Spectroscopy*, vol. 26, no. 6, pp. 398–405, 2018.

[14] Z. Dong, J. Wei, X. Chen, and P. Zheng, "Face detection in security monitoring based on artificial intelligence video retrieval technology," *IEEE Access*, vol. 8, no. 99, pp. 63421–63433, 2020.

[15] V. P. Hai and P. Moore, "A proposal for information systems security monitoring based on large datasets," *International Journal of Distributed Systems and Technologies*, vol. 9, no. 2, pp. 16–26, 2018.

[16] N. Mix, F. Gist, R. Lynn, and A. Lai, "Technology advances expand water system security options," *Opflow*, vol. 44, no. 5, pp. 10–14, 2018.

[17] P. S. Mann and S. Singh, "Improved metaheuristic-based energy-efficient clustering protocol with optimal base station location in wireless sensor networks," *Soft Computing*, vol. 23, no. 3, pp. 1021–1037, 2019.

[18] F. Ahmed, Z. Wadud, N. Javaid, N. Alrajeh, M. Alabed, and U. Qasim, "Mobile sinks assisted geographic and opportunistic routing based interference avoidance for underwater wireless sensor network," *Sensors*, vol. 18, no. 4, p. 1062, 2018.

[19] L. Kong, J. S. Pan, V. Snášel, P. W. Tsai, and T. W. Sung, "An energy-aware routing protocol for wireless sensor network based on genetic algorithm," *Telecommunication Systems*, vol. 67, no. 3, pp. 451–463, 2018.

[20] Y. Han, J. Zhang, and D. Sun, "Error control and adjustment method for underwater wireless sensor network localization," *Applied Acoustics*, vol. 130, pp. 293–299, 2018.

[21] X. Fang, Z. Jiang, L. Nan, and L. Chen, "Optimal weighted K-nearest neighbour algorithm for wireless sensor network fingerprint localisation in noisy environment," *IET Communications*, vol. 12, no. 10, pp. 1171–1177, 2018.

[22] B. Shi, V. Sreeram, D. Zhao, S. Duan, and J. Jiang, "A wireless sensor network-based monitoring system for freshwater fishpond aquaculture," *Biosystems Engineering*, vol. 172, pp. 57–66, 2018.

[23] A. G. Baydin, B. A. Pearlmutter, and A. A. Radul, "Automatic differentiation in machine learning: a survey," *Journal of Machine Learning Research*, vol. 18, no. 153, pp. 1–43, 2018.

[24] M. Raissi and G. E. Karniadakis, "Hidden physics models: machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, 2018.

[25] J. H. Thrall, X. Li, Q. Li et al., "Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success," *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 504–508, 2018.

[26] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.

[27] I. Goodfellow, P. Mcdaniel, and N. Papernot, "Making machine learning robust against adversarial inputs," *Communications of the ACM*, vol. 61, no. 7, pp. 56–66, 2018.

[28] R. Hejazi, A. Grime, M. A. Randolph, and M. Efthymiou, "A Bayesian machine learning approach to rapidly quantifying the fatigue probability of failure for steel catenary risers," *Ocean Engineering*, vol. 235, no. 4, Article ID 109353, 2021.

[29] M. Di Napoli, D. Di Martire, G. Bausilio et al., "Rainfall-induced shallow landslide detachment, transit and runout susceptibility mapping by integrating machine learning techniques and GIS-based approaches," *Water*, vol. 13, no. 4, p. 488, 2021.

[30] L. Xu, L. Hou, Z. Zhu et al., "Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm," *Energy*, vol. 222, no. 1, Article ID 119955, 2021.

[31] E. Duarte and J. Wainer, "Empirical comparison of cross-validation and internal metrics for tuning SVM hyperparameters," *Pattern Recognition Letters*, vol. 88, no. 5, pp. 6–11, 2017.

[32] D. K. Jain, S. B. Dubey, R. K. Choubey et al., "An approach for hyperspectral image classification by optimizing SVM using self organizing map," *Journal of Computational Science*, vol. 25, pp. 252–259, 2018.

[33] R. Tamilarasi and S. Prabu, "Automated building and road classifications from hyperspectral imagery through a fully convolutional network and support vector machine," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 13243–13261, 2021.

[34] Q. Chen and Y. Wang, "Key-performance-indicator-related state monitoring based on kernel canonical correlation analysis," *Control Engineering Practice*, vol. 107, no. 1, Article ID 104692, 2021.

[35] X. Gao, S. Niu, and Q. Sun, "Two-directional two-dimensional kernel canonical correlation analysis," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1578–1582, 2019.

[36] J. Chen, G. Wang, Y. Shen, and G. B. Giannakis, "Canonical correlation analysis of datasets with a common source graph," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4398–4408, 2018.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8692865, 9 pages
https://doi.org/10.1155/2022/8692865

Research Article

The Impact of Hearing Aids on Speech Perception in Mandarin-Speaking Children

Yuan Zhang, Yun Zheng, and Gang Li

Department of Otolaryngology & Head and Neck Surgery, West China Hospital of Sichuan University, No. 37 Guo Xue Xiang, Chengdu, Sichuan 610041, China

Correspondence should be addressed to Yun Zheng; shirleyzyun@163.com

Received 14 June 2022; Revised 11 July 2022; Accepted 16 July 2022; Published 12 August 2022

Academic Editor: Wenming Cao

Background. Severe hearing loss can affect speech perception in children, and hearing aids as a medical device may help improve speech perception in children. *Objective*. To explore the effects of fitting hearing aids (HAs) on speech perception in children with severe hearing loss (60–70 dB HL). *Methods*. Ninety-five children with bilateral severe hearing loss who were fitted bilaterally with HAs before the age of 3 years were followed up. The subjects were grouped according to their age at the time of fitting, i.e., <1, 1–2 , and 2–3 years groups. The Mandarin Early Speech Perception test was used to evaluate speech perception of Mandarin monosyllabic words at 12, 24, and 36 months after fitting. *Results*. There were significant improvements in vowel, consonant, and tone perception scores from 12 to 36 months after fitting HAs in the three age groups, and the mean score at 36 months after fitting was significantly improved at >85%. The mean speech pattern and spondee perception scores averaged at >90% at 12 months after fitting and were comparable to the scores of 2-year-old children with normal hearing. *Conclusions*. HA helps with speech perception in children with severe hearing loss.

1. Introduction

Learning to speak strongly depends on speech perception, which is accurately perceiving and decoding speech signals [1]. The improvement of speech recognition indicates whether children with hearing loss (HL) have received effective intervention. The materials of speech perception are presented in open and closed sets. Closed-set speech perception tests require the children to choose from limited options, whereas there are unlimited choices in open-set tests. As open-set speech perception tests are closer to the real-world listening conditions to some extent, they are preferred over closed-set tests for assessing speech perception in children with HL.

In previous studies, for children with the same degree of HL, the difference in speech perception test results between children with hearing aids (HAs) and children with cochlear implants (CIs) formed the basis for determining the target population for CIs [2, 3]. Indications for CIs have expanded from profound to severe HL. Unlike children with profound bilateral HL for whom CIs can be confidently recommended based on an unaided audiogram, children with severe HL need a comprehensive evaluation of their speech and language development to make such a recommendation. Therefore, dynamic monitoring of the auditory and speech development of children with severe HL who have been fitted with HAs are key to making appropriate recommendations. The first 3 years of life are essential for acquiring a spoken language. During this time, the auditory system perceives speech signals in the environment and transmits them to specific areas of the brain, thus saving the development of head-to-head communication skills. Theoretically, it is also the best age window for children to receive implants [3, 4]. As the tools to measure speech and language development objectively in Mandarin-speaking children aged <3 years are very limited, making recommendations without missing the optimal implantation age based on limited data are clinically challenging.

The administration of open-set speech perception tests is not possible for children aged <3. Therefore, early speech

perception (ESP) information in young children is mainly derived from the closed-set test. The Mandarin Early Speech Perception (MESP) test is a closed-set measure of speech perception tasks using words; it is the only test available to assess speech perception of Mandarin-speaking children as young as 2 years of age [5]. The MESP test was developed based on the English ESP test [6]. Due to the differences between Mandarin and English, only the first three categories (speech sound detection, speech pattern, and spondee perception) are similar between the MESP test and the English ESP test, and category 4 of the English ESP test, which includes 12 monosyllabic words, was revised to form categories 4 (vowel perception), 5 (consonant perception), and 6 (tone perception) of the MESP test. The two languages differ greatly in terms of syllables, stress patterns, vowels, and consonants. Most Mandarin syllables comprise three parts: an initial consonant, a syllable vowel, and a tone. Mandarin uses tone or pitch variations to convey lexical information, unlike Western languages. Therefore, assessment of tone perception is essential when evaluating Mandarin monosyllabic word recognition.

In the past 10 years, many studies have reported on the development of early speech perception in children with normal hearing and children with CIs [5, 7–16]. Unfortunately, the development of early speech perception in Mandarin-speaking children with HAs remains unexplored. Certainly, theoretically speaking, children with severe HL fitted with HAs can be considered implantation candidates if their speech perception lags behind that of those with CIs. However, many studies have shown concern about the ability of current CI speech processors to provide cues for the acquisition of Mandarin, and the results of most studies have shown that tone perception is challenging for Mandarin-speaking children with CIs [7, 9, 11–14]. In these studies, the accuracy of tone perception in Mandarin-speaking children with CIs was <80% regardless of the age at the time of implantation and the duration of CI use. It is well known that the basic working principles of HAs and CIs are different. HA amplifies the sound and senses it through the user's own hair cells. CIs bypass the damaged hair cells and sense sound directly by stimulating the auditory nerve and converting the sound into electrical stimulation. Suppose there is an objective gap between CIs and HAs in improving tone perception in children with HL. In that case, tone perception results will not be suitable as a data reference for identifying candidates for CIs.

The current research aimed to examine the development of speech perception of Mandarin monosyllabic words using the MESP test in children with severe HL fitted with HAs before 3 years of age. To avoid the impact on speech perception due to insufficient auditory compensation [2, 17], this study selected children with severe HL whose degree of HL ranged from 60 to 70 dB·HL in the better ear as the subjects.

2. Materials and Methods

2.1. Design and Sampling. This study comprised 95 children from the Department of Otolaryngology, Head and Neck Surgery Hearing Center at West China Hospital of Sichuan University. The inclusion criteria were as follows: children with the pure tone average (PTA, at 0.5, 1, 2, and 4 kHz) of the better ear ranging from 60 to 70 dB·HL; children who were fitted bilaterally with HAs before the age of 3 years; children with no other known developmental disabilities; children with parents and caregivers who were native speakers of Mandarin; and children with speech being the main mode of communication within their families. The exclusion criteria were as follows: children with a history of otorhinolaryngological surgery; children who abandoned the treatment; children who received CIs; children with the presence of syndromes (e.g., Down's syndrome) and associated neurological alterations; and children with other developmental abnormalities. The study protocol was approved by the medical ethics committee of our hospital, and written informed consent was obtained from the children's parents. All patients in this study completed this study, and no one dropped out of the study.

The age and gender distribution for the sample are shown in Table 1. The mean unaided PTA in the better ear for all subjects was 64.8 dB HL (SD = 3.4), and the aided PTA was 25.1 dB HL (SD = 5.4). The average duration of daily HA use was over 10 h for each age group. Subjects were assessed at 12, 24, and 36 months after fitting. Table 2 shows the number of children who participated in each assessment. Based on the fitted age, the samples were stratified at 12 months as an interval, allowing for analysis of assessment data for each age group. Stratification for the current study comprised three groups based on the age at the time of fitting: <1 , 1–2, and 2–3 years. Since the minimum age of a child for using existing Mandarin speech perception materials is 2 years and children aged <1 years could not reach the assessment age until 2 years after fitting, the speech perception of children aged <1 year was reported only at the second and third year after fitting.

2.2. Speech Perception Assessment. Zheng et al. developed the MESP test [5], which is a tool commonly used to measure the earliest evidence of speech perception in a child for standard Mandarin. Zheng et al. [5] and Zhang et al. [16] used the MESP test to study the speech perception ability of children with normal hearing and established normal reference values. The MESP test has successfully evaluated early speech perception development in young Mandarin-speaking children after cochlear implantation [8, 10, 12–15]. This test includes the following six hierarchical categories of speech perception with increasing difficulty levels: 1: speech sound detection; 2: speech pattern; 3: spondee perception; 4: vowel perception; 5: consonant perception; and 6: tone perception. All the test words were presented in the form of pictures. Category 1 is used only to determine whether the subject can perform the MESP test, and no scores are calculated. Each of the categories 2–5 comprises only one test picture, and each picture has 12 test words. In category 6 (tone perception), there are six pictures because there are four tones in standard Mandarin, and six tone pairs can be formed. Each picture includes eight test words.

Table 1: General characteristics of subjects.

Gender	Age at the time of fitting			Totals
	<1 year	1–2 years	2–3 years	
Male	22	14	12	48
Female	23	13	11	47
Total	45	27	23	95
Mean age	0.5	1.6	2.5	1.5
Percent of sample	47.4%	28.4%	24.2	100%

Table 2: The number of children who were tested in each category of the MESP test and the percentages of children who achieved each category.

Age at the time of fitting	<1 year		1–2 years			2–3 years		
	24 months	36 months	12 months	24 months	36 months	12 months	24 months	36 months
Number								
Category 2	42	38	23	18	14	19	14	12
Category 3	41	36	22	18	14	17	14	12
Category 4	37	36	19	15	13	16	13	12
Category 5	36	35	17	13	13	16	12	11
Category 6	35	34	17	12	13	15	11	11
Proportion (%)								
Category 2	100	100	86.4	100	100	84.2	100	100
Category 3	100	100	87	100	100	88.2	100	100
Category 4	86.5	91.7	47.1	80	92.3	43.7	92.3	100
Category 5	77.8	94.2	52.9	76.9	92.3	56.3	83.3	91.6
Category 6	77.1	91.2	41.2	83.3	92.3	43.8	81.8	91.6

The MESP test software captured every response on every trial and used the responses to determine the overall speech perception of each individual. All these test results were presented in percentages. Since each measure was obtained from a closed set of response alternatives, the chance performance level and threshold scores required for significance above chance ($P < 0.05$) for each measure were used in the analyses. The percentages of the threshold scores for speech pattern, spondee, vowel, and consonant perception were 54.5%, 30.3%, 63.6%, and 63.6%, respectively. The threshold scores for overall tone perception, individual tone pair discrimination, and individual tone recognition were 68.3%, 75.0%, and 67.5%, respectively. As the categories advanced, the children not meeting a threshold score were dropped out from being tested for the next, more challenging category. Table 2 displays the percentage of the sample from each age group that reached each category as well as the percentage with scores significantly better than chance.

2.3. Data Analyses. The data analysis focused on determining (1) the percentage of subjects whose performance scores were significantly above the threshold scores for each category, (2) the mean scores for each category, and (3) the accuracy of discriminating each of the six tone pairs and recognizing each of the four individual tones. Note that the mean scores are from the subjects whose scores were significantly above the threshold scores. The MESP test scores were compared for the present participants and their normal-hearing peers. A P-value of <0.05 was considered significant.

3. Results

3.1. Category 2: Speech Pattern Perception. Although the percentage of children who scored significantly above the threshold score improved from 84.2% (2–3 years group) and 86.4% (1–2 years group) at 12 months after fitting to 100% (both groups) at 36 months after fitting, the improvement was statistically not significant ($P > 0.05$). Category 2 scores of children in all age groups were above the threshold score after 2 years of HA use (Table 2). When a child's score was higher than the threshold score of category 2, the child was considered to have achieved category 2.

The mean scores for category 2 in each age group for every test interval are shown in Figure 1(a). This category seemed easy for the children as the mean score of 1–2 and 2–3 years groups at 12 months after fitting showed >90.0% accuracy. The improvement from 12 to 36 months after fitting was not significant in any age group ($P > 0.05$), and the age at the time of fitting did not significantly influence the performance of children in category 2 ($P > 0.05$).

3.2. Category 3: Spondee Perception. The results for category 3 are similar to those for category 2. Although the percentage of children who scored significantly above the threshold score improved from 87.0% (1–2 years group) and 88.2% (2–3 years group) at 12 months after fitting to 100% (both groups) at 36 months after fitting, the improvement was statistically not significant ($P > 0.05$). Category 3 scores of children in all age groups were above the threshold score after 2 years of HA use (Table 2). The mean score of the 1–2

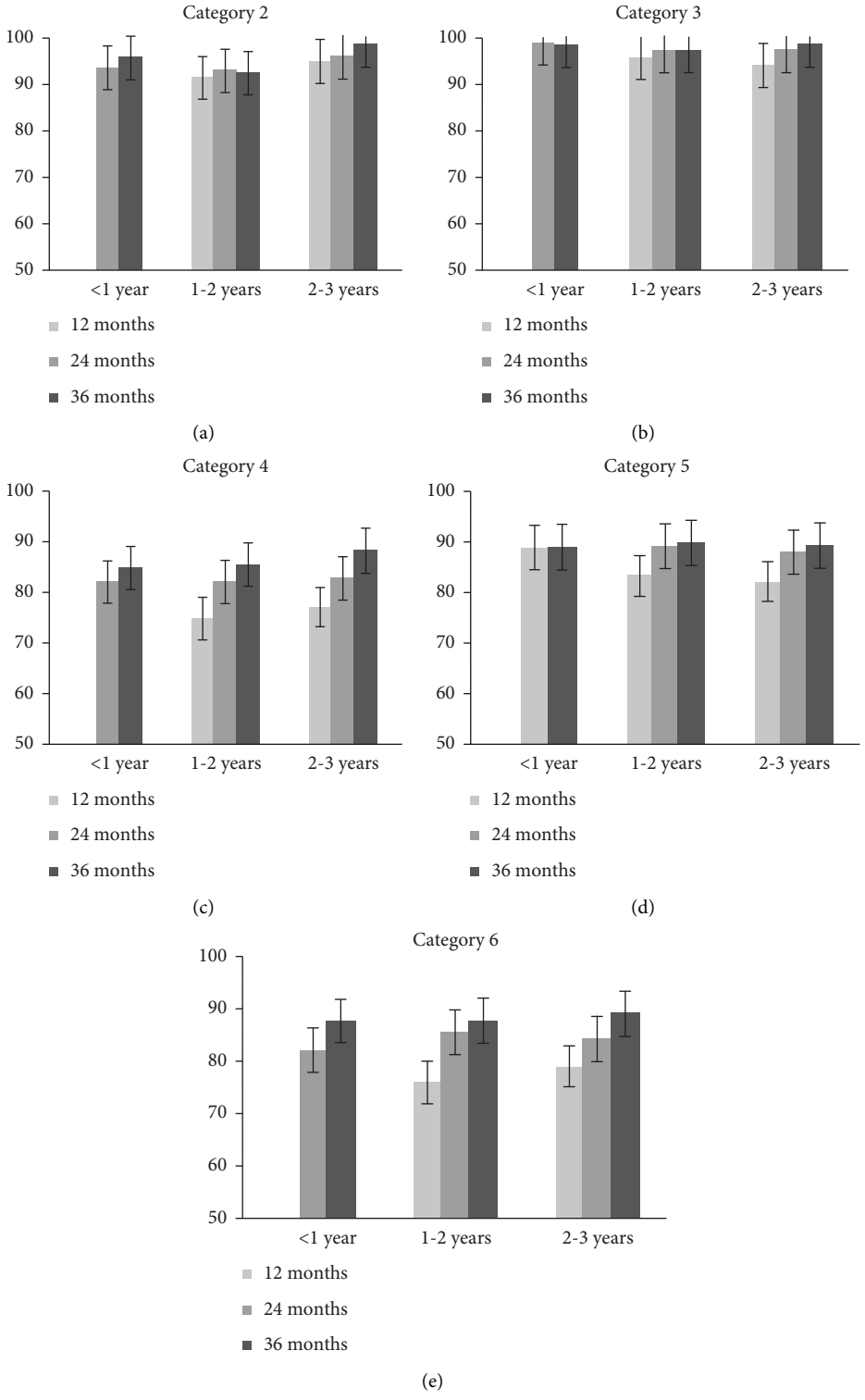


Figure 1: The mean scores in categories 2–6 in three age groups at each test interval.

and 2–3 years groups revealed an accuracy of >94.0% at 12 months after fitting (Figure 1(b)). The improvement from 12 to 36 months after fitting was not significant in any age group ($P > 0.05$), and the age at the time of fitting did not significantly influence the performance of children in category 3 ($P > 0.05$).

Previously, Zheng et al. [5] reported that all children with normal hearing successfully achieved categories 2 and 3, and the scores of children with normal hearing aged 2–3 years in categories 2 and 3 were 99.4%, and 100.0%, respectively. Therefore, our results indicate that children with severe HL can achieve better speech pattern and spondee

perception after 12 months of HA use and that the age minimally influences the performance in these two categories at the fitting time.

3.3. Category 4: Vowel Perception. The percentage of children who scored significantly above the threshold score improved significantly from 47.1% (1–2 years group) and 43.7% (2–3 years group) at 12 months after fitting to 91.7% (<1 year group), 92.3% (1–2 years group), and 100% (2–3 years) at 36 months after fitting ($P < 0.05$; Table 2). The mean score of the three age groups significantly improved from 75.0% (1–2 years group) and 77.0% (2–3 years group) at 12 months after fitting to 84.8% (<1 year group), 85.3% (1–2 years group), and 88.1% (2–3 years group) at 36 months after fitting ($P < 0.05$; Figure 1(c)). For 1–2 and 2–3 years groups, the scores at 12 months after fitting were significantly different from those at 24 and 36 months after fitting, and the scores at 24 months after fitting were significantly different from those at 36 months after fitting ($P < 0.05$). For <1 year group, the scores at 24 months after fitting were significantly different from those at 36 months after fitting ($P < 0.05$). However, the differences among the three age groups at the same evaluation interval were insignificant ($P > 0.05$). The average score of children with normal hearing aged 2–3 years in category 4 was 90.8% [5]. Thus, after 36 months of HA use, the vowel perception of children with severe HL fitted with HAs before 3 years of age was comparable to that of children with normal hearing aged 2–3 years.

3.4. Category 5: Consonant Perception. The percentage of children who scored significantly above the threshold score improved significantly from 52.9% (1–2 years group) and 56.3% (2–3 years group) at 12 months after fitting to 94.2% (<1 year group), 92.3% (1–2 years group), and 91.6% (2–3 years group) at 36 months after fitting ($P < 0.05$; Table 2). The mean score of the three age groups improved significantly from 83.5% (1–2 years group) and 82.0% (2–3 years group) at 12 months after fitting to 89.0% (<1 year group), 89.7% (1–2 years group), and 89.5% (2–3 years group) at 36 months after fitting ($P < 0.05$; Figure 1(d)). For all age groups, the improvement from 24 to 36 months was not significant ($P > 0.05$). For 1–2 and 2–3 years groups, there was a significant improvement from 12 to 36 months after fitting ($P < 0.05$). However, the difference among the three age groups at the same evaluation interval was insignificant ($P > 0.05$). The average score of children with normal hearing aged 2–3 years in category 5 was 90.1% [5]. Thus, after 36 months of HA use, the consonant perception of children with severe HL fitted with HAs before 3 years of age was comparable to that of children with normal hearing aged 2–3 years.

3.5. Category 6: Tone Perception. The percentage of children who scored significantly above the threshold score improved significantly from 41.2% (1–2 years) and 43.8% (2–3 years) at 12 months after fitting to 91.2% (<1 year group), 92.3% (1–2 years group), and 91.6 (2–3 years group) at 36 months after fitting ($P < 0.05$; Table 2). The mean score of the three age groups improved significantly from 76.0% (1–2 years group) and 79.0% (2–3 years group) at 12 months after fitting to 87.6% (<1 year group), 87.7% (1–2 years group), and 88.9% (2–3 years group) at 36 months after fitting ($P < 0.05$; Figure 1(e)). For the 1–2 and 2–3 years groups, the scores at 12 months after fitting were significantly different from those at 24 and 36 months after fitting, and the scores at 24 months after fitting were significantly different from those at 36 months after fitting ($P < 0.05$). For the <1 year group, the scores at 24 months after fitting were significantly different from those at 36 months after fitting ($P < 0.05$). However, the difference among the three age groups at the same evaluation interval was not significant ($P > 0.05$). In category 6, the average score was 79.3% for children with normal hearing aged 2–3 years, 84.7% for those aged 3–4 years, and 90.6% for those aged 4–5 years [16]. The mean score of children in the <1 year group in the present study was higher than that of age-matched children with normal hearing reported by Zhang et al. [16]. For 1–2 and 2–3 years groups, the mean score at 12 months after fitting was comparable to that of children with normal hearing aged 2–3 years. Furthermore, the mean score at 24 months after fitting was comparable to that of children with normal hearing aged 3–4 years, and the mean score at 36 months after fitting was comparable to that of children with normal hearing aged 4–5 years.

3.6. Tone Pair Discrimination. For each of the six tone pairs, the mean percent correct discrimination scores of children at different test intervals for the three age groups are shown in Figure 2. For the < 1 year group (Figure 2(a)), the scores for only tone pair 1–4 improved significantly from 24 to 36 months after fitting ($P < 0.05$). For the 1–2 years group (Figure 2(b)), the scores for tone pairs 1–2 and 1–3 improved significantly from 12 to 24 months after fitting ($P < 0.05$). The scores for only tone pair 2–3 improved significantly from 24 to 36 months after fitting ($P < 0.05$). Finally, the scores for tone pairs 1–2, 2–3, and 2–4 improved significantly from 12 to 36 months after fitting ($P < 0.05$). For the 2–3 years group (Figure 2(c)), the scores for tone pairs 1–2 and 3–4 improved significantly from 12 to 24 months after fitting. The scores for only tone pairs 2–3 significantly enhanced from 24 to 36 months after fitting ($P < 0.05$). Finally, the scores for tone pairs 1–3, 1–4, 2–3, and 3–4 improved significantly from 12 to 36 months after fitting ($P < 0.05$). Notably, the scores for tone pair 2–3 did not exceed chance performance levels even at 36 months after fitting in any age group.

3.7. Tone Recognition. For each of the four tones, the mean percent correct recognition scores of children at different test intervals for the three age groups are shown in Figure 3. For the <1 year group (Figure 3(a)), the scores for tones 2 and 3 improved significantly from 24 to 36 months after fitting ($P < 0.05$). For the 1–2 years group (Figure 3(b)), the scores for tones 1 and 3 improved significantly from 12 to 24 months after fitting ($P < 0.05$). None of the tones' scores

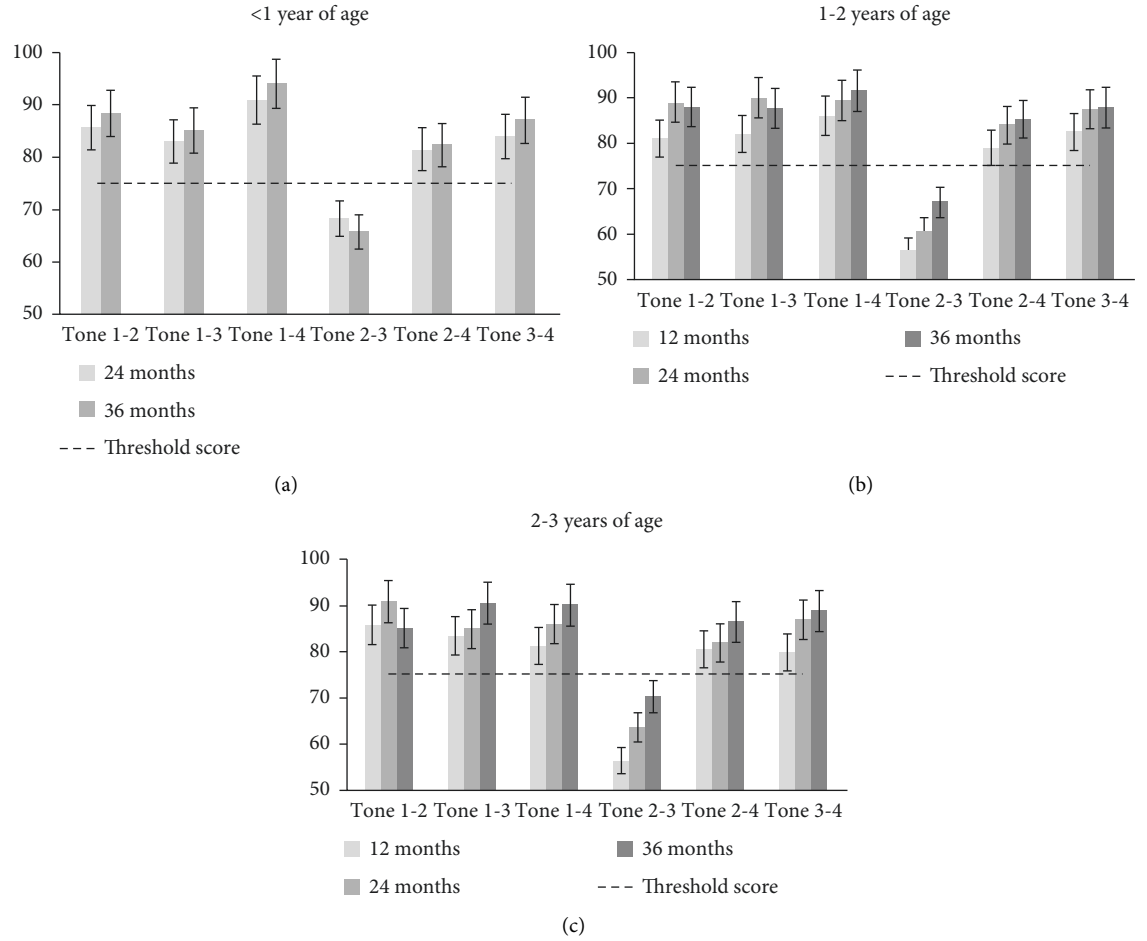


Figure 2: The mean scores for the discrimination of six tone pairs in three age groups at each test interval. The dashed line displays the threshold score for individual tone pair discrimination.

improved significantly from 24 to 36 months after fitting ($P > 0.05$). Finally, the scores for tones 1, 3, and 4 improved significantly from 12 to 36 months after fitting ($P < 0.05$). For the 2–3 years group (Figure 3(c)), the scores for only tone 3 improved significantly from 12 to 24 months after fitting ($P < 0.05$). None of the tones' scores improved significantly from 24 to 36 months after fitting. Finally ($P > 0.05$), the scores for tones 2 and 3 improved significantly from 12 to 36 months after fitting ($P < 0.05$). Notably, the scores for tone 3 in 1–2 and 2–3 years groups did not exceed chance performance levels even at 12 months after fitting.

4. Discussion

The main purpose of this study was to investigate early speech perception in Mandarin-speaking children with severe HL who had been fitted with HAs before the age of 3 years. The speech perception skills of these children significantly improved during the first 36 months of HA use, particularly for the aspects of vowel, consonant, and tone perception. The three age groups exhibited similar patterns of speech perception development. Although the three age groups showed similar trajectories of speech perception, the earlier intervention was associated with a smaller developmental gap between children with severe HL and age-matched children with normal hearing.

In the current study, Mandarin-speaking children with HA had significant improvements in speech perception. When comparing the speech perception performance of children with HAs and children with normal hearing [5], a very large gap in speech perception between the two was noted during the first year after fitting. This gap gradually narrowed with the increased duration of HA use. Zheng et al. [5] established the MESP test by investigating 2- to 5-year-old children with normal hearing ($N = 92$); they reported that all children successfully achieved categories 2–5 by 2 years of age; they performed equally well in categories 2–5, and their mean scores in these categories exceeded 90%. All children successfully achieved category 6 by 3 years of age, and their mean scores in this category approached 90%. In the current study, over 80% of subjects achieved categories 2 and 3 at 12 months after fitting, whereas nearly half of the subjects did not achieve categories 4, 5, and 6 over the same period. At 24 months after fitting, all subjects achieved categories 2 and 3. At 36 months after fitting, over 90% of subjects in the three age groups achieved categories 4, 5, and 6. The mean scores in categories 2 and 3 exceeded 90% at 12 months after fitting, whereas the

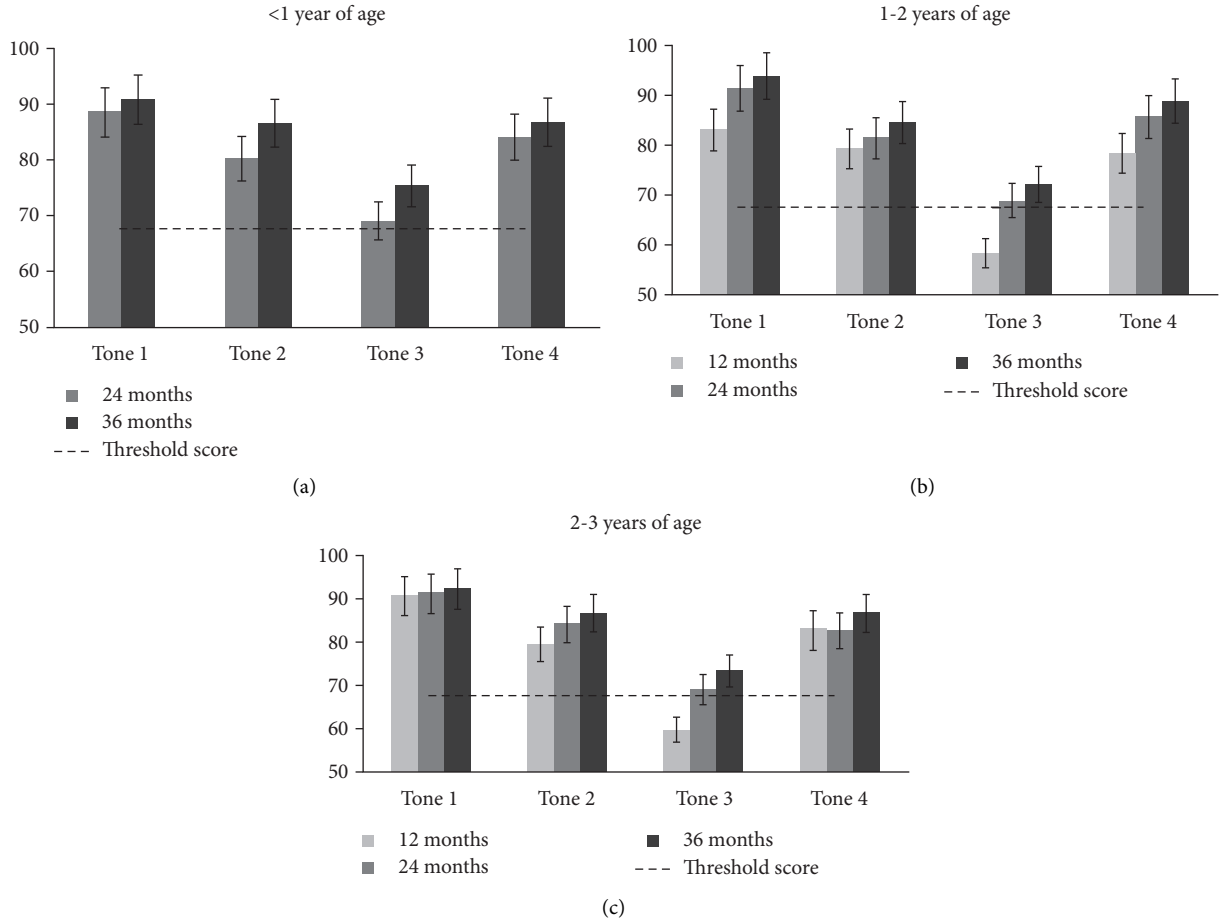


Figure 3: At each test interval, the mean scores for the recognition of four tones in three age groups. the dashed line displays the threshold score for individual tone recognition

mean scores in categories 4, 5, and 6 approached 90% at 36 months after fitting, regardless of the age at the time of fitting. Speech pattern and spondee perception subtests were evidently easy for children with HAs. As these two categories are designed for children with profound HL, the focus is on subjects' identification of lexical rhythms rather than on recognizing true monosyllabic words [5]. However, vowel, consonant, and tone perception results suggested that children with severe HL found it challenging to acquire the perception of Mandarin monosyllabic words in these aspects. We noted that most children with severe HL needed at least 36 months of HA use to acquire good vowel, consonant, and tone perception of Mandarin monosyllabic words.

Acquiring the ability to recognize monosyllabic words is similarly challenging for children with CIs [8, 13, 15, 18]. For example, Lu et al. [15] used the MESP test to report speech perception developments in 132 children who received CIs between the ages of 8 months and 7 years during the 24 months after receiving CIs. The results showed that <50% of pediatric CI users achieved tone perception category within the first year of implantation, and 70% of subjects achieved tone perception category 2 years after implantation. Eisenberg et al. [18] used the English version of the ESP test and reported that 20/42 children (48.0%) who received an implant at 2.3 years of age achieved category 4 (monosyllabic word identification) within the first year after implantation. However, Baumgartner et al. [19] investigated the speech perception performance of 33 German-speaking children with CIs during 36 months after implantation and found that the scores of the closed-set monosyllabic word tests were 100% at 12 months after implantation and remained steady over the 36-month evaluation period. This suggests that the process of monosyllabic word perception differs among children with HL belonging to different language backgrounds.

Although vowel, consonant, and tone perception subtests evaluate segmental perception of Mandarin monosyllabic words, the tone perception subtest is more challenging for children with normal hearing [5]. Zhang et al. [16] found that tone perception ability in children with normal hearing continues to develop from 2 to 6 years of age. However, children with severe HL performed equally well on vowel, consonant, and tone perception at each test interval. In the present study, children with HAs performed as well as or better than children with normal hearing in the tone perception subtest. For the 1–2 and 2–3 years groups, the mean tone perception scores were 76% and 79% at 12 months after fitting, respectively, which show that these children performed as well as 2-year-old children with normal hearing [16]. The mean tone perception score for the < 1 year group was 82.1% at 24 months after fitting, indicating better performance than that of 2-year-old

children with normal hearing. Meanwhile, the patterns of tone pair discrimination and tone identification for children with HAs in the current study were similar to those of children with normal hearing [5, 16]. The mean score for tone pairs 2–3 at each test interval was significantly lower than that for other tone pairs in all three age groups. This may be because tones 2 and 3 have similar characterizations of tone height; tone 2 has a rising contour and tone 3 has a falling dip and then a rising contour, therefore, tones 2 and 3 show similar rising contours [16]. It can be inferred that children with severe HL fitted with HAs can rapidly acquire tone perception in the early stage of speech development, and the duration of HA use was a significant predictor of tone perception.

A review of previous studies on tone perception in children with CIs found that the mean score of tone perception did not exceed 80%, except in case reports. For example, Peng et al. [7] reported the tone perception of 30 children who received CIs at 5.8 years of age and had used CIs for 3.7 years, and results showed that the mean score of tone perception was 72.9%. Xu et al. [9] reported that the mean score of tone perception was 71.0% in 25 children who received CIs at 6.4 years of age and had used them for 3.1 years. Zhou et al. [11] investigated the tone perception ability of 110 children with CIs at 3.9 years of age who had used them for 1.3 years. The results showed that the mean score of tone perception was 67.3%. Chen et al. [12] used the MESP test to evaluate tone perception in children who received implants at the age of 2.7 years and whose duration of CI use was 1.5 years. The results showed that the mean score of tone perception was 77%. Li et al. [13] also used the MESP test to evaluate tone perception in 12 children who received implants at the age of 35 months, and the results showed that the mean score of tone perception was 79.9% after 3 years of CI use.

Recently, Li et al. [14] conducted a retrospective cross-sectional study that also used the MESP test to evaluate the development of tone perception during the 48 months after implantation in children who received the implant at 1–4 years of age. The tone perception scores ranged from 70% to 80% for all age groups, and neither the age at the time of implantation nor the duration of CI use was significantly related to the tone perception score. In all of the above studies, the scores of tone perception showed great individual differences, and the scores ranged from near-chance-level scores to near-perfect scores. However, regarding the mean scores of tone perception, children with severe HL fitted with HAs in the present study performed better than children with CIs in previous studies.

There are also some shortcomings in this study. The patients in this study are all from the same hospital, not representative of the overall level, and sample size of this study is small; these problems can lead to some bias in the results of this study.

5. Conclusions

Our findings provide the first evidence that HAs can effectively promote the development of speech perception in Mandarin-speaking children with severe HL. The developmental trajectories of speech perception after fitting were similar for children who were fitted with HAs within 3 years of age. Furthermore, HAs can synchronously improve their vowel, consonant, and tone perception. Most Mandarin-speaking children with severe HL fitted with HAs can master speech pattern and spondee perception skills within 12 months of fitting Has, and master vowel, consonant, and tone perception within 36 months of fitting HAs. The outcomes of tone perception were better for children with HAs than for children with CIs reported in previous studies.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Ethical Approval

The research protocol for the current study satisfied the appropriate ethics review policies of the West China Hospital of Sichuan University.

Conflicts of Interest

The authors report no conflicts of interest.

Acknowledgments

This research was supported by Fund of West China Hospital of Sichuan University (fund number 2018HXFH043). The authors gratefully acknowledge the contributions of the families of our pediatric subjects who often traveled long distances to participate in the study. The authors also acknowledge Qian Zhu and Lei Jin for their great efforts in data collection in this study.

References

[1] J. H. Lowenstein and S. Nittrouer, "Perception–Production links in children's speech," *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 4, pp. 853–867, 2019.

[2] J. Leigh, S. Dettman, R. Dowell, and J. Sarant, "Evidence-based approach for making cochlear implant recommendations for infants with residual hearing," *Ear and Hearing*, vol. 32, no. 3, pp. 313–322, 2011.

[3] J. R. Leigh, S. J. Dettman, and R. C. Dowell, "Evidence-based guidelines for recommending cochlear implantation for young children: audiological criteria and optimizing age at implantation," *International Journal of Audiology*, vol. 55, no. 2, pp. 9–18, 2016.

[4] A. Sharma, A. A. Nash, and M. Dorman, "Cortical development, plasticity and re-organization in children with cochlear implants," *Journal of Communication Disorders*, vol. 42, no. 4, pp. 272–279, 2009.

[5] Y. Zheng, Z. L. Meng, K. Wang, Y. Tao, K. Xu, and S. D. Soli, "Development of the Mandarin early speech perception test: children with normal hearing and the effects of dialect exposure," *Ear and Hearing*, vol. 30, no. 5, pp. 600–612, 2009.

[6] A. Geers, "Techniques for assessing auditory speech perception and lipreading enhancement in young deaf children," *Volta Review*, vol. 96, pp. 85–96, 1994.

[7] S.-C. Peng, J. B. Tomblin, H. Cheung, Y.-S. Lin, and L.-S. Wang, "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," *Ear and Hearing*, vol. 25, no. 3, pp. 251–264, 2004.

[8] Y. Zheng, S. D. Soli, Z. Meng et al., "Assessment of Mandarin-speaking pediatric cochlear implant recipients with the Mandarin Early Speech Perception (MESP) test," *International Journal of Pediatric Otorhinolaryngology*, vol. 74, no. 8, pp. 920–925, 2010.

[9] L. Xu, X. Chen, H. Lu et al., "Tone perception and production in pediatric cochlear implants users," *Acta Oto-Laryngologica*, vol. 131, no. 4, pp. 395–398, 2011.

[10] Y. Zheng, S. D. Soli, Y. Tao et al., "Early prelingual auditory development and speech perception at 1-year follow-up in Mandarin-speaking children after cochlear implantation," *International Journal of Pediatric Otorhinolaryngology*, vol. 75, no. 11, pp. 1418–1426, 2011.

[11] N. Zhou, J. Huang, X. Chen, and L. Xu, "Relationship between tone perception and production in prelingually deafened children with cochlear implants," *Otology & Neurotology*, vol. 34, no. 3, pp. 499–506, 2013.

[12] Y. Chen, L. L. N. Wong, F. Chen, and X. Xi, "Tone and sentence perception in young Mandarin-speaking children with cochlear implants," *International Journal of Pediatric Otorhinolaryngology*, vol. 78, no. 11, pp. 1923–1930, 2014.

[13] Y. Li, R. Dong, Y. Zheng et al., "Speech performance in pediatric users of Nurotron® Venus™ cochlear implants," *International Journal of Pediatric Otorhinolaryngology*, vol. 79, no. 7, pp. 1017–1023, 2015.

[14] G. Li, S. D. Soli, and Y. Zheng, "Tone perception in Mandarin-speaking children with cochlear implants," *International Journal of Audiology*, vol. 56, no. sup2, pp. S49–S59, 2017.

[15] X. Lu and Z. Qin, "Auditory and language development in Mandarin-speaking children after cochlear implantation," *International Journal of Pediatric Otorhinolaryngology*, vol. 107, pp. 183–189, 2018.

[16] Y. Zhang, Y. Zheng, and G. Li, "Dialect effects on Mandarin tone perception development," *Language and Speech*, vol. 65, no. 3, pp. 681–696, 2022.

[17] E. M. Fitzpatrick, L. Crawford, A. Ni, and A. Durieux-Smith, "A descriptive analysis of language and speech skills in 4- to 5-yr-old children with hearing loss," *Ear and Hearing*, vol. 32, no. 5, pp. 605–616, 2011.

[18] L. S. Eisenberg, K. C. Johnson, A. S. Martinez et al., "Speech recognition at 1-year follow-up in the childhood development after cochlear implantation study: methods and preliminary findings," *Audiology and Neurotology*, vol. 11, no. 4, pp. 259–268, 2006.

[19] W. D. Baumgartner, S. M. Pok, B. Egelierler, P. Franz, W. Gstoettner, and J. Hamzavi, "The role of age in pediatric cochlear implantation," *International Journal of Pediatric Otorhinolaryngology*, vol. 62, no. 3, pp. 223–228, 2002.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9761741, 1 page
https://doi.org/10.1155/2023/9761741

Retraction

Retracted: Research on Kinetic Energy Recovery of Energy Vehicle ABS Solenoid Valve Based on the ELM Deep Learning Model

Computational Intelligence and Neuroscience

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Tu and L. Zhang, "Research on Kinetic Energy Recovery of Energy Vehicle ABS Solenoid Valve Based on the ELM Deep Learning Model," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 6571085, 7 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6571085, 7 pages
https://doi.org/10.1155/2022/6571085

Research Article

Research on Kinetic Energy Recovery of Energy Vehicle ABS Solenoid Valve Based on the ELM Deep Learning Model

Chaoqun Tu and Lingli Zhang

Guangzhou Nanyang Polytechnic Vocational College, Guangzhou 510925, Guangdong, China

Correspondence should be addressed to Chaoqun Tu; 20160588@ayit.edu.cn

Received 21 June 2022; Accepted 13 July 2022; Published 9 August 2022

Academic Editor: Wenming Cao

Aiming at the energy vehicle ABS kinetic energy recovery, this study optimizes the ABS system through the IPSO-ELM model, so that the energy vehicle can recover the energy generated by the ABS system to the greatest extent, so as to achieve the purpose of kinetic energy recovery and reduce energy consumption and vehicle cost. Based on the PSO-ELM model, a linear decreasing weighting method is introduced, and then, an IPSO-ELM model is proposed for the optimization analysis of ABS brake kinetic energy recovery. The results show that compared with the simple ELM model and PSO-ELM model, the simulation mean square error and relative error are significantly smaller, the generalization ability and prediction accuracy are higher, and the maximum relative error of the prediction result is 5.43% and the average relative error is 2.72%. The results confirm that the use of IPSO-ELM for ABS kinetic energy recovery optimization is extremely effective, and the study of ABS kinetic energy recovery for energy vehicles based on IPSO-ELM model optimization has strong application prospects and application potential.

1. Introduction

Although various countries are vigorously developing electric vehicles, the development momentum is still relatively slow. The main reason is that it is difficult to break through core technologies, such as solving energy problems. Energy source limitations increase the service life of the energy source and reduce the manufacturing cost of electric vehicles [1]. Under the current technical conditions, only the method of braking energy recovery can be used to increase the mileage of electric vehicles [2]. During the braking process, the ABS (Anti-lock Brake System) controls the brake pressure to keep the slip ratio near an optimal value when the wheels tend to lock, which can not only exert the maximum braking force but also ensure the handling and solidity of the vehicle [3]. When the energy vehicle is braked urgently, the wheel is subjected to the braking torque generated by the regenerative braking system and the braking torque generated by the friction brake at the same time [4]. How to distribute the relationship between friction braking and regenerative braking has become a key issue for regenerative braking systems [5]. If the braking energy regeneration system and the ABS system are reasonably coordinated and controlled, the economy of the vehicle can be improved, and driving safety can be enhanced [6]. Therefore, it is extremely necessary to study the kinetic energy recovery of ABS for electric new energy vehicles.

Unlike conventional vehicles, electric vehicle motor braking systems can provide regenerative braking force to recover energy. The goal of braking energy optimization is to recover as much energy as possible and improve vehicle economy on the premise of ensuring safe and stable braking of the vehicle. More energy can be recovered if more use of the motor is used for regenerative braking. Japan's research on in-wheel motor electric vehicles started relatively early. In the late 1990s, Japan's Toyota Motor Corporation began research on four-wheel in-wheel electric vehicles. In terms of control, by controlling the braking force of the four wheels, the ESC (Electronic Stability Control), the ASR (Anti-Slip Regulation), the ABS (Anti-lock Brake System), and related research studies on vehicle vertical vibration control [7, 8]. General Motors Corporation of the United States launched a

hydrogen fuel cell four-wheel electric vehicle Sequel at the 2005 North American International Auto Show. Compared with traditional vehicles, the vehicle not only recovers vehicle solidity and traction presentation but also has a braking energy recovery function. Based on this idea, some scholars have designed a purely motor FE-RBS (Fully Electric Regenerative Braking System), which uses the motor as the only brake actuator [9]. However, because the regenerative braking system is limited by the external features of the motor and the state of responsibility of the battery, its braking torque working range is limited, and it may not be able to meet the driving braking requirements, and there are certain safety hazards, which limits the Wide application of FE-RBS. Therefore, most of the current braking energy recovery systems of electric vehicles are hydraulic and motor composite braking systems [10–12]. With the advancement of technology, ABS technology has been greatly improved, with higher control precision, stronger adaptability, smaller and smaller size, and lower and lower prices, ABS has become the standard of respective cars. Nowadays hybrid vehicles have begun to attract people's attention, and the world's major auto manufacturers have proposed regenerative braking and anti-lock braking strategies in the ABS technology of such vehicles, which not only improves the vehicle's performance. The dynamic performance can also be recovered as energy [13]. At the same time, with further research on the vehicle, the integration of ABS technology and driving anti-skid control device ASR has appeared, and ESP (Electronic Stability Control) has also been added to confirm the solidity of the vehicle on the basis of the application of ABS/ASR. With the development of technology, expanding the braking effect of the car, it is used together with the emerging EHB (Electronic Hydraulic Braking) and EMB (Electronic Mechanical Braking) to make the ABS respond faster [14]. At the same time, it is integrated with power electronic technology to better control the braking of the vehicle, such as adding EBD (brake force distribution device) to the ABS system, by calculating the adhesion on each wheel, effectively distributing the braking force of each wheel, thereby increasing the braking stability of the vehicle [15–17].

With the development of computer technology, ELM (Extreme Learning Machine) is widely used in various fields [18]. However, there are also some disadvantages. Randomly generating the weights and thresholds of hidden layer nodes may cause invalid hidden layer nodes, resulting in insufficient generalization ability [19, 20]. In order to search for higher precision and faster, this research proposes the adaptive weight method to improve the basic particle swarm algorithm, which effectively increases the diversity of particles and avoids falling into the local optimum. The IPSO was used to adjust the initial input weights and then establish the IPSO-ELM model, and apply the model to the research on kinetic energy recovery of the ABS solenoid valve of energy vehicles. The study has shown that the model has strong learning ability and high prediction accuracy. This method is very effective for optimizing the kinetic energy recovery of the ABS solenoid valve of energy vehicles.

2. Models and Methods

2.1. ELM Model. ELM is a machine learning method developed by Ding et al. [21] based on feedforward neural networks. The model contains input, hidden, and output layers. During the training process, it is no need to adjust the initial weights, only the need to set the number of nodes between the hidden layers, and the training can be completed by calculating the output weights by the least square method [22]. Related to the old-style neural network, this model has the rewards of faster training speed, higher accuracy, simpler parameter setting, and better generalization ability.

Assuming that there are N arbitrary samples (X_i, Y_i), among them, $X_i = \$ \in \$[x_{i1}, x_{i2}, x_{i3}, \cdots, x_{in}]^T \in R^n$, $Y_i = [y_{i1}, y_{i2}, y_{i3}, \cdots, y_{in}]^T \in R^m$,

$$\sum_{i=1}^{K} \beta_i g\left(W_i \times X_i + B_i\right) = t_i. \tag{1}$$

Here, K is the number of nodes; $g(x)$ is the activation function; $W_i = [w_{i,1}, w_{i,2}, w_{i,3}, \cdots, w_{i,n}]^T$ is the connection weight course between the input layer and the hidden layer; $\beta_i = [\beta_{i,1}, \beta_{i,2}, \beta_{i,3}, \cdots, \beta_{i,n}]^T$ is the joining mass vector among the output layer; B_i is the bias value of the i-th node in the neuron nodes. The approximate sample with zero error, which can be expressed as

$$\sum_{i=1}^{N} \left\| t_i - Y \right\|_i = 0. \tag{2}$$

That is, there are β_i, W_i, B_i, so that the following formula holds.

$$\sum_{i=1}^{K} \beta_i g\left(W_i \times X_i + B_i\right) = Y_i. \tag{3}$$

Equation (3) can be expressed as Equation (4) through a matrix.

$$H\beta = T. \tag{4}$$

Here, H is the output matrix of the hidden layer; β is the output weight; T is the expected output. When the selected hidden layer excitation function $g(x)$ in the model is infinitely differentiable, the input weights ω and hidden layer biases B can be randomly initialized, and the output weight β could be obtained by the minimum norm of the function as shown in the following formula:

$$\beta = H^+ \times T. \tag{5}$$

Here, H^+ is the generalized inverse of Moore-Penrose of matrix H.

2.2. Improved Particle Swarm Optimization for Extreme Learning Machines (IPSO-ELM)

2.2.1. Basic Particle Swarm Optimization (PSO). The basic PSO procedure is resulting from the reproduction of simplified social models [23]. Pretentious that in the d dimensional

pursuit planetary, there is a particle population of size n, and as (6) and (7), respectively [24].

$$V_{id}(t+1) = V_{id}(t) + c_1 r_1 \left(P_{id}(t) - X_{id}(t)\right) + c_2 r_2 \left(P_{gd}(t) - X_{id}(t)\right), \quad (6)$$

$$X_{id}(t+1) = X_{id}(t) + V_{id}(t+1). \quad (7)$$

In the formula, t is the number of iterations; $P_{id}(t)$ is the different optimum explanation; $P_{gd}(t)$ is the global optimal solution; c_1 and c_2 are the knowledge issues; r_1 and r_2 are accidental figures consistently dispersed among $[0, 1]$.

2.2.2. Improved Particle Swarm Optimization (IPSO). Aiming at the phenomenon that the PSO algorithm is disposed too early adulthood, a linear decreasing weight method is proposed. Decreasing weight method, even if the apathy weight reductions linearly from small to large, its change formula is [25]

$$\omega = \omega_{\max} - \frac{t \times (\omega_{\max} - \omega_{\min})}{t_{\max}}. \quad (8)$$

Here, $\omega_{\max}$ is the maximum apathy weight; $\omega_{\min}$ is the minimum apathy weight; t is the existing repetition ladder. The calculation steps of the linear decreasing weight method are as follows.

(1) Arbitrarily generate the velocity and location of individual particles

(2) Assess the suitability value of individual particles, stock the place and appropriateness value of the particle in the individual dangerous value P_{best} of the element, and accept the separate location and suitability value of the optimal suitability value in all P_{best}

(3) Update the velocity and position of the particle, and the control is exposed in the following equations.

$$V_{i,j}(t+1) = \omega V_{i,j}(t) + c_1 r_1 \left(P_{i,j}(t) - X_{i,j}(t)\right) + c_2 r_2 \left(P_{g,j}(t) - X_{i,j}(t)\right), \quad (9)$$

$$X_{i,j}(t+1) = X_{i,j}(t) + V_{i,j}(t+1),\ j = 1, 2, \ldots, n. \quad (10)$$

(4) Update the weight, the calculation is shown in the following formula.

$$\omega = \omega_{\max} - \frac{t \times (\omega_{\max} - \omega_{\min})}{t_{\max}}. \quad (11)$$

(5) When the model meets the disorder and reaches the rest, stop the search for corresponding consequence, else go back to step 3) to endure the study

2.2.3. IPSO-ELM Model. The IPSO-ELM model aims to globally optimize the connection weights and hidden layer by relying on the advantages of the improved particle group procedure with strong worldwide exploration skill and fast junction speed [26] and then optimize the performance of the IPSO-ELM model. Prediction accuracy and prediction speed [27]. The specific operation steps are as follows, and the diagram of the IPSO-ELM model is revealed in Figure 1 [28].

(1) Among the selected 35 groups of sample data

(2) Initialize particle swarm parameters and generate particles randomly

(3) Determine the ELM network and initialize the w and threshold of an extreme learning machine

(4) Calculate the output weights to obtain the MSE

(5) Take the MSE of the training output as the PSO fitness value

(6) Initialize P_{best} and g_{best}, update the speed and position of each particle according to P_{best} and g_{best}, and calculate the fitness value corresponding to the current particle

(7) Compare the current fitness value of the calculated particle swarm with the previously calculated P_{best} and g_{best}, update the individual extremum and the global extremum, and finally get the position of the optimal particle

(8) Assign the obtained optimal w and b to the extreme learning machine, predict, and output the test sample

2.3. Evaluation Standard. In this paper, NRMSE (Normalized Root Mean Square Error) [29] and NMAPE (Normalized Mean Absolute Percentage Error) [30] are used as prediction evaluation indicators. In addition, this study also used the mean squared error MSE to compare the three models. Its formula is as follows:

$$e_t = y_t - \widehat{y}_t,$$

$$\text{NMAPE} = \frac{1}{N} \sum_{t=1}^{N} |e_t| \times \frac{100}{C}\%, \quad (12)$$

$$\text{NRMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} e_t^2} \times \frac{100}{C}\%.$$

Among them, e_t is the difference between the sample and the real value; N is the entire quantity of examples.

3. Test Analysis

The stability of ABS is affected by many factors, among which the main factors are the rotor speed N_{ABS} of the automobile anti-lock brake system, the speed of the automobile V_{VEHICLE}, and the wheel radius r_{wheel}, and the transmission ratio of the main reducer i_0. These factors work together to affect the stability of ABS. and energy recovery performance, but there is a complex relationship between N_{ABS} and the transmission ratio of the final reducer, and it is difficult to make accurate judgments with traditional

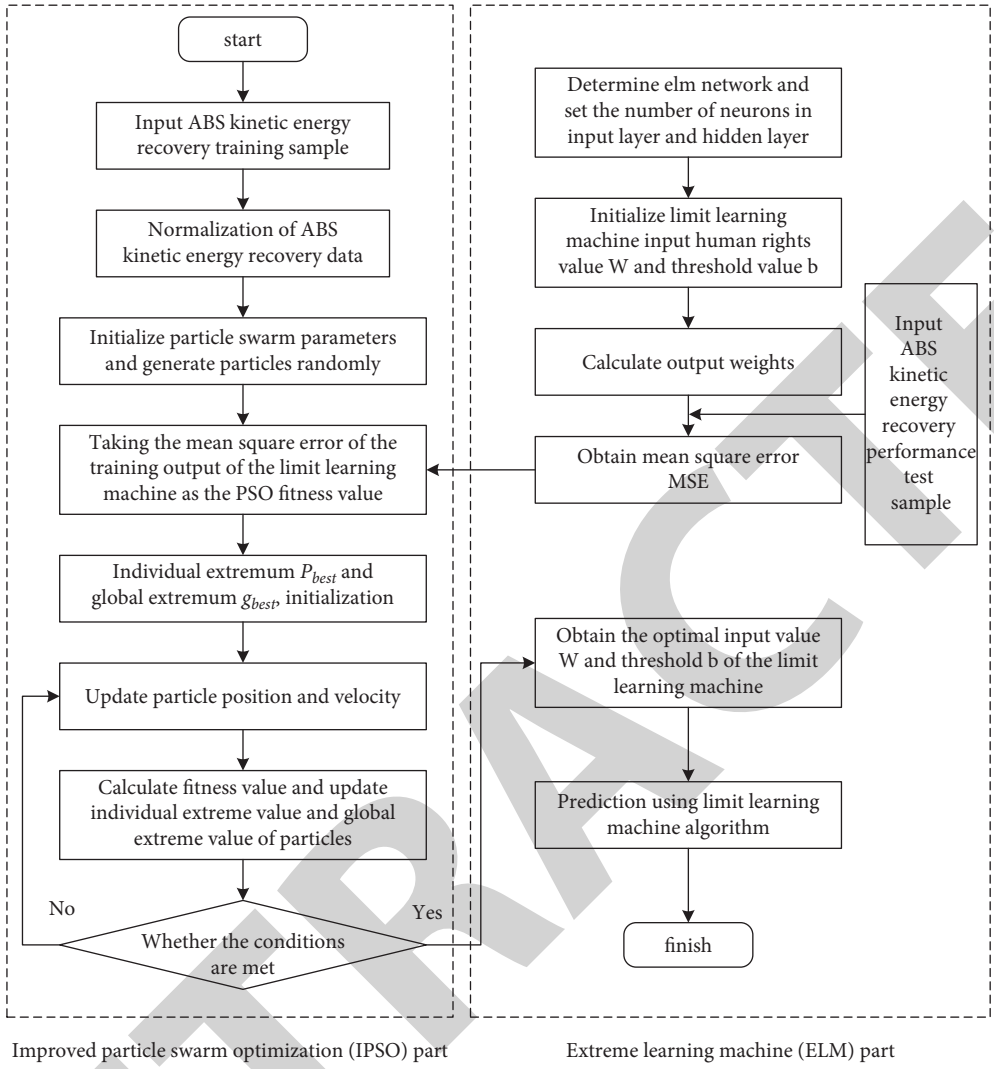


Figure 1: IPSO-ELM model diagram.

methods. The IPSO-ELM model could predict it, and four main influencing factors are made as the input restrictions model and output parameters model. 35 representative groups of sample data were selected, the first 30 groups were training data, and parameters were optimized, and the last 5 groups were test data. The detailed data are shown in Table 1.

To confirm the advantage of the IPSO-ELM, IPSO-ELM is associated with pure ELM and IPSO-ELM, as shown in Figure 2. It can be realized from Figure 2 that the IPSO-ELM has higher prediction accuracy. The predicted value of this model is nearby to the factual value, which is far better than ELM and IPSO-ELM. It can be realized from the comparison that the improved particle swarm algorithm is used to optimize the ELM model, which can advance the forecast accuracy of the ELM model and achieve a better prediction effect.

Based on MATLAB software, the MSE of each model can be obtained after running, and the MSEs of different models are compared, as exposed in Table 2. It can be realized from Table 2 that the MSE of the ELM model is 0.52376; the MSE of the PSO-ELM model is 0.45573; the MSE of the IPSO-ELM model is 0.09866. It can be realized that the IPSO-ELM model has high prediction accuracy, and has certain feasibility and effectiveness in the prediction of ABS stability and energy recovery performance.

Through MATLAB software, the iteration times of the IPSO-LEM model and PSO-ELM model are set to 100 times, and the error of each iteration is recorded and compared, as exposed in Figure 3. It can be gotten from Figure 3 that the iterative error of the IPSO-LEM model decreases successively. In the first 20 iterations, the error reduction range is large, and then the error reduction range is reduced, but it has been in a state of error reduction, and finally dropped to 0.043, indicating that the IPSO-LEM model has an ideal optimization effect. However, after 16 iterations of the PSO-LEM model, the iteration error does not change, and its value is stable at 0.13456, indicating that the forecast effect of the forecast model is not ideal and the prediction error is large.

The relation errors of the prediction results are compared in Table 3. The $\text{Error}_{\max}$ of the prediction results of the ELM model is 24.12%, and $\text{Error}_{\text{average}}$ is 16.12%; the maximum relative error of the PSO-ELM model prediction

Table 1: Training data and test data.

Serial number	N_{ABS}	$V_{VEHICLE}$	r_{wheel}	i_0	Serial number	N_{ABS}	$V_{VEHICLE}$	r_{wheel}	i_0
1	13.00	11.98	25	0.45	19	20.40	24.90	14	0.36
2	15.44	15.34	29	0.38	20	21.49	23.21	18	0.34
3	18.03	23.18	27	0.12	21	22.33	28.96	33	0.24
4	19.00	20.11	35	0.51	22	19.01	5.11	31	0.30
5	21.42	9.99	31	0.34	23	18.85	14.37	26	0.45
6	22.41	10.01	34	0.40	24	19.55	22.21	29	0.59
7	15.91	19.11	18	0.23	25	21.39	10.01	31	0.26
8	17.92	50.31	26	0.15	26	22.01	20.11	35	0.27
9	19.06	11.81	27	0.12	27	29.44	32.33	35	0.41
10	21.12	33.72	30	0.14	28	19.84	9.10	22	0.46
11	27.2	18.81	29	0.26	29	20.51	34.22	15	0.22
12	18.85	9.21	17	0.25	30	21.22	24.81	14	0.36
13	17.6	19.33	26	0.17	31	22.41	39.51	15	0.14
14	19.52	11.98	21	0.41	32	18.91	15.33	24	0.39
15	20.21	17.21	24	0.31	33	21.41	33.52	31	0.21
16	21.52	6.94	31	0.36	34	22.82	8.63	12	0.51
17	18.02	24.01	32	0.12	35	23.41	21.29	33	0.18
18	18.93	16.71	9	0.41					

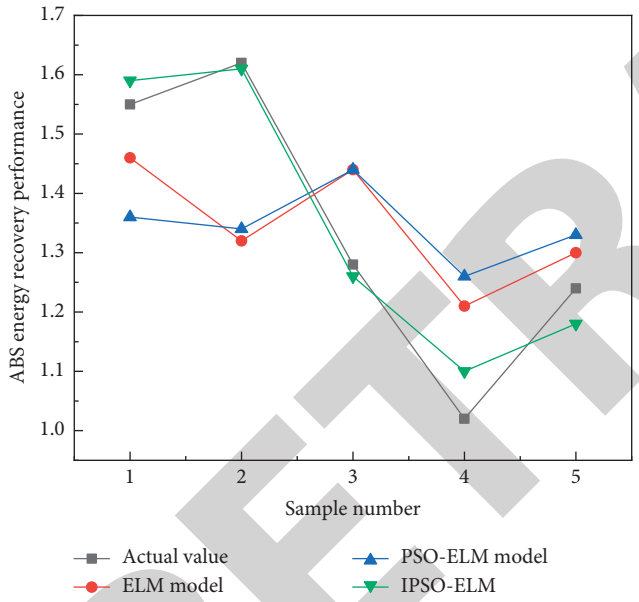


Figure 2: Comparison graph of predicted value and true value of different models.

Table 2: MSEs of different models.

Model	ELM	PSO-ELM	IPSO-ELM
MSE	0.52376	0.45573	0.09866

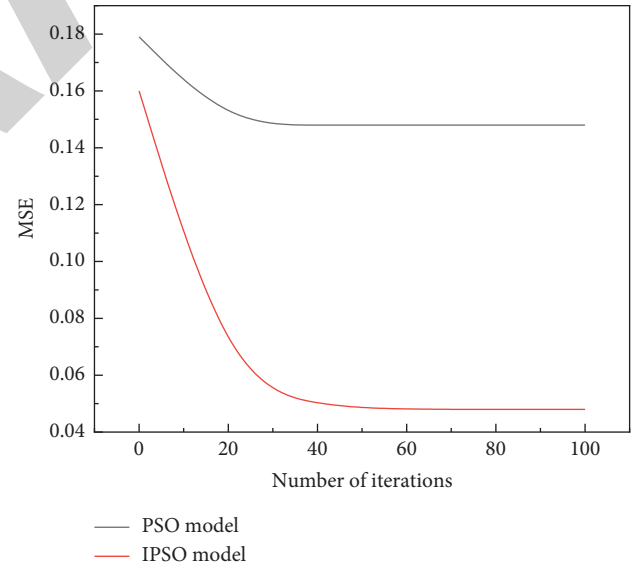


Figure 3: Error change diagram of 100 iterations for different models.

results is 24.55%, and $Error_{average}$ is 15.19%. $Error_{max}$ of IPSO-ELM model prediction results is 5.43% and $Error_{average}$ is 2.72%. It can be realized that the IPSO-ELM model has high prediction accuracy, the prediction results are relatively stable, the error among the forecasted actual is small, the model construction is reasonable, and the applicability is strong.

For the data of this study, comparing the ELM model and the classical method, it can be realized from the prediction consequences in Table 4 that when the prediction time is 24 h, the IPSO-ELM model has the highest accuracy. The NRMSE values decreased by 51.6% and 17.7%, respectively, and the NMAPE values decreased by 54.9% and 12.3%, respectively. When the prediction time was 48 h, the NRMSE values of the IPSO-ELM model decreased by 44.6% and 1.9%, and the NMAPE values decreased by 51.8% and 4.4%, respectively, compared with the PSO-ELM and ELM models. For the 72-hour power prediction, the NRMSE values of the IPSO-ELM model decreased by 31.6% and 7.1%, and the NMAPE values were reduced by 37.8% and 1.2%, respectively. For other classical models, the NRMSE and NMAPE values of the proposed model decreased the most within 24 h, followed by the second within 48 h.

Table 3: Relative errors of prediction results of different models.

Sample number	Actual value	ELM	PSO-ELM	IPSO-ELM	Relative error (ELM and true value) (%)	Relative error (PSO-ELM and true value) (%)	Relative error (IPSO-ELM and true value) (%)
1	1.560	1.484	1.361	1.596	4.9	12.74	2.32
2	1.630	1.239	1.341	1.615	24.12	17.67	0.91
3	1.280	1.491	1.447	1.277	16.51	13.07	0.27
4	1.030	1.261	1.283	1.086	22.36	24.55	5.43
5	1.250	1.413	1.350	1.192	13.06	8.02	4.66
Average value	1.350	1.377	1.357	1.353	16.12	15.19	2.72

Table 4: Experiment results with other models.

Evaluation criteria	Model	Prediction duration (h)		
		24	48	72
NRMSE	ELM	48.56	56.72	46.43
	PSO-ELM	28.53	32.02	34.17
	IPSO-ELM	23.47	31.39	31.73
NMAPE	ELM	41.33	50.65	39.71
	PSO-ELM	21.23	25.52	24.99
	IPSO-ELM	18.60	24.39	24.68

4. Conclusion

Under the current technical conditions, only the method of braking energy recovery can be used to increase the mileage of electric vehicles. During the braking process, the ABS controls the brake pressure to keep the slip ratio near an optimal value when the wheels tend to lock, which can not only exert the maximum braking force but also ensure the handling and stability of the vehicle. If the braking energy regeneration system and the ABS system are reasonably coordinated and controlled, the economy of the vehicle can be improved, and the safety of driving can be enhanced. This study optimizes the ABS system through the IPSO-ELM model so that energy vehicles can maximize the energy generated by the ABS system, so as to achieve the purpose of kinetic energy recovery and reduce energy consumption and vehicle costs.

The conclusions are as follows. (1) On the basis of the PSO-ELM model, a linear decreasing weight method is introduced, and then, an IPSO-ELM model is proposed for the optimization analysis of ABS braking kinetic energy; (2) By comparing the prediction results of the three models with the actual values, the results show that compared with the pure ELM model and the PSO-ELM model, the IPSO-ELM model has significantly smaller simulation MSE and relative error and has higher generalization ability and prediction accuracy. The maximum relative error of IPSO-ELM model prediction results is 5.43%, and the average relative error is 2.72%. Compared with ELM and PSO-ELM, the error is lower. It is extremely effective to use IPSO-ELM for ABS kinetic energy recovery optimization.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by The Characteristic Innovation Projects of Colleges and Universities in Guangdong Province (2020KTSCX379) and Guangzhou Science and Technology Project (201904010118).

References

[1] M. Gavas, "Increasing the drawing height of conical square cups using anti-lock braking system (ABS)," *Journal of Mechanical Science and Technology*, 2009.

[2] H. Schramm and W. Stumpe, *ANTI-LOCK BRAKING SYSTEM (ABS) FOR VEHICLES WITH AN ADDITIONAL AXLE*, US, 2003.

[3] M. Gavas and M. Izciler, "Deep drawing with anti-lock braking system (ABS)," *Mechanism and Machine Theory*, vol. 41, no. 12, pp. 1467–1476, 2006.

[4] C. K. Huang and M. C. Shih, "Design of a hydraulic anti-lock braking system (ABS) for a motorcycle," *Journal of Mechanical Science and Technology*, vol. 24, no. 5, pp. 1141–1149, 2010.

[5] H. C. Chen, A. Wisnujati, A. M. Widodo, Y. L. Song, and C. W. Lung, "Antilock Braking System (ABS) Based Control Type Regulator Implemented by Neural Network in Various Road Conditions," *Advances in Networked-Based Information Systems,Lecture Notes in Networks and Systems*, 2022.

[6] M. Gavas, "Increasing the drawing height of conical square cups using anti-lock braking system (ABS)," *Journal of Mechanical Science and Technology*, vol. 23, no. 11, pp. 3079–3087, 2009.

[7] J. Rauh and D. Ammon, "System dynamics of electrified vehicles: some facts, thoughts, and challenges," *Vehicle System Dynamics*, vol. 49, no. 7, pp. 1005–1020, 2011.

[8] Y. Jang, M. Lee, I. S. Suh, and K. Nam, "Lateral handling improvement with dynamic curvature control for an independent rear wheel drive EV," *International Journal of Automotive Technology*, vol. 18, no. 3, pp. 505–510, 2017.

[9] G. Xu, K. Xu, C. Zheng, X. Zhang, and T Zahid, "Fully electrified regenerative braking control for deep energy recovery and maintaining safety of electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1186–1198, 2016.

[10] F. Z. Ji, X. X. Zhou, and W. B. Zhu, "Coordinate control of electro-hydraulic hybrid brake of electric vehicles based on carsim," *Applied Mechanics and Materials*, vol. 490-491, pp. 1120–1125, 2014.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1169438, 9 pages
https://doi.org/10.1155/2022/1169438

Research Article

Using Convolutional Neural Network as a Statistical Algorithm to Explore the Therapeutic Effect of Insulin Liposomes on Corneal Inflammation

Yukun Liao,[1,2] **Huiting Jiang**,[1,2] **Yangrui Du**,[1] **Xiaojing Xiong**,[1,2] **Yu Zhang**,[1] **and Zhiyu Du**[1]

[1]*Department of Ophthalmology, Second Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, 400010, China*
[2]*Chongqing Key Laboratory of Ultrasound Molecular Imaging, Second Affiliated Hospital of Chongqing Medical University, Chongqing Medical University, Chongqing, 400010, China*

Correspondence should be addressed to Zhiyu Du; dr.duzhiyu@163.com

Received 6 June 2022; Accepted 28 June 2022; Published 31 July 2022

Academic Editor: Man Fai Leung

Aiming at the disadvantages of easy recurrence of keratitis, difficult eradication by surgery, and easy bacterial resistance, insulin-loaded liposomes were prepared, and convolutional neural network was used as a statistical algorithm to build SD rat corneal inflammation model and study insulin-loaded liposomes, alleviating effect on corneal inflammatory structure in SD rats. The INS/PFOB@LIP was developed by means of thin-film dispersive phacoemulsification, its structure was monitored using a transmission electron microscope, particle size and appearance potential were monitored using a Malvern particle sizer, and ultraviolet consumption spectrum was monitored using a UV spectrophotometer. The encapsulation rate, drug loading, and distribution of insulin liposomes in rat corneal inflammatory model were measured and calculated. The cytotoxicity of liposome materials was evaluated by CCK-8 assay, and the toxic effects of insulin and insulin liposomes on cells were detected. The cornea of SD rats was burned with NaOH solution (1 mol/L), and the SD rat corneal inflammation model was created. The insulin liposome was applied to the corneal inflammation model, and the therapeutic effect of insulin liposome on corneal inflammation was evaluated by slit lamp, corneal immunohistochemistry, corneal HE staining, and corneal Sirius red staining. Insulin-loaded liposomes were successfully constructed with an average particle size of (130.69 ± 3.87) nm and a surface potential of (-38.24 ± 2.57) mV. The encapsulation rate of insulin liposomes was $(48.89 \pm 1.24)\%$, and the drug loading rate was $(24.45 \pm 1.24)\%$. The SD rat corneal inflammation model was successfully established. After insulin liposome treatment, the staining area of corneal fluorescein sodium was significantly reduced, the corneal epithelium was significantly thickened, the content of corneal collagen was increased, the expression of inflammatory factors was significantly reduced, and new blood vessels (corneal neovascularization, CNV) growth was inhibited.

1. Introduction

The cornea is a kind of crystal clear tissue, which plays a pivotal role in the improvement of visual function. Considering that the cornea has direct contact with the surrounding space, it also replaces the mechanical protective layer and has an important barrier effect on microorganisms, poisons, and injuries [1]. The cornea is located in the outermost layer of the eye and is highly susceptible to infection. The infection rate of the cornea accounts for 15% of eye injuries, and corneal inflammation is the most common [2]. Various external factors may cause keratitis. It is one of the common eye diseases and one of the major blinding eye diseases in the world. The cornea is located at the front of the eyeball and is in direct contact with the outside world. The clinical manifestations are blurred vision, pain, photophobia and lacrimation, and other stimuli symptoms and obvious vision loss. Ophthalmological examination showed the disappearance of corneal luster, reduced transparency, and ulceration. In severe cases, corneal perforation, intraocular infection, and even blindness may occur. At present, keratitis is divided into infectious, immune, dystrophic, nerve

paralysis, and exposure according to the pathogenic cause. Infectious keratitis mostly occurs in the central area of the cornea, while immune keratopathy tends to occur in the peripheral area of the cornea [3].

Insulin (INS) is a small molecule protein with anti-inflammatory, antiapoptotic, and cell growth-promoting effects. It can be used to treat chronic inflammation and promote wound healing. Due to the barrier effect of the tear film and the cornea, the efficiency of drug entry into the eye is extremely low. The liposome (LIP) drug-delivery system can improve drug bioavailability [4]. Liposomes are artificial spherical vesicles composed of single or multiple phospholipid bilayers, with a hydrophilic structure inside and a lipophilic structure outside, similar to cell membranes. Liposomes have complete biodegradability, low toxicity, and high stability, and can also prolong the residence time of encapsulated drugs on the corneal surface through sustained release. They have been used in ophthalmology as drug carriers in the 20th century [5].

Because of its good corneal penetration and biocompatibility, as well as the advantages of targeting and sustained release, it is suitable for applications in the field of ophthalmology. Eye instillation is a commonly used administration method in ophthalmology. Compared with ocular injection, eye drops can reduce the infection risk and economic burden of patients and improve patient compliance. Using liposomes as a carrier can improve the stability, bioavailability, and targeting of insulin on the ocular surface, reduce systemic adverse reactions, and eye irritation, and achieve a sustained release effect [6]. In order to better detect the effect of liposomes on drug loading and cell activity, the method of deep learning is used for relevant statistical analysis. As a common form of deep learning, convolutional neural network is now processing experimental data [7–10] and it is widely used. A convolutional neural network is a multilayer neural network with a deep supervised learning architecture that can be viewed as a combination of two parts: an automatic feature extractor and a trainable classifier. The feature extractor includes convolution filtering and downsampling [11, 12]. The convolution filter kernel on the feature map is usually $3 * 3$, and the downsampling operation after filtering usually has a ratio of 2. The classifier is composed of a fully connected network, and the features learned by the feature extractor will be output by the fully connected layer [13–15]. The purpose of this study was to use convolutional neural network as a statistical algorithm to construct a model of corneal inflammation in SD rats and to explore the therapeutic effect of insulin liposomes on corneal inflammation.

2. Materials and Methods

2.1. Experimental Drugs. Recombinant human insulin (Beijing Soleibao Technology Co., Ltd.), soybean lecithin (purity ≥99%, Shanghai McLean Biochemical Technology Co., Ltd., China), cholesterol (China Xianruixi Biotechnology Co., Ltd.), perfluorooctane (PFOB, purity ≥99%, USA, Sigma), chloroform (CHCL3, Chongqing Chuandong Chemical Group), methanol (CH3OH, Chongqing Chuandong Chemical Group), and methoxy polyethylene glycol 2000-distearoyl phosphatidylethanolamine were used.

2.2. Experimental Apparatus. The experimental instruments used in this experimental study are shown in Table 1.

2.3. Experimental Animals. Forty SPF healthy SD rats aged 10 to 12 weeks were selected, and the right eye was taken as the experimental eye, with a body weight of 300 g to 450 g (provided by the Animal Experiment Center of the Chongqing Medical University). Before the experiment, the animals were observed in the animal experiment center for one week, and routine slit lamp examination was performed to exclude diseases related to the ocular surface.

2.4. Preparation of Insulin Liposomes. Insulin liposomes were prepared by thin-film dispersion phacoemulsification method. We weighed appropriate amount of soybean lecithin, DSPE-PEG2000, and cholesterol. All three were dissolved in chloroform at a temperature of 50°C and a rotational speed of 130 rpm. The organic solvent was evaporated to dryness under the swirl, so that a uniform film was formed in the round-bottomed flask, and then the buffer reagent was added, and the treatment was performed by shaking and washing the film, and ultrasonically emulsified for 30 W 1 min to form a W/O emulsion. The emulsion was added to the film eluent, and sonicated for 45 W 3 min to obtain nanoparticle liposomes. Then, centrifuge at 6000 rpm/min for 5 min, remove the supernatant, collect the precipitate, store at 4°C, low temperature, and protect from light for later using.

2.5. Characterization Assay of Insulin Liposomes. Particle size and potential: take 100 μL of insulin diluted hydrochloric acid solution, dissolve it in triple distilled water, and measure the particle size and Zeta potential of INS/PFOB@LIP with a Malvern particle size. Encapsulation efficiency and drug loading: prepare a certain amount of insulin liposomes, and measure the encapsulation efficiency and drug loading of the drug by the demulsification method. The calculation formula is as follows:

$$\phi = \frac{P_{\text{in}}}{P_{\text{all}}} \times 100\%, \tag{1}$$

$$\eta = \frac{P_{\text{in}}}{Z_{\text{all}}} \times 100\%, \tag{2}$$

where Φ represents the insulin encapsulation rate; P_{in} represents the drug loading; P_{all} represents the total input amount; η represents the insulin drug loading rate; and Z_{all} represents the total amount of liposomes.

2.6. Liposome Distribution and Animal Model Establishment. In a dark environment, the corneal inflammation model rats were instilled with DiI-labeled insulin liposomes. After 2 h, the corneas were removed, and 10 μm frozen sections were made. Insulin immunofluorescence staining was observed and photographed under a fluorescence microscope.

Table 1: Instruments required for the experiment.

Experimental instrument name	Company
Rotary evaporator	Shanghai Yarong Biochemical Instrument Factory
Sonicator	United States, Sonics
37°C cell incubator and light microscope	Olympus Corporation
Dragonfly 200 high-speed confocal imaging platform	Oxford Instruments Technology Ltd
4°C centrifuge	Thermofisher Corporation
Transmission electron microscope	TEM, Japan, HiTachi

40 SD rats (provided by the Animal Experiment Center, Chongqing Medical University, on behalf of them), weighing 300 g to 450 g, were randomly divided into normal saline group (NS group), liposome empty sphere group (PFOB@LIP), insulin group (INS), and insulin liposome group (INS/PFOB@LIP), 10 in each group. Three days before modeling, rats were given 30 μL/time of levofloxacin eye drops in cornea and conjunctival sac of both eyes, three times a day for 3 days. The rats were anesthetized by intraperitoneal injection of 4 mL/kg ketamine solution (100 mg/mL), and 30 μl of lidocaine hydrochloride eye drops was administered. A circular filter paper with a diameter of 3 mm infiltrated with NaOH solution was placed in the center of the cornea of the right eye of the rat for 40 s, the cornea and conjunctival sac were washed with a large amount of normal saline for 3 minutes, and the left eye was left without any treatment to establish a control experiment of corneal epithelial cell inflammation model. We used NS, PFOB@LIP, INS, and INS/PFOB@LIP (the concentration of liposomes in PFOB@LIP and INS/PFOB@LIP is 400 μg/mL, and the concentration of INS is 200 μg/mL) in the right eye, respectively, 30 μL, once in the morning, noon, and night, for 14 days.

2.7. Evaluation of the Therapeutic Effect of Keratitis. At 1, 3, 7, and 14 days after the corneal epithelial cells appeared obvious inflammation, the cornea and new blood vessels were observed and photographed under white light and no red light using a slit lamp. 2 μL of 1% fluorescein sodium dye solution was given to stain the cornea, and the excess dye solution is rinsed with normal saline. The cornea was observed and photographed under cobalt blue light using a slit lamp. Referring to the literature [6], the cornea is divided into 4 parts by the horizontal and vertical axes, and Image-*J* software is used to analyze and calculate the area of corneal neovascularization of each part and sum up 3 rats were randomly selected from each group at each time point, sacrificed by cervical dislocation, and corneal tissue was harvested for tissue fixation, embedding, and sectioning. To detect the uptake of liposomes by cells, the red cell membrane fluorescent probe DiI was loaded into the liposomes as a fluorescent dye. After 24 h of incubation, paraformaldehyde was fixed and HE staining was used to observe epithelial cells and fibrous tissue and the number of inflammatory cells. The arrangement and content of collagen fibers were observed by sirius scarlet staining. Immunohistochemical staining was used to detect the protein expressions of CD31, IL-1β, and TNF-α. Image-J software was used to analyze the positive reaction sites in the photographs, and the relative optical density was measured.

2.8. Statistical Analysis Based on Convolutional Neural Network Algorithm. In order to better detect the effects of insulin liposomes on drug loading, cell activity, and the therapeutic effect of keratitis, a deep learning method was used to conduct relevant statistical analysis [16, 17]. The convolutional neural network algorithm is one of the most important frameworks in deep learning. It can make good use of the prior information of "space-time locality" in the data, so as to achieve the purpose of enabling the network to remember past time information, so that the network can not only realize the connection from bottom to top (input-output), but also realize the information transmission and recording from left to right (time t to time $t+1$).

Convolutional neural network is constructed by imitating the visual perception mechanism of biology, which can perform supervised learning and unsupervised learning. Small computational effort for grid-like topology features, such as modeling insulin absorption and depletion, with stable effects and no additional feature engineering requirements on the data [18–21]. Therefore, the use of the convolutional neural network algorithm can more accurately analyze the data collected from the experiment. The statistical data are expressed as the mean ± standard deviation, which is statistically significant. The specific algorithm flow is as follows.

2.8.1. Forward Propagation

(1) Convolutional Layer. The network search data are two-dimensional feature data. After the input data and the filter are convolved, the local features of the data are extracted. In this paper, X_j^l is used to represent the output of the 1th channel of the convolutional layer, W_{ij}^l is the weight vector of the *1*th channel of the convolutional layer, b_j^l is the bias vector of the layer, and M_j is the receptive field of the convolution kernel on the input data, " $*$ " represents the convolution operation, $f(*)$ is the activation function, the process of extracting features is shown in [22, 23]

$$z_j^l = \sum_{i \in M_j} X_i^{l-1} * W_{ij}^l + b_j^l,$$
$$X_j^l = f\left(z_j^l\right). \tag{3}$$

In the convolution layer, a convolution kernel is only locally connected to the input data and has nothing to do

with other data, reducing irrelevant connections, which is the local receptive field; at the same time, this convolution kernel uses the same volume on one channel of the data. The product kernel performs feature mapping and reduces the number of parameters, which is called weight sharing. These two characteristics are one of the cores of the simplified operation of convolutional neural network.

(2) Downsampling Layer. Downsampling the data can reduce the amount of data processing while retaining the useful information of the data, which is another major feature of the convolutional neural network to reduce parameter calculation [24–26]. The downsampling generally uses a $2 * 2$ window. If it is a maximum pooling layer, the maximum value of the $2 * 2$ numbers is taken; if it is an average pooling layer, the average of these 4 numbers is taken. X_j^l represents the feature of the j-th channel in layer l, β_j^l is the weight, and b_j^l is the bias. *down* ( ) represents downsampling, f ( ) is the activation function, then the principle of the downsampling layer is as follows:

$$\begin{aligned} z_j^l &= \beta_j^l \operatorname{down}\left(X_j^{l-1}\right) + b_j^l, \\ X_j^l &= f\left(z_j^l\right). \end{aligned} \tag{4}$$

(3) Fully Connected Layer. The data passing through the convolutional layer is flattened into a one-dimensional vector, and then enters the fully connected layer for classification or regression. The calculation of the fully connected layer is shown in (2). In the sales forecast, it is a regression problem, and the output is a real number, so the loss function adopts MSE (mean square error). The calculation formula of MSE is shown in (4), where $\hat{y}_i$ represents the predicted value, y_i represents the real value, and m is the number of samples [27, 28].

$$\begin{aligned} X^l &= f\left(W^l X^{l-1} + b^l\right), \\ \mathrm{MSE} &= \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2. \end{aligned} \tag{5}$$

2.8.2. Backpropagation. In the forward propagation phase, the loss of the model is obtained in each round of iteration. In the back propagation phase, the loss is backpropagated through the gradient descent method to update the weights and biases of each layer of the convolutional neural network. Through parameter update, the loss of the model no longer decreases, then the requirement is met. The main reference for backpropagation of convolutional neural networks is [29, 30].

(1) Convolutional Layer. Since the next layer of the convolutional layer is usually a downsampling layer, the sensitivity of the downsampling layer needs to be used to represent the sensitivity of the convolutional layer during backpropagation. For example, if the maximum pooling layer is used, the size of the pooling area is $2 * 2$, each value of the $l + 1$ layer corresponds to the $2 * 2$ area of the lth layer, and the up-operation puts each value of the $l + 1$ layer into the Return to the position of the original maximum value of this $2 * 2$ area and adds 0 to other positions. δ represents the Hadamard product, representing the multiplication of the corresponding elements of the matrix, and then the sensitivity of the convolutional layer is [31, 32]

$$\delta_j^l = \beta_j^{l+1}\left(f'\left(z_j^l\right) \circ \operatorname{up}\left(\delta_j^{l+1}\right)\right). \tag{6}$$

When deriving the parameters of the convolution kernel, it is necessary to consider the input features multiplied by the convolution kernel. $x_{u+i-1,\,v+j-1}^{l-1}$ is the part that is multiplied by W_{ij}^l input elements, η is the learning rate, and then the weights of the convolutional layers are updated as follows:

$$W_{ij}^l = W_{ij}^l - \eta \sum_u \sum_v \left(\delta_{uv}^l x_{u+i-1,\,v+j-1}^{l-1}\right). \tag{7}$$

Since each output value of the channel is related to the bias term, the gradient of the bias term is the sum of the sensitivity elements of the channel and the bias term is updated as follows [33]:

$$b_j^l = b_j^l - \eta \sum_{u,v} \left(\delta_j^l\right)_{u,v}. \tag{8}$$

(2) Downsampling Layer. When the lth layer is a downsampling layer, assuming that the $l+1$th layer is a convolutional layer, the sensitivity of the next layer is used to represent the sensitivity of the previous layer, and the calculation is as follows [34]:

$$\delta_j^l = \delta^{l+1} * \operatorname{rot}180\left(W_j^l\right) \circ f'\left(z_j^l\right). \tag{9}$$

$rot180\,(W_j^l)$ means that the convolution kernel is rotated by 180°, and δ^{l+1} needs to be filled with 0 s on its boundary, so that the size is the same as z_j^l. The update of the bias term and weight is shown in the following equations [35]:

$$b_j^l = b_j^l - \eta \sum_{u,v} \left(\delta_j^l\right)_{u,v}, \tag{10}$$

$$\beta_j^l = \beta_j^l - \eta \sum_{u,v} \left(\delta_j^l \circ \operatorname{down}\left(X_j^{l-1}\right)\right)_{u,v}. \tag{11}$$

3. Results

3.1. Effects of Insulin and Insulin Liposomes on Corneal Cell Viability. The cytotoxicity of liposome materials was evaluated by CCK-8 assay, and the toxic effects of insulin and insulin liposomes on cells were detected. The cytotoxic effects of liposomal materials were investigated using corneal cells. Figure 1 shows the viability of cells after insulin, blank liposome, and insulin liposome treatment. The cells treated with blank liposomes still had a cell viability of more than 90% after being placed for 400 h, indicating that the

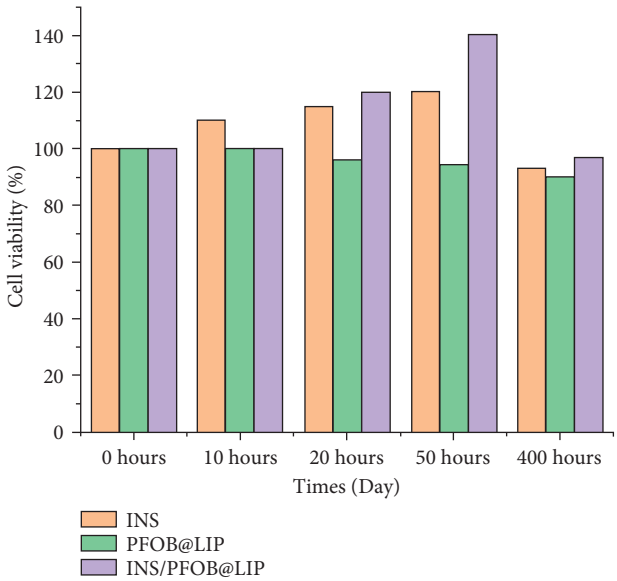


Figure 1: The effects of INS, PFOB@LIP, and INS/PFOB@LIP on corneal cell viability.

liposome material has high biocompatibility and no obvious cytotoxicity.

The results show that INS/PFOB@LIP can significantly increase the cell activity of corneal cells for 20 h, so this study selects insulin liposomes for follow-up experiments.

3.2. Physicochemical Properties, Encapsulation Efficiency, and Drug Loading Capacity of Liposomes. We dissolve PFOB@LIP and INS/PFOB@LIP in PBS to prepare 400 μg/mL 2 mL. The INS solution (200 μg/mL 2 mL) is colorless and transparent, the PFOB@LIP and INS/PFOB@LIP suspensions are milky white, and the INS/PFOB@LIP suspension is dark. Transmission electron microscopy showed that INS/PFOB@LIP had a relatively uniform and consistent spherical appearance, the liposome was a shell-core structure, and a phospholipid bilayer could be observed (Figure 1(a)). The average particle size of PFOB@LIP is (105.39 ± 2.83) nm, and the average particle size of INS/PFOB@LIP is (130.69 ± 3.87) nm (Figures 2(b) and 2(c). The Zeta potentials of PFOB@LIP and INS/PFOB@LIP were (−10.03 ± 1.27) mV and (−38.24 ± 2.57) mV, respectively (Figure 1(d)) ($P < 0.001$, the difference was statistically significant). According to the physical and chemical properties of INS, it can be expressed by the following linear regression equation: $Y = 0.01476 \times X - 0.1203$, $R^2 = 0.9854$; the encapsulation efficiency and drug loading of insulin INS are calculated as (48.89 ± 1.24)% and (24.45 ± 1.24)%.

3.3. Distribution of Liposomes in Rats. After 4 h, the frozen sections of the cornea were taken out in batches for testing. The results showed that the cornea had a strong fluorescent signal, indicating that the insulin liposomes had entered the eye and were diffusely distributed. To more accurately measure the distribution of liposomes in rats, the uptake of fluorescent liposomes by corneal epithelial cells was observed by confocal laser microscopy. It can be seen that the fluorescent liposomes in the nuclei and corneal epithelial cells gradually increased, as shown in Figure 3. This indicated that the uptake of insulin liposomes by corneal epithelial cells increased with time, indicating that corneal epithelial cells had strong uptake ability to liposomes.

3.4. Evaluation of the Therapeutic Status of Rat Corneal Inflammation. In order to more accurately analyze the therapeutic effect of insulin liposome on corneal inflammation, statistical analysis was done using convolutional neural network. First, the percentage of corneal epithelial defect area is extracted separately, the percentage of corneal epithelial defect area, NS, PFOB@LIP, INS, and INS/PFOB@LIP is brought into the convolutional neural network statistical analysis model, and the results are shown in Figure 4(a). As can be seen from Figure 4(a), in 1 d–7 d, the corneal epithelial defect area gradually decreased, and the percentage of corneal epithelial defect area at 7 days was (34.49 ± 0.98)% in the NS group; (32.43 ± 0.18)% in the PFOB@LIP group; and (0.54 ± 0.4)%; INS/PFOB@LIP group was (0 ± 0)%, and INS group and INS/PFOB@LIP group basically healed. INS/PFOB@LIP compared with each group, the difference was statistically significant ($P < 0.01$). On the 7-th day, the new blood vessels in each group exceeded the cornea and tended to invade the center of the cornea. The new blood vessels were small, pale, sparse, and brush-shaped. On the 14-th day, the new blood vessels in each group grew towards the burn area, the blood vessels gradually thickened and denser, and the growth of INS/PFOB@LIP blood vessels was the slowest. From 7 d to 14 d, the new blood vessels increased in each group, and the area of new blood vessels in the INS and INS/PFOB@LIP groups was significantly smaller than that in the other groups.

Secondly, the recalculated area percentage of new blood vessels in each group after 7 days and 14 days, NS, PFOB@LIP, INS, and INS/PFOB@LIP were brought into the convolutional neural network statistical analysis model, and the results are shown in Figure 4(b). As shown in Figure 4(b), the area of new blood vessels in the INS/PFOB@LIP group was smaller than that in the other groups, and the difference was statistically significant ($P < 0.05$). Through daily inspection, it was found that the corneal epithelium of each group had different degrees of defects on the 1st and 3rd day after the cornea of the rats appeared inflammation, which did not reach the normal cell coverage of 6–8 layers, the arrangement of collagen fibers was loose and disordered, and a large number of inflammatory cells were infiltrated. At 7 d, the corneal epithelium of NS and PFOB@LIP was still defective and the tissue structure was disordered. The epithelial cells of INS and INS/PFOB@LIP basically reached 5–6 layers. In the four groups, the inflammatory cells and corneal thickness reached the peak value, and the formation of neovascular cavity was seen in the stromal layer. On the 14th day, the inflammatory cells in the four groups were reduced, the corneal epithelial cells reached 6–7 layers, and a large number of new blood vessels were formed, red blood cells infiltrated in the lumen, and the corneal tissue structure

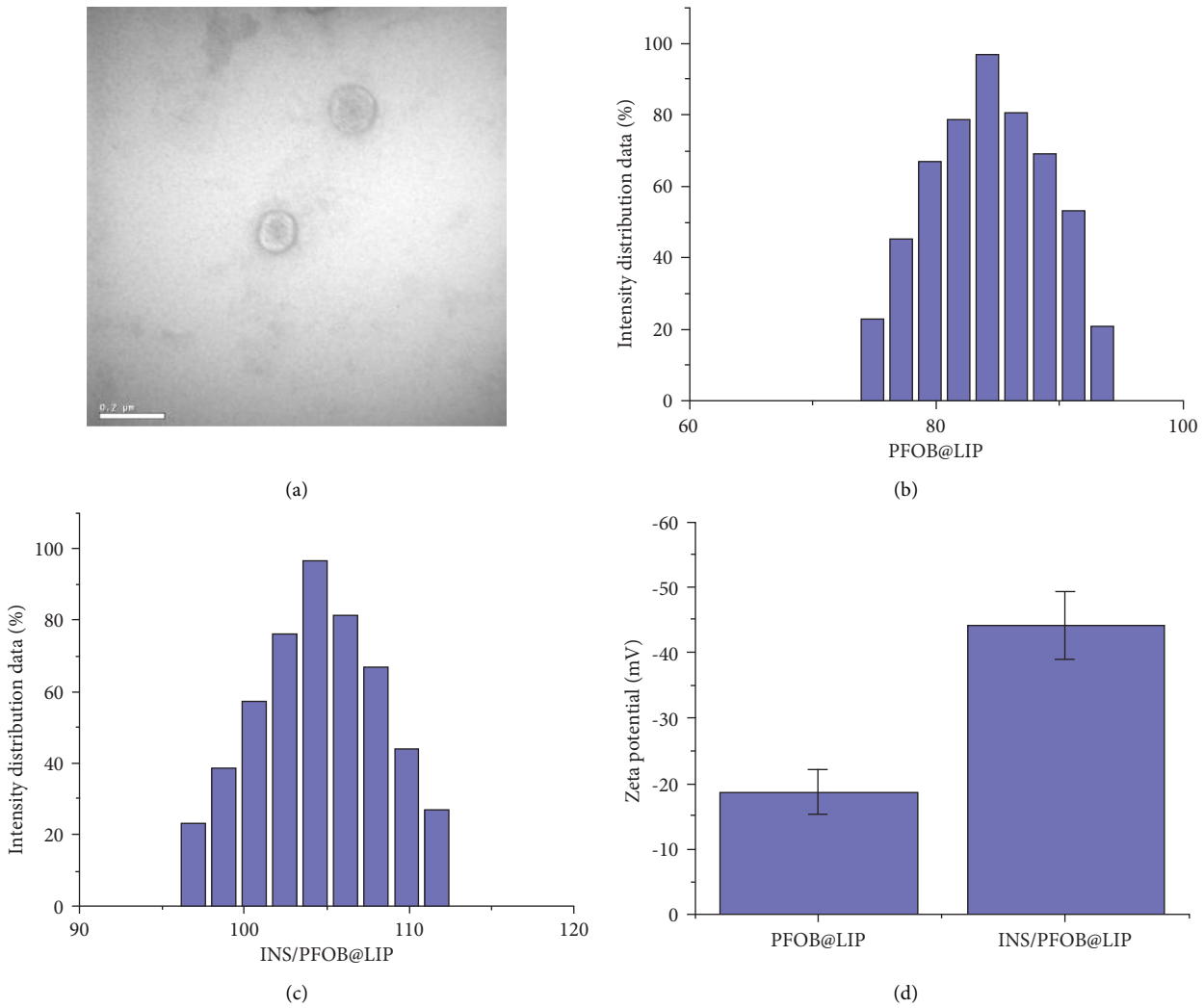


Figure 2: (a) TEM of INS/PFOB@LIP (50 μg/mL). (b) Particle size distribution of PFOB@LIP. (c) Particle size distribution of INS/PFOB@LIP. (d) Zeta potential maps of PFOB@LIP and INS/PFOB@LIP.

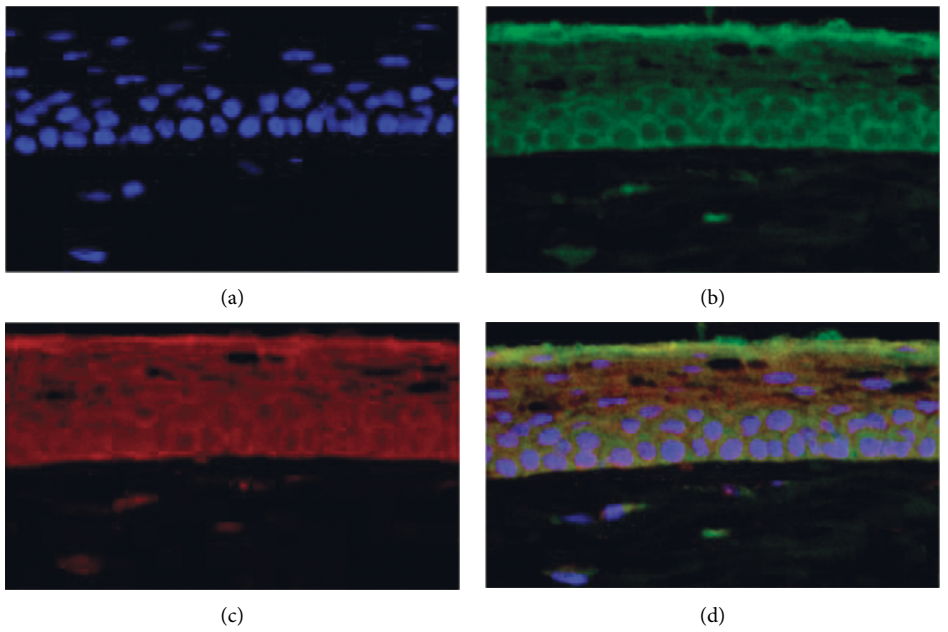


Figure 3: (a) Blue represents DAPI. (b) Green represents insulin. (c) Red represents liposomes. (d) Fluorescence image of cornea after 2 hours (×400).

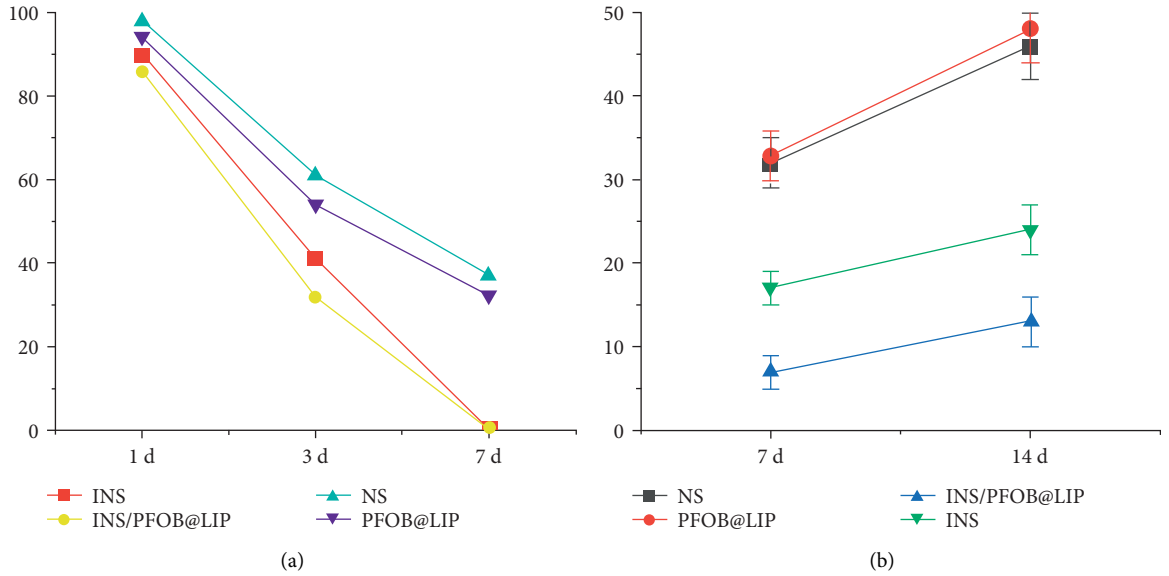


Figure 4: (a) 7 days comparison of corneal inflammation area; (b) 14 days comparison of new blood vessel area.

tended to be neat. After 14 days, the rat corneal cells were detected, and it was found that in the INS/PFOB@LIP group, the collagen fibers in the cornea were densely arranged and not neatly arranged, while the collagen fibers in the other three groups were thick, abundant, criss-cross, loose, and disordered. Histochemical staining was used to observe the growth of corneal neovascularization and inflammatory infiltration by vascular endothelial cell marker CD31, inflammatory cell markers IL-1β, and TNF-α. On the 14-th day, NS, PFOB@LIP, and INS showed a large amount of angiogenesis and a large number of inflammatory cells infiltration. The relative optical density of each cytokine was detected by immunohistochemical method on 14 days, the average optical density of the blank control group was used as the reference, and the parameter was 1. The relative optical density of each group is shown in Figure 4, and the difference between each group was statistically significant ($p < 0.05$):

(a) $P < 0.01$, compared with INS/PFOB@LIP group. In each group, the area of corneal epithelial inflammation gradually decreased for 7 days, INS and INS/PFOB@LIP healed completely, and INS/PFOB@LIP recovered the fastest.

(b) $P < 0.05$, compared with the INS/PFOB@LIP group. The area of neovascularization in each group increased at 14 days, and the area of neovascularization in the INS/PFOB@LIP group was the least.

4. Conclusion

We successfully prepared liposomes loaded with PFOB and INS by thin-film dispersion phacoemulsification. The prepared liposomes had a typical shell-core structure, and insulin and perfluorooctane were encapsulated in the phospholipid membrane layer. Through DiI staining of liposomes, it can be seen that liposomes can enter and diffusely distribute in the eye, so as to achieve better therapeutic effect, establish an SD rat model of corneal inflammation, and observe the therapeutic effect of INS/PFOB@LIP on alkali burns. In order to more accurately analyze the therapeutic effect of insulin liposomes on corneal inflammation, the convolutional neural network statistical analysis method was used. The percentage of corneal epithelial defect area, percentage of new blood vessel area, NS, PFOB@LIP, INS, and INS/PFOB@LIP in each group were included in the convolutional neural network statistical analysis model, and the convolutional neural network statistical analysis algorithm showed the treatment effect. After 7 d of treatment, compared with NS, PFOB@LIP, and INS, the corneal epithelium of INS/PFOB@LIP was almost completely healed with minimal neovascularization. After 14 days of treatment, compared with NS, PFOB@LIP, and INS, the growth rate of new blood vessels in INS/PFOB@LIP group was the slowest. It can be seen that INS/PFOB@LIP has good anti-inflammatory and antiangiogenesis effects.

At different time intervals, the occurrence and development of corneal epithelial defect, corneal edema, corneal opacity, and corneal neovascularization of NS, PFOB@LIP, INS, and INS/PFOB@LIP were observed under slit lamp. After 7 days of treatment, compared with NS, PFOB@LIP and INS, the corneal epithelium of INS/PFOB@LIP was almost completely healed with minimal neovascularization. After 14 days of treatment, compared with NS, PFOB@LIP, and INS, the growth rate of new blood vessels in the INS/PFOB@LIP group was the slowest. The results of HE staining, Sirius red, and immunohistochemistry showed that after topical application of INS/PFOB@LIP eye drops, the inflammatory factors in the INS/PFOB@LIP group were lower than those in the NS, PFOB@LIP, and INS groups, and the collagen fibers were densely arranged and there were followed by the other three groups. Using CD31 to label neovascular endothelial cells, it could be observed that the number of neovascularization in INS/PFOB@LIP was

significantly lower than that in NS, PFOB@LIP, and INS groups.

In conclusion, insulin liposome INS/PFOB@LIP can significantly improve the cell activity of corneal cells for 20 h. The convolutional neural network algorithm was used to carry out the relevant statistical analysis, and the experimental data was analyzed by the algorithm. It can be seen that after the insulin liposome INS/PFOB@LIP treatment, the symptoms of keratitis were alleviated, the INS/PFOB@LIP corneal epithelium was basically completely healed, and the new angiogenesis was the least, and compared with other experimental control groups, the INS/PFOB@LIP group had lower inflammatory factors than NS, PFOB@LIP, and INS groups, and the differences were statistically significant ($P < 0.01$), demonstrating that insulin liposomes can be used as potential therapeutics for corneal inflammation.

Data Availability

The dataset can be accessed upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yukun Liao and Huiting Jiang contributed equally to this work.

Acknowledgments

The authors thank the Department of Ultrasound at the Second Affiliated Hospital of Chongqing Medical University and Chongqing Mingda Ophthalmology for the support of experiments. This work was supported by the Natural Science Foundation of Chongqing City (CQYC20200303112).

References

[1] K. Baba, Y. Tanaka, A. Kubota et al., "A method for enhancing the ocular penetration of eye drops using nanoparticles of hydrolyzable dye," *Journal of Controlled Release*, vol. 153, no. 3, pp. 278–287, 2011.

[2] J. Cheng, H. Zhai, J. Wang, H. Duan, and Q. Zhou, "Long-term outcome of allogeneic cultivated limbal epithelial transplantation for symblepharon caused by severe ocular burns," *BMC Ophthalmology*, vol. 17, no. 1, 2017.

[3] F. Bo, "Advances in the treatment of corneal alkaline burns," *Electronic Journal of Clinical Medicine Literature*, vol. 6, no. 40, pp. 196-197, 2019.

[4] G. Ryman and B. E. Ryman, "Liposomes as carriers of enzymes or drugs: a new approach to the treatment of storage diseases," *Biochemical Journal*, vol. 124, no. 5, p. 58P, 1971.

[5] D. D. Lasic, "Novel applications of liposomes," *Trends in Biotechnology*, vol. 16, no. 7, pp. 307–321, 1998.

[6] R. J. D'Amato, M. S. Loughnan, E. Flynn, and J. Folkman, "Thalidomide is an inhibitor of angiogenesis," *Proceedings of the National Academy of Sciences*, vol. 91, no. 9, pp. 4082–4085, 1994.

[7] W. Huang, Y. Chen, and L. Xiang, "An experimental study on the changes of inflammatory cytokines in aqueous humor of corneal alkali burns," *Journal of Ophthalmic Ophthalmology in Ophthalmic Trauma. Attached to Ophthalmic Surgery*, vol. 10, pp. 654-655, 2003.

[8] S. K. Park, M. Ha, E. J. Kim et al., "Hyaluronic acid hydrogels crosslinked via blue light-induced thiol-ene reaction for the treatment of rat corneal alkali burn," *Regenerative Therapy*, vol. 20, pp. 51–60, 2022.

[9] Y. Zhong, "Ocular Wnt/β -catenin pathway inhibitor XAV939-loaded liposomes for treating alkali-burned corneal wound neovascularization," *Frontiers in Bioengineering and Biotechnology*, vol. 9, 2021.

[10] A. Al-Qaissi, M. Papageorgiou, H. Deshmukh et al., "Effects of acute insulin-induced hypoglycaemia on endothelial microparticles in adults with and without type 2 diabetes," *Diabetes, Obesity and Metabolism*, vol. 21, no. 3, pp. 533–540, 2019.

[11] A. M. Jingi, A. T. Tankeu, N. A. Ateba, and J. J. Noubiap, "Mechanism of worsening diabetic retinopathy with rapid lowering of blood glucose: the synergistic hypothesis," *BMC Endocrine Disorders*, vol. 17, no. 1, p. 63, 2017.

[12] A. L. Wang, E. Weinlander, B. M. Metcalf et al., "Use of topical insulin to treat refractory neurotrophic corneal ulcers," *Cornea*, vol. 36, no. 11, pp. 1426–1428, 2017.

[13] S. Liu, V. Romano, B. Steger, S. B. Kaye, K. J. Hamill, and C. E. Willoughby, "Gene-based angiogenic applications for corneal neovascularization," *Survey of Ophthalmology*, vol. 63, no. 2, pp. 193–213, 2018.

[14] V. Galvis, C. A. Niño, A. Tello, J. Grice, and M. Gómez, "Topical insulin in neurotrophic keratopathy after resection of acoustic neuroma," *Archivos de la Sociedad Espanola de Oftalmologia*, vol. 94, no. 2, pp. 100–104, 2019.

[15] H. Qin, W. Zhang, Yi Luo, Y. Wei, S. Sun, and J. Chen, "The performance of composite nanoparticles based on Fe3O4@SiO2/PLGA/PFOB in magnetic resonance imaging and photoacoustic imaging," *Science of Advanced Materials*, vol. 11, no. 12, pp. 1816–1824, 2019.

[16] X. Cai, Y. Ju, and J. Dai, "Research progress of tumor microenvironment-modulating nanomaterials," *Acta Pharmaceutics*, vol. 56, no. 04, pp. 1016–1024, 2021.

[17] Y. Zhao, M. Pu, Y. Wang, L. Yu, X. Song, and Z. He, "Application of nanotechnology in acute kidney injury: from diagnosis to therapeutic implications," *Journal of Controlled Release*, vol. 336, pp. 233–251, 2021.

[18] M. M. Merlo-Mas, T. M. Tomsen-Melero, C. Corchero et al., "Application of Quality by Design to the robust preparation of a liposomal GLA formulation by DELOS-susp method," *The Journal of Supercritical Fluids*, vol. 173, Article ID 105204, 2021.

[19] F. Nienhaus, D. Colley, A. Jahn et al., "Phagocytosis of a PFOB-nanoemulsion for 19F magnetic resonance imaging: first results in monocytes of patients with stable coronary artery disease and ST-elevation myocardial infarction," *Molecules*, vol. 24, no. 11, 2018.

[20] Z. Zhu, T. Hu, Z. Wang et al., "Anti-inflammatory and organ protective effect of insulin in scalded MODS rats without controlling hyperglycemia," *The American Journal of Emergency Medicine*, vol. 36, no. 2, pp. 202–207, 2018.

[21] Y. W. Chang, L. C. Hung, Y. C. Chen et al., "Insulin reduces inflammation by regulating the activation of the NLRP3 inflammasome," *Frontiers in Immunology*, vol. 11, Article ID 587229, 2020.

[22] J. Liu, X. Wang, Z. Peng et al., "The effects of insulin preadministration in mice exposed to ethanol: alleviating hepatic

oxidative injury through anti-oxidative, anti-apoptotic activities and deteriorating hepatic steatosis through SRBEP-1c activation," *International Journal of Biological Sciences*, vol. 11, no. 5, pp. 569–586, 2015.

[23] J. Wang and J. Xu, "Effects of topical insulin on wound healing: a review of animal and human evidences," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 13, pp. 719–727, 2020.

[24] M. M. Momin and S. D. Afreen, "Nanoformulations and highlights of clinical studies for ocular drug delivery systems: an overview," *Critical Reviews in Therapeutic Drug Carrier Systems*, vol. 38, no. 4, pp. 79–107, 2021.

[25] W. He, J. Zhou, and X. Guo, "Preparation and in vitro drug release of insulin nanoliposomes coated with N-trimethyl chitosan for ophthalmic use," *Journal of Guangdong University of Pharmacy*, vol. 29, no. 1, pp. 1–5, 2013.

[26] A. Bakunowicz-Łazarczyk and B. Urban, "Assessment of therapeutic options for reducing alkali burn-induced corneal neovascularization and inflammation," *Advances in Medical Sciences*, vol. 61, no. 1, pp. 101–112, 2016.

[27] J. Dai, J. Shen, Y. Chai, and H Chen, "IL-1β impaired diabetic wound healing by regulating MMP-2 and MMP-9 through the p38 pathway," *Mediators of Inflammation*, vol. 2021, pp. 1–10, 2021.

[28] C. Yan, N. Gao, H. Sun et al., "Targeting imbalance between IL-1β and IL-1 receptor antagonist ameliorates delayed epithelium wound healing in diabetic mouse corneas," *American Journal Of Pathology*, vol. 186, no. 6, pp. 1466–1480, 2016.

[29] M. L. C. Bastion and K. P. Ling, "Topical insulin for healing of diabetic epithelial defects: a retrospective review of corneal debridement during vitreoretinal surgery in Malaysian patients [J]," *Medical Journal of Malaysia*, vol. 68, no. 3, pp. 208–216, 2013.

[30] L. J. Shanley, C. D. McCaig, J. V. Forrester, and M Zhao, "Insulin, not leptin, promotes in vitro cell migration to heal monolayer wounds in human corneal epithelium," *Investigative Opthalmology & Visual Science*, vol. 45, no. 4, p. 1088, 2004.

[31] S. Yang, Y. Zhang, Z. Zhang et al., "Insulin promotes corneal nerve repair and wound healing in type 1 diabetic mice by enhancing wnt/β-catenin signaling," *American Journal Of Pathology*, vol. 190, no. 11, pp. 2237–2250, 2020.

[32] C. Xiang, "New progress of ocular surface drug delivery system," *Chinese Journal of Experimental Ophthalmology*, vol. 38, no. 12, pp. 1091–1096, 2020.

[33] M. A. Hossain, A. Adithan, M. J. Alam et al., "IGF-1 facilitates cartilage reconstruction by regulating PI3K/AKT, MAPK, and NF-kB signaling in rabbit osteoarthritis," *Journal of Inflammation Research*, vol. 14, pp. 3555–3568, 2021.

[34] S. C. Cheng, W. C. Huang, J. H. Pang, Y. H. Wu, and C. Y. Cheng, "Quercetin inhibits the production of IL-1β-induced inflammatory cytokines and chemokines in ARPE-19 cells via the MAPK and NF-κB signaling pathways," *International Journal of Molecular Sciences*, vol. 20, no. 12, p. 2957, 2019.

[35] S. Qiu, *Correlation of Type 2 Innate Lymphocytes on the Expression of Fungal Keratitis and Changes in Inflammatory Cells in Mice*, Zhengzhou University, Zhengzhou, China, 2018.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3692984, 9 pages
https://doi.org/10.1155/2022/3692984

Research Article

Performance Optimization of Surface Plasmon Resonance Imaging Sensor Network Based on the Multi-Objective Optimization Algorithm

Zhiyou Wang,[1,2] **Maojin Wang**,[1,2] **Ying Chen**,[1,2] **and Fangrong Hu**[2,3]

[1]*School of Electronic Communication and Electrical Engineering, Changsha University, Kaifu District, Changsha, China*
[2]*Hunan Engineering Technology Research Center of Optoelectronic Health Detection, Changsha, China*
[3]*Guangxi Key Laboratory of Automatic Detecting Technology and Instrument, Guilin University of Electronic Technology, Guilin 541004, China*

Correspondence should be addressed to Zhiyou Wang; zywang@ccsu.edu.cn

Received 15 June 2022; Revised 28 June 2022; Accepted 7 July 2022; Published 31 July 2022

Academic Editor: Wenming Cao

In this work, we report performance optimization of a wireless sensor network (WSN) based on the plain silver surface plasmon resonance imaging (SPRi) sensor. At the sensor node level, we established the refractive index-thickness models for both gold and silver in the sensor and calculated the depth-width ratio (DWR) and penetration depth (PD) values of the sensor of different gold and silver thicknesses by the Jones transfer matrix and Kriging interpolation. We optimized the DWR and PD simultaneously by using the multi-objective optimization genetic algorithm (MOGA). In the following performance optimization of WSN, we simultaneously optimized the transmission success rate and information dimension with the number of nodes and transmission failure rate of the sensor node as variables by the same algorithm. By calculating the information dimension and the transmission success rate of each Pareto optimal solution, we obtained the number of nodes and transmission failure probability of the node available for practical deployment of WSN. The above results indicate that the Pareto optimal solution set obtained from MOGA can help to provide the best solution for the optimization of some certain performance parameters and also assist us in making the trade-off decision in the structure design and network deployment if optimal values of all the performance parameters can be obtained simultaneously.

1. Introduction

Surface plasmon resonance imaging (SPRi) detection is one type of label-free and contact-free assays which do not alter or impair the function of biomolecules. Because of its ability to acquire the endpoint signal and binding intensity information simultaneously during drug screening and protein sequencing, the SPRi technique has been widely applied in cancer immunotherapy, drug development, food safety, and other important fields [1–4]. To achieve further applications in early diagnosis and point of care testing (POCT) cases, the performances of the SPRi sensors need to be promoted in the following ways. First, the sensors should be configured as miniaturized devices of high detection sensitivity in mobile and noisy scenes. Among varieties of metals in the SPRi configuration, gold and silver have been the main choices in the reported works [5–7]. Compared with gold film-based SPRi sensors, silver film-based sensors demonstrated around 1-time sensitivity enhancement. Although the chemical properties of silver are more active than gold, the reported plain silver SPRi sensors prepared by using a gold island film as the adhesion enhancing layer between the silver film and the chromium adhesion layer show the same level of stability as the gold thin film-based sensors. The working life of plain silver SPRi sensors can reach more than 8 hours in chemical reagents such as strong acids, strong bases, and highly permeable salt solutions [8, 9]. To further improve the sensitivity performance of the plain silver SPRi sensor, two factors need to be considered comprehensively. The first is the depth-width ratio (DWR),

which is defined as the ratio between the depth and full width at half maximum (FWHM) of the SPR curve. According to Homola et al. work, the SPRi sensors with higher DWR values feature higher sensitivity and lower resolution in detection [10]. The second is evanescent wave penetration depth (PD) into the measured biomolecules, which is defined as the distance from the silver surface where the electromagnetic field intensity drops to $1/e^2$ of the value at the surface. In our previous work, a larger PD value was reported to be helpful for the detection of weak binding signals between biomolecules at the immobilized cell surface [8]. Second, the SPRi sensors should be deployed as portable devices or wireless sensor networks (WSN) to cover a large area in buildings [11–15]. Among the portable SPRi sensing devices, Guner et al. reported a smartphone-based SPRi biodetection platform. The device was composed of a green LED, a fiber optical cable, a bimetallic SPR sensor chip, a smartphone with a camera collecting the reflection from the sensor chip, and other beam shaping elements [15]. Though a resolution of 4.12×10^{-5} refractive index unit (RIU) was reported in this work, further application of the platform was hindered by the camera design updates in the smartphone and arbitrary bending of the fiber. Different from the reported portable devices, deployment of WSN is not dependent on the performance of the sensor, and does not even require the sensor itself to be mobile. Besides, WSN allows information to be collected and presented over a larger area, and the network transmission performance can be optimized by adjusting configuration parameters [11–14]. To deploy WSN in practice, we need to consider two factors. The first is the complexity of WSN, which is highly related to resource allocation and energy consumption [16]. In our previous work, the fractal dimension was used to evaluate the structural complexity of WSN [17]. The second is the transmission success rate, which is the key parameter in the evaluation of the network performance and user experience [18].

It is noteworthy that two factors need to be considered comprehensively in the performance optimization of both the plain silver SPRi sensor design and WSN deployment, which is multi-objective discrete optimization problems [19–21]. To solve such problems, various multi-objective optimization (MOO) algorithms have been reported, which can be divided into two main ways [22, 23]. The first type of the MOO algorithm can be categorized as the scalarization method. In this type of method, we can either combine all the individual objective functions into a composite function by weighting each objective or move all but one objective to a constraint set by setting a constraint value for each objective. In both of the choices, we can pursue a single solution considering all the objectives as the output of the optimization [24]. The second type of the MOO algorithm is known as the Pareto method. The output of the method is a set of single solutions, i.e., the Pareto optimal solution set. When we move from one solution to another in the set, a gain of one objective can be obtained at the cost of loss in another object [25]. Based on our previous works, we found that all the aforementioned performance factors of both the plain silver SPRi sensor and WSN cannot be expressed in

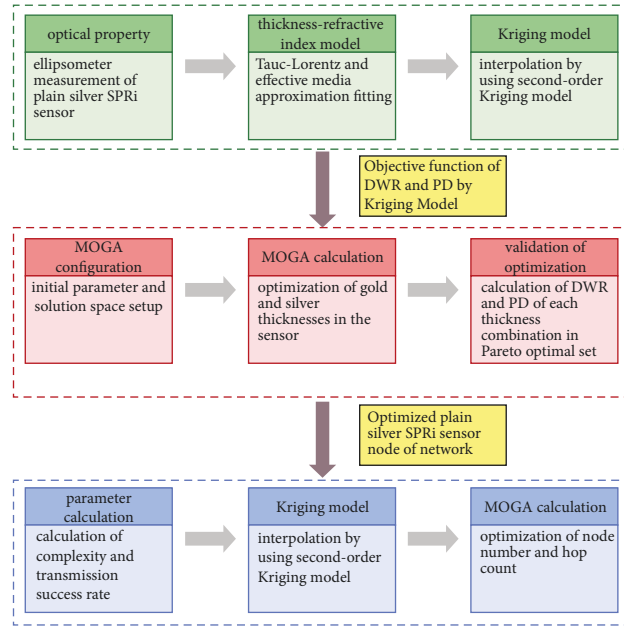


Figure 1: Flowchart of the plain silver SPRi sensor wireless network design and implementation.

individual objective functions, thus the MOO algorithms in the category of the Pareto method are the preferred choices in the performance optimization of both the sensor and the network. Due to its ability to find several solutions of the Pareto optimal solution set in one run with less dependency on the shape or continuity of the Pareto front in the whole process, the multi-objective genetic algorithms (MOGAs) have been widely used in mechanical engineering, industrial processes, and other intelligent decision-making fields [26, 27]. Genetic algorithms are search algorithms inspired by natural evolutionary theory. By mimicking the process of natural selection, crossover, and mutation, genetic algorithms can provide high-quality solutions in non-convex, discontinuous, and multimodal spaces to a variety of problems involving search, optimization, and learning [22]. MOGAs have been reported in the optimization of plasmonic structures and WSN performances [28,29]. In the optimization of plasmonic structures, Kim and Jung maximized the figure of merit (FOM) of the plasmonic curve and peak transmission power simultaneously by optimizing the slit width and height [28]. In the performance optimization of WSN, key factors, such as coverage, lifetime, energy consumption, and reliability, can be optimized simultaneously according to Fei et al. survey [29].

To enhance the DWR and PD value of the plain silver SPRi sensor, and improve the complexity and transmission success rate of the sensor-based WSN, we reported optimization of the silver SPRi sensor node-based WSN in this work. At the node level, we measured ellipsometer data of the silver SPRi sensor and calculated thickness-refractive index relations of different metals in the sensor. We then established two objective functions of DWR and PD of the sensor at different combinations of gold and silver thicknesses by using Kriging models. Afterward, we carried out

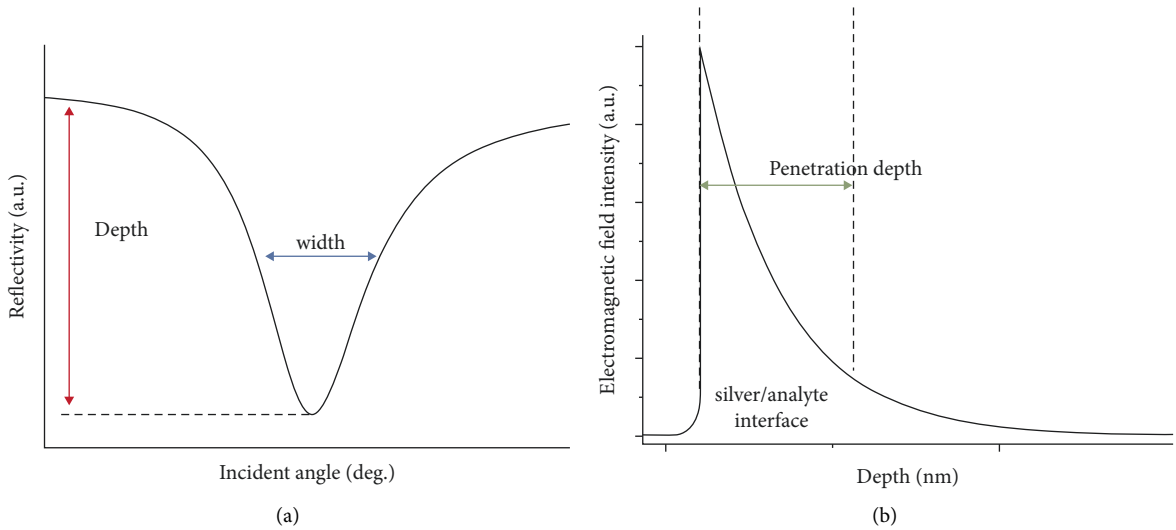


Figure 2: Principal illustration of (a) DWR and (b) PD in the plain silver SPRi sensor.

the MOGA using the Pareto optimality. At the network level, we calculated the information dimension as complexity and the transmission success rate of WSN at different combinations of the number of nodes and transmission failure probability of nodes. We also established two objective functions of the two factors by Kriging models and carried out a similar MOGA algorithm. The results show that the performance parameters of both the plain silver SPRi sensor and WSN can be optimized by the Pareto optimal solution set obtained by MOGA. This optimization can not only help us to pursue the optimal choice of some certain performances of the sensor and the network but also assist us in making optimal decisions in practical sensor production and WSN deployment.

2. Methods

The work of the plain silver SPRi sensor-based WSN design and implementation includes three parts as in Figure 1: establishing objective functions of DWR and PD with the gold and silver thicknesses as variables by using the Kriging model, optimizing the film thicknesses by the MOGA algorithm, and optimizing the number of nodes and the hop count of WSN after setting up the objective functions of complexity and the transmission success rate. The calculation program was developed on the Python 3.7 platform on a personal computer with an Intel Core i7 2.8 GHz CPU and 16 GB RAM running the Windows 10 Ultimate operating system.

2.1. Optimization of Gold and Silver Thicknesses by MOGA. We used a spectroscopic ellipsometer (Sentech SE850DUV, SENTECH Instruments GmbH, Germany) to measure the intensity and phase of reflection from the sensor surface, and fitted the refractive indices of gold and silver films under different thickness conditions using the effective media approximation (EMA) compensated the Tauc-Lorentz (TL) model in commercial software WINELII according to the different elemental ratios of the gold and silver film layers [30]. Afterward, polynomial fitting was used to establish the correspondence between the refractive indices and thicknesses of gold and silver films at a wavelength of 660 nm, respectively. We then used the second-order universal Kriging (UK) interpolation method to establish the relationship between DWR, PD (Figure 2), and the gold and silver thicknesses based on the Jones transmission matrix, respectively, and used MOGA to find the Pareto optimal solution for gold and silver thicknesses with the optimization of the two indexes as the objective [28].

2.2. Improvement of Complexity and the Transmission Success Rate of WSN. Based on our previous work, we calculated both the information dimension and the transmission success rate of plain silver SPRi sensor-based WSN by the data transmission probability fractal model (PFM) with the number of nodes and transmission failure probability of nodes as variables [11,31]. Then, we established the objective functions of the two parameters by applying the second-order UK interpolation to the calculated data. After the MOGA optimization, we obtained a Pareto optimal solution set of the number of nodes and the hop count and validated the improvement of complexity and the transmission success rate.

3. Results and Discussion

3.1. Establishment of the Refractive Index-Thickness Models of Gold and Silver Films. According to our previous work, we fabricated the plain silver SPRi sensor of gold thickness ranging from 0.5 nm to 2.5 nm, and silver thickness ranging from 40 nm to 60 nm for spatial elemental distribution under transmission electron microscopy (TEM) [30]. A thicker silver film provides more atoms as the diffusion source and a longer fabrication process allowing the silver atoms to cross the gold-silver boundary, resulting in a higher elemental ratio in the gold layer. Thus, as shown in

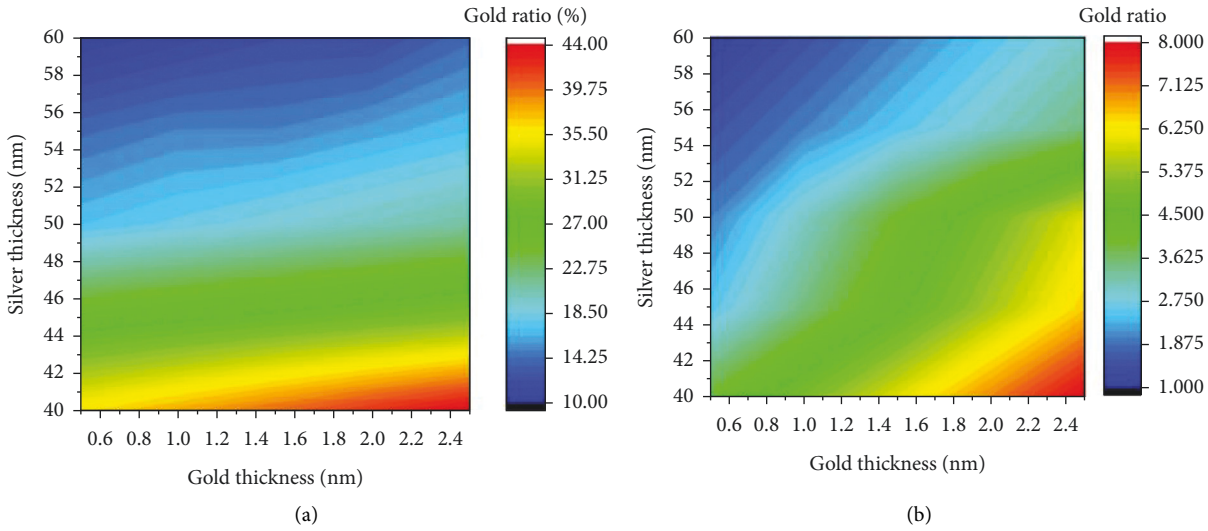


FIGURE 3: Gold element ratio in the (a) gold film side and (b) the silver film side of the gold-silver boundary under different gold and silver thicknesses.

Figure 3(a), when the gold film becomes thinner and the silver film becomes thicker, the gold element ratio in the gold film side of the gold-silver boundary decreases dramatically. In the contrast, the gold film is much thinner compared with the silver film, which cannot provide sufficient gold atoms diffusing across the gold-silver boundary, resulting in an elemental ratio of less than 10% in the silver film as shown in Figure 3(b).

In the ellipsometer measurement, intensity and phase data of reflection from the silver surface of the sensor were collected for data fitting. Considering the thin gold film was covered by the thick silver film, we did not fit the refractive index of the gold film, but calculated the index value by the EMA method instead [32]. During fitting the refractive index of the silver film, the EMA compensated TL model can be described as in equations (1) and (2). In equation (1), ε_i and ε_r represent the real part and the imaginary part of the dielectric constant, respectively. E, A, E_0, Eg, C, P, and $\varepsilon_r(\infty)$ represent the energy of a photon, amplitude, peak transition energy, band gap energy, spreading correction parameter, Cauchy body value, and fitting correction parameter of the silver element, respectively. In equation (2), G represents the depolarization parameter of the silver element, and ε is the dielectric constant calculated by equation (1). f_f and f_v are the silver and gold element ratios in the silver film observed under TEM, and n_{EMA} is the effective refractive index of the silver film for the fitting work [33–35].

$$\varepsilon_i(E) = \frac{AE_0C\left(E - E_g\right)^2}{\left(E^2 - E_0^2\right)^2 + C^2E^2} \cdot \frac{1}{E}, \varepsilon_r(E) = \varepsilon_r(\infty) + \frac{2}{\pi}P\int \frac{\xi\varepsilon_i(\xi)}{\xi^2 - E^2}d\xi\,, \tag{1}$$

$$f_f \frac{\varepsilon - n_{\mathrm{EMA}}^2}{n_{\mathrm{EMA}}^2 + G\left(\varepsilon - n_{\mathrm{EMA}}^2\right)} + f_v \frac{1 - n_{\mathrm{EMA}}^2}{n_{\mathrm{EMA}}^2 + G\left(1 - n_{\mathrm{EMA}}^2\right)} = 0. \tag{2}$$

Owing to the diffusion of gold atoms in silver film, the fitted imaginary part of the refractive index is smaller than the theoretical value of the silver bulk material in Figure 4(b). The small imaginary part of the refractive index can help increase the reflectivity of the metal surface, thus the measured reflection is higher than the fitted data, while the measured phase data can be fitted well in Figure 5. Besides, the real part of the refractive index is smaller than that of the theoretical value at the wavelength ranges from 500 nm to 750 nm in Figure 4(a), while the measured reflection is around 0.04 higher than the fitted data in the same range (Figure 5(a)). This phenomenon indicates that high reflectivity can affect the fitting accuracy of the real part of the refractive index [35].

We repeated the same ellipsometer measurement and curve-fitting work to the plain silver SPRi sensor of other gold and silver thicknesses, and calculated the relation between the refractive index at a wavelength of 600 nm and thickness of both the gold and silver films by polynomial fitting as in equation (3) [36,37]. For gold and silver, f_f (%) in equation (3) can be expressed as functions of gold thickness d_g (nm) and silver thickness d_s (nm) in equations (4) and (5), respectively. Because of the mutual diffusion between the gold and silver films during fabrication, the refractive index expressions of the metals contain the thicknesses of both films. It is noteworthy that the gold refractive index has a much stronger dependence on its thickness than silver, thus it is necessary to decrease the step of the gold thickness in the following second-order UK interpolation.

$$n_{\mathrm{EMA}}^2 = \varepsilon \cdot \left[(1-\varepsilon)^2 f_f^2 - 5(1-\varepsilon)f_f + \varepsilon^2 - 3\varepsilon + 3\right], \tag{3}$$

$$f_f = 17.64d_g - 0.052d_s - 3.94, \tag{4}$$

$$f_f = -2.00d_g + 0.16d_s + 91.2. \tag{5}$$

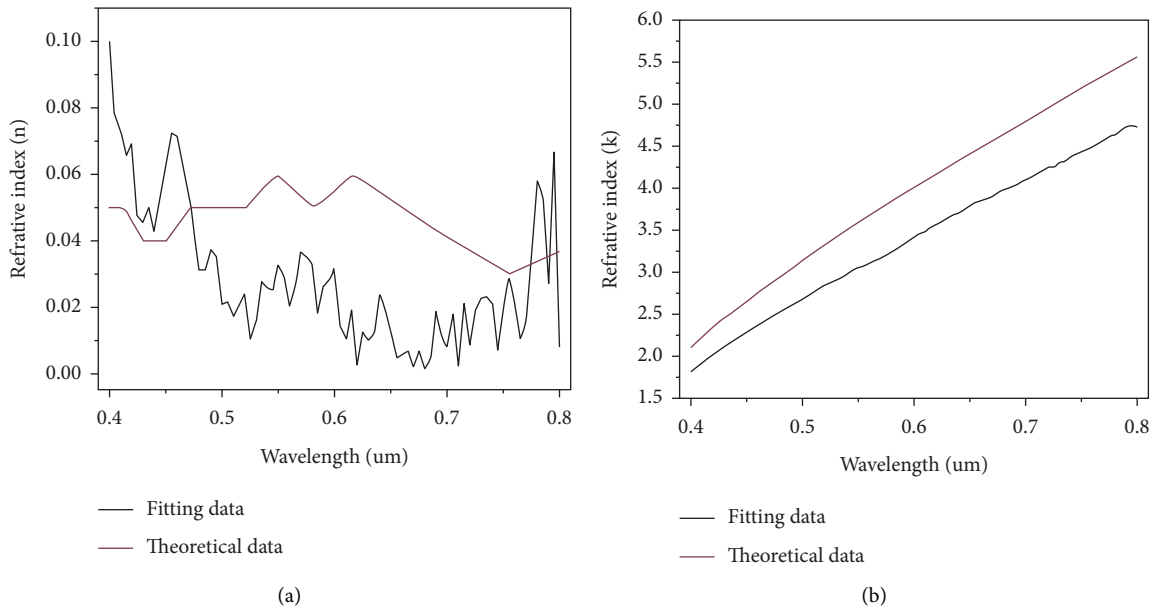


Figure 4: Fitting and theoretical real part (a) and imaginary part (b) of the silver refractive index in the plain silver SPRi sensor.

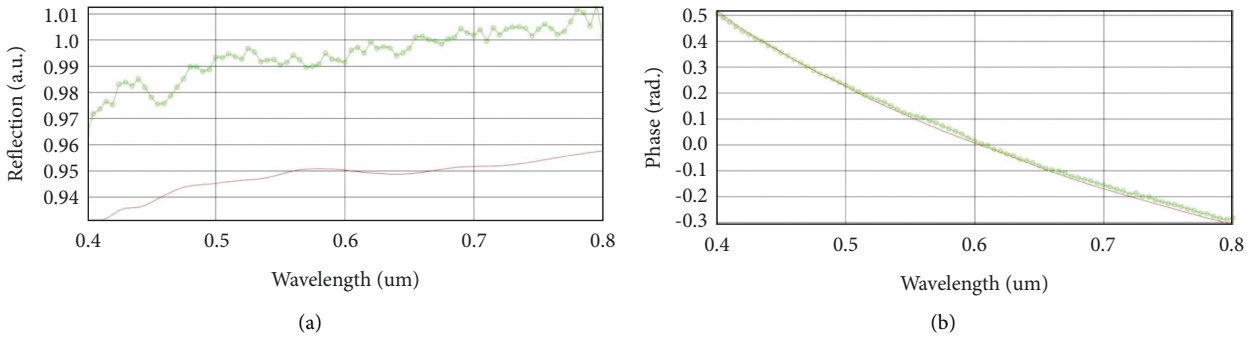


Figure 5: Experimental (green) and fitting (red) results of (a) refraction and (b) phase data under ellipsometer measurement (gold: 2 nm, silver: 50 nm).

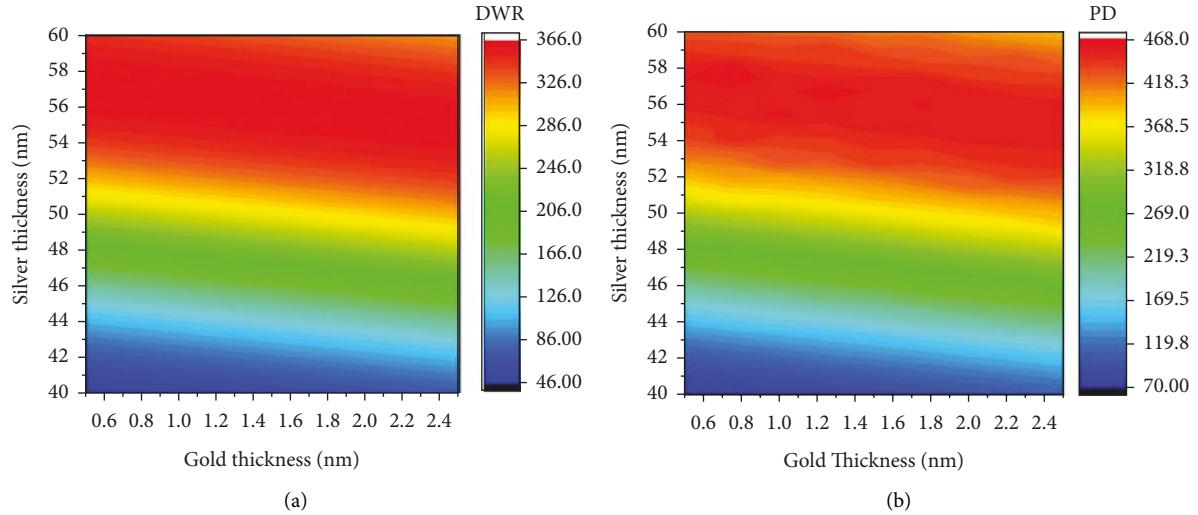


Figure 6: (a) DWR and (b) PD of the plain silver SPRi sensor after second-order UK interpolation.

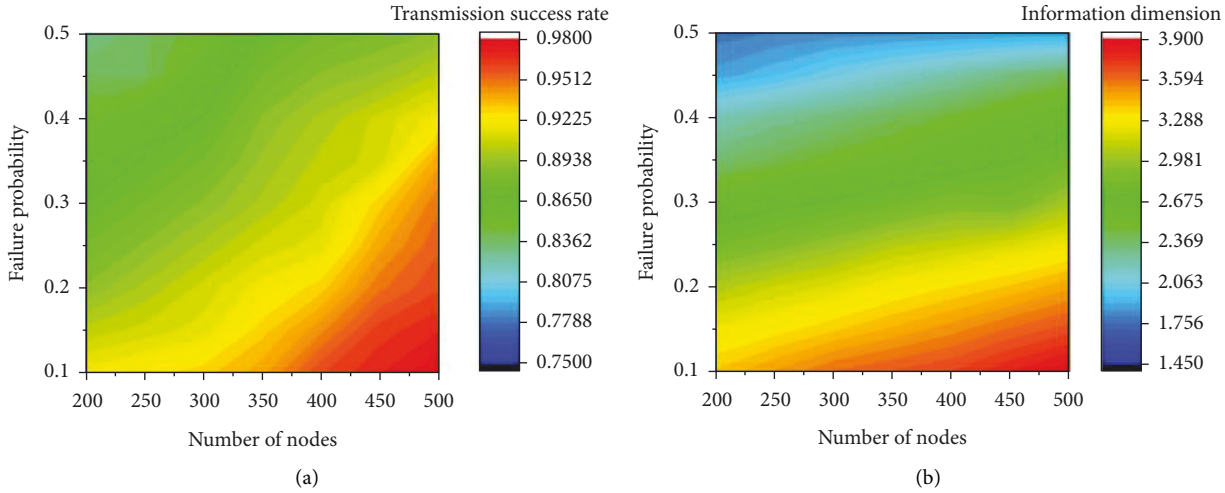


Figure 7: (a) The transmission success rate and (b) the information dimension of WSN after second-order UK interpolation.

3.2. MOGA of Silver and Gold Thicknesses by Using Kriging Interpolation. Based on the aforementioned thickness-refractive index model of gold and silver films, we calculated the DWR and PD values at 660 nm according to the Jones transmission matrix at different gold and silver thicknesses [35]. Since DWR and PD can only be calculated according to Figure 2, the relation between the two parameters and these thicknesses are unknown functions and can be approximated by the second-order UK model according to reference [28]. In the approximation, the thickness steps of gold and silver films are 0.25 nm and 0.5 nm, respectively.

In Figure 6, we can see that the silver thickness is the key factor determining the incremental trend of DWR and PD rather than the gold thickness. Besides, the variations of DWR and PD do not appear to match each other well with the gold and silver thicknesses. Thus, both objective functions cannot be optimized by a single objective function, but need to be optimized for the Pareto optimal solution set for DWR and PD. Here, our problem can be expressed as equation (6), where x is a two-dimensional vector with components as the gold and silver thicknesses, and the x_i^L and x_i^U are the lower and upper bound of each variable, respectively. The multi-objective optimization problems are solved with a population size of 320, a crossover rate of 0.8, and a generation of 100 according to reference [28].

$$\begin{aligned} &\text{Maximize DWR}(x) \text{ and PD}(x), \\ &\text{subject to } x_i^L \le x_i \le x_i^U, \quad i = 1, 2. \end{aligned} \tag{6}$$

The optimized thicknesses of metallic films in the Pareto optimal solution set and corresponding objective function values listed in Table 1 show a similar trend to the data in Figure 6. For the different value of the gold thickness, the optimal silver thickness is around 58 nm, and the calculated DWR and PD varies less than 1%. This trend tells that the silver thickness is the most determinant of the performance of the silver SPRi sensor. We also notice that DWR and PD cannot be optimized at a single combination of gold and silver thicknesses. The combination no. 1 is desirable if we pursue a high DWR, while the combination no. 4 can be adopted if we need a large PD in the detection of bio-interactions occurring on the cell surface. Combination no. 9 may be chosen for practical manufacturing because a thick gold film may provide strong adhesion to the silver film [30].

3.3. Improvement of Complexity and the Transmission Success Rate of WSN. In previous work, we reported that increasing the number of nodes and decreasing the communication transmission failure probability of nodes can help to improve the transmission success rate [11]. However, the adjustment of the two factors has an impact on the information dimension of the network. The information dimension will become large if we increase the number of nodes or decrease the transmission failure probability of nodes [38]. Thus, we need to establish objective functions of the information dimension and the transmission success rate with the number of nodes and transmission failure probability of nodes as variables [39, 40]. Since the information dimension can only be calculated by a polynomial approximation, the aforementioned objective functions are unknown functions, which need to be approximated by the second-order UK model. In approximation with the hop count as 6, the range of the number of nodes is from 200 to 500 with 30 as a step, while the range of transmission failure probability is from 0.1 to 0.5 with 0.05 as a step [41, 42].

In Figure 7, the variations in the transmission success rate and the information dimension show an opposite trend with the variations in the number of nodes and transmission failure probability of nodes. Thus, both of the objective functions of the two performance parameters need to be optimized for the Pareto optimal solution set. Different from equation (6), our problem can be expressed as equation (7), where x is a two-dimensional vector with components as the number of nodes and the transmission failure probability of nodes. The multi-objective optimization problems are solved with a population size of 121, a crossover rate of 0.8, and a generation of 40.

Table 1: Thicknesses of films and objective function values in the Pareto optimal set.

No.	Gold thickness (nm)	Silver thickness (nm)	DWR (%/°)	PD (nm)
1	0.5	58	365.503	459.314
2	0.75	58	364.880	465.498
3	1	58	363.922	455.690
4	1.25	57.5	364.661	466.833
5	1.5	57	363.475	458.975
6	1.75	58.5	364.500	460.193
7	2	58.5	363.944	458.526
8	2.25	57.5	363.239	459.284
9	2.5	58	363.591	459.689

Table 2: WSN performance factors and objective function values in pareto optimal set.

No.	Number of nodes	Transmission failure probability of nodes (%)	Information dimension	Transmission success rate (%)
1	260	0.45	1.95	84.5
2	320	0.35	2.53	88.1
3	350	0.3	2.81	90.8
4	380	0.2	3.12	93.0
5	440	0.15	3.60	96.3

$$\begin{aligned}&\text{Maximize Success\_rate}\,(x) \text{ and Minimize Dimension}\,(x),\\ &\text{subject to } x_i^L \le x_i \le x_i^U, \quad i = 1, 2.\end{aligned} \tag{7}$$

The optimized number of nodes and transmission failure probability of nodes in the Pareto optimal solution set and corresponding objective function values listed in Table 2 show a similar trend to the data in Figure 7. The information dimension and the transmission success rate cannot be optimized at a single combination of the two performance factors. The combination no. 1 is desirable if we pursue a small complexity of WSN, while the combination no. 5 can be adopted if a high transmission success rate is in urgent need. Combinations no. 2 and 3 may be chosen for deployment because both complexity and reliability are sufficient for practical usage.

4. Conclusion

In this work, we reported performance optimization of plain silver SPRi sensor-based WSN by using MOGA. In the optimization of the plain silver SPRi sensor, we chose the maximum values of DWR and PD as the optimization objective, and the thicknesses of gold and silver as variables to obtain the Pareto optimal solution set. In the optimization process, we established the refractive index-thickness models for both gold and silver, and calculated the DWR and PD values of the sensor with different gold and silver thicknesses by the Jones transfer matrix and Kriging interpolation. We found that DWR and PD cannot reach the maximum value simultaneously, but a trade-off can be made to get ambient results for both performance parameters. In the performance optimization of WSN, we chose the maximum value of the transmission success rate and the minimum value of the information dimension as the optimization objective, and the number of nodes and the transmission failure probability of nodes as variables. We also found that the aforementioned maximum and minimum values cannot be obtained simultaneously. By calculating the information dimension and the transmission success rate of each Pareto optimal solution, we obtained the number of nodes and transmission failure probability of nodes available for practical deployment of WSN. Besides, we can also provide the best solution for the optimization of some certain aforementioned performance parameters. The above results indicate that MOGA can be applied in simultaneous optimization of multiple parameters of sensors and networks in practical design, production, and deployment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the National Natural Science Foundation of China (Grant no. 61905027), the Natural Science Foundation of Hunan Province (Grant no. 2019JJ50685), and Key Research and Development Program of Hunan Province (Grant no. 2020SK2111).

References

[1] R. Liu, Q. Wang, Q. Li, X. Yang, K. Wang, and W. Nie, "Surface plasmon resonance biosensor for sensitive detection of microRNA and cancer cell using multiple signal amplification strategy," *Biosensors and Bioelectronics*, vol. 87, pp. 433–438, 2017.

[2] J. Homola, "Surface plasmon resonance sensors for detection of chemical and biological species," *Chemistry Review*, vol. 108, no. 2, pp. 462–493, 2008.

[3] J. E. Bradner, O. M. McPherson, and A. N. Koehler, "A method for the covalent capture and screening of diverse small molecules in a microarray format," *Nature Protocols*, vol. 1, no. 5, pp. 2344–2352, 2006.

[4] L. He, Q. Pagneux, I. Larroulet et al., "Label-free femtomolar cancer biomarker detection in human serum using graphene-coated surface plasmon resonance chips," *Biosensors and Bioelectronics*, vol. 89, pp. 606–611, 2017.

[5] L. Xia, S. Yin, H. Gao, Q. Deng, and C. Du, "Sensitivity enhancement for surface plasmon resonance imaging biosensor by utilizing gold-silver bimetallic film configuration," *Plasmonics*, vol. 6, no. 2, pp. 245–250, 2011.

[6] S. Chen and C. Lin, "High-performance bimetallic film surface plasmon resonance sensor based on film thickness optimization," *Optik*, vol. 127, no. 19, pp. 7514–7519, 2016.

[7] C. T. Li, T. J. Yen, and H. F. Chen, "A generalized model of maximizing the sensitivity in intensity-interrogation surface plasmon resonance biosensors," *Optics Express*, vol. 17, no. 23, p. 20771, 2009.

[8] Z. Cheng, Z. Wang, D. E. Gillespie et al., "Plain silver surface plasmon resonance for microarray application," *Analytical Chemistry*, vol. 87, no. 3, pp. 1466–1469, 2015.

[9] Z. Wang, Z. Cheng, V. Singh et al., "Stable and sensitive silver surface plasmon resonance imaging sensor using trilayered metallic structures," *Analytical Chemistry*, vol. 86, no. 3, pp. 1430–1436, 2014.

[10] J. Homola and M. Piliarik, *Springer Series on Chemical Sensors and Biosensors*, O. S. Wolfbeis, Ed., vol. 1, p. 45, Springer-Verlag, Berlin Heidelberg, 2006.

[11] Z. Wang and M. Wang, "Wireless network of silver film lysozyme aptasensor based on fractal measurement," *Security and Communication Networks*, vol. 2021, pp. 1–9, 2021.

[12] C. Zhan, Y. Zeng, and R. Zhang, "Energy-efficient data collection in UAV enabled wireless sensor network," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 328–331, 2018.

[13] B. Wang, X. Gu, L. Ma, and S. Yan, "Temperature error correction based on BP neural network in meteorological wireless sensor network," *International Journal of Sensor Networks*, vol. 23, no. 4, p. 265, 2017.

[14] S. Jain, M. Nehra, R. Kumar et al., "Internet of medical things (IoMT)-integrated biosensors for point-of-care testing of infectious diseases," *Biosensors and Bioelectronics*, vol. 179, Article ID 113074, 2021.

[15] H. Guner, E. Ozgur, G. Kokturk et al., "A smartphone based surface plasmon resonance imaging (SPRi) platform for on-site biodetection," *Sensors and Actuators B: Chemical*, vol. 239, pp. 571–577, 2017.

[16] M. Ndiaye, G. P. Hancke, and A. M. Abu-Mahfouz, "Software defined networking for improved wireless sensor network management: a survey," *Sensors*, vol. 17, no. 5, p. 1031, 2017.

[17] Z. Wang, Y. Zhou, A. S. Bajenaid, and Y. Chen, "Design of wireless sensor network using statistical fractal measurements," *Fractals*, vol. 30, no. 02, Article ID 2240092, 2022.

[18] S. Nayak, N. R. Blumenfeld, T. Laksanasopin, and S. K. Sia, "Point-of-Care diagnostics: recent developments in a connected age," *Analytical Chemistry*, vol. 89, no. 1, pp. 102–123, 2017.

[19] G. A. López-Muñoz, M. C. Estévez, M. Vázquez-García et al., "Gold/silver/gold trilayer films on nanostructured polycarbonate substrates for direct and label-free nanoplasmonic biosensing," *Journal of Biophotonics*, vol. 11, no. 8, Article ID e201800043, 2018.

[20] D. C. Christodouleas, B. Kaur, and P. Chorti, "From point-of-care testing to eHealth diagnostic devices (eDiagnostics)," *ACS Central Science*, vol. 4, no. 12, pp. 1600–1616, 2018.

[21] Q. Liu, X. Li, H. Liu, and Z. Guo, "Multi-objective metaheuristics for discrete optimization problems: a review of the state-of-the-art," *Applied Soft Computing*, vol. 93, Article ID 106382, 2020.

[22] N. Gunantara and Q. Ai, "A review of multi-objective optimization: methods and its applications," *Cogent Engineering*, vol. 5, no. 1, Article ID 1502242, 2018.

[23] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: a tutorial," *Reliability Engineering & System Safety*, vol. 91, no. 9, pp. 992–1007, 2006.

[24] J. Jia, G. W. Fischer, and J. S. Dyer, "Attribute weighting methods and decision quality in the presence of response error: a simulation study," *Journal of Behavioral Decision Making*, vol. 11, no. 2, pp. 85–105, 1998.

[25] K. Deb and R. Datta, "Hybrid evolutionary multi-objective optimization and analysis of machining operations," *Engineering Optimization*, vol. 44, no. 6, pp. 685–706, 2012.

[26] J. L. J. Pereira, G. A. Oliver, M. B. Francisco, S. S. Cunha, and G. F. Gomes, "A review of multi-objective optimization: methods and algorithms in mechanical engineering problems," *Archives of Computational Methods in Engineering*, vol. 29, no. 4, pp. 2285–2308, 2022.

[27] S. C. Cerda-Flores, A. A. Rojas-Punzo, and F. Napoles-Rivera, "Applications of multi-objective optimization to industrial processes: a literature review," *Processes*, vol. 10, no. 1, p. 133, 2022.

[28] K. Y. Kim and J. Jung, "Multiobjective optimization for a plasmonic nanoslit array sensor using Kriging models," *Applied Optics*, vol. 56, no. 21, p. 5838, 2017.

[29] Z. Fei, B. Li, S. Yang, C. Xing, H. Chen, and L. Hanzo, "A survey of multi-objective optimization in wireless sensor networks: metrics, algorithms, and open problems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 550–586, 2017.

[30] Z. Wang, Y. Wang, M. Wang, and Q. Zheng, "Analysis of adhesion strength between silver film and substrate in plain silver surface plasmon resonance imaging sensor," *Sensors and Materials*, vol. 34, no. 5, p. 1629, 2022.

[31] C. Fan, T. Dong, Z. Wen, and Q. Wu, "A low energy algorithm of wireless sensor networks based on fractal dimension," in *Proceedings of the 2016 IEEE International Conference on Network Infrastructure and Digital Content ICNIDC*, pp. 455–459, Beijing, China, September 2016.

[32] R. W. Christy, "Optical constants of the noble metals," *Physical Review B*, vol. 6, no. 12, pp. 4370–4379, 1972.

[33] C. Zhong, K. E. Ballantine, C. Kervick et al., "Mapping of surface plasmon dispersion in thin Ag–Au layered composite films," *Journal of the Optical Society of America B*, vol. 33, no. 4, p. 566, 2016.

[34] Y. Jiang, S. Pillai, and M. A. Green, "Grain boundary effects on the optical constants and Drude relaxation times of silver films," *Journal of Applied Physics*, vol. 120, no. 23, Article ID 233109, 2016.

[35] O. Peña-Rodríguez, M. Caro, A. Rivera, J. Olivares, J. M. Perlado, and A. Caro, "Optical properties of Au-Ag alloys: an ellipsometric study," *Optical Materials Express*, vol. 4, no. 2, p. 403, 2014.

[36] J. Dong and R. Lu, "Characterization of weakly absorbing thin films by multiple linear regression analysis of absolute unwrapped phase in angle-resolved spectral reflectometry," *Optics Express*, vol. 26, no. 9, p. 12291, 2018.

[37] Z. Wang, J. J. Diamond, R. Hou et al., "An accurate and precise polynomial model of angular interrogation surface plasmon resonance data," *Sensors and Actuators B: Chemical*, vol. 151, no. 2, pp. 309–319, 2011.

[38] J. Wang, Y. Gao, W. Liu, A. Sangaiah, and H. J. Kim, "An improved routing schema with special clustering using PSO algorithm for heterogeneous wireless sensor network," *Sensors*, vol. 19, no. 3, p. 671, 2019.

[39] C. Song, S. Havlin, and H. A. Makse, "Self-similarity of complex networks," *Nature*, vol. 433, no. 7024, pp. 392–395, 2005.

[40] Y. W. Kuo, C. L. Li, J. H. Jhang, and S. Lin, "Design of a wireless sensor network-based IoT platform for wide area and heterogeneous applications," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5187–5197, 2018.

[41] D. Wei, B. Wei, Y. Hu, H. Zhang, and Y. Deng, "A new information dimension of complex networks," *Physics Letters A*, vol. 378, no. 16-17, pp. 1091–1094, 2014.

[42] A. Khasawneh, M. S. B. A. Latiff, O. Kaiwartya, and H. Chizari, "A reliable energy-efficient pressure-based routing protocol for underwater wireless sensor network," *Wireless Networks*, vol. 24, no. 6, pp. 2061–2075, 2018.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3740634, 7 pages
https://doi.org/10.1155/2022/3740634

Research Article

Research on Blended Teaching of Flipped Classroom Based on CNN-SSA-Bi-LSTM Deep Learning Model Computer Media

Feng Lu

Chang Chun Radio and TV University, Changchun 130051, Jilin, China

Correspondence should be addressed to Feng Lu; 1551140006@xzyz.edu.cn

Received 6 June 2022; Accepted 1 July 2022; Published 30 July 2022

Academic Editor: Wenming Cao

Aiming at the problem that the influencing factors of computer media flipped classroom hybrid teaching lead to the teaching effect not reaching the expected, this study proposes an ultra-short-term prediction model based on CNN-SSA-Bi-LSTM. CNN-SSA-Bi-LSTM is used to flip the study of mixed teaching in the classroom. This method constructs a one-dimensional convolutional neural network, performs data fusion and feature transformation on multiple key variables, and then constructs a two-way long-term short-term memory network prediction model, which realizes a 45-minute classroom for ultra-short-term prediction of the future. In addition, data optimization is performed through SSA to improve the predictive effect of the CNN-Bi-LSTM model. Experimental results show that compared with the traditional machine learning method, the proposed prediction model can effectively improve the prediction accuracy of the ultra-short-term classroom effect, and the relative variance of the continuous model is increased by 16.22%. High prediction accuracy and low error prove that CNN-SSA-Bi-LSTM deep learning model has strong application prospects in the research of flipped classroom hybrid teaching.

1. Introduction

With the rapid development of modern information technology and the arrival of the era of big data and the Internet+ [1], "University Computer Fundamentals" has become a public basic course for all majors in colleges and universities, and it is an essential part of the knowledge system for professional personnel training [2]. Its teaching goal is to comprehensively cultivate students' information literacy, cultivate students' good computational thinking ability, and enable students to apply computers to learn related courses in various disciplines, so as to improve the computer application level of college students and use computers to solve problems [3]. Blended teaching is a combination of traditional teaching methods and information technology methods [4]. The core idea of flipped classroom design is to flip the two links of knowledge acquisition and knowledge internalization in traditional teaching activities, completely changing the relationship, status, and role of teachers and students in traditional classrooms, thereby improving the quality of teaching. Therefore, it is extremely necessary to conduct research on the flipped classroom and blended teaching of university computer-based teaching.

As early as 1868, Harris, an American educator at that time, proposed a plan of "activity grouping system." After teaching new content, teachers should provide knowledge guidance to low-level students when they are required to practice relevant skills. After the low-level students have basically mastered the knowledge system, they can explain and teach the next content [5]. From the 1930s to the end of World War II, countries were mired in economic crises and world wars and had no time to take care of the development of education. At the same time, people began to criticize and reflect on the previously hot Winnetka and Dalton systems. The hierarchical teaching experiment completely ignores the role of the teacher and the classroom and moves towards flipped teaching [6]. From the 1960s to 1979, stratified teaching in the world fell into a period of silence, most notably in the UK [7]. Since the 1960s, people have begun to pursue educational equity. Therefore, during this period, the UK underwent educational reforms, shifting from the previous structural stratification to mixed competency

stratification, beginning to divide students of different abilities into one class. Middle-class students are stratified in the same class, pay attention to the differentiated development of students in the same class, and begin to move towards modern flipped teaching [8]. In 2013, the British government made changes to the standards of student achievement and progress, from selective to comprehensive, giving schools more freedom to implement blended teaching. Today, blended teaching presents more diverse characteristics, and sufficient educational resources are used to support students' all-round development of personality under blended teaching [9]. However, some researchers believe that it is difficult to achieve the expected educational effect with a single educational model, so they began to look for an optimization method, which can optimize the teaching mode according to the actual situation, so as to achieve the expected teaching effect [10]. Restricted Boltzmann Machine (RBM) is a generative network that can not only learn effective features from data but also reconstruct the original data using the learned features. In the education classroom, the results show that RBM can better predict the effect based on the original data [11]. Some studies [12] extracted the eigenvalues of classroom teaching by singular value decomposition method, and then used Bi-LSTM to effectively predict the effect of classroom teaching. Some researchers [13] proposed a Bi-LSTM-based cyclic neural network structure for the prediction of English reading classroom effects and designed a Bi-LSTM network that received random length samples for training to further improve the generalization ability of the model. Some researchers [14, 15] combined one-dimensional CNN and Bi-LSTM for aero-engine RUL prediction and used Bi-LSTM to predict the performance degradation amount in time series to obtain the future trend of performance degradation, a more accurate RUL prediction result is obtained by setting the performance degradation threshold.

2. Correlation Model Theory

The structure of the hybrid teaching research model of the flipped classroom of computer media proposed is shown in Figure 1. It consists mainly of two parts: CNN and Bi-LSTM, in which the one-dimensional CNN network consists of a convolutional layer and a pooling layer [16–18]. The Sparrow Search Algorithm (SSA) calculates fitness values based on input values and updates data to increase prediction accuracy. The time series of various factors affect the teaching effect. Various factors are extracted and transformed through a one-dimensional CNN and then input to the Bi-LSTM layer; finally, the fully connected layer outputs the predicted effect of formic acid and media in the flipped classroom blended teaching. In addition, SSA optimization [19] is also added to the CNN data analysis.

2.1. Data Preprocessing and Input Variable Ordering. This study uses multisample analysis, variable selection, and data analysis from actual data, constructs a hybrid teaching prediction model for flipped classrooms, and obtains the

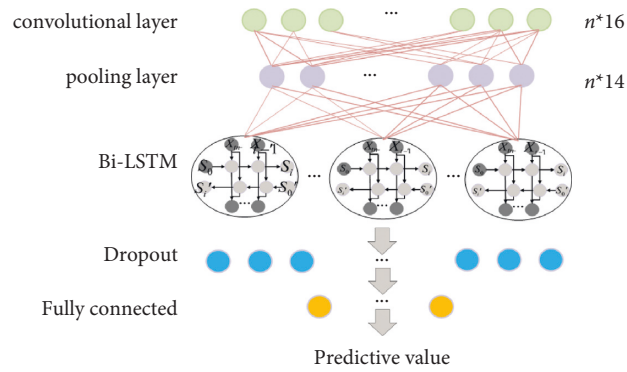


Figure 1: The composition of the CNN-Bi-LSTM model.

parameters of the prediction model through training. In order to facilitate the comparison of the experimental results and ensure the reliability of the experimental data, the data in the front and rear directions are uniformly selected for prediction in this paper.

The Bayesian information criterion (BIC) [20] proposed by Schawrtz based on Bayesian theory can effectively compensate for the subjectivity of order determination of autocorrelation graphs and partial autocorrelation graphs. optimal model. Therefore, this paper chooses the BIC discriminant criterion to analyze the prediction target. The information content of the BIC criterion is defined as follows:

$$B_{\mathrm{BIC}} = k\ln(N) - 2\ln(L). \tag{1}$$

In the formula, k is the number of model parameters; N is the number of samples; L is the likelihood function. For the selection of the optimal model, the maximization of the likelihood function should be considered.

According to the class requirements, the ultra-short-term prediction is a 45-minute prediction point, so the measurements of each class are collected, the average value of each 45 minutes is used as a data sample, and the BIC is used to determine the order of the models. Since the data has an obvious daily cycle, the first-order difference is performed on the 45-min time series data to make the series more stationary. Finally, four factors, such as teacher, class time, student grouping, and flipping situation, are used, and the step size is set to 6, which constitutes the input of the model, and the output of the model is the predicted value of the effect in the next 45 minutes.

2.2. Input Variable Fusion and Feature Extraction. As a multilayer neural network, a convolutional neural network is usually used to deal with image-related problems [21]. Continuously reduce the dimension of the image with a huge amount of data, finally, the recognition or classification of image objects is completed through several fully connected layers [22]. With the help of the two-dimensional convolutional neural network's characteristics of image feature extraction and data dimensionality reduction, this paper 1D-CNN to perform data fusion and feature on the four

different types of data sequences selected above. First, the time series of the four influencing factors are used as the input of 1D-CNN, and the four classroom variables are fused through the convolution layer.

2.3. Bidirectional Long Short-Term Memory Network. The long short-term memory network [23] is an improvement to the RNN, which not only has the connection between the hidden layer nodes in the RNN structure but also increases the filtering of the past state. The more influential historical state is used as input. Compared with ordinary RNN models, it is more suitable for time series prediction problems, so it is more suitable for the prediction of mixed teaching in the flipped classroom [24]. As a deep learning neural network model, each neuron of LSTM is a memory cell with 3 gates, namely, the forgetting gate f_t, input gate i_t, and output gate o_t. First, the forgetting gate f_t is multiplied by the previous cell state $C_{(t-1)}$ to determine which information needs to be discarded in the memory cell at the previous moment, Then, the new information is obtained by multiplying the input gate i_t by the alternative content $C_{\mathrm{r}t}$ that needs to be updated, and C_t can be obtained by discarding the information and updating the information. Finally, push the state value of Ct to between −1 and 1 through $\tan_h$ in the hyperbolic function, and multiply it with o_t to obtain the output h_t at the current moment. The whole process is shown in equations (2) to (7) [25], which is

$$f_t = \sigma\left(W_f \times \left[h_{(t-1)} \;\; x_t\right] + b_f\right), \tag{2}$$

$$i_t = \sigma\left(W_i \times \left[h_{(t-1)} \;\; x_t\right] + b_i\right), \tag{3}$$

$$\widetilde{C}_t = \tan h\left(W_C \times \left[h_{(t-1)} \;\; x_t\right] + b_C\right), \tag{4}$$

$$C_t = \left(f_t \times C_{(t-1)}\right) + \left(i_t \times \widetilde{C}_t\right), \tag{5}$$

$$o_t = \sigma\left(W_o \times \left[h_{(t-1)} \;\; x_t\right] + b_o\right), \tag{6}$$

$$h_t = o_t \times \tan h\left(C_t\right). \tag{7}$$

In the formula, $h_{(t-1)}$ is the output at time $t-1$; x_t is the input at the current time t; σ is the sigmoid function; W is the model weight parameter; b is the model bias parameter. Bi-LSTM is based on LSTM, changing the unidirectional LSTM layer into a bidirectional backward layer and forward layer. Through Bi-LSTM, classroom information in two directions is comprehensively considered, thereby improving the prediction performance of the model.

2.4. Sparrow Search Algorithm (SSA). SSA [26] is a new type of intelligent optimization algorithm, and its idea originates from the foraging behavior and anticapture behavior of sparrows. The specific description of the SSA optimization process is as follows:(1) The update formula of the finder position $X_{i,j}^{(t+1)}$ is

$$X_{i,j}^{(t+1)} = \begin{cases} X_{i,j}^{t} \exp\left[\dfrac{-i}{\alpha i_{\mathrm{ter}_{\max}}}\right], & R_2 < S_T, \\ X_{i,j}^{t} + QL, & R_2 \geqslant S_T. \end{cases} \tag{8}$$

In the formula, i_{termax} is the maximum number of iterations; $x_{i,j}$ is the position of the i-th sparrow in the j-th dimension; $\alpha \in (0, 1]$ is a random number; $R2 \in [0, 1]$ is a warning value; $ST \in [0.5, 1]$ is a safety value; Q is a random number that obeys a normal distribution; L is a matrix of $1 \times d$, each element in the matrix is 1. (2) Joiner position $Xt + 1\, i, j$ update

$$X_{i,j}^{(t+1)} = \begin{cases} Q \exp\left[\dfrac{\left(X_{\mathrm{worst}} - X_{i,j}^{t}\right)}{i^2},\right] & i > \left(\dfrac{n}{2}\right), \\ X_p^{(t+1)} + \left|X_{i,j}^{t} - X_p^{(t+1)}\right| \mathbf{A}^{+}\mathbf{L}, & \text{Other.} \end{cases} \tag{9}$$

Among them, $A^{+} = A^{T}\left(AA^{T}\right)^{-1}$.

In the formula, X_p is the optimal position of the current finder; X_{worst} is the current global worst position; A is the matrix of $1 \times d$, in which each element is randomly assigned 1 or −1. (3) Assume that 10% ~ 20% in this group of sparrows is aware of the danger. Aware of the danger, the mathematical expression of its warning position $X_{i,j}^{(t+1)}$ is

$$X_{i,j}^{t+1} = \begin{cases} X_{\mathrm{best}}^{t} + \beta\left|X_{i,j}^{t} - X_{\mathrm{best}}^{t}\right|, & f_i > f_g, \\ X_{i,j}^{t} + K\left(\dfrac{\left|X_{i,j}^{t} - X_{\mathrm{worst}}^{t}\right|}{\left(f_i - f_w\right) + \varepsilon}\right), & f_i = f_g. \end{cases} \tag{10}$$

In the formula, X_{best} is the current global optimal position; β is the random number of the step size control parameter; $K \in [-1, 1]$ is a random number; f_i is the fitness value of the current sparrow individual; f_g is the current global optimal fitness value; f_w is the current global worst fitness value; ε is a constant [27].

The steps of the SSA are as follows:

Step 1: initialize the population, the proportion of predators and joiners, and the number of iterations

Step 2: calculate the fitness value, and then sort from large to small

Step 3: update the finder location

Step 4: update the joiner location

Step 5: update the vigilante position (sparrow aware of danger)

Step 6: calculate the fitness value and update the sparrow position

Step 7: if the requirements are met, output the result; otherwise, repeat steps 2–6

2.5. CNN-Bi-LSTM Prediction Model. Finally, an ultra-short-term prediction model based on CNN-Bi-LSTM is established [28], whose input X is 4 variables, the input step

size of each variable is 6, and the output y is the predicted value in the next 45 minutes, which is defined as follows:

$$X = \begin{bmatrix} x_{1,(t-6)} & \cdots & x_{10,(t-6)} \\ \vdots & \ddots & \vdots \\ x_{1,(t-1)} & \cdots & x_{10,(t-1)} \end{bmatrix}, \quad y = x_{1,t}. \tag{11}$$

In the formula, $x_{1,t}$ is the predicted value at time t; $x_{1,(t-1)}$ is the real value at the previous time (with a time interval of 45 min); X_{i-4} is the sample data of the variable.

The input X is input to the Bi-LSTM layer through a one-dimensional convolutional network, and the output of the Bi-LSTM layer is connected to the Dropout layer to prevent overfitting, and finally y is output through the fully connected layer, that is, the forecast value for the next 45 minutes.

Select classroom practice data, and obtain data samples sampled for 45 minutes according to the previous data processing method to form a training set. After many experiments, the final convolution layer number of the proposed CNN-Bi-LSTM prediction model is set to 1, the convolution kernel is set to 16, the pooling layer adopts the maximum pooling method, and the number of Bi-LSTM layers is 1, the neural network. The number of elements is 50.

3. Test and Result Analysis

3.1. Model Performance Evaluation Indicators. The evaluation indicators used in this paper include R, the relative mean error n*MBE*, the relative mean absolute error n*MAE*, and the relative variance n*RMSE* [29, 30]. The calculation formulas are shown in formulas (13)–(16), respectively.

$$R = \frac{\sum_{i=1}^{N}\left(I_{ci} - I_e\right)\left(I_{\mathrm{mi}} - I_m\right)}{\sqrt{\sum_{i=1}^{N}\left(I_{ci} - \overline{I}_e\right)^2 \sum_{i=1}^{N}\left(I_{\mathrm{mi}} - \overline{I}_m\right)^2}}, \tag{12}$$

$$nMBE = \frac{1}{N}\frac{\sum_{i=1}^{N}\left(I_{ci} - I_{mi}\right)}{I_m} \times 100\%, \tag{13}$$

$$nMAE = \frac{1}{N}\frac{\sum_{i=1}^{N}\left|I_d - I_{\mathrm{mi}}\right|}{I_m} \times 100\%, \tag{14}$$

$$nRMSE = \sqrt{1/N\sum_{i=1}^{N}\left(I_{ci} - I_{mi}\right)^2}/I_m \times 100\%. \tag{15}$$

In the formula, I_{ci} is the model predicted value; I_{mi} is the measured value; N is the overall sample size; I_c^- is the average of all model predicted values; I_m^- is the average of all measured values. In addition, for time series forecasting problems, continuous models are often used to measure the forecasting effect of the model. The continuous model uses the measured value at the previous moment as the predicted value at the next moment, and the prediction effect Fs is used to evaluate the performance of the prediction model relative to the continuous model, which is defined as follows:

Table 1: Performance statistics of different models for predicting classroom teaching.

Model	R	nMBE (%)	nMAE (%)	nRMSE (%)
SVM	0.93	0.60	19.05	26.15
BPNN	0.92	0.52	15.48	25.11
ELM	0.93	−0.15	14.75	24.90
LSTM	0.94	3.04	14.41	24.34

$$Fs = \frac{n\mathrm{RMSE}_{\mathrm{pet}} - n\mathrm{RMSE}_f}{n\mathrm{RMSE}_{\mathrm{pet}}} \times 100\%. \tag{16}$$

In the formula, n$RMSE_{\mathrm{pet}}$ is the relative variance of the persistent model and n$RMSE_{\mathrm{f}}$ is the relative variance of the predictive model.

3.2. CNN-SSA-Bi-LSTM Model Performance Evaluation. First, the first group of experiments was designed; using univariate time series as input, to predict the classroom effect in the next 45 minutes, and compare the performance of LSTM and three traditional machine learning methods in ultra-short-term prediction of classroom teaching. The experimental results are shown in Table 1. The relative average error of the LSTM model is 3.04 higher than that of the other three traditional machine learning methods, and the relative average absolute error and relative variance are 14.41 and 24.34, respectively, which are smaller than the three models of SVM, BPNN, and ELM. Therefore, the ultra-short overall performance prediction of the LSTM model for flipped classroom hybrid teaching has certain advantages.

The second group analyzes the effects of the single input variable, multiple input variables, and LSTM model structure on the prediction performance. The experimental results are shown in Table 2. The single variable model uses only a single course teaching history sequence as the model input, and the multivariate model uses the four variables shown in Figure 1 as model inputs. By comparing univariate and multivariate LSTM models, it was found that adding variables that affect the teaching of the course can improve the prediction accuracy to a certain extent, especially the systematic bias (nMBE) of the predictive model. Comparing the unidirectional LSTM and bidirectional LSTM prediction models, it can be seen that adjusting the structure of the LSTM model can further improve the prediction effect. The proposed CNN-SSA-Bi-LSTM model combines CNN for data fusion and feature extraction of variables, taking advantage of Bi-LSTM's good at processing time series, the prediction performance is better than other LSTM models in Table 2.

Figure 2 shows the prediction effect of different LSTM models. It can be seen that the proposed CNN-Bi-LSTM improves nRMSE by about 14% compared with the continuous model, while the CNN-SSA-Bi-LSTM model reaches 16.22%, slightly higher than CNN. -Bi-LSTM model, is much higher than other prediction models. Therefore, among the five LSTM-type models used in this study, the LSTM optimized by the CNN and SSA models has higher accuracy and greater applicability.

Table 2: Statistical table of performance predicted by different LSTM models.

Model	R	nMBE (%)	nMAE (%)	nRMSE (%)
Continuous model	0.93	1.01	15.50	25.51
Univariate LSTM	0.94	3.02	14.41	24.34
Multivariate LSTM	0.94	−1.86	13.40	23.27
Multivariate Bi-LSTM	0.94	−1.31	13.21	23.14
Multivariate CNN-Bi-LSTM	0.95	−0.31	12.39	21.96
Multivariate CNN-SSA-Bi-LSTM	0.96	−0.18	11.28	20.09

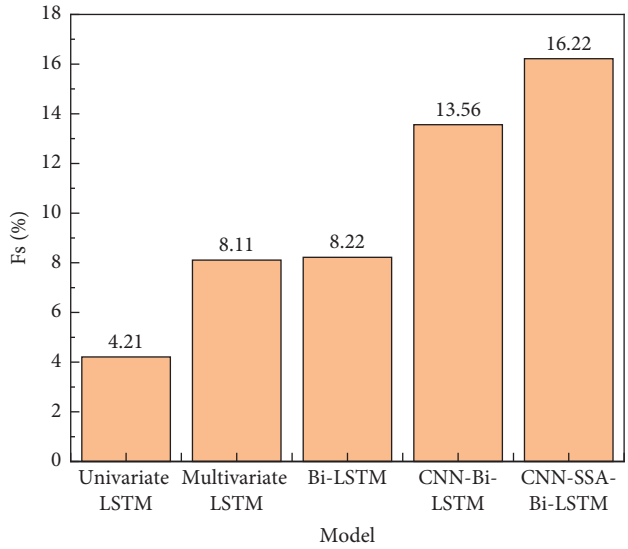


Figure 2: Forecast skills of different LSTM models.

Combining Tables 1 and 2, we can know that using multiple influencing factors as model input and constructing the CNN-SSA-Bi-LSTM model to predict classroom teaching effect, the errors nMBE, nMAE, and nRMSE are all the smallest, which are −0.18, 11.28, 20.09, respectively. The performance improves the most compared to the continuous model.

In addition, considering that the actual course teaching is carried out in separate batches, the 4 variables (see Figure 1) may not be obtained at the same time, to this end, the third set of experiments analyzes the impact of the absence of different input variables on the predictive performance of the proposed A model. Figure 3 express the experimental results. Teacher, class time, student grouping, flip situation, and student grouping indicate that the missing input variable is a, and the remaining three variables on the horizontal axis are used as model input. The same as the expression a, b, c, and d indicate that the missing variables are teacher, class time, and flipping situation, respectively. All indicate that the above four variables are used as model input.

Figure 3 shows that when the student grouping is missing in the input variable, the nRMSE of the prediction result of the CNN-SSA-Bi-LSTM model is the largest, and the prediction effect is the worst, with an nRMSE of 24.1%.

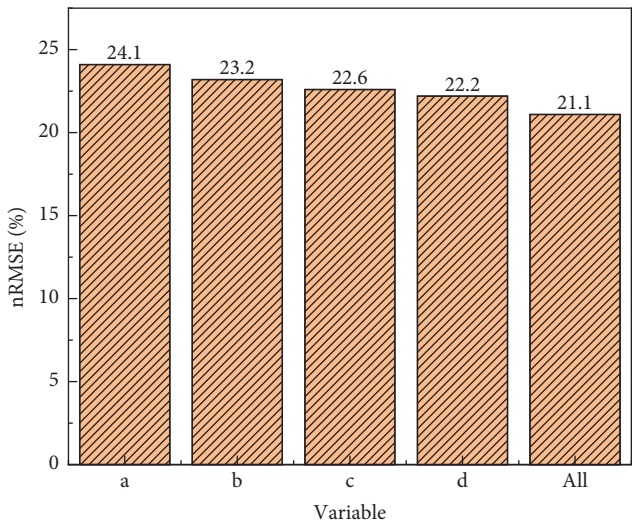


Figure 3: nRMSE of CNN-SSA-Bi-LSTM model without one input variable.

Therefore, among these 4 variables, student grouping is the most important for prediction 45 minutes ahead. When all the above four variables are used as model input, the nRMSE of the model prediction result is the smallest and the performance is optimal; if any one of the variables is missing, the nRMSE of the model will increase, that is, the prediction performance will deteriorate. This shows that increasing the input variables helps improve the prediction performance.

3.3. CNN-SSA-Bi-LSTM Model Ultra-Short-Term Prediction. According to the normal course teaching, the ultra-short-term forecast is a week of teaching time, and the predicted value of the teaching effect every 45 minutes, with a total of 5 forecast points. Figure 4 shows that CNN-SSA-Bi-LSTM uses the above 4 variables as model input, the step size is 6, and it predicts the teaching method of the experimental data for one week, takes a predicted value every 45 minutes, and calculates the nRMSE of the model at each predicted value point. Figure 3 shows that as the prediction time interval becomes larger, the prediction effect of the model gradually deteriorates. The first 45-minute prediction effect is the best, and the fifth 45-minute prediction performance is the worst, and its nRMSE is as high as 30.2%. This is because the classroom effect is better. With the increase of courses, the longer it takes to digest, the average effect increases slowly. The nRMSE calculated i is relative to the average of all target values, which is different from the maximum value of the target value in the test set divided by the partially published literature.

In summary, the poor prediction results of the LSTM model may be attributed to the fact that the model belongs to the "black box model," which is directly predicted based on a single time series, while the normal teaching effect has many influencing factors, and the influencing mechanism is more complicated. As a result, the prediction accuracy of a single time series model is limited, and even the prediction results cannot meet the actual needs. The CNN-SSA-Bi-LSTM

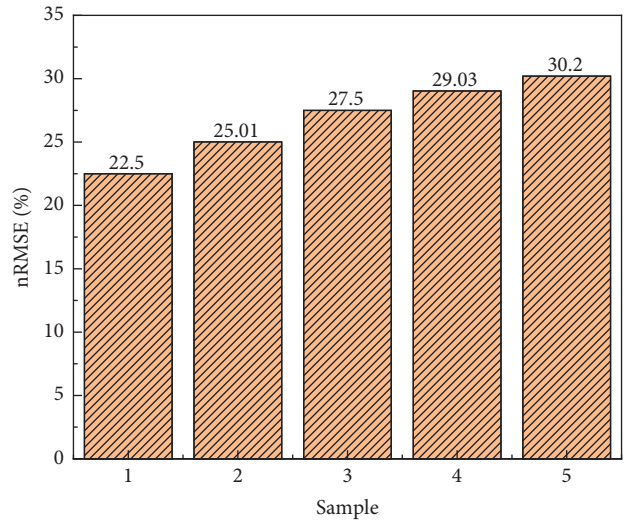


Figure 4: nRMSE of CNN-SSA-Bi-LSTM for predicting every 45 min.

model can achieve high prediction accuracy in both the training period and the verification period. The CNN model can denoise the data sequence and extract the complex and effective information hidden in the data, it can better reflect the internal mechanism that affects flipped classroom teaching, and SSA optimizes the optimal model parameters of LSTM and improves the efficiency of model parameter selection. In addition, Bi-LSTM has higher prediction accuracy than LSTM. It shows that the CNN-SSA-Bi-LSTM model is feasible for monthly runoff prediction.

4. Conclusion

This paper proposes an ultra-short-term prediction model based on CNN-SSA-Bi-LSTM for flipped classroom hybrid teaching research. The model uses one-dimensional convolutional neural network fusion to extract 4 variable data, takes advantage of the two-way long short-term memory network to process time series problems, and combines the Dropout layer and the fully connected layer to achieve ultra-short-term prediction of classroom effect 45 minutes in advance. Through multiple sets of experiments, the rationality and prediction performance advantages of the proposed CNN-SSA-Bi-LSTM model are verified. The following conclusions are obtained: (1) compared with the continuous model, the proposed model increased the relative variance by about 16.22% in the GHI prediction 45 minutes in advance, which is better than many other comparative models; (2) all four variables discussed in this paper help to improve prediction performance, with the student grouping of historical measurements having the greatest impact on the predictive performance of the proposed model. When it is missing as input, the model predicts the worst, with a predictor nRMSE of 24.1%; (3) when making ultra-short-term forecasts, the longer the forecast interval, the worse the model prediction performance. When predicting the classroom effect on the fifth day in the future, the nRMSE value of the proposed model is 30.2%.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. A. Chapelle, *Computer Application in Second Language Acquisition: Computer-Assisted Language Testing*, 2001.

[2] M. Feng, "Research on the construction of student ability evaluation system based on computer application," *Journal of Physics: Conference Series*, vol. 1915, no. 2, Article ID 022037, 2021.

[3] Q. Zhu, P. M. Kumar, and M. Alazab, "Computer application in game map path-finding based on fuzzy logic dynamic hierarchical ant colony algorithm," *International Journal of Fuzzy Systems*, pp. 1–12, 2021.

[4] J. Wang, B. J. Liu, W. He, J. K. Xue, and X. Y. Han, "Research on computer application software monitoring data processing technology based on NLP," *IOP Conference Series: Materials Science and Engineering*, vol. 1043, no. 3, Article ID 032021, 2021.

[5] M. Bemer and U. Grossniklaus, "Dynamic regulation of Polycomb group activity during plant development," *Current Opinion in Plant Biology*, vol. 15, no. 5, pp. 523–529, 2012.

[6] M. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori, "A hierarchical deep temporal model for group Activity recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June 2016.

[7] H. Oyama, N. Watanabe, Y. Ono, and T. Sakashita, "Community-based suicide prevention through group activity for the elderly successfully reduced the high suicide rate for females," *Seishin shinkeigaku zasshi = Psychiatria et neurologia Japonica*, vol. 110, no. 9, pp. 731–738, 2008.

[8] R. A. Mittermeier, "Group activity and population dynamics of the howler monkey on Barro Colorado Island," *Primates*, vol. 14, no. 1, pp. 1–19, 1973.

[9] R. A. Ellis, A. F. Steed, and A. C. Applebee, "Teacher conceptions of blended learning, blended teaching and associations with approaches to design," *Australasian Journal of Educational Technology*, vol. 22, no. 3, pp. 312–335, 2006.

[10] M. P. Gagnon, J. Gagnon, M. Desmartis, and M. Njoya, "The impact of blended teaching on knowledge, satisfaction, and self-directed learning in nursing undergraduates: a randomized, controlled trial," *Nursing Education Perspectives*, vol. 34, no. 6, pp. 377–382, 2013.

[11] I. Sutskever, G. E. Hinton, and G. W. Taylor, *The Recurrent Temporal Restricted Boltzmann Machine*, pp. 1601–1608, 2008.

[12] J. C. Cronjé, "Using Hofstede's cultural dimensions to interpret cross-cultural blended teaching and learning," *Computers & Education*, vol. 56, no. 3, pp. 596–603, 2011.

[13] M. Ma, "Multimedia Emergency Event Extraction and Modeling Based on Object Detection and Bi-LSTM Network," in *Proceedings of the 2021 IEEE International Conference on Consumer Electronics and Computer Engineering*, China, January 2021.

[14] X. Wan, J. Wan, M. Xu et al., "Enteromorpha coverage information extraction by 1D-CNN and Bi-LSTM networks considering sample balance from GOCI images," *Ieee Journal*

of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, no. 99, pp. 9306–9317, 2021.

[15] L. Du, D. Cao, J. Li, and H. Ye, "Classification and extraction of medical clinical trial screening standard texts based on Bi-LSTM and Attention mechanism," *IOP Conference Series: Earth and Environmental Science*, vol. 632, no. 5, Article ID 052088, 2021.

[16] S. Shao, T. Wang, C. Song, Y. Su, Y. Wang, and C. Yao, "Fine-grained and multi-scale motif features for cross-subject mental workload assessment using Bi-lstm," *Journal of Mechanics in Medicine and Biology*, vol. 21, no. 05, Article ID 2140020, 2021.

[17] Z. Xueqing, Z. Zhansong, and Z. Chaomo, "Bi-LSTM deep neural network reservoir classification model based on the innovative input of logging curve response sequences," *IEEE Access*, vol. 9, no. 99, Article ID 19902, 2021.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[19] K. Tzoumpas, A. Estrada, and P. Miraglio, *A Data Filling Methodology for Time Series Based on CNN and (Bi)LSTM Neural Networks*, ArXiv abs/2204.09994, 2022.

[20] F. Wikipedia and T. F. Encyclopedia, *Bayesian Information Criterion*, Ricksawatzky Com, 2014.

[21] T. Roska and L. O. Chua, "The CNN universal machine: an analogic array computer," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, no. 3, pp. 163–173, 1993.

[22] T. Roska and L. O. Chua, "The CNN universal machine," *Journal of Circuits, Systems, and Computers*, vol. 12, no. 04, pp. 377–388, 2008.

[23] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to Forget: Continual Prediction with LSTM, Neural Computation," vol. 12, no. 10, pp. 2451–2471, 2000.

[24] X. Ma and E. Hovy, *End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF*, pp. 1064–1074, 2016.

[25] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1333–1340, 2001.

[26] S. Nema, J. Y. Goulermas, G. Sparrow, and P. Helman, "A hybrid cooperative search algorithm for constrained optimization," *Structural and Multidisciplinary Optimization*, vol. 43, no. 1, pp. 107–119, 2011.

[27] C. Song, L. Yao, C. Hua, and Q. Ni, "Comprehensive water quality evaluation based on kernel extreme learning machine optimized with the sparrow search algorithm in Luoyang River Basin, China," *Environmental Earth Sciences*, vol. 80, no. 16, 521 pages, 2021.

[28] İ. Özer, S. B. Efe, and H. Özbay, "CNN/Bi-LSTM-based deep learning algorithm for classification of power quality disturbances by using spectrogram images," *International Transactions on Electrical Energy Systems*, vol. 31, no. 12, 2021.

[29] H. Chen, W. Wu, and H. B. Liu, "Assessing the relative importance of climate variables to rice yield variation using support vector machines," *Theoretical and Applied Climatology*, vol. 126, no. 1-2, pp. 105–111, 2015.

[30] H. Hannah Inbarani, P. K. Nizar Banu, and A. T. Azar, "Feature selection using swarm-based relative reduct technique for fetal heart rate," *Neural Computing & Applications*, vol. 25, no. 3-4, pp. 793–806, 2014.

Hindawi
Computational Intelligence and Neuroscience
Volume 2023, Article ID 9845814, 1 page
https://doi.org/10.1155/2023/9845814

Retraction

Retracted: Prediction and Analysis of Dynamic Changes of College Students' Ideological and Political Changes Based on Multiple Regression

Computational Intelligence and Neuroscience

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. Wang, "Prediction and Analysis of Dynamic Changes of College Students' Ideological and Political Changes Based on Multiple Regression," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5323699, 8 pages, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5323699, 8 pages
https://doi.org/10.1155/2022/5323699

Research Article

Prediction and Analysis of Dynamic Changes of College Students' Ideological and Political Changes Based on Multiple Regression

Minhong Wang

College of Marxism, Chongqing Industry Polytechnic College, Chong Qing 401120, China

Correspondence should be addressed to Minhong Wang; wangmh@cqipc.edu.cn

Received 30 May 2022; Revised 20 June 2022; Accepted 28 June 2022; Published 30 July 2022

Academic Editor: Man Fai Leung

The construction of a harmonious society requires college students to coordinate their ideological, political, and moral qualities with social development and the needs of the times. Through the investigation and analysis of the ideological, political, and moral qualities of college students, on the one hand, we can see that the ideological, political, and moral qualities of college students are generally positive and healthy. On the other hand, it also exposes the outstanding problems in the ideological and political aspects of college students and the shortcomings of the ideological and political work in colleges and universities. This paper analyzes the dynamic changes of college students' ideological and political changes and further studies the relationship between various indicators and students' ideological and moral qualities through multiple linear regression analysis.

1. Introduction

Colleges and universities are an important stage for ideological and political education for young students, and the ideological and political quality of college students is also a major criterion for evaluating the effectiveness of college education. With the continuous improvement of the socialist market economic system, the rapid development of the economy, society and science, and culture, and the integration of Chinese and foreign cultures, many new changes have taken place in the ideological values of current members of society, and some even have contradictions. As the successors and builders of the cause of socialism with Chinese characteristics, the ideological and moral conditions of contemporary college students have also appeared some new changes, new characteristics, and new conditions. In some college students, there are situations such as loss of ideals and beliefs, lack of social responsibility, a weakened spirit of hard work, and a weak concept of teamwork. At the same time, due to the continuous expansion of colleges and universities in my country in recent years and the influence of the global economic recession, new requirements have also been put forward for the ideological and political education work in colleges and universities. In this environment, it is very important and urgent to further strengthen the ideological and political education work of college students. For any society, college students are the most valuable human resources and the hope and future of a country and nation. Do a good job in the ideological and political education of college students, improve their ideological and political level, and train them to be builders and successors of the socialist cause with firm political beliefs and lofty political ideals. This will ensure the healthy and sustainable development of the socialist cause with Chinese characteristics. It has long-term and important strategic significance [1, 2]. The university stage is the last stage for a person to move from school to society, and it is also an important period for the formation of young people's values, outlook on life, and moral outlook. If there is a lack of correct theoretical guidance and education in such a period, it is bound to have a negative impact [3, 4]. General Secretary Hu Jintao pointed out "College education must adhere to people-oriented, take the all-round development of college students as the goal, and guide the majority of students to learn not only to do things but also to learn to be people."

In recent years, with the continuous development of my country's reform and opening up and the gradual improvement of the market economy system, many new situations, new characteristics, and new changes have emerged in the

ideological and political quality of contemporary college students [5, 6]. On the one hand, college students' self-improvement and self-reliance, innovation and development awareness, and patriotism awareness have been significantly strengthened; at the same time, some college students also have problems such as loss of ideals and beliefs, lack of social responsibility, the weakened spirit of hard work, and teamwork concept. Under this circumstance, the task of actively promoting the ideological and political education of college students is very urgent and important.

All in all, the ideological and political quality of college students is related to the future and destiny of the country and the nation. Whether they can cultivate qualified and excellent builders and successors who meet the requirements of the cause of socialism with Chinese characteristics, it can be said that the ideological and political work of colleges and universities plays a vital role in it. [7]. Under the background of the new era, ideological and political education in colleges and universities should closely grasp the new changes in the development of the situation, start from multiple angles and levels, and actively promote and improve the ideological and political education of college students. At present, the most commonly used prediction methods are regression analysis method and decision tree method. Multiple regression analysis is one of the scientific prediction methods, and this paper will apply this method to establish multiple regression models [8–10].

2. Related Work

Stepwise regression: the basic idea of the stepwise regression analysis method is to automatically select the most important variables from a large number of available variables and establish a prediction or interpretation model for regression analysis. The basic idea is that the independent variables are introduced one by one, and the condition of the introduction is that the partial regression sum of squares is significant after testing. At the same time, each time a new independent variable is introduced, the old independent variables should be tested one by one, and the independent variables with insignificant partial regression sum of squares should be eliminated. In this way, it has been introduced and eliminated until neither new variables are introduced nor old variables are deleted. Its essence is to establish the "optimal" multiple linear regression equation. According to the above idea, stepwise regression can be used to screen and eliminate variables that cause multicollinearity. The specific steps are as follows: first is to perform a simple regression on each explanatory variable considered with the explanatory variable and then use the explanation that contributes the most to the explanatory variable. Based on the regression equation corresponding to the variable, the remaining explanatory variables are gradually introduced. After a stepwise regression, the explanatory variables that remained in the model were both significant and did not have severe multicollinearity.

2.1. MATLAB. MATLAB is commercial mathematical software produced by MathWorks in the United States. It is used in data analysis, wireless communication, deep learning, image processing and computer vision, signal processing, quantitative finance and risk management, robotics, control systems, and other fields. MATLAB is a combination of the words matrix and laboratory, which means matrix factory (matrix laboratory). The software mainly faces the high-tech computing environment of scientific computing, visualization, and interactive programming. It integrates many powerful functions such as numerical analysis, matrix calculation, scientific data visualization, and modeling and simulation of nonlinear dynamic systems in an easy-to-use window environment for scientific research, engineering design, and many sciences that must perform the effective numerical calculation. Realm provides a comprehensive solution and is largely free from the editing mode of traditional noninteractive programming languages [11, 12] (e.g., C, Fortran). MATLAB, Mathematica, and Maple are known as the three major mathematical software. It is second to none in numerical computing among mathematical technology applications. As the same time, we use the Row matrix operations, plotting functions, and data implementing algorithms, creating user interfaces, interfacing programs in other programming languages, etc. The basic data unit of MATLAB is a matrix, and its instruction expression is very similar to the form commonly used in mathematics and engineering. Therefore, it is much simpler to use MATLAB to solve problems than to use C, FORTRAN, and other languages to complete the same thing, and MATLAB also absorbs the advantages of software such as Maple, making MATLAB as powerful mathematical software. Also we added support for C, FORTRAN, C + +, and JAVA in the new version. MATLAB is a collection containing a large number of computational algorithms. It has more than 600 mathematical operation functions used in projects, which can easily realize various calculation functions required by users. The algorithms used in the function are the latest research results in scientific research and engineering computing and have undergone various optimizations and fault tolerance processing [13]. In general, it can be used in place of low-level programming languages such as C and C++. With the same computational requirements, the programming effort using MATLAB is greatly reduced. These functions of MATLAB range from the simplest and most basic functions to complex functions such as matrices, eigenvectors, and fast Fourier transforms. The problems that functions can solve generally include matrix operations and the solution of linear equations, differential equations and partial differential equations, symbolic operations, Fourier transform and statistical analysis of data, optimization problems in engineering, sparse matrix operations, Various operations on complex numbers, trigonometric functions and other elementary mathematical operations, multidimensional array operations, and modeling dynamic simulations. In the development environment, it makes it easier for users to control multiple files and graphics windows; in terms of programming, it supports function nesting and conditional interruption; in terms of graphics, there are more

Table 1: The influence of different social settings on the ideological and political education of college students.

	The first institution	The second body	The third agency
Moral and ethical	Family (45.2%)	Books, newspapers, and magazines (21.5%)	School (17.1%)
The ideal faith	Family (28.5%)	Books, newspapers, and magazines (28.4%)	School (21.9%)
Love marriage	Family (37.4%)	Others (26.6%)	School (13.8%)
Handle affairs	Family (44.6%)	School (30.1%)	Books, newspapers, and magazines (10.0%)
Up employment	School (35.7%)	Family (19.3%)	Books, newspapers, and magazines (16.6%)
Interpersonal communication	School (35.7%)	Family (30.7%)	Books, newspapers, and magazines (7.6%)
Interests and hobbies	School (49.3%)	Books, newspapers, and magazines (19.8%)	TV (16.5%)
Entertainment consumption	TV (23.9%)	Internet (15.9%)	School (15.9%)
Leisure life	Internet (21.8%)	School (19.4%)	TV (17.9%)

Table 2: The degree of influence of different social subjects on the ideological and political education of college students.

	The first object	The second object	The third object
Moral and ethical	Parents (52.2%)	Others (13.4%)	Class teacher or counselor (11.6%)
The ideal faith	Others (27.5%)	Parents (24.9%)	Friends or classmate (22.2%)
Love marriage	Parents (33.7%)	Others (23.1%)	Friends or classmate (19.5%)
Handle affairs	Parents (47.2%)	Friends or classmate (20.9%)	Class teacher or counselor (12.6%)
Up employment	Parents (27.2%)	Class teacher or counselor (23.8%)	Friends or classmate (15.0%)
Interpersonal communication	Friends or classmate (43.2%)	Parents (34.9%)	Others (8.1%)
Interests and hobbies	Friends or classmate (55.2%)	Others (24.0%)	Lovers (8.1%)
Entertainment consumption	Friends or classmate (56.6%)	Others (14.8%)	Parents (11.5%)
Leisure life	Friends or classmate (55.3%)	Others (15.7%)	Lovers (15.2%)

powerful graphics annotation and processing functions, including, in terms of input and output, you can directly connect to Excel and HDF5[14].

3. Our Model

3.1. Analysis of the Significant Influencing Factors of College Students' Ideological and Political Changes. The personal socialization environment of college students is a direct factor affecting the ideological and political education of college students [15]. It mainly includes two aspects, one is the "social setting", that is, social institutions closely related to college students, mainly including families, schools, television, books, newspapers, the Internet, and radio; other social subjects other than college students who have a close relationship with life and study mainly include parents, lovers, relatives of the same age, friends or classmates, class teachers or counselors, and classroom teachers. During the implementation of the survey, various main factors included in the above-mentioned "social settings" and "social subjects" were measured in college students' "morality and ethics", "ideal beliefs", "love and marriage", "handling affairs", "talent employment", "interpersonal communication", "hobbies", "entertainment consumption," and "leisure life". Tables 1 and 2 show the main results of the survey [16].

3.1.1. Parents and Families Are Extremely Important Influencing Factors of College Students' Values, Morality, and Ethics. A person's life is inseparable from the family, and the family and its parents play a very important role in the growth of young people. The data in Table 1 show that among the "social setting" factors that affect the ideological and political status of college students, the influence of family on college students' "morality and ethics", "ideal beliefs", "love and marriage", and "handling of affairs" is all in the First, the proportions are 45.2%, 28.5%, 37.4%, and 44.6%, respectively [6]. It is also in second place in "successful employment" (19.3%) and "interpersonal communication" (30.7%). For college students who have left their families to study abroad, family and parents are still the main factors affecting their ideological and political status. The results in Table 2 also show that college students are most willing to listen to their parents' opinions in terms of "morality and ethics," "handling affairs," "love and marriage," and "successful employment." The proportions are 52.2%, 47.2%, and 33.7%, and 27.2%, respectively; in terms of "ideal beliefs" (24.9%) and "interpersonal communication" (34.9%), the importance of parents' opinions also ranks second. These phenomena fully demonstrate the position and weight of parents in the minds of college students. When making decisions about life events, they first consider the opinions of their parents. It is worth noting that the "ideal beliefs" of college students are affected by many factors, and the proportion of choosing "other

people" occupies the first place with 27.5% [17]. This shows that the factors affecting college students' "ideological beliefs" are diverse. Except for parents, friends, and classmates, the factors affecting college students are not the same [18].

3.1.2. School Is the Main Factor Affecting the Employment and Interpersonal Communication of College Students. From the data in Table 1, it can be seen that schools play a leading role in "interpersonal communication", "talent employment", and "hobbies", accounting for 49.3%, 35.7%, and 20.6%, respectively. It has the second-largest influence on "dealing with affairs" (30.1%) and "leisure life" (19.4%); in "ideals and beliefs" (21.9%), "morality and ethics" (17.1%) and "love and marriage" (13.8%) play the third influential role [19]. These situations show that schools have an important influence on the development of college students' ideological and political quality. From the data in Table 2, it can be found that from the perspective of the ideological and political education workforce, counselors and class teachers are in the "successful employment" (23.8%), "handling affairs" (12.6%), and "morality and ethics" (11.6%). It also has a direct and important influence on the ideological and political quality of college students.

In addition, social practice activities as important teaching content and teaching links are not only an important influencing factor of college students' ideological and political education but also an effective carrier of college students' ideological and political education. The survey shows that "participating in public welfare social practice" is listed as the first effective way for ideological and political education by college students, accounting for 65.1%. At the same time, among the activities carried out by the Communist Youth League in colleges and universities, campus cultural activities, youth volunteer activities, community activities, "Challenge Cup" activities, and "Three Going to the Countryside" activities are the most popular among college students, with 48.4% of them being selected, and 48.4% of them being selected. 47.6%, 30.0%, 29.8%, and 29.1%. These activities belong to the category of social practice and play a positive role in promoting college students to contact the society, understand society, and enhance their hands-on ability. However, the lack of social practice bases and positions makes it difficult to meet the needs of most students. The survey shows that 66.1% of college students have not engaged in on-campus work-study work, and 40.8% have not engaged in any economic activities. Among the economic activities they have engaged in, tutoring (37.2%) and part-time work (29.7%) are their main social practices. It can be seen from this that contemporary college students have few opportunities for social practice in ideological and political education, and their social practice ability has not been effectively improved, which leads them to place great expectations on social practice. At the same time, most colleges and universities believe that there are problems in the base construction, time arrangement, funding guarantee, and system establishment of social practice. Therefore, it is necessary to pay more attention to and strengthen the social practice work of college students, strive for local support, strengthen the construction of social practice bases, and guide students' social practice activities to develop in a long-term, positional, project-based, and hierarchical direction. Efforts should be made to explore the daily form of social practice activities and use weekends or other spare time to choose the community near the school as a fixed social practice base to carry out social practice activities.

Table 3: The main content of communication between college students and their roommates.

Communication content	Percentage (%)
Learning	51.3
Daily chores	32.3
Employment prospects	31.7
Current news	31.3
What's going on in class and school	29.3
Men's and women's emotion	18.7
Personal mood	17.8
Television and movies	16.1
Sports news	15.4
Evaluation of the opposite sex	12.7
Consumption	12.5
Entertainment news	10.5

3.1.3. Peer Groups Play an Important Role in College Students' Hobbies and Daily Behaviors. The peer groups of college students mainly include classmates, roommates, and members of clubs. Because college students get along with their peers day and night, have the same hobbies, similar life experiences, and have the same troubles and needs, they influence each other infrequent interactions, and especially in specific behaviors, they play a dominant role. The survey data in Table 2 show that in terms of "entertainment consumption" (56.6%), "leisure life" (55.3%), "hobbies" (55.2%), and "interpersonal communication" (43.2%), their influences are all listed No. 1. The influence of "handling affairs" (20.9%) ranks second, and the influence of "ideal and belief" (22.2%), "love and marriage" (19.5%), and "talent employment" (15.0%) and the other three influence of each aspect ranks third. It can be seen that peer group plays a very important role in the life and study of college students.

The influence of peer groups is mainly through good friends. The survey shows that middle school classmates (85.1%), college classmates (76.0%), and roommates (73.5%) are the main objects within the scope of college students' good friends. At the same time, the good friends of college students have also included some new objects, such as fellow villagers (32.7%), members of group organizations or student unions (17.6%), members of clubs (15.0%), netizens (11.7%), class teachers (9.8%), classroom teachers (9.8%), and counselors (9.1%). Once you understand the main scope of good friends for college students, you can start with these objects in a targeted manner and achieve the goal of ideological and political education through the familiarity and example of good friends. At the same time, the dormitory atmosphere is also an important factor affecting the

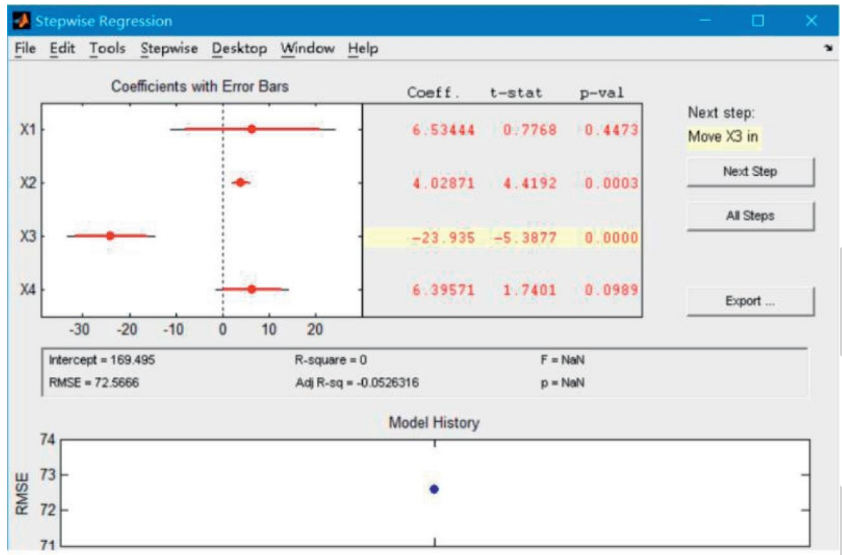


Figure 1: Stepwise regression diagram.

ideological and political education of college students. According to the survey, "study" (51.3%) is the eternal theme discussed by college students in the dormitory. "Daily trifles" (32.3%), "employment prospects" (31.7%), "current affairs news" (31.3%), and "things happening in the class and school" (29.3%) were also discussed in the "squatting meeting" main content (see Table 3). Therefore, in a certain sense, ideological and political education is not only preaching political theory but also achieves its goal through comprehensive interpersonal interaction.

3.2. Multiple Regression Analysis. Stepwise regression is a common mathematical method for selecting regressors in multiple linear regression models [5, 10]. The basic idea is, according to the size of the effect of factors $x_1, x_2, \ldots, x_m$ on y, the factors are introduced into the regression equation one by one from large to small. For the factors that have been introduced into the equation, after the introduction of new factors, they may be eliminated from the equation at any time because they have no significant effect on y. After the variable is introduced, it can also be put back in order to obtain a regression equation with some optimal properties [12] and study the regression between a dependent variable and two or more independent variables. Also known as multiple linear regression, it reflects the law that the number of a phenomenon or thing changes correspondingly with the change of the number of a variety of phenomena or things. We have used a statistical method for establishing the quantitative relations of linear or nonlinear mathematical models among multiple variables. In regression analysis, if there are two or more independent variables, it is called multiple regressions. In fact, a phenomenon is often associated with multiple factors. It is more effective and practical to predict or estimate dependent variables by the optimal combination of multiple independent variables than to predict or estimate only one independent variable. Therefore, multivariate linear regression is more practical than univariate linear regression. When all the variables introduced into the equation have reached a significant level, and at the same time, no new variables can be introduced, the stepwise regression is announced, and the optimal equation obtained at this time is between the least independent variables and the best fitting effect. The optimal equilibrium is reached [20]. It should be pointed out that the reason why it is emphasized here that the independent variables included in the regression equation should be as few as possible without affecting the fitting effect of the regression equation is that the more independent variables, the higher the requirements for data sources [21]. Many variables must be measured during prediction, which undoubtedly increases the difficulty of implementation. The effect of prediction will also decline with the inaccuracy of some variables, thereby increasing the instability of the system [12]. The following is a stepwise regression analysis of the above 16 variables that significantly affect the duration of events to determine the best combination of variables for predicting the dynamic changes in college students' ideological and political changes and establish a multiple linear regression model. In this study, the stepwise regression function [22] in the MATLAB statistical toolbox is used to select variables for the multiple regression equation, and the optimal variable combination is shown in Figure 1. The graph shows the regression coefficients and confidence intervals for each variable, and the terms shown in white (confidence intervals are indicated by solid lines) are variables that have been selected for the model; confidence intervals shown in grey are indicated by dashed lines of terms are variables that are not in the model. Building a harmonious society requires college students to coordinate their ideological, political, and moral qualities with social development and the needs of the times. Through the investigation and analysis of college students' ideological and political moral qualities, on the one hand, we can see that college students' ideological and political moral quality is positive, healthy, and upward on the whole; on the other

		Ideological change	
Column#	Parameter	Lower	Upper
2	9322	6244	124
3	1942	1615	2268
4	513	1835	8388
5	8415	4771	1206
6	5481	3452	7511
7	2636	2262	3009
8	4499	7382	826
9	5868	4592	7123
RMSE	R-square	F	P
1325	0.8356	413.5	0

Figure 2: Stepwise regression diagnosis table.

hand, it also exposes the outstanding problems in college students' ideological and political work and the deficiencies in college students' ideological and political work.

The parameters of the regression equation are shown in Figure 2. The figure shows the regression coefficient and confidence interval of each variable, the root mean square of the current model error, the amount of the corresponding variable explained by the model, the statistic F of the regression model, and the significant probability P associated with the analysis.

The 16 variables in Figures 1 and 2 are interpersonal communication, talent and employment, handling affairs, morality, participation in public welfare social practice, laws and regulations, love and marriage, talent and employment, ideals and beliefs, social responsibility, school uniforms, and friends' influence, teacher teaching, parental requirements, self-discipline, and moral restraint. As can be seen from Figure 1, after step-by-step regression analysis, variables 2 to 9, namely career and employment (x_1), handling affairs (x_2), morality and ethics (x_3), participation in public welfare social practice (x_4), laws and regulations (x_5), love marriage (x_6), successful employment (x_7), and ideal beliefs (x_8), were selected into the regression equation, while other variables were eliminated from the original regression equation. Using y to represent the event duration, the established multiple linear regression model is

$$y = 9.3x_1 + 19.4x_2 + 5.1x_3 + 8.4x_4 + 5.5x_5 + 26.4x_6 + 4.5x_7 + 58.7x_8 + 28.9. \quad (1)$$

It can be seen from the above formula that among the eight independent variables of the equation, ideal beliefs have the greatest impact on ideological and political changes. In addition, two variables, dealing with affairs and love and marriage, also have a large effect on the duration.

The constant term in the regression equation is large, which makes the equation predict ideological and political changes, even if the values of all variables are 1, the minimum predicted value is 28.9. That is to say, for small events with small changes, the prediction of the equation will generally be too large.

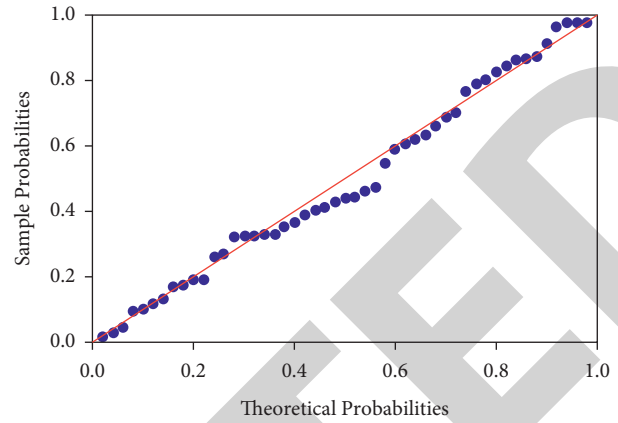


Figure 3: Comparison of predicted and actual values of multiple regression equations.

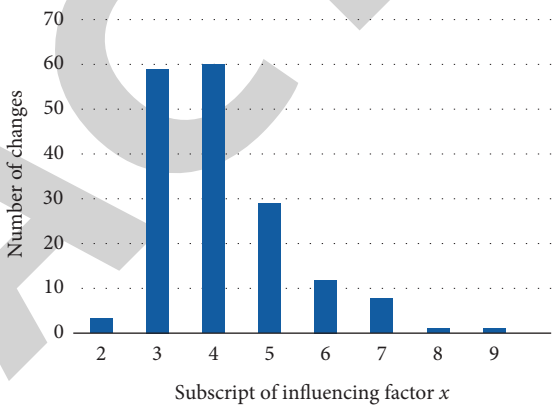


Figure 4: Statistical chart of test error.

As can be seen from Figure 2, the optimal regression equation obtained by stepwise regression is significant ($P = 0$).

4. Multiple Regression Model Effect Test

In order to examine the validity of the regression model, 170 groups of actual event data are used to test the prediction effect of the regression equation. The results are shown in Figure 3.

Figure 3 shows that the predicted value of most events is still relatively close to the actual value. The correlation analysis shows that the predicted value has a good correlation with the actual value, and the correlation coefficient is 0.8573. The results predicted by the multiple linear regression model can basically reflect the real dynamic changes. In order to illustrate the prediction effect of the regression model more clearly and intuitively, the difference between the predicted value and the actual value is counted below, and the statistical result is given in a histogram, as shown in Figure 4.

As can be seen from Figure 4, among the 170 test samples, 1 has a prediction error of 0, 59 are between 0 and 10, 59 are between 10 and 20, and 59 are between 20 and 30. There are, 21 more than 30. That is to say, about 35% of ideological and political changes have a prediction error of

less than 10, and 70% of events have a prediction error of less than 20.

In order to verify the significance of the prediction of the developed multiple regression model, this study took the significance level $\alpha = 0.0$ and performed a nonparametric one-way ANOVA. The results show that $P = 0.0058 < \alpha$, which shows that the multiple linear regression equation has significant statistical significance, and the prediction can basically obtain a satisfactory prediction effect. Finally, we use the above data to prove the feasibility of our model.

5. Conclusion

This paper conducts a step-by-step regression analysis of the ideological and political change group data of 660 students in a university and establishes the ideological and political dynamics with 8 variables of handling affairs, morality, participation in public welfare social practice, laws and regulations, love and marriage, successful employment, and ideals and beliefs. The multiple linear regression model of change prediction was used, and the prediction accuracy of the prediction model was tested with another 170 sets of data. The correlation coefficient between the predicted value and the actual value was as high as 0.8573. This shows that the multiple linear regression model proposed in this paper can more realistically reflect the duration characteristics of events and can basically obtain satisfactory prediction results. If more and more detailed data can be obtained for the establishment of multiple regression prediction models, it is believed that the prediction accuracy will be further improved. However, due to the extensiveness, randomness, and complexity of the factors that affect the duration of events, the author believes that even the best algorithms cannot predict the duration of events very accurately. It is hoped that the research in this paper can provide some guidance and reference for the development of college students' ideological and political development.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] X. Chen, R. Tu, L. Ming, and Y. Xu, "Prediction models of air outlet states of desiccant wheels using multiple regression and artificial neural network methods based on criterion numbers," *Applied Thermal Engineering*, p. 204, 2022.

[2] F. Zhang, "Psychological features and formation mechanism of ideological and political identity among Chinese college students," *Revista Argentina de Clinica Psicologica*, vol. 29, pp. 782–789, 2020.

[3] M. Zhao, Y. Ning, and Y. Yang, "A new thought analysis on the ideological and political education of college students under the socialist core values[P]," *Annual Conference of Education, Teaching and Learning*, vol. 2020, 2020.

[4] L. Wen, "The practice of "curriculum ideology and politics" concept in the ideological and political education of contemporary Chinese college students," in *Proceedings of the 2020 International Conference on Economics, Education and Social Research*, 2020.

[5] X. Liu, "Application of emotional education in college students' ideological and political work," *SOCIAL SCIENCE, EDUCATION and HUMAN SCIENCE*, 2019.

[6] Y. Liu, "A study of identity psychological analysis and formation mechanism of contemporary college Students'Ideology and politics," *Annual Conference of Education, Teaching and Learning*, vol. 2020, 2020.

[7] H. Su, "The Socialist Core Price System in the New Media Environment Integrates into the Process of College Students' Ideological and Political Education," in *Proceedings of the 2020 International Conference on Education, Management, Business and Economics*, 2020.

[8] X. Chen, L. Pan, and N. Xiu, "Solution sets of three sparse optimization problems for multivariate regression," *Journal of Global Optimization*, (prepublish), 2022.

[9] B. Li, S. Zhang, S. Wang, L. Ning, and L. Wang, "Research on the evaluation model of the ecological impact of the Saihanba based on multiple regression," *Environment, Resource and Ecology Journal*, vol. 5, no. 4, 2021.

[10] K. Ashish, K. Dwivedi Ravi, and S. Jain, "Optimization of structural health monitoring attributes under variable failure rate condition using teaching learning based optimization and multiple regression," *IOP Conference Series: Materials Science and Engineering*, vol. 1136, no. 1, 2021.

[11] Y. He, F. Kou, X. Wang et al., "Hybrid model combining multivariate regression and machine learning for the rapid prediction of interior temperatures affected by thermal diodes and solar cavities," *Building and Environment*, vol. 211, p. 108723, 2022.

[12] E. Gunawardena, "Forecasting Students' Final Exam: Results Using Multiple Regression Analysis in an Undergraduate Business Statistics Course," *Asian Journal of Economics, Business and Accounting*, vol. 21, no. 14, pp. 30–40, 2021.

[13] M. van Heel, G. Dikta, and R. Braekers, "Bootstrap based goodness-of-fit tests for binary multivariate regression models," *Journal of the Korean Surgical Society*, vol. 51, pp. 308–335, 2021, (prepublish).

[14] P. Harnpon and U. Phairat, "Environmental problem shifting analysis of pollution control units in a coal-fired powerplant based on multiple regression and LCA methodology," *Sustainability*, vol. 13, no. 9, 2021.

[15] Y. Hao, "Problems and countermeasures in the political and ideological education of college students in the new media," in *Proceedings of the 2020 Annual Conference of Education, Teaching and Learning*, Hungary, 10th December 2020.

[16] Y. Han and R. Varatharajan, "Research on influencing factors of stock returns based on multiple regression and artificial intelligence model," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 4, pp. 6457–6467, 2021.

[17] G. He, "Research on the construction and optimization of college students' ideological and political education based on mobile Internet platform," in *Proceedings of the 2020 3rd International Workshop on Advances in Social Sciences*, 2020.

[18] A. Jamiu Oyekan, B. Hadi, and A. Bera, "Investigations on the relationship among the porosity, permeability and pore throat size of transition zone samples in carbonate reservoirs using multiple regression analysis, artificial neural network and

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9274737, 7 pages
https://doi.org/10.1155/2022/9274737

Research Article

Analysis of Regional Financial Risk in Guangdong Province Based on the DCN Deep Learning Model

Yan Yuan

Guangdong Mechanical & Electrical Polytechnic, School of Economics and Trade, Guangzhou 510515, China

Correspondence should be addressed to Yan Yuan; 2009010017@gdmec.edu.cn

Received 3 June 2022; Revised 2 July 2022; Accepted 5 July 2022; Published 20 July 2022

Academic Editor: Wenming Cao

In the free flow of financial factors oriented to capital, returns will be accompanied by the concentration and diffusion of financial resources to form regional financial spatial differences, which is an objective phenomenon of regional financial practice. Localized regional financial risks may appear in the process of regional financial practice in each region. To address the abovementioned problems, we propose a model for regional financial risk analysis based on the DCN deep learning model. The main contents are as follows: elaborating the financial risk transmission mechanism involving intra- and interregional financial risks, sorting out the relationship between sectors as clues; the designing process of regional financial risk index as well as the measurement method, and the regional financial risk index for typical regions is measured and found to be at peak in 2017 with a risk index of 0.58; and the construction of an early warning model based on the value of the regional financial risk index and the expansion of the RNN network applied to the construction of the regional financial risk early warning system. Based on the construction of the RNN network application risk early warning system, the three types of risks, payment risk, loan loss risk, and market risk with the percentages of 49.62%, 26.82%, and 23.56%, respectively, are derived, and the focus is on their supervision and management in the follow-up work.

1. Introduction

The cyclical movements and adjustment fluctuations of the global economy in all major regions can be seen as a combination of financial risk phases erupting and subsiding in time [1–3]. The international financial crisis that erupted in 2008 was a major cross-regional financial risk event in recent years, which plunged the world economy into the longest, most extensive and deepest downward adjustment since the Second World War. Different from the previous adjustment, the path of economic stagnation, and natural liquidation towards recovery, after this adjustment, each major economy successively relied on the power of government policy, using extremely loose macro policy to implement a major bottoming out for the briefly stagnant economy. At present, China's reform and opening up is developing in depth, and the regional economy is also in the stage of continuous development. At the same time, regional economic operation is an indispensable part of the national economic development process, is an intermediate level between macroeconomics and microeconomics, is to ensure the stable development of the regional economy and the sustainable development of the national economy as a whole, and will play an important role of stability and coordination. Regional economic development in the sustained development of the national economy is to have an important role: first of all, there is innovation and breakthrough role. Due to the multiple economic development models developed in each region [3, 4], which are the result of proactive economic development and strengthening of regional economic management, there is an important contribution to the economic development of the region and other regions. Second, it has a stabilizing and coordinating effect. On the one hand, because of the large differences in the resources and production capacities possessed by each region, regional economic management can promote the exchange of products and strengthen the economic ties between different regions, thus ensuring the economic stability of each region

and realizing the complementary advantages of different regions. On the other hand, regional economic management can self-regulate faster and implement the mechanism of national macromanagement and regional hierarchical management, which makes the whole management system perform more flexible and effective. For the study of regional economic development, we analyzed the regional economic research theme direction through the database, as shown in Figure 1, in the regional economic research, and financial risk research is the most critical [1, 2, 5], accounting for 29.28%, followed by regional economy and economic growth and financial investment. In order to ensure stable growth of regional economy, considering financial risks in the context of actual situation and based on regional perspective, it is a necessary path for regional economic development.

The analysis of financial risks for regional economic development is gathered in the relationship between financial risks and regional characteristics which are the specific triggers of financial risks and the evolution of financial risks for regional economic development afterward. In recent years, there have been many regional financial risks in China, such as local private lending, usury, and other financial risks with insufficient control [3, 4] and lack of effective isolation, all of which can easily lead to the collapse of financial risks. Strategiesand characteristics of regional economic development in China's regional economic development strategy refers to the planning and decision-making on issues related to the overall, long-term and critical economic and social development in a certain region. To be more specific, it refers to the total planning and decision-making of the guiding ideology of regional economic development, the goals to be achieved, the priorities to be solved, and the stages to be experienced as well as the countermeasures that must be taken in a longer period of time, based on the estimation of the regional economic and social development, taking into account all aspects of the regional economic [6–9] and social development. Financial risk is mainly due to the uncertainty of financial activities causing losses of participants' subjects, which are latent accumulation and suddenly dramatic and have a strong contagious diffusion. Feng and Yuan [10] studied the interaction between the financial structure and financial risk, its multifractal asymmetric version and time delay based detrended intercorrelation, and market-oriented financial structure which increases financial risk; moreover, the financial structure and financial risk and their interaction is characterized by multifractal, asymmetric, and bidirectional transmission. Song et al. [11] proposed an early warning method for corporate financial risk. The results show that the random forest algorithm can more objectively and intelligently obtain the basic probability assignment of evidence theory. Gonzalez and Quast [12] studied the relationship between sovereign risk and financial risk, and a large part of the linkage between sovereign spreads is caused by changes in global financial risk. Regional financial risk issues are related to the economic development of the region, diverse and complex, and at the same time contagious. In addition, regional financial risk is related to regional finance, and the

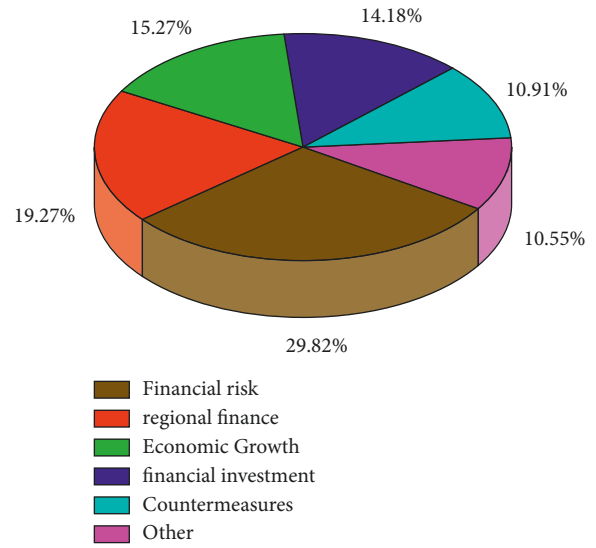


Figure 1: Regional economic research topics proportion chart.

current state of financial development in different regions makes the regional tolerance for financial risk different. Due to the inconsistency of regional economies, financial risk indices are not at all consistent. Therefore, regional financial risk is also the focus of research efforts. Huang et al. [13] performed correlation analysis from regional, financial, and global stock indices, and in their research work, they proposed a model for measuring systematic risk based on dynamic topological indicators and introduced a topological parametric model to measure systematic risk and compare it with traditional measurement models. The results show that the new model can provide more detailed and accurate information about the systematic risk of the stock market. The method can be used for investment decision recommendation and systematic risk warning. Tao [14] set the explanatory variables of financial stress index; through the analysis of the research of various scholars, we found that the deep learning model algorithm plays a vital role in the analysis of regional economic and financial risk.

Regional financial risks are distinctive because of the characteristics of regional economic development. For regional financial risks, common inducing factors include payment risk due to illiquidity, loan loss risk, and market risk. Guangdong Province is the leading province in China's regional economic development. Since 1989, Guangdong has been ranked first in GDP in China for 33 consecutive years, and the economy of Guangdong Province is pivotal in the national economy. The economic structure and regional economic characteristics of Guangdong Province are like a smaller version of China. Although Guangdong has a huge volume, the economic growth rate is still stable at 6.3% at the medium to high speed level. Therefore, this paper analyzes the financial risks related to the regional economy of Guangdong Province, hoping to contribute to the rapid and stable growth of the regional economy of Guangdong Province. This paper analyzes and predicts the financial risks related to the regional economy of Guangdong Province by

considering three main factors: payment risks, loan loss risks, and market risks related to the regional economy through deep learning algorithms.

2. Regional Financial Risk-Related Concepts

Financial liquidity is subject to regional limitations [15–17], and there are large objective differences between different regional economic environments and financial systems. This inter-regional economic and geographical limitation and inter-regional differences in factors make financial risks occur in a horizontal spatial perspective with a regional basis and can exhibit regional characteristics. Regional financial risk is an integral risk in the region and contains regional and integral risk characteristics. Regional financial risk is a dynamic and continuous change, which is a brittle process of accumulation and development in the region. The characteristics of the main influencing factors for regional financial risk are shown in Table 1.

For the generation of regional financial risk, it mainly stems from two main reasons: the fragility and instability of finance itself and the distortion of the real environment in the region. The fragility of finance is a state shown by the gradual loss of its own ability to resist risks in the process of operation and is an important reason for the generation of financial risks and financial instability. The micro perspective, or financial vulnerability in the narrow sense [18, 19], refers to the nature of financial institutions that are more prone to failure and that failure will have significant consequences due to the industry characteristics of operating with leveraged liabilities. The macro perspective, that is, financial vulnerability in the broad sense, refers to a financial state in which the financial system often tends to be highly risky in its self-operation and in its daily interactions with the outside world. For financial vulnerability, it is often caused by information asymmetry, homogeneity of market participants, too much homogeneity in the process of interaction between the government's real sector for macro policies, too much improper government intervention, the distorted reality of the regional passing environment, serious local government debt problems, high leverage of real enterprises resulting in unstable regional economic development, structural imbalance, imbalance of local industrial structure, local economic development relying solely on a single industry, insufficient development of new high-tech multifaceted industries, overcapacity in traditional industries, etc.

2.1. Feasibility of Recurrent Neural Networks for Financial Risk Analysis. Recurrent neural network (RNN) is developed based on the neural network and introduces directional loops, that is, the input of the hidden layer at each moment, thus reflecting the backward and forward sequence relationship between the input data. The regional economy is a microcosm of the national economy and is comprehensive and regional in nature. In recent years, convolutional neural networks such as residual networks and dense structured networks have been proposed one after another continuously refreshing the accuracy of classification tasks on data sets, and the recognition ability of some categories even surpasses that of humans. Therefore, the development of convolutional neural networks has driven the rapid development of computerized financial risk analysis, which has become a mainstream method.

2.2. Structure of Recurrent Neural Network. In this paper, a recurrent neural network (RNN) is used for speech recognition. Recurrent networks can process sequences of arbitrary lengths to prevent the computational results from overfitting or underfitting the dataset [20, 21]. The network calculation error can be reduced by dropout, regularization, and orthogonalization. The following is a detailed introduction to its operation:

(1) Dropout. It refers to randomly freezing a portion of the neurons in the hidden layer from the fully connected neural network in each iteration of training by reducing the correlation between neurons and the complexity of the model in such a way that the effect of regularization can be achieved

(2) Regularization. It reduces the test set error at the cost of increasing the training set error to improve the generalization ability of the model. Currently, the three main regularization methods include parametric penalty, dropout, and early stopping

(a) Parametric penalty. By limiting the number of nonzero parameters, the L_0 parametric number (the number of nonzero elements in the vector) is added to the cost function as a penalty term. The specific calculation results are as follows:

$$\overline{L}(w,b) = L(w,b) + \frac{\lambda}{2m}\Omega(w), \tag{1}$$

where $L(w,b)$ is the function before parametric penalty, m refers to the number of samples, λ represents a hyperparameter to control the degree of regularization, and $\Omega(w)$ denotes the regularization term

The L_1 parametrization is the optimal convex approximation of the L_0 parametrization, and the minimization problem of the L_0 parametrization can be transformed into the minimization problem of the L_1 parametrization under certain conditions. The specific calculation results are as follows:

$$\overline{L}(w,b) = L(w,b) + \frac{\lambda}{2m}\|w\|_1, \tag{2}$$

where w_1 is the sum of the absolute values of the elements in the L_1 parametrization vector

L_2 regularization is the change of the cost function penalty term from the sum of the absolute values of the elements of the vector in L_1 regularization to the square of each element squared sum of the elements, that is, the square

Table 1: Characteristics of the nature of the main influencing factors of regional financial risk.

Characteristic features	Specific explanations
Negative externality	Regional financial risks may affect the geographical area several times and its origin and negative externalities may increase exponentially
Cryptogenic cumulative	Financial risks are the result of a long evolutionary build-up and it is difficult to predict when they will arrive
Complexity	Financial systems are increasingly complex and subject to multiple influences at the same time, making it difficult to anticipate risks
Multiple spaces depend on each other	Regional financial risks are interdependent with multiple spaces and there is a risk of spillover from the next external space
Multidimensional contagion	Regional financial risk correlation mechanisms are multidimensional and have broad contagion

of the mode of the vector. The specific calculation results are as follows:

$$\overline{L}(w,b) = L(w,b) + \frac{\lambda}{2m}\|w\|_2^2, \tag{3}$$

where w_2 is the square root of the sum of the squares of the elements of the vector

(b) Dropout. Randomly freezes a portion of the implicit layer neurons from the fully connected neural network at each iteration of training, reducing the correlation between neurons and the complexity of the model in such a way that the effect of regularization is achieved

(c) Early stopping. For a specific early stop rule, a hyperparameter patience is usually set to stop training when the test error does not decrease in consecutive patience iterations

2.3. Learning Algorithms for RNN Networks. RNNs are a class of recurrent neural networks and are recursive in the evolutionary direction of the sequence. In recurrent neural networks, the weight parameters W, U, and V are shared at each time step, thus greatly reducing the number of parameters to be learned in the network. The training of recurrent neural networks is the process of continuously correcting the weight parameters W, U, and V, using a back-propagation algorithm over time. The specific computational procedure is as follows:

$$s_t = f\left(Ux_t + Ws_{t-1}\right), \tag{4}$$

where s_t is the state of the implied layer at time t and x_t is the input at moment t.

$$o_t = f\left(Vs_t\right), \tag{5}$$

where o_t is activation functions, including sigmoid, tanh, and softmax functions. The specific calculation process is as follows:

$$\begin{aligned} &\text{Sigmoid forward propagation } f(z) = \frac{1}{1+e^{-z}}, \\ &\text{Sigmoid back propagation } f(z)' = f(z)(1-f(z)). \end{aligned} \tag{6}$$

The sigmoid function maps the input value to the interval between [0, 1], and the function is more sensitive to the signal change in the central area and less sensitive to the signal change in the two sides.

$$\begin{aligned} &\text{tanh forward propagation } f(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}, \\ &\text{tanh back propagation } f(z)' = 1 - f(z)^2. \end{aligned} \tag{7}$$

The tanh function can be seen as a variant of the sigmoid function, which maps the input value to the interval between [−1, 1]. The derivative of the function is generally larger than the sigmoid function, resulting in faster convergence during training.

$$\begin{aligned} &\text{softmax forward propagation } f\left(z_j\right) = \frac{e^{z_j}}{\sum_j z_j}, \\ &\text{softmax back propagation } \frac{\delta \alpha_j}{\delta z_j} = \begin{cases} \alpha_i\left(1-\alpha_j\right), \\ -\alpha_i \alpha_j. \end{cases} \end{aligned} \tag{8}$$

Softmax function's output is mapped to the [0, 1] interval, and the sum of the inputs of all output layer neurons is 1. After applying the activation function, the output value of the output layer represents the relative probability of belonging to a certain category.

2.4. Deformable Convolution Module. Therefore, this paper adds a deformable convolution to the output layer to extract more robust features, allowing the network to learn from the rectangular box to learn the data changes autonomously. Features are extracted from the data through different convolution kernels, and these features can be used to preserve the spatial relationship between the data, as shown in Figure 2.

It mainly includes two steps: sampling the input feature X using a specified step size S and a rectangular grid R; multiplying the data in the sampling area by the corresponding weight matrix, and then summing and outputting the corresponding data y, and the formula is as follows:

$$y(p_0) = \sum_{p_0 \in R} w(p_n) \cdot x(p_0 + p_n), \tag{9}$$

where p_0 is the output point on the output y and p_n is the set of positions on the convolution kernel R.

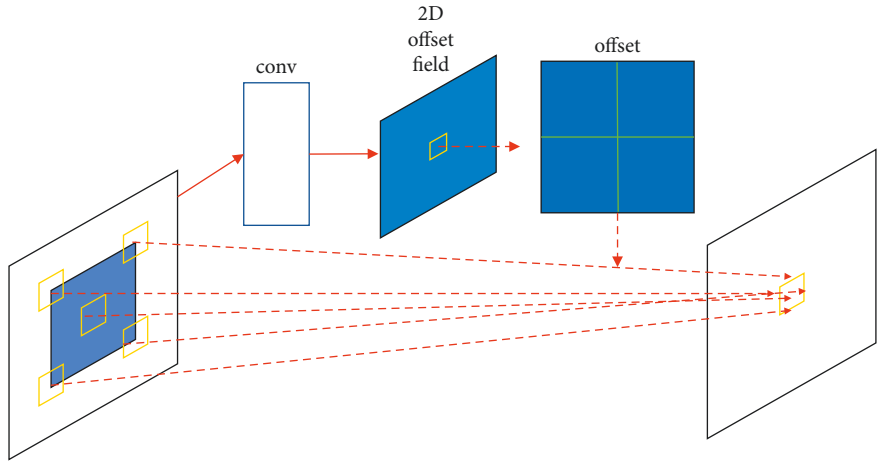


Figure 2: Variable convolutional architecture.

For the value of each position p_0 on the output feature map y, the standard convolution is regularly sampled within the grid, while the deformable convolution adds an offset to the regular sampling:

$$y(p_0) = \sum_{p_0 \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n), \tag{10}$$

where Δp_n is offset.

Add the offset Δp_n to the original sampling position and then the new sampling position becomes $p_n + \Delta p_n$, but the offset is usually a floating-point number, so it needs to be solved by bilinear interpolation. The bilinear interpolation formula is as follows:

$$x(p) = \sum_{q} G(q, p) \cdot x(q), \tag{11}$$

where p is $p_0 + p_n + \Delta p_n$, $G(q, p)$ is a two-dimensional bilinear interpolation kernel function. The two one-dimensional multiplication formulas are disassembled as follows:

$$G(q, p) = g(q_x, p_x) \cdot g(q_x, p_x). \tag{12}$$

3. Experimental Verification and Comparative Analysis

For the experimental parameters, this paper trains the model by migration learning, and the backbone network uses a network pretrained on the dataset. The pretraining can stabilize the gradient change in the initial stage of the network and help to improve the accuracy rate. In this paper, the Adm descent method is used as the optimization algorithm of the model, the momentum value is set to 0.92 and the canonical factor is set to 0.0007, and the canonical term is set for overfitting in this paper. The base learning rate is 0.01, the batch size is set to 32, and the decay coefficient of the base learning rate is 0.1. When the network is close to the optimal solution, the learning rate needs to be reduced by the decay coefficient. Through the statistics of the number of training samples, the network trained according to different input data is more robust.

3.1. Empirical Analysis. The more common approach to financial risk measurement is the composite index model, which has the advantage of simplicity and clarity and can be used in combination with many complex models, expanding the scope of the method. The reality of the relatively late development of China's financial markets makes this study lack sufficient sample data and historical extreme scenarios to accurately estimate the model parameters.

This paper goes to determine whether the principal component analysis method is applied to determine the weight of each dimensional indicator. The test is more applied in factor analysis of multivariate statistics mainly by analyzing the sum of squared simple correlation coefficients; when the former is greater than the latter, the number is close to 1, which means that the correlation among variables is very strong and factor analysis can be performed. The KMO and Bartlett test results obtained in this study are shown in Table 2, which shows a KMO test value of $0.738 > 0.6$. Meanwhile, the Bartlett spherical test value is 1078.215 and Sig is 0.000, indicating that the original hypothesis is rejected and the original variables can be correlated by principal component analysis, which can be used to construct a regional financial risk index model.

3.2. Model Optimization. On top of the traditional indicators, this paper then considers the search index risk measurement submodel constructed on the basis of RNN to reconstruct the index analysis. Four major regions are selected for this study to carry out the analysis of regional financial risk. The general situation is generally considered as a period of high stress when the degree of deviation is greater than twice its standard deviation, and the specific formula and determination criteria are as follows.

$$\mathrm{FSII}_{it} = \frac{\mathrm{FSII}_{it}\mathrm{MEAN}(\mathrm{FSI})}{2 \times \mathrm{SD}(\mathrm{FSI}_{it})} - 1, \tag{13}$$

where FSII_{it} is the regional financial risk stress identification index of the i-th region in period t, and FSI_{it} is the regional financial risk index in period t.

Table 2: KMO and Bartlett test results table.

Kaiser-Meyer-Olkin metrics		0.738
Bartlett's sphericity test	Approximate chi-square	1078.215
	df	98
	Sig.	0.000

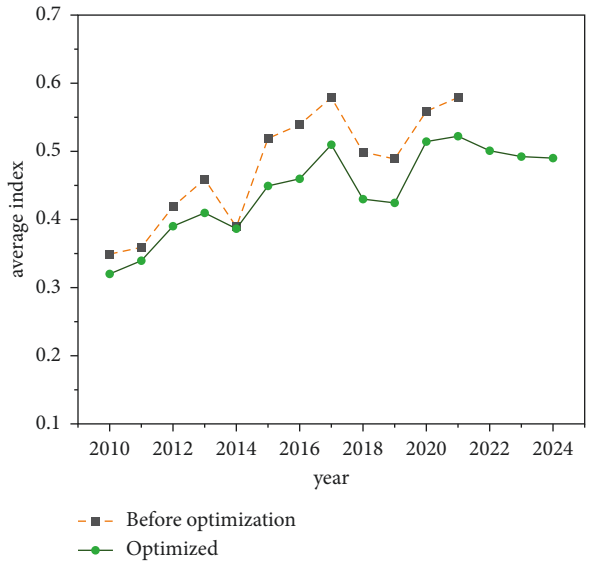


Figure 3: Evolution of the average regional financial risk index in Guangdong Province.

The study of this paper explores in detail the evolution of the actual regional financial risk index in Guangdong Province, which is elevated as shown in Figure 3, and finds that, in general, the regional financial risk in the country and in each region is on an upward trend compared with 2010 with different and fluctuating changes in the process. The average regional financial risk index of Guangdong Province peaked at 0.47, 0.58, and 0.57 in 2013, 2017, and 2021, respectively, indicating that there is still significant risk pressure on regional financial risk in China in recent years. By region, the regional financial risk in the central and western regions of Guangdong Province has been greater than that in the eastern and northeastern regions, requiring this paper to focus on prevention and mitigation. In summary, this is inextricably linked to the good economic and social conditions in these regions, the complete financial infrastructure and the strong local financial supervision as well as the ability to mitigate risks, and it is worthwhile to summarize the successful experience of all other regions in China.

This paper compares the three types of risks, such as payment risk, loan loss risk, and market risk, through the model software as shown in Figure 4. It can be seen that the loan risk accounts for nearly half of the risk, which indicates that personal lending is very risky and has a great impact on the economic risk of Guangdong region, so we focus on its supervision and management in the follow-up work. In addition, payment risk and market risk account for 26.82%

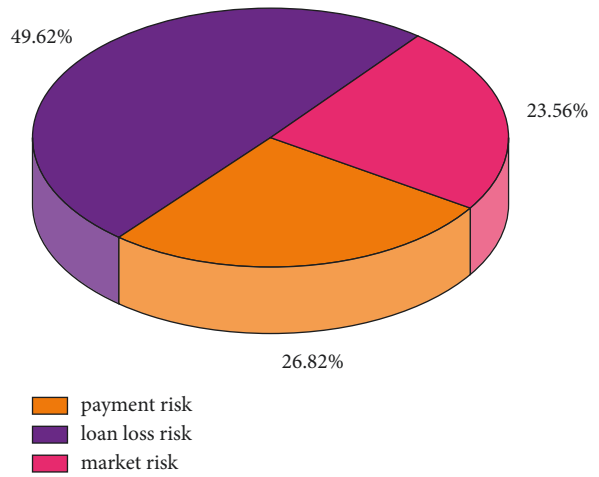


Figure 4: Evolution of the average regional financial risk index in Guangdong Province.

and 23.56%, respectively. They also need to be further optimized for research and analysis.

4. Conclusion

This paper systematically summarizes the existing comprehensive regional financial risk measurement system, including the construction methods of financial risk measurement and financial risk early warning system, in order to provide a theoretical reference system for subsequent research. This paper systematically summarizes the existing regional financial risk measurement system, including the construction methods of financial risk measurement and financial risk early warning system, in order to provide a theoretical reference system for subsequent studies.

(1) We elaborate on the financial risk transmission mechanism involving intra and interregional financial risks and sort out the relationship between sectors as a clue to discover a complete and systematic regional financial risk transmission mechanism.

(2) The regional financial risk index design process and the measurement method and the regional financial risk index measurement for typical regions fully exploit and use such data to facilitate the analysis of the real-time dynamic advantages of financial risk, making up for the shortcomings of traditional basic indicators.

(3) The early warning model based on regional financial risk index values is constructed to extend the application of RNN networks to the construction of regional financial risk early warning system. It both extends the early warning system for building regional financial risks and provides an effective means for risk monitoring by relevant regulatory authorities.

(4) Based on the construction of the RNN network application risk warning system, three types of risks, namely, payment risk, loan loss risk, and market risk, with the percentages of 49.62%, 26.82%, and 23.56%, respectively, were derived, which will be focused on the supervision and management in the follow-up work.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The work described in this paper was supported by the Guangdong Province Philosophy Social Science Fund Project (GD20CYJ21, provincial level).

References

[1] W. Zhu, T. Zhang, Y. Wu, S. Li, and Z. Li, "Research on optimization of an enterprise financial risk early warning method based on the DS-RF model," *International Review of Financial Analysis*, vol. 81, Article ID 102140, 2022.

[2] C. Liu, Y. Fan, Q. Xie, and C. Wang, "Market-based versus bank-based financial structure in China: from the perspective of financial risk," *Structural Change and Economic Dynamics*, vol. 62, pp. 24–39, 2022.

[3] P. Danisewicz, C. H. Schaeck, and K. Schaeck, "Private deposit insurance, deposit flows, bank lending, and moral hazard," *Journal of Financial Intermediation*, Article ID 100967, 2022.

[4] A. de Jong, T. Veld, and C. Veld, "Legal risk and information spillover through private lender reports," *Journal of Financial Markets*, Article ID 100706 2022.

[5] J. Boukhatem, "How does financial risk affect sukuk market development? empirical evidence from ARDL approach," *Heliyon*, vol. 8, no. 5, Article ID e09453, 2022.

[6] Y. Bai, M. Zhao, R. Li, and P. Xin, "A new data mining method for time series in visual analysis of regional economy," *Information Processing & Management*, vol. 59, no. 1, Article ID 102741, 2022.

[7] B. Zhou, Z. Wen, and Y. Yang, "Agglomerating or dispersing? spatial effects of high-speed trains on regional tourism economies," *Tourism Management*, vol. 87, Article ID 104392, 2021.

[8] P. Zając and A. Avdiushchenko, "The impact of converting waste into resources on the regional economy, evidence from Poland," *Ecological Modelling*, vol. 437, Article ID 109299, 2020.

[9] S. K. Chowdhury and M. L. Endres, "The influence of regional economy- and industry-level environmental munificence on young firm growth," *Journal of Business Research*, vol. 134, pp. 29–36, 2021.

[10] W. Feng and H. Yuan, "Haze pollution and economic fluctuations: an empirical analysis of Chinese cities," *Cleaner Environmental Systems*, vol. 2, Article ID 100010, 2021.

[11] M. Song, X. Ma, Y. Shang, and X. Zhao, "Influences of land resource assets on economic growth and fluctuation in China," *Resources Policy*, vol. 68, Article ID 101779, 2020.

[12] F. Gonzalez and T. Quast, "The relationship between abortion rates and economic fluctuations," *Economics and Human Biology*, vol. 46, Article ID 101120, 2022.

[13] G. Huang, X. Yao, and X. Yao, "Dynamics of China's regional carbon emissions under gradient economic development mode," *Ecological Indicators*, vol. 51, pp. 197–204, 2015.

[14] Z. Tao, "Strategy of city development in low-carbon economic mode-a case study on qingdao," *Energy Procedia*, vol. 5, pp. 926–932, 2011.

[15] Y. Zhang, T. Qian, and W. Tang, "Buildings-to-distribution-network integration considering power transformer loading capability and distribution network reconfiguration," *Energy*, vol. 244, 2022.

[16] X. Jia, W. Zhang, and C. Zhang, "Commodity financialization and funding liquidity in China," *The North American Journal of Economics and Finance*, vol. 60, Article ID 101674, 2022.

[17] W. Zhou, G. Y. Zhong, and J. C. Li, "Stability of financial market driven by information delay and liquidity in delay agent-based model," *Physica A: Statistical Mechanics and Its Applications*, vol. 600, Article ID 127526, 2022.

[18] I. Faiella, L. Lavecchia, V. Michelangeli, and A. Mistretta, "A climate stress test on the financial vulnerability of Italian households and firms," *Journal of Policy Modeling*, vol. 44, no. 2, pp. 396–417, 2022.

[19] T. Qian, X. Chen, Y. XinXin, W. Tang, and L Wang, "Resilient decentralized optimization of chance constrained electricity-gas systems over lossy communication networks," *Energy*, vol. 239, Article ID 122158, 2022.

[20] T. Qian, Y. Liu, W. Zhang, W. Tang, and M. Shahidehpour, "Event-triggered updating method in centralized and distributed secondary controls for islanded microgrid restoration," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1387–1395, 2020.

[21] C. Fang, Y. Tao, J. Wang et al., "Mapping relation of leakage currents of polluted insulators and discharge arc area," *Frontiers in Energy Research*, vol. 9, Article ID 777230, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6120511, 12 pages
https://doi.org/10.1155/2022/6120511

Research Article

Study on Tourism Consumer Behavior and Countermeasures Based on Big Data

Jing Li[1] **and Bin Cao**[2]

[1]*Jinzhong University, Jinzhong 030619, Shanxi, China*
[2]*Universiti Sains Malaysia, Gelugor 11800, Penang, Malaysia*

Correspondence should be addressed to Jing Li; fyt20170601@163.com

Received 10 May 2022; Revised 20 June 2022; Accepted 1 July 2022; Published 19 July 2022

Academic Editor: Man Fai Leung

In our study, through consulting, summarizing, and analyzing a large number of related literature studies on tourism consumer behavior, tourism big data, text data analysis, and so on, a framework of research ideas on tourism consumption was constructed. The train browser, NLPIR, and other software packages are used to crawl, preprocess, and mine the travel sample data, and the word frequency analysis, co-occurrence analysis, content analysis, sentiment analysis, network analysis, and other methods are used to analyze the characteristics and decision-making behavior of tourists. Based on the results of behavioral analysis, we proposed tourism development strategies from three aspects: reforming and promoting tourism marketing strategies, improving tourism product and service quality, and improving tourism destination management methods. The results show that (1) for the tourist characteristics, taking into account the factors of climate and geographical location, the domestic market is divided into four grades of markets, and different marketing strategies are adopted according to different market characteristics; (2) for the tourism decision-making behavior, a "push-pull resistance" tourism decision-making model was established through word frequency analysis, co-occurrence analysis, and content analysis; (3) for the tourism consumption preferences, through network analysis of scenic spots, it is found that there are three tourist routes preferred by tourists; and (4) for the tourism perception evaluation behavior, based on the "cognitive-emotional" model, this study describes the tourism image from the two dimensions of the cognitive image and emotional image. Generally speaking, tourists show a positive perception state. The research on tourism consumer behavior based on UGC (user-generated content) data can help scenic spots and other tourism companies to understand the characteristics and rules of tourists' behavior, understand the consumption preferences of different tourism groups, develop diversified tourism products, improve the quality of tourism services, and further cater to market segments. This research provides a new idea for tourist attractions and tourism management departments to monitor tourist behavior through big data analysis.

1. Introduction

Tourism consumer behavior involves many disciplines such as marketing [1]. Tourism consumption behavior refers to the process that tourism consumers choose and purchase tourism products to meet the needs of tourism pleasure and other experiences. This process includes the generation of needs before travel, the decision-making process, consumption in scenic spots, and post-purchase evaluation. Influenced by many factors such as economy, society, and cultural environment, it is an experience activity with comprehensive, marginal, and extraordinary characteristics [2].

The American Marketing Association defines consumer behavior as follows: "The dynamic interaction process among emotional, cognitive, behavioral, and environmental factors is the behavioral basis for human beings to perform the exchange function in life" [3]. This definition regards consumer behavior as a process, including multiple stages of selection, purchase, and disposal, and the process is dynamic. Tourism consumption behavior is more complex because it involves the whole process from leaving home to returning home [4]. To reveal the rules of tourist behavior, the relationship between various behaviors, and the core influencing factors of behavior, academia has proposed

many different models. Despite the differences among different models, these models collectively emphasized these psychological activities and behavioral manifestations, such as travel motivation, travel decision-making, choice preference, destination image, and satisfaction, before, during, and after travel [5].

Tourism motivation is an internal factor that prompts potential tourists to take a certain tourism action, and it is also a key factor for a series of tourism behaviors in the future. Unlike consumption demand, consumption motivation is related to products, so tourism motivation is the link between tourism demand and specific tourism destinations [6]. Scholars at home and abroad have made a lot of research on tourism motivation. The most popular theoretical models are the distribution center model, push-pull theory, optimal arousal theory, leisure incentive method, and travel career ladder (TCL) model, among which push-pull theory is the most widely used [7].

Based on scholars' cognition that "tourism behavior is a series of rational decision-making activities," most of the tourism decision-making theories have evolved from the decision-making model of consumer behavior, which has led to the emergence of many tourism decision-making behavior models [8]. Sirakaya and Woodside, Smallman, Cohen, and other scholars have made a more detailed summary of the tourism decision-making model, mainly focusing on the model research of destination selection and various subdecisions (such as accommodation and tourism activities) [9]. Summarizing previous research on decision-making models, Sirakayaet al. believe that the final tourism consumption decision will depend on the interaction between the following variables: intrinsic variables, including travel attitude, motivation, information search behavior, and personality characteristics; extrinsic variables, including constraints, destination pull factors, the influence of family and reference groups, marketing mix, social class, and culture; the nature of the scheduled trip, including travel distance and date and playtime; and travel experience, including emotions and feelings during the trip and post-tour evaluation [10].

Motivation initiates action and guides desirable behavior, but policymaker preferences help filter choices precisely; preferences are more specific than motivation, revealed by visitor whereabouts and visitor behavior. Many scholars' research on tourism consumption preference focuses on destination choice preference (such as Tzu-Kuang Hsu), shopping preference (such as Samuel), accommodation preference (such as Isabel), food preference (such as Richard), etc. Destination image is "an attitude or psychological structure that reflects the sum total of tourists' personal views, trust and impressions of a tourist destination" [11]. During the entire tourism process, tourists' perception of the destination image is dynamic. Destination image theory is an important part of tourism research and is widely regarded as an important factor affecting tourists' decision-making, destination selection, post-tour evaluation, and future behavior [12]. The advent of the Travel 2.0 era (i.e., new ways for real travelers to interact and share information or content through Web 2.0) has changed the way travel information is searched, viewed, and evaluated. In this context, the tourism destination image has been reshaped and widely disseminated through Web 2.0 and UGC [13]. In the study of destination image perception, the most widely used theoretical model is the "cognitive-affective" model; that is, the destination image is divided into two interrelated components: cognitive image and affective image, where the cognitive image is an individual knowledge or belief about a destination. Emotional image refers to an individual's subjective feelings about the destination. The interaction of cognition and emotion forms a unique overall image of each destination, including positive or negative evaluations. Compared with emotional image, cognitive image has a stronger impact on the overall image [14].

At present, scholars at home and abroad have not reached a clear consensus on consumer satisfaction. Some scholars (such as Oliver) believe that satisfaction is an emotional state derived from tourism experience [15]. In this view, tourists' emotion has become an important indicator of satisfaction evaluation. Some scholars (such as Chon) believe that satisfaction is a function of goodness of fit between tourists' expectations of the destination before travel and the perceived evaluation of their visit experience after travel, and the expectation uncertainty model is the most commonly used measurement model.

With the rapid development of the Internet, structured and unstructured data are generated, recorded, stored, and accumulated, forming big data with the characteristics of volume, variety, and velocity, bringing amazing changes to tourism research, which has entered the big data era [16]. According to different data sources, tourism big data are mainly divided into three categories: UGC data, equipment data, and transaction data. Among the three types of data, UGC data have the most applications due to its low cost and easy access [17]. Device data applications have problems such as high cost, low accuracy, and small coverage. Since most of the transaction data are private information, only tourism organizations or government departments are used. Owned, its use level is the lowest.

In the digital age, the proliferation of the Internet and social media has dramatically changed travel patterns, providing a platform to share UGC data. User-generated content (UGC), also known as user-generated content (UCC) and consumption-generated media (CGM), was first proposed by KPCB chief analyst Mary Meeker in 2005, thinking that UGC is a new type of content in the Web 2.0 environment. There is no unified and clear definition of the creation, release, and organization mode of information resources. UGC data have been widely used in tourism research and mainly include two types: (1) online text data, such as online reviews, online travel notes, and blogs posted on social media; (2) online photo data, such as in photo sharing online photo data published by websites or online travel journals. However, due to various limitations, UGC data such as audio and video have not been widely used by researchers at home and abroad. The UGC data discussed in this study are mainly text data [18]. Online text data mainly include two types: one is review data, that is, the evaluation data of online tourism products or tourism destinations,

expressing tourists' attitudes towards tourism products (such as accommodation facilities, restaurants, and scenic spots), mainly used to measure satisfaction, electronic word of mouth, etc. [19]. The second is online travel notes or blogs, that is, travel note modules on online travel websites or travel blogs on social media, which record travel feelings and experiences, mainly involving travel behavior, travel sentiment analysis, and travel recommendations. Research on UGC tourism data, at both home and abroad, focuses on empirical research, but the research focus is slightly different. Foreign research focuses on tourist satisfaction/service quality, destination image, tourism behavior, and tourist segmentation. In China, the purpose of land image and tourism behavior is the focus of research [20].

Co-occurrence refers to the phenomenon that specific keywords co-occur in a text collection. Co-occurrence analysis is a quantitative study of co-occurrence phenomena to reveal the content correlation of information and the knowledge implied by feature items [21]. The specific process of co-word analysis mainly has four stages: one is to determine the analysis data set; the other is to determine the analysis object. High-frequency words related to the research object are extracted from the text collection; the third is to construct a binary co-word matrix to obtain quantitative information such as co-occurrence frequency; the fourth is to conduct co-word analysis, which is generally combined with social network analysis for visual display [22].

Throughout the research on tourism consumer behavior at home and abroad, it is more inclined to conduct research on tourists' tourism motivation, tourism decision-making, spatial behavior, consumption preference, tourism experience, destination image, tourist satisfaction, and other subfields from an individual and micro-perspective [20]. However, they all ignore a problem: tourists are an organic whole, and their behavior runs through the entire tourism process before, during, and after tourism, but the relevant research literature seldom regards pre-tourism, tourism, and post-tourism as the study as a whole. Previous studies usually divide tourism consumer behavior into multiple subfields and use questionnaire surveys and other survey methods to conduct in-depth, detailed, and enlarged research on a subfield, but it is easily affected by questionnaire design structure, sample sampling, and investigators. Due to the limiting factors such as the subjectivity of the survey, it is impossible to objectively and completely understand the behavioral characteristics and laws of tourists. However, with the widespread of social media and Web 2.0, an online search for travel information, purchase and evaluation of travel products, and sharing travel experiences have become one of the important behavioral characteristics of tourists. UGC data with other characteristics have also become one of the important data sources for tourism behavior research [23]. In addition, in the research on tourism consumption behavior based on UGC data, domestic and foreign researchers have focused on empirical research on tourist satisfaction, destination image, tourism spatial behavior, tourism decision-making, etc., lacking the overall research on tourist behavior [24].

Therefore, this study combines the characteristics and content of online travel data to conduct an overall analysis of tourist behavior from the aspects of tourism motivation, tourism decision-making, consumption preference, destination image perception, and tourist satisfaction. This research proposes a research framework of tourism consumption behavior based on online travel notes, which is mainly divided into two levels: the first level is a three-stage dynamic model of tourism consumption behavior; the second level is to establish tourism decision-making, tourism consumption preference, and behavioral models such as perceptual evaluation after swimming.

In our study, we first conducted the data collection and mining and introduced the general situation of the tourist area and introduced the data collection and preprocessing of online travel notes, the selection basis of text mining tools, and the selection criteria of online travel note samples in Section 2 and then introduced the methods of word frequency analysis, co-occurrence analysis, network analysis, sentiment analysis, etc., and conducted mining and research on the characteristics of tourist consumption behavior from four aspects: tourist characteristics, decision-making behavior, consumption choice preference, and perception evaluation, which can be found in Section 2; in Section 3, through in-depth research on the behavior characteristics extracted by text data mining, the influencing factors of tourism consumption behavior are summarized, and combined with the results of field research, the corresponding strategies for tourism development planning in tourist areas are put forward.

2. Methods and Data Collection

2.1. Methods

2.1.1. Literature Research. The literature research method is a basic research method adopted in this article. A large number of related disciplines and professional books and relevant predecessors' research results have been consulted. On the basis of sorting out and summarizing previous research, the research direction and content of this study are determined; at this stage, by consulting a large number of related literature on summer tourism, consumer behavior, tourism big data, etc., the theoretical basis and methods of this research are summarized [3].

2.1.2. Content Analysis. Content analysis is a method that quantitatively analyzes the content of nonquantitative literature to find out the core content and laws contained in the literature. This research uses NLPIR big data semantic intelligent analysis software to analyze the content of the collected online travel text data and obtained valuable and meaningful data, then explores the behavioral characteristics and laws of summer tourism consumers, and finally explains the research phenomenon.

2.1.3. Sentiment Analysis. Text sentiment analysis refers to the process of analyzing, processing, summarizing, and reasoning on subjective texts with emotional color [20]. According to the different granularity of text, it can be divided into multiple levels such as word level, sentence level, and chapter level. This study uses sentiment analysis to

extract emotional words or emotional information from the online travel text data set and uses the chapter as a unit to identify the emotional color of travel notes, so as to obtain tourists' emotional tendencies towards four places in southern Xinjiang.

2.1.4. Network Analysis [12]. Network analysis is a method to study the characteristics of the entire network and the individuals in the location-based network structure by describing the relationship structure between given entities (represented by nodes) and applying quantitative techniques to generate relevant indicators or results [15]. Based on the idea that the relationship between tourist attractions is not isolated but related to each other, this study analyzes the network structure between tourist attractions by calculating the co-occurrence frequency of different tourist attractions and proposes a proposal for tourism attraction cooperation strategy based on summarizing the spatial flow characteristics of tourists between tourist attractions.

2.1.5. Mathematical Statistical Analysis [5]. This study uses SPSS and Excel software to carry out statistical analysis on the crawled tourist sources, days of play, travel companions, and other indicators, as well as the number of articles and frequency of use of related subject words, to obtain the basic characteristics of tourists, and to study summer tourism consumption. Behavioral characteristics provide corresponding data support [25].

This study mainly uses NLPIR (natural language process and information retrieval) software to achieve co-occurrence analysis. The calculated quantitative information includes co-occurrence frequency, binary probability, and binary word pair information entropy. Co-occurrence frequency refers to the number of co-occurrences of two words before and after, which is used to reflect the strength of association between them. It is generally believed that the higher the co-occurrence frequency of a phrase, the closer the relationship between the phrases. Binary probability refers to the probability of occurrence of co-occurring word pairs, and the information entropy of a binary word pair represents the breadth of information contained in the phrase [6]. The calculation formula of the information entropy is as follows:

$$H(U) = -\sum_{i}^{n} P_i \log(P_i), \tag{1}$$

where $H(U)$ is information entropy and P_i is the probability of variable i.

To extract and utilize useful information hidden in online text data, methods such as text mining and content analysis are widely used in tourism research. Text mining mainly includes three typical stages: data collection, data mining, and result output, in which the data mining stage includes two sub-steps of data preprocessing and pattern discovery. The first step of data collection is as follows. There are two main ways to collect UGC data: open API access and Web crawler. The second step is data mining. Collected online text data are analyzed to extract useful information through two substages: data preprocessing and pattern discovery. Pattern discovery is another key stage of text mining, aiming to explore interesting information in documents, and typical techniques in existing tourism research are LDA analysis, sentiment analysis, statistical analysis, clustering and classification, text summarization, and dependency modeling. The third step results in output [8]. Interesting information extracted through data mining is transformed into useful knowledge to further serve tourism research. According to relevant research, valuable knowledge covers tourist satisfaction, consumption preference, tourist destination image, tourist route, review characteristics, etc., which is of great help in improving tourism management and providing tourism advice [19].

2.2. Data Collection. The four prefectures in southern Xinjiang include Aksu Region, Kizilsu Kirgiz Autonomous Prefecture, Kashgar Region, and Hotan Region. It is an important section of the central route of the Silk Road, the southern route, and the church. The two routes of the Silk Road pass through the four prefectures in southern Xinjiang. It is located on the northwestern border of the motherland, the western edge of the Taklimakan Desert, and the southwestern part of the autonomous region; it borders six countries including Kyrgyzstan and India. There are 5A and 4A tourist attractions such as the ancient city of Kashgar, the Populus euphratica scenic spot in Zepu County, the Stone City scenic spot, and the Daolang portrait scenic spot, the Honghai Bay scenic spot, and the Etigar Mosque. It is a first-class port in 5 countries including Kashgar International Airlines and has a rich border and ethnic characteristic tourism resources.

2.2.1. Sample Selection of Online Travel Data. This research mainly uses the travel notes related to the actual tour itinerary shared by tourists on the Internet after the trip. By searching for websites with tourism community functions (based on the comprehensive ranking of tourism websites on top.chinaz.com), and according to the research object and research purpose of this study, Ctrip, Mafengwo, Baidu Travel, and Qunar were selected as the main data sources. In the crawling of online travel note text data, the Web crawler software (Train Browser 8.2 version) was mainly used to crawl the online travel note data related to summer travel in southern Xinjiang from four selected online travel websites, and the original online travel note database was formed, namely Database.mdb. Then, the crawled travel database according to the following criteria is filtered: online travel notes that do not take southern Xinjiang as a tourist destination are excluded; travel notes that only have pictures, less text (less than 500 words), incomplete travel information, and are not suitable for text content analysis are eliminated; travel notes that are not for the purpose of sharing travel experiences are eliminated, such as advertisements and propaganda; and duplicate travel notes and merge serial travel notes of the same author are removed. After screening, 500 online travel notes (including 250 on

Mafengwo, 110 on Qunar, 90 on Baidu Travel, and 40 on Ctrip) were selected as the initial text database for this study.

Due to the limitation of HTML structure, some indicator data need to be further sorted out. For example, Excel's "replace" function is used to clean up redundant text information such as "departure time/" in indicators such as "departure date"; blank data are filled according to the content of travel notes; the CLEAN function is used to remove characters such as "newline" and "space" in the text of the travel note; and individual vacancies are manually extracted from the text to supplement and improve them. In addition, the body text data of the 500 travel notes are copied into an "original corpus.txt" document to form a travel note one-line format.

2.2.2. Selection of Chinese Text Mining Tools. In this study, the NLPIR big data semantic intelligent analysis platform developed by Zhang Huaping was selected for text mining analysis. The NLPIR analysis platform has the following advantages: (1) the word segmentation tagging technology based on the cascading hidden horse model has a word segmentation accuracy of close to 98.23% and has the advantages of high accuracy, and fast speed, and strong adaptability. (2) The entity extraction system based on role annotation can intelligently identify the names of people, places, institutions, media, authors, and the subject keywords of articles that appear in the text. (3) The word frequency statistics function based on the patented perfect double array TRIE tree algorithm has high efficiency and is more than ten times faster than the conventional algorithm. (4) The keyword extraction technology based on context conditional entropy can extract several words or phrases representing the semantic content of the article on the basis of comprehensively grasping the central idea of the article. In addition, it uses cross-information entropy to calculate the context of each candidate word. Conditional entropy identifies the most recent new words and assigns them weights. (5) Text sentiment analysis is based on a deep neural network, and NLPIR sentiment analysis provides two modes: sentiment discrimination of full text and sentiment discrimination of specified objects [26, 27].

2.2.3. Preprocessing of Travel Note Text Data Set. In this study, the travel note text data are preprocessed as follows: (1) new words are discovered and user dictionaries are built. Using the new word discovery function of NLPIR, words with new connotations and new concepts are mined from the original corpus and added to the user dictionary to improve the accuracy of the word segmentation system. In addition, the user dictionary is manually checked to fill in the gaps, and the content of the scenic spots and tour routes in the research area in the user dictionary is further improved; (2) building stop word lists for data cleaning: data cleaning is used to detect and remove inaccurate or useless records in text data, such as misspellings and stop words, to leave valuable travel information; and (3) word segmentation and word frequency statistics are circularly performed, and the user dictionary is continuously improved, and word list is stopped until accurate word segmentation and word frequency statistics are achieved. Finally, the more accurate word segmentation results are analyzed by Chinese word frequency statistics, and the phrases that are irrelevant or meaningless to the research content are eliminated.

3. Results and Discussion

3.1. Analysis of the Basic Characteristics of Tourists. For the consideration of personal privacy protection, online travel notes rarely show the demographic characteristics of tourists but can capture module information such as tourist source, travel time, travel days, travel partners, and travel methods and obtain basic characteristics of tourists through descriptive statistical analysis [20].

Distribution characteristics of customer sources: the distribution of source areas shows obvious geographical differences between east-west and north-south. North China, Northeast China, and East China accounted for 82.6% of the tourists [21-23]. Affected by the principle of travel distance attenuation and the difference in summer climate, Northeast China accounted for the largest proportion, followed by North China and East China. There are many high-temperature weather in summer and economic development factors, and there is a strong demand for going north to escape the summer heat. The tourism resources in the southwest and northwest regions are relatively concentrated, which can meet the tourism needs of tourists in the local and surrounding areas, so there is a big difference between the east and west source of tourists and affected by the travel distance, and southern Xinjiang is more attractive to northern tourists than southern tourists. Therefore, there are differences in the distribution of tourists from north to south. However, due to the law of regional differentiation, there are differences in tourism resources between the north and the south. For the motivation of seeking innovation, the central and southern regions and the southwest regions will become potential tourist sources (Table 1).

Figure 1 shows the results of statistics of departure date data in 500 travel notes based on the year. The number of travel notes generally maintains a continuous upward trend, with a larger increase after 2015, which is not only closely related to the tourism development of the four southern Xinjiang regions but also related to the popularity of online travel notes. With the gradual improvement of tourism-sharing community-related websites, the creation of online travel journals has gradually become a part of the tourism process.

Regarding the number of days to visit (Figure 2), due to the abundant resources of scenic spots in southern Xinjiang (three tour routes have been developed), the tickets are valid for three days, and the summer vacation is long, so more tourists choose the itinerary within 1 week, accounting for 84%, of which 3-4 days of in-depth tours are the most popular, followed by 1-2 weeks of itineraries, accounting for 13.3%, and the least number of tourists who choose to play for more than 2 weeks. After calculation, the average playing days of the travel note sample is 5 days, which shows that tourists in southern Xinjiang spend a long time playing.

Table 1: Statistical results of customer source distribution.

Distribution of tourists	Occur number	Frequency (%)
Northeast	160	32
North China	140	28
East China	102	20.4
Central South	58	11.6
Western South	12	2.4
Western North	8	1.6
Others	20	4
Total	500	100

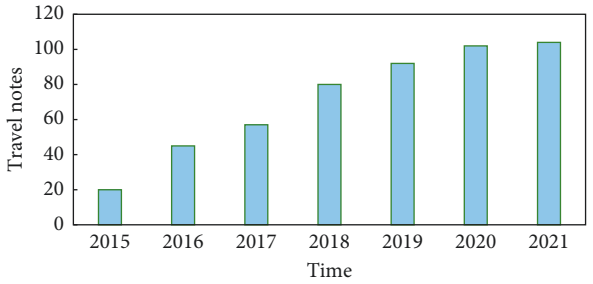


Figure 1: 2015–2021 statistical statistics of online travel notes in southern Xinjiang.

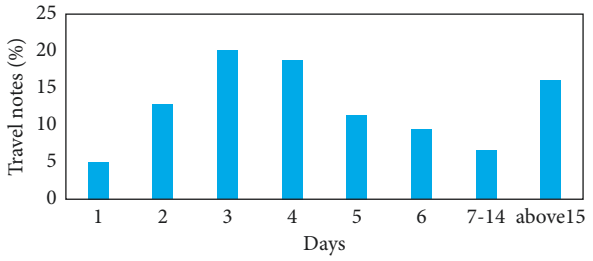


Figure 2: Statistics of tourist travel days.

The per capita consumption of tourists has a relatively large span, with a minimum per capita consumption of 200 yuan and a maximum of 10,000 yuan, but in general, the per capita consumption is concentrated in the range of 1,000–3,999 yuan (57.4%). According to the content of the travel notes, the reasons for the low consumption are as follows: first, the distance to the destination and the number of days of play are few, so the consumption is not high; second, there are friends or relatives in the local area; it is to choose low-cost group tours. Combined with the number of days of tourist travel, it is calculated that the per capita daily consumption is between 100 and 2550 yuan (see Figure 3).

3.2. Analysis of Tourists' Decision-Making Behavior. Travel motivation is a key factor in all travel behaviors and an important driver of travel decisions [24-25]. 59 related entries were extracted through word frequency analysis articles, and entries were classified into three types of factors: push force, pull force, and resistance force according to the word meaning (see Table 2).

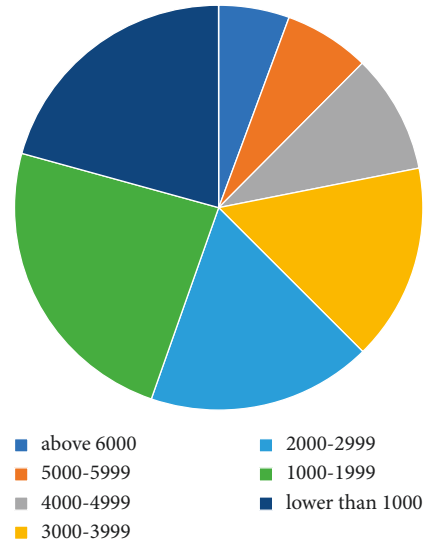


Figure 3: Statistics on per capita consumption of tourists.

Iso-Ahola divides push factors into two dimensions: escape (such as escape from everyday life or interpersonal circumstances) and seeking (such as adventure or friendship building), which explain why tourists travel and what type of destination or type of destination they need. Drawing on Iso-Ahola's division of dimensions, after content analysis, this study divides the thrust factors into two dimensions: escape and seek [26]. Through the co-occurrence analysis, tourists are not only eager to escape the high temperature in summer but also want to escape the hustle and bustle of the city, and they want to go on a trip. Seeking: through the cluster analysis of the words of seeking dimension, tourists' travel motives are mainly to accompany family members or friends, to seek novelty, to go on vacation, to pursue freedom and excitement, etc. Among them, accompanying family members and friends to experience different sceneries and perceive different worlds is an important factor driving tourists to travel abroad. Summer tourism is not only a leisure and vacation way for people to escape from high-temperature cities but also an important way for people to seek companionship and novelty.

Tourism decision-making is the external manifestation of tourism motivation and the core behavior in the process of tourism [27]. A decision-making process contains multiple subdecisions of information, and each subdecision exists in two stages of the tourism process: one is information search and decision-making before travel, such as which information acquisition channel to choose and what information to search for; the second is tourism instant decisions. According to the content analysis and word frequency analysis of travel notes, this section mainly analyzes the characteristics of tourism decision-making behavior from three aspects: tourism decision-making behavior, decision-making information needs, and information acquisition channels.

After keyword analysis and co-occurrence analysis (see Table 3), tourism information channels mainly come from

Table 2: Word frequency statistics of tourist motivation.

	Weight	Frequency of words
Pull factor		
Kashgar Old Town	455.01	3259
Bayanbulak Grassland	279.5	1317
Tianshan Mysterious Grand Canyon	263	1438
Pamirs	185	730
Wensu Tomur Grand Canyon	124	340
Fragrant Concubine Garden	110	395
Xinjiang Robu Village Scenic Spot	108	400
Lop Nor	104	360
Kuqa Temple	90	344
Daolang Tribe	86	280
Kezier Thousand Buddha Caves	85	220
Loulan Ancient Town	83	89
Aksu Tianshan Sacred Wood Garden	69	230
Thousands of Miles of Grape Corridor	66	170
Bosten Lake	57	170
Taklimakan Desert	42	80
Kuqa Old Town	34	130
Muztagh	34	40
Karakul Lake	30	90
Tarim Populus Euphratica Park	23	200
Datongxiang Apricot Flower	22	180
Karakoram Highway	20	140
Taxkorgan Baisha Lake	18	60
Kashgar Ferris Wheel	18	130
Edgebrook Swan Lake	17	150
Muji Crater	14	90
Niya Ruins	12	130
Taxkorgan Town	10	100
Iron Gate	7	70
Bayanbulak Reserve	6	40
Thrust factor		
Friends	127	300
Children	107	320
Experience	101	220
Summer	85	270
Hot summer	40	120
High temperature	35	50
Climb mountains	33	210
Fascinated	27	60
Escape	24	50
Entertainment	23	70
Go away	22	70
Visiting	20	130
Feel	19	140
Appreciate	18	140
Relax	18	60
Parents	18	110
Seek	16	50
Enjoy	16	130
Vacation	15	320
Stimulate	14	100
Freedom	9	150
Resistance factor		
Weather	165	560
Traffic	95	290
Can't see the most wanted scenery	33	80
Bad weather	16	10
Changeable weather	13	20
Pity	10	120
Inconvenient traffic	4	22
Black car	4	15

the Internet, including travel notes, strategies, and reviews of travel websites such as Ctrip, official websites, or APPs of scenic spots and search engines such as Baidu, Taobao, and Dianping. Due to third-party platforms information needs, tourists have different preferences in choosing information channels. For example, tourists mainly obtain information on destinations and attractions, tourist routes, itineraries, and other information through the strategies, travel notes, and reviews of tourism community websites; book and buy tickets through the official website or APP of the scenic spot, online travel website, Taobao, etc.; through the official website of the scenic spot or Ctrip, etc.; booking accommodation on travel websites; and inquiring about recommended food through Dianping and other platforms.

3.3. Analysis of Tourists' Consumption Preference. Online travel notes are travel experiences and experiences recorded by tourists after their trips based on their memories. The sights and travel activities frequently mentioned by tourists in travel notes indicate that they have left a deep impression on tourists or unforgettable travel experiences. Such experiences are often positive. Therefore, the greater the frequency of tourist attractions or tourist activities, the higher the popularity of the tourist attractions or tourist activities, and the more tourists prefer them. Based on this, this study filters out the tourist attractions that tourists prefer based on the word frequency of the names of the attractions, as shown in Figure 4.

The preference of tourist attractions based on word frequency screening shows a typical long-tailed curve distribution from high to low. The trend of the curve changes can reflect the uniformity of the heat distribution of tourist attractions. The faster the curve changes and the greater the amplitude, the more uneven the heat distribution of tourist attractions and vice versa. It can be seen from Figure 4 that only a few high-fat spots are distributed in the head of the curve, and most of the low-fat spots are distributed in the long tail of the curve, of which Kashgar Old Town has the highest weight.

3.4. An Analysis of Tourists' Perception and Evaluation Behavior. Tourism image refers to an individual's overall perception or overall impression of a tourist destination, which not only affects the destination selection stage but also affects the behavior of tourists, including sharing behavior, revisit intention, and recommendation behavior. It is widely accepted that the image of a tourist destination is generalized into two interrelated dimensions: the cognitive dimension and the emotional dimension, of which the cognitive

Table 3: Keyword extraction and analysis of tourism information channels.

Keywords	Weight	Frequency of words
Raiders	104	250
Travel APP	58	80
Ctrip	51	90
Where to	43	90
Public comment	27	35
Online/Web	22	370
Travel notes	20	500
Baidu map	19	22
Hornet's nest	11	25
Website/official website	15	40
Baidu	7	30
Taobao	5	20

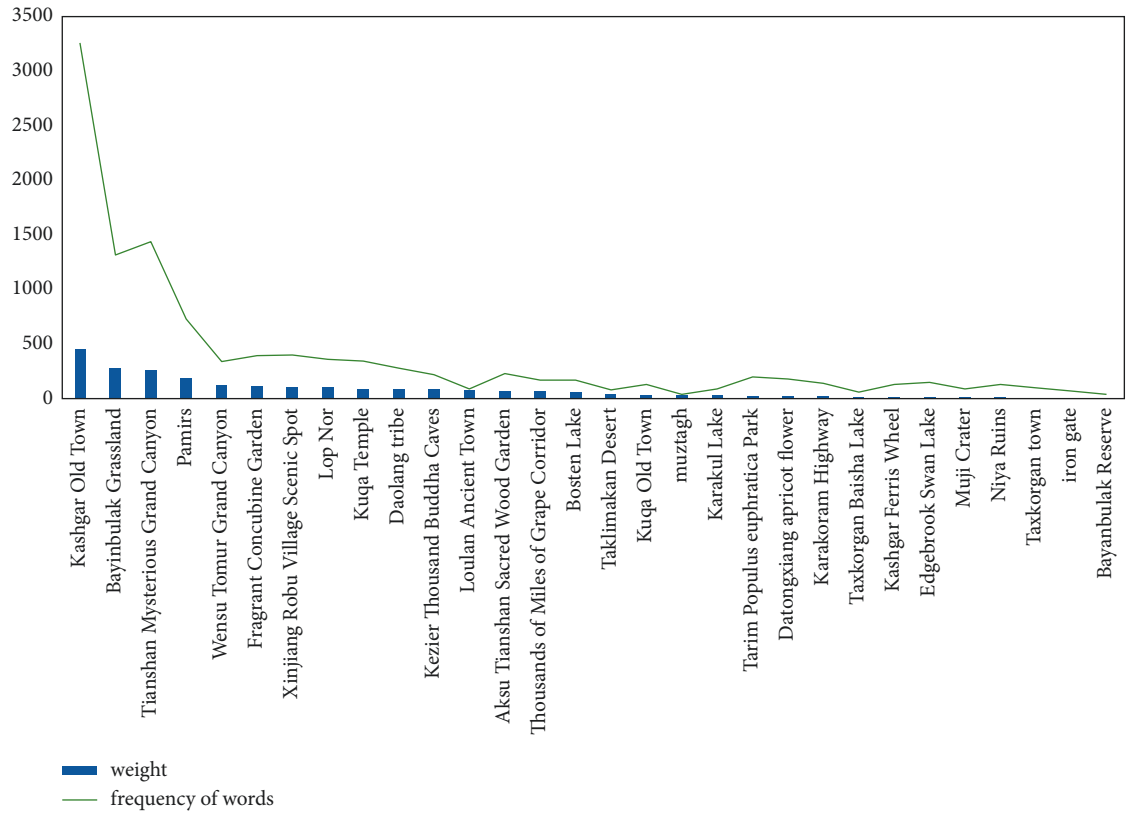


Figure 4: Tourist attraction preference statistics based on word frequency and weight.

dimension is the premise of the emotional dimension, both of which have a direct impact on the overall image of the destination.

This section analyzes the perception of tourists' destination image from the dimensions of intuition, positive emotion, and negative emotion. The specific dimensions are shown in Table 4.

Tourists' perception of the cognitive image of southern Xinjiang mainly focuses on natural tourism resources, special catering, climate environment, social environment, etc., and more of the praise of natural scenery and special food, which tends to be positive and emotional.

Through the extraction of emotional comment words and co-occurrence analysis, the positive and negative factors perceived by summer tourists are analyzed. Among the 30 emotional comment words extracted, high-frequency words such as "good," "like," "worthy," "beautiful," and "happy" reflect tourists' positive evaluation of the trip. "Enjoyment," "stimulation," "comfort," and "comfortable" are the positive emotional expressions of tourists who escape from the usual environment to the tourist destination for vacation and leisure and enjoy the novel natural scenery and comfortable climate environment. However, due to weather factors, tourists are worried that

Table 4: Analysis of intuition dimension, positive affective dimension, and negative affective dimension.

	Dimension	Keyword weight and word frequency	Co-occurrence word pair and word frequency
Tourism perception	Cognitive dimension (total word frequency 2488)	Beautiful, delicious, beautiful, spectacular, lung washing, clean, natural, beautiful, cool, mysterious, enthusiasm, blue sky and white clouds, open, beautiful, magical, full view of Tianchi, beautiful, fairyland, holy, spectacle, beautiful, fresh air, natural oxygen bar	Korean cuisine/northeastern cuisine—delicious (72); northeasterner/service staff/staff—enthusiastic (50); Grand Canyon—spectacular (45); mountain top—holy (46); environment—quiet (25)
	Positive (total word frequency 2806)	Feeling, good, like, worthy, lucky, enjoy, hope, comfortable, exciting, beautiful, comfortable, happy, anticipation, shock, comfort, excitement, gratitude, surprise	Worth—recommended/experience/share/see (134); taste/taste—good (68); scenery/environment—good (50); air/environment—comfort (26); nature/travel/food (26); weather—nice (9)
	Negative (total word frequency 1087)	average, bad, regret, worry, can't see the scenery, missed, disappointment, disappointment, regret, tired, dangerous, crowded	Weather/luck/road condition—bad (44); no view/stay—regret (29); worry—no view/weather/time (27); scenic/crowd—crowded (24)

they "cannot see what they want to see," so they have emotional expressions of "regret," "disappointment," "missing," and "gap." In addition, due to the excessive passenger flow in the peak season, tourists also feel "dangerous," "crowded," and "tired."

Tourism consumption is a typical hedonic consumption, experience consumption, the pursuit of feeling, fun, and pleasure, and tourism satisfaction mainly comes from experience, especially emotional experience. Rojas and Camarero believe that emotion is the determinant of satisfaction, and emotion has a moderating effect on satisfaction; Bosque believes that positive emotion and negative emotion are the antecedent variables of satisfaction, and loyalty is the dependent variable of satisfaction. Therefore, this study measures tourists' post-tour satisfaction evaluation from the two aspects of emotional tendency and behavioral willingness (i.e., loyalty).

To understand each tourist's sentiment evaluation of tourism, this section conducts sentiment analysis on text data based on document granularity and word granularity. First, the positive or negative sentiment words in each document are identified and scored through the sentiment dictionary, and then, the scores of all sentiment words in the document are aggregated and summed to obtain the final sentiment score. After the sentiment score statistics, the average sentiment score of the 500 travel notes is 46.73 points, of which the average positive (i.e., positive) sentiment score is 87.04 points, and the negative (i.e., negative) sentiment score is −40.31 points. Different emotional intensities are measured according to the sentiment score: positive affect (5, +∞), negative affect (−∞, −5), and neutral affect [−5, 5], where positive affect can be divided into high (25, +∞), moderate (15, 25], and moderate (5, 15); negative affect can be divided into high (−∞, −25), moderate [−25, −15), and moderate [−15, −5).

The statistical results in Figure 5 show that 91% of tourists have a positive emotional orientation towards the travel experience, of which 65% of tourists have the highest degree of positive emotion, 17% are moderate, and 18% are average. Only 1% of tourists have negative emotional inclinations, and all of them are highly negative, which shows

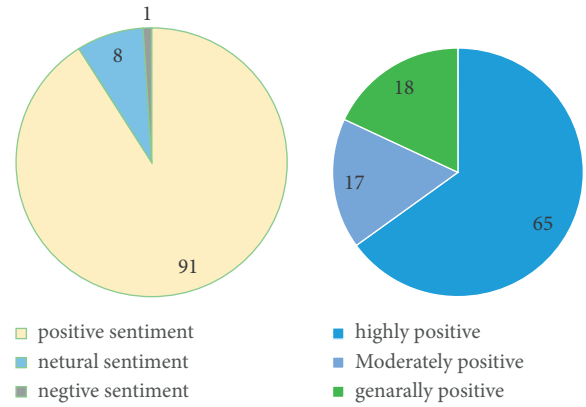


Figure 5: Statistical chart of emotional tendency analysis of travel notes.

that most tourists have positive and positive emotional inclinations towards the tourism experience in southern Xinjiang.

3.5. Behavioral Analysis-Based Tourism Development Strategies. Based on the two factors of geographic location and climate, the domestic primary market is the Bohai Rim Region with Beijing-Tianjin-Hebei as the core. This region has a developed economy, long high-temperature period in summer, and strong demand for summer vacation. The secondary market is the provinces and cities around scenario spots, and the travel distance in this area is short and suitable for short-term self-driving tourism products; the tertiary market is the Pearl River Delta region with Shanghai as the core, and travel in this area is long distance, developed economy, long high-temperature period, suitable for multi-destination in-depth tourism, and "airplane + other" combined transportation products; the fourth-level market is other domestic provinces and cities.

Relying on geographical advantages, "geo-tourism" is developed. Geo-tourism is developed, and characteristic border tourism products are created. Infrastructure construction is improved, tourism products are innovated and developed, the

tourism industry chain is extended, and the "trade zone + tourism zone" construction is used to reduce travel barriers for tourists and promote the development of tourism.

Unique attractions are relied to develop "themed tourism." Relying on the unique attractive elements of tourist destinations, we should develop characteristic-themed tourism products that meet the needs of different tourist groups: first is relying on the advantages of climate and environment and the second is to develop "food tourism" routes based on local catering characteristics. Finding, experiencing, and enjoying food have increasingly become an important tourist motivation.

Local cultural resources are relied to develop "cultural tourism". Cultural tourism resources should be developed in the development of tourism resources, and cultural tourism forms such as historical exploration and folk customs should be developed using the unique local Korean culture, Kanto culture, national culture, and cultural resources and landscape resources. While satisfying the basic needs of tourists for sightseeing and taking pictures of beautiful scenery, it also develops tourism products with high participation and strong experience of tourists such as folk experience and adventure exploration, prolonging the length of tourists' stay, and satisfying tourists for summer vacation and vacation leisure purposes.

We should focus on making up for the shortcomings of infrastructure services, improving transportation service facilities, and improving the level of tourist reception services during peak seasons. First of all, in terms of tourism transportation, the accessibility of transportation is improved, especially the transportation accessibility outside the scenic spots, and the management of the order of the tourism transportation market is strengthened. Secondly, in terms of tourism commodities, there are many types of special products in tourist areas, but only blueberries and ginseng are mentioned in the travel notes. This shows that other special products cannot attract tourists' attention and desire to buy. The development and packaging of special products should be strengthened, as well as publicity. Finally, a connection mechanism for tourists in peak seasons is established to improve the quality and level of reception services for tourists in peak seasons, thereby reducing the safety risk of crowd congestion and the negative emotions of tourists.

Cooperation between cities and counties in tourist attractions in tourist areas is strengthened. Scenic spots and their management departments should strengthen the cooperation between core scenic spots and surrounding scattered scenic spots, establish a scenic spot cooperation mechanism, integrate scenic spots in tourist areas, extend outward from the core scenic spots tour routes, and implement a combined ticket or pass system in large tourist areas. The integrated development of regional tourism is promoted. Tourism cooperation outside tourist areas is strengthened. The tourism management department should strengthen the tourism linkage and cooperation with the tourism departments of other cities and jointly establish summer escape and border tourism routes or tourism strategies across the three northeastern provinces, to achieve a win-win situation for all parties.

Classified Internet marketing based on tourist segmentation: according to the degree of network involvement, tourists can be divided into high involvement and low involvement; according to the degree of experience, tourists can be divided into deep experience, moderate experience, and shallow experience; and according to age, tourists can be divided into teenagers, youth, middle-aged, and elderly; and according to the place of residence, tourists can be divided into the urban type and rural type. Different market segments have different motivations for online consumption, online consumption preferences, and online consumption perceptions and evaluations. Classification of online marketing strategies should be formulated according to the consumption habits and travel attitudes of different groups. For example, travel groups with high Internet involvement should cooperate with well-known and popular travel websites or online opinion leaders to increase the popularity of scenic spots in this area.

The factors that cause tourists' "dissatisfaction" are the concentrated congestion of tourists in peak season, which not only brings adverse effects on tourists' travel experience but also on the ecological environment of scenic spots. The best way to solve the problem of congestion is to accurately predict the flow of tourists, understand the spatial behavior of tourists, and implement tourist diversion.

There are still many deficiencies in this study, which need to be improved and further studied in future research. It is in terms of data: due to the limitation of data acquisition, 500 travel notes were obtained after screening, and the amount of data is not particularly large. In the future, more channels and methods will be explored to collect data, so that the reference value of the data will be more perfect; due to the influence of the personal privacy protection policy, personal information such as the age and knowledge level of the travel journal publisher cannot be accurately obtained, so the representativeness of the travel journal sample cannot be accurately measured; the travel journal includes text and image data. This article only studies the form of text data. The next step is to conduct research on the form of image data. In the future, data mining algorithms such as cluster analysis, topic identification and classification, and dependency modeling can be further studied and applied to in-depth study of tourist behavior patterns and mechanisms, as well as the establishment of an intelligent analysis system for tourist behavior based on UGC data.

4. Conclusions

By consulting, summarizing, and analyzing a large number of related literature studies on tourism consumer behavior, tourism big data, text data analysis, etc., this study constructs a thinking framework for the research of tourism consumption behavior based on UGC text data and conducts research on the basis of this thinking framework. We use train browser, NLPIR, and other software packages to crawl, preprocess, and travel sample data and use the word frequency analysis, co-occurrence analysis, content analysis, sentiment analysis, network analysis, and other

methods to analyze the characteristics of tourists and decision-making. According to the behavior analysis, tourism management departments, and tourism enterprises, the reform and promotion of tourism marketing strategies and the improvement of tourism product and service quality were done. This study proposes the development strategy of southern Xinjiang summer vacation tourism in three aspects to improve the management mode of tourism destination. Finally, the conclusions of this study are as follows:

(1) This study uses ten days as the unit to count the travel time and proposes to use ten days to months as the tourism cycle to develop tourism, refine product design, and adjust and allocate tourism resources.

(2) Based on the "push-pull" theory, this study establishes a "push-pull resistance" tourism decision-making model. Through word frequency analysis, co-occurrence analysis, and content analysis, it is concluded that "escape" and "seek" are the internal driving forces that affect tourists' travel decisions. Among them, companionship and experience are the main travel motives. Tourism enterprises should develop and design in-depth and diversified experience products to create unforgettable travel experiences, thereby stimulating tourists' travel motivation: "climate suitability," "attraction of scenic spots," and "scenic spots." "Uniqueness" is the pulling factor for tourists to choose the four destinations in southern Xinjiang. Scenic spots can use the pulling factor to create tourism IP and create "selling points" to attract tourists; in addition, the changeable weather and inconvenience of transportation have become obstacles for tourists to travel, so scenic spots should focus on making up for shortcomings and improving the quality of tourism services.

(3) Through network analysis of scenic spots, it is found that there are three tourist routes preferred by tourists, namely short-distance travel and long-distance travel, and long-distance border travel. Tourism management departments and scenic spots should strengthen tourism cooperation between cities and scenic spots and develop theme tourism and diversified tourism routes such as geo-tourism and cultural tourism to achieve regional coordinated development.

(4) Based on the "cognitive-emotional" model, this study describes the tourism image from two dimensions, cognitive image and emotional image. Generally, tourists present a positive perception state. In addition, this study measures tourist satisfaction from the analysis of emotional tendencies and future behavior intentions (i.e., loyalty, including willingness to revisit and recommendation). Emotional state but the revisit rate is not high, indicating that the tendency of post-tour memories and evaluations to be effectively transformed into re-consumption behavior is weak. Scenic spots or research communities can conduct in-depth research on the effective needs of tourists, and at the same time, the positive electronic word-of-mouth effect is used to adopt "tourism". + new media," "tourism + UGC," and other marketing methods.

(5) Based on the research on tourism consumer behavior based on UGC data, this study can help scenic spots and other tourism enterprises understand the consumption preferences of different tourism groups, optimize tourism routes, develop diversified tourism products, improve the quality of tourism services based on the characteristics and laws of tourists' behavior, and further cater to market segments.

(6) This research provides a new idea for tourism scenic spots and tourism management departments to monitor tourists' behavior through big data analysis. A tourism big data analysis platform based on UGC data can be established to help relevant management entities to trace back the characteristics, changes in tourists' consumption behavior, and the impact on scenic spots. It can also help tourism websites to establish a personalized scenic spot recommendation system according to tourists' consumption preferences.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding this study.

Acknowledgments

This work was supported by the Key Project of Shanxi Federation of Social Sciences, Study on integrated development of rural tourism, SSKIZDKT2020166; Philosophy and Social Project of Higher Education Department of Shanxi Province, Study on the integration of rural tourism industry in resource-based areas, 2020W189; and Jinzhong University Maker Team Project, Study on the path of improving the quality and efficiency of rural tourism in Shanxi Province under rural revitalization strategy, jzxycktd2019015.

References

[1] H. Y. Shih, "Network characteristics of drive tourism destinations: an application of network analysis in tourism[J]," *Tourism Management*, vol. 27, no. 5, pp. 1029–1039, 2006.

[2] M. Caber and T. Albayrak, "Push or pull? Identifying rock climbing tourists' motivations," *Tourism Management*, vol. 55, pp. 74–84, 2016.

[3] M. Y. Wu and P. L. Pearce, "Chinese recreational vehicle users in Australia: a netnographic study of tourist motivation," *Tourism Management*, vol. 43, pp. 22–35, 2014.

[4] D. B. Klenosky, "The "pull" of tourism destinations: a means-end investigation," *Journal of Travel Research*, vol. 40, no. 4, pp. 385–395, 2002.

[5] S. S. Kim, C.-Ki Lee, and D. B. Klenosky, "The influence of push and pull factors at Korean national parks," *Tourism Management*, vol. 24, no. 2, pp. 169–180, 2003.

[6] M. J. Stone, "Deciding not to choose: d," *Tourism Management*, vol. 57, pp. 168–179, 2016.

[7] E. Sirakaya and A. G. Woodside, "Building and testing theories of decision making by travellers," *Tourism Management*, vol. 26, no. 6, pp. 815–832, 2005.

[8] A. Schroeder and L. Pennington-Gray, "The role of social media in international tourist's decision making," *Journal of Travel Research*, vol. 54, no. 5, pp. 584–595, 2015.

[9] J. Berbegal-Mirabent, M. Mas-Machuca, and F. Marimon, "Antecedents of online purchasing behaviour in the tourism sector," *Industrial Management & Data Systems*, vol. 116, no. 1, pp. 87–102, 2016.

[10] C. H. Chen, B. Nguyen, P. P. Klaus, and M. S. Wu, "Exploring electronic word-of-mouth (eWOM) in the consumer purchase decision-making process: the case of online holidays - evidence from United Kingdom (UK) consumers," *Journal of Travel & Tourism Marketing*, vol. 32, no. 8, pp. 953–970, 2015.

[11] Y. Zhang and Y. Peng, "Understanding travel motivations of Chinese tourists visiting Cairns, Australia," *Journal of Hospitality and Tourism Management*, vol. 21, pp. 44–53, 2014.

[12] T.-K. Hsu, Yi-F. Tsai, and H.-H. Wu, "The preference analysis for tourist choice of destination: a case study of Taiwan," *Tourism Management*, vol. 30, no. 2, pp. 288–297, 2009.

[13] S. S. Kim, D. J. Timothy, and J. Hwang, "Understanding Japanese tourists' shopping preferences using the Decision Tree Analysis method," *Tourism Management*, vol. 32, no. 3, pp. 544–554, 2011.

[14] I. P. Albaladejo-Pina and M. T. Díaz-Delfa, "Tourist preferences for rural house stays: e," *Tourism Management*, vol. 30, no. 6, pp. 805–811, 2009.

[15] R. C. Y. Chang, J. Kivela, A. H. Mak, and N. Mak, "Food preferences of Chinese tourists," *Annals of Tourism Research*, vol. 37, no. 4, pp. 989–1011, 2010.

[16] J. L. Crompton, "An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image," *Journal of Travel Research*, vol. 17, no. 4, pp. 18–23, 1979.

[17] H. Zhang, Y. Wu, and D. Buhalis, "A model of perceived image, memorable tourism experiences and revisit intention," *Journal of Destination Marketing & Management*, vol. 8, pp. 326–336, 2018.

[18] S. Baloglu and K. W. McCleary, "A model of destination image formation," *Annals of Tourism Research*, vol. 26, no. 4, pp. 868–897, 1999.

[19] C. F. Chen and D. C. Tsai, "How destination image and evaluative factors affect behavioral intentions?" *Tourism Management*, vol. 28, no. 4, pp. 1115–1122, 2007.

[20] E. Marine-Roig and S. Anton Clavé, "Perceived image specialisation in multiscalar tourism destinations," *Journal of Destination Marketing & Management*, vol. 5, no. 3, pp. 202–213, 2016.

[21] S. Choi, X. Y. Lehto, and A. M. Morrison, "Destination image representation on the web: content analysis of Macau travel related websites," *Tourism Management*, vol. 28, no. 1, pp. 118–129, 2007.

[22] E. Çakmak and R. K. Isaac, "What destination marketers can learn from their visitors' blogs: an image analysis of Bethlehem, Palestine," *Journal of Destination Marketing & Management*, vol. 1, no. 1-2, pp. 124–133, 2012.

[23] J. M. Hernández-Mogollón, P. A. Duarte, and J. A. Folgado-Fernández, "The contribution of cultural events to the formation of the cognitive and affective images of a tourist destination," *Journal of Destination Marketing & Management*, vol. 8, pp. 170–178, 2018.

[24] R. L. Oliver, "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of Marketing Research*, vol. 17, no. 4, p. 460, 1980.

[25] J. Wirtz, A. S. Mattila, and R. L. Tan, "The moderating role of target-arousal on the impact of affect on satisfaction-an examination in the context of service experiences," *Journal of Retailing*, vol. 76, no. 3, pp. 347–365, 2000.

[26] K. Chon and M. Olsen, "Functional and symbolic congruity approaches to consumer satisfaction/dissatisfaction in tourism," *Journal of the International Academy of Hospitality Research*, vol. 3, pp. 2–22, 1991.

[27] Y. Jin and X. Li, "Research on the development of Changbai mountain summer tourism from the perspective of consumer behavior," *E3S Web of Conferences*, vol. 292, Article ID 03038, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6065435, 7 pages
https://doi.org/10.1155/2022/6065435

Research Article

Application of Big Data Analysis on the Relationship between Career Delayed Gratification and Organizational Socialization Outcomes for New Generation Employees

Gaoxi Hu,[1] Ting Nie,[2] Tenfeng Qiu,[3] Guifeng Tian,[4] and Yaozhong Liu[5]

[1]*Guangzhou Institute of Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510023, Guangdong, China*
[2]*School of Business, Macau University of Science and Technology, Macau 999078, China*
[3]*School of Public Administration, Guangdong University of Finance, Guangzhou 510521, Guangdong, China*
[4]*Guangdong Construction Polytechnic, Guangzhou 510440, Guangdong, China*
[5]*Guangzhou Huashang College, Guangzhou 511399, Guangdong, China*

Correspondence should be addressed to Yaozhong Liu; liuyaozhong@jnu.edu.cn

Received 19 May 2022; Revised 19 June 2022; Accepted 23 June 2022; Published 19 July 2022

Academic Editor: Man Fai Leung

The effect of occupational delayed gratification is that it can prompt individuals to postpone short-term instant gratification in order to pursue more valuable long-term career goals when they encounter a choice of interests, which is very beneficial to the career development of the new generation of employees. Based on extensive literature review and interviews, this paper compiles a pretest questionnaire for investigation and proposes models and hypotheses. For the data of the pretest questionnaire, big data item discrimination analysis, Cronbach's alpha internal consistency reliability test, and exploratory factor analysis were used to eliminate unreasonable items. From the four aspects of self-evaluation, self-efficacy, self-esteem, and social support, it explains how to cultivate and improve the career delayed gratification of the new generation of employees, so that they can make the right choice between short-term and long-term benefits, setbacks, and success, thereby improving the organizational socialization outcome relationship.

1. Introduction

With the advent of the new economic era dominated by knowledge and information, the global financial crisis has intensified, and knowledge capital has gradually replaced financial capital and has risen to the status of strategic resources. Enterprises pay more attention to human capital investment and long-term returns. Therefore, humanity management in enterprise management is more and more dominant [1–3]. As the core of human capital investment, career management has become an important tool to achieve a win-win situation for enterprises and individuals coordinated development. Modern enterprises must also expand from the incentive mechanism centered on interests to the dual management of people-oriented and high satisfaction. As an important part of self-career management, occupational delayed gratification occupies a major position in enterprise management. Therefore, in order to occupy a seat in the fiercely competitive market, enterprises must not only compete for talents, but more importantly, strive to make organizations and talents develop together [4–6].

Undoubtedly, we should not only focus on satisfying the current development, but also need to look to the future and consider the long-term development goals of higher value and sustainable development. Therefore, occupational delayed gratification is particularly important from both an organizational and an individual perspective. Therefore, fundamentally understanding the psychological quality and career maturity of employees plays an important role in the effective management of enterprise employees.

At present, post-80s employees enter the workplace and gradually become the backbone of enterprises. Most of them

are well-educated and knowledge-based employees. The value they create for enterprises in terms of productivity and creativity is immeasurable [7–10]. Knowledge-based talents are a group who are particularly focused on their career growth and job achievement, and they attach great importance to their career development. However, the post-80s employees pay more attention to personal feelings and subjective perceptions in their work, with strong subjective will, weak self-selection ability, and lack of a clear sense of belonging. In addition to the endless temptations in the current Internet age, work pressure is gradually increasing. The turnover rate is also increasing, which poses new challenges to the management of enterprises [11–13]. How to effectively improve employees' satisfaction and loyalty to the enterprise, stimulate employees' work enthusiasm and initiative, cultivate and retain talents, and finally achieve a win-win situation for the enterprise and employees is one of the problems that enterprises urgently need to solve at present, a topic worthy of study and discussion.

This paper has carried out in-depth research on the occupational delayed gratification of knowledge-based talents, which has certain theoretical and practical significance. For individuals, through appropriate occupational delayed gratification, employees can be guided by more valuable long-term goals in the future in their career choices or work, and at the same time, it can also help the continuous development of personal comprehensive ability, so that the ultimate goal can be achieved. The possibility is greatly improved. For an organization, hiring employees with a higher level of occupational delayed gratification is conducive to achieving common goals between the organization and employees, improving the loyalty and core competitiveness of employees, and promoting the organization's future sustainable and high-speed development.

2. Theoretical Background

2.1. Dimensions of Organizational Socialization. Fisher (1986) proposed a four-dimensional structure of organizational socialization through empirical research [14]. He believed that the content of organizational socialization includes four aspects: the first aspect is to learn organizational-related knowledge. Fisher believes that there are many aspects of the organization that new employees need to learn, some of which are explicit, such as the specified work standards, reward and punishment standards, and personnel promotion standards; some are implicit, such as employees attitudes towards work, personal beliefs and work values, interpersonal expressions, and corporate culture. The second aspect is learning how to function in group collaboration. Similar to the previous description, what new employees need to learn after entering a new organization or team is also divided into two categories: explicit and implicit. The explicit content includes the names of members, the responsibilities and obligations to be performed in the work, and teamwork and communication skills; the implicit content includes the group's hidden standards and codes of conduct, and team culture. The third aspect is the study of work-related knowledge. The fourth area is personal learning and improvement. Organizational socialization can prompt individuals to face up to their true professional skills, work potential, and deficiencies.

Since then, Ostroff and Kozlowski in 1992 also further integrated the social content dimension of employee organization into four parts [15]: task, role, group, and organization. This is very close to Fisher's four-dimensional structure theory. Chao et al. conducted an empirical study on 594 American college graduates through a large number of literature review and analysis and used factor analysis to explore the content dimension of organizational socialization [16]. Chao et al. concluded that the content structure of employee organizational socialization covers six dimensions: job performance standardization, interpersonal relationships, organizational politics, language, organizational goals/values, and history. Taormina (1997) limited the research object of organizational socialization to employees in mainland China, Hong Kong, and Singapore. Through empirical research, the organizational socialization content dimension is constructed into four dimensions [17], which are the degree of training, the degree of organizational understanding, the degree of support from colleagues, and the expectations for the future. The four-dimensional theory proposed by Taormina et al. is more universal in the world, so it is widely accepted by scholars. Thomas and Anderson (1998) also agreed that organizational socialization is composed of four dimensions. Based on Taormina's research, they further proposed that the components of organizational socialization are role orientation, social environment, interpersonal resources, and organizational knowledge. The organizational social structure is shown in Figure 1.

2.2. Organizational Socialization Strategy and Employee Active Socialization Behavior. The organization's socialization strategy mainly refers to the organization as the main body of action, in order to assist new employees to integrate into the organization, speed up the process of adapting to the organizational environment, increase employees' sense of job security, reduce anxiety, and understand and master the necessary skills as members of the organization, methods of work attitude, and code of conduct. In the early research on organizational socialization by foreign scholars, the process of employee organizational socialization is usually regarded as the process of employees passively adapting to the new organizational environment, so the socialization strategy adopted by the organization plays a major role in this process. Even the research in this field has developed to the present, and the organizational socialization strategy is also one of the focuses in this research field. In recent domestic and foreign scholars' research on organizational socialization, experts and managers have further concluded the following.

The organizational socialization process of employees is not completely passive; on the contrary, new employees also assume an important position in the process of entering the organization. They actively adapt to organizational norms and culture by learning organizational culture,

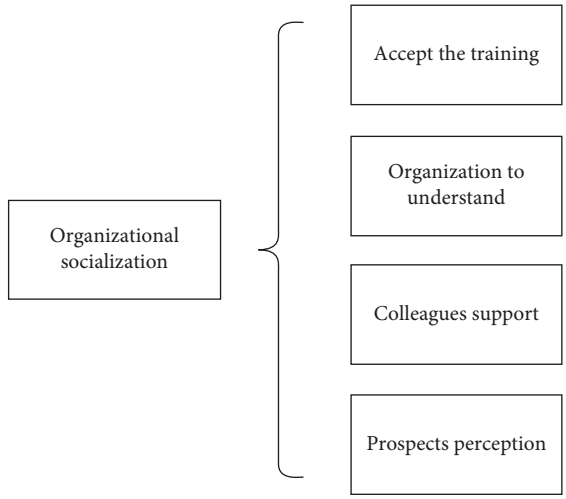


Figure 1: The organizational social structure.

understanding organizational information, and building good relationship networks with other members of the organization.

Louis (1980) regarded this active socialization behavior of employees as the behavior of new employees actively understanding organizational information in the process of adjusting their own positioning and changing roles, trying to improve their speed of adapting to the organizational environment, and mastering the work content. On this basis, Mitchell believes that the active socialization behaviors that employees take are generated by actively seeking opportunities for interaction between organizational and team members. The content of the behavior mainly covers consultation between members of the organization, talking in other members' offices or work areas, and joining parties and other social activities.

2.3. Measures of Organizational Socialization. For the measurement and evaluation criteria of organizational socialization strategies, the Organizational Socialization Strategy Scale, which integrates six strategies of organizational socialization, compiled by Jones (1986) has generally been supported by scholars. The scale integrates six representative strategies of organizational socialization, and each strategy contains five items, covering a total of thirty items.

Cable and Parsons (2001) improved and revised the Organizational Socialization Strategy Scale developed by Jones on the basis of empirical evidence. In 2005, Kim et al. agreed with the above point of view and used the revised scale in a subsequent empirical study with Korean employees as the subject sample. The study confirmed that the scale revised by Cable et al. has high measurement reliability. They constructed the other twenty-six items as a set (Ashforth et al., 1995), and the higher and lower mean of the total score of the whole set confirms the institutional degree of the organization's adoption of socialization strategies and can be achieved by providing systematic and sophisticated activities for the organization, reducing insecurity awareness among new recruits. The focus of the employee active social behavior measurement questionnaire developed by Ashford and Black includes collecting information (including four items, reliability 0.78), seeking feedback (including four items, reliability 0.92), socialization (including three items, 0.84)building a relationship network (three items, reliability 0.82), evaluation of superior-subordinate relationship (three items, reliability 0.82), work communication and work adjustment (three items, reliability 0.90), and so on.

3. Big Data Analysis Organization Socialization

3.1. Questionnaire Design. In order to improve the survey questionnaire and try to control the scientificity and accuracy of the data, we selected three different enterprises in Dalian to conduct a small-scale survey on the basis of preliminary qualitative analysis and purification of the items. The questionnaire items were adjusted or deleted through statistical analysis of small-scale sample data, which further improved the setting of the questionnaire items and formed a formal questionnaire and applied it to the later large-scale data research.

The formal questionnaire is divided into four parts: the first part, the basic information of the subjects, including gender, age, education, job type, position level, and company nature. The second part is the entry expectation part, which includes 4 items of job expectations, 4 items of team expectations, and 3 items of enterprise expectations, for a total of 11 items. The third part, psychological capital, includes 6 items of optimism, 4 items of hope, 4 items of self-efficacy and 2 items of resilience, a total of 16 items; the fourth part, organizational socialization, includes received training on 7 items, organization to understand 6 items, colleagues support 4 items, and prospect perception 4 items, a total of 21 items.

3.2. Statistical Analysis. The survey method adopts the form of questionnaire to collect data. A total of 600 paper questionnaires were distributed in the formal survey, 432 valid questionnaires were recovered, and the recovery rate of valid questionnaires was about 72%. At the same time, 300 electronic questionnaires were distributed, 160 valid questionnaires were recovered, and a total of 592 valid questionnaires were recovered.

In this survey, male employees surveyed accounted for 55.07% of the total number of employees, and female employees accounted for 44.93%, and the ratio of male and female employees was basically the same. Since this survey is not conducted for a certain gender group, and ideally the number of male and female employees accounts for half of the total number of employees, this gender distribution is basically consistent with the preconceived idea.

The majority of those surveyed were under the age of 30, accounting for 98.82% of all respondents, and 31–35. The age-old subjects accounted for 1.18% of the total number of subjects, which showed that most of the subjects in the study were young people. Most of the surveyed sample groups are college graduates or college graduates, accounting for 87.5% of the respondents, and 3.72% of the graduates. In Figure 2,

the points determined by the abscissa and ordinate represent the correlation between different variables.

This survey mainly involves state-owned enterprises, foreign-funded enterprises, Sino-foreign joint ventures, and private enterprises. Among them, the number of private enterprises accounted for 46.79% of the total number, ranking first. Sino-foreign joint ventures and state-owned enterprises followed, accounting for 25.84% and 16.39% of the total sample, respectively.

Among the industries involved in the sample group of this survey, the number of people in the service industry is the majority, accounting for 43.58%; the number of people in manufacturing ranked second, accounting for 31.08%.

Most of the job types are technical personnel, accounting for 63.68% of the total sample population, and the proportions of functional personnel and marketing personnel are similar.

Since the selected samples belong to new employees, grass-roots employees account for the vast majority in terms of position level, reaching 81.76%, and 13.85% of the subjects are middle-level employees, and no samples of senior employees have been obtained.

3.3. Reliability Analysis of Big Data. The Organizational Socialization Scale in the large sample test has a total of 21 items. We conducted reliability analysis on each item of the training organization's understanding of colleague support and prospect prediction and evaluated all the questions under the four dimensions of organizational socialization. The reliability analysis was carried out, and the results are shown in Table 1. In Table 1, the Cronbach's coefficient of the organizational socialization measurement scale is 0.947, and the Cronbach's coefficient of the training dimension is 0.946, and the Cronbach's coefficient of the organizational understanding dimension is 0.951. The Cronbach's coefficient of the colleagues support dimension is 0.872, and the Cronbach's coefficient of the prospect prediction dimension is 0.855, both above 0.7, indicating that the questionnaire has good reliability.

3.4. Correlation Analysis Based on Big Data. In the formal large sample test, Pearson's correlation coefficient was used for correlation analysis. Choose a two-tailed test to judge the correlation between variables. The correlation coefficient between the variables is positive or negative. The correlation coefficient is greater than 0, indicating that there is a positive correlation between the variables, and the correlation coefficient is less than 0, indicating that there is a negative correlation between the variables. The correlation coefficients between variables are different, indicating that the degree of correlation between variables is different. The correlation coefficient analysis is shown in Table 2.

The experiment tested the remaining 16 measurements, variables Kaiser–Meyer–Olkin Measure of Sampling Adequacy value of 0.865, greater than 0.8, suitable for factor analysis, and the Bartlett Sphericity test X2 is 465.408 (55 degrees of freedom, $P < 0.001$), which is significant, indicating that there are common factors among correlation

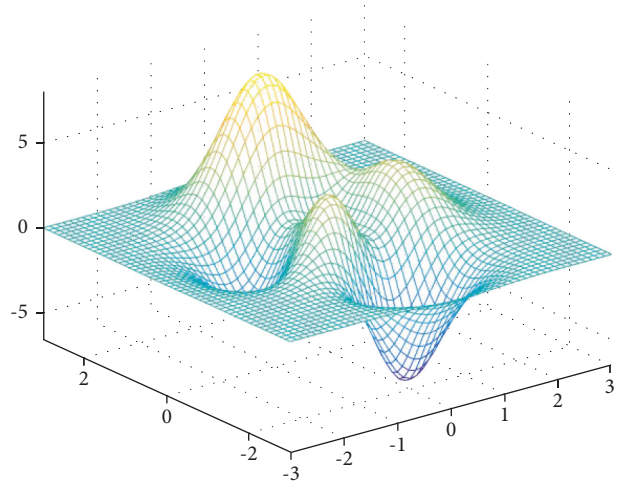


Figure 2: Correlation of factors.

matrices of the parent population, which is suitable for factor analysis on this basis. Above, the principal component factor analysis method is adopted to extract factors. As shown in Figure 3, 16 factors meet the requirements.

4. Occupational Delayed Gratification Optimizes Organizational Social Relationships

The effect of occupational delayed gratification is that it can prompt individuals to postpone short-term instant gratification in order to pursue more valuable long-term career goals when they encounter a choice of interests, which is very beneficial to the career development of employees. This section describes how to cultivate and improve employees' career delayed gratification from four aspects: self-evaluation, self-efficacy, self-esteem, and social support, so that employees can make the right choice between short-term and long-term benefits, setbacks, and success, thereby improving the entrepreneurial energy efficiency of employees.

4.1. Cultivating Employees to Have a Positive Core Self-Evaluation. Core self-evaluation is the corresponding evaluation and conclusion of self-worth given by an individual at the subconscious level. Positive core self-evaluation often focuses on the attention and discovery of one's own strengths. Most individuals who are accustomed to positive self-evaluation will face setbacks positively. To cultivate employees' positive core self-evaluation, it should be done from all aspects of education and teaching.

Expectation effect tells us that when we give more positive attention to the individual, give timely positive feedback to his achievements, and always face him with upward and positive force, the individual will perceive our expectations and will move towards him. We work hard in the desired direction and finally achieve self-realization. In order to improve the core self-evaluation power of employees in the process of starting a business in the future, the society should make good use of the expectation effect,

Table 1: Reliability tables for organizational socialization tables.

Name of scale	Dimension	Deleted value	Cronbach's alpha	Standardized Cronbach's alpha
Accepting the training	OP1	0.943	0.946	0.947
	OP2	0.941		
	OP3	0.941		
Organization to understand	HO1	0.941	0.951	
	HO2	0.940		
Colleagues support	SE1	0.942	0.872	
	SE2	0.944		
	SE3	0.944		
	SE4	0.942		
Outlook	RE1	0.941	0.855	
	RE2	0.942		
	RE3	0.943		

Table 2: Pearson correlation coefficient matrix of onboarding expectation, psychological capital, and organizational socialization variables.

	Organization to understand	Colleagues support	Prospects perception	Optimistic	Self-efficacy	Companies expect
Organization to understand	1	0.747	0.803	0.561	0.675	0.509
Colleagues support	0.747	1	0.849	0.721	0.726	0.632
Prospects perception	0.803	0.849	1	0.478	0.751	0.632
Optimistic	0.561	0.721	0.478	1	0.690	0.989
Self-efficacy	0.675	0.726	0.751	0.690	1	0.815
Companies expect	0.509	0.508	0.632	0.989	0.815	1

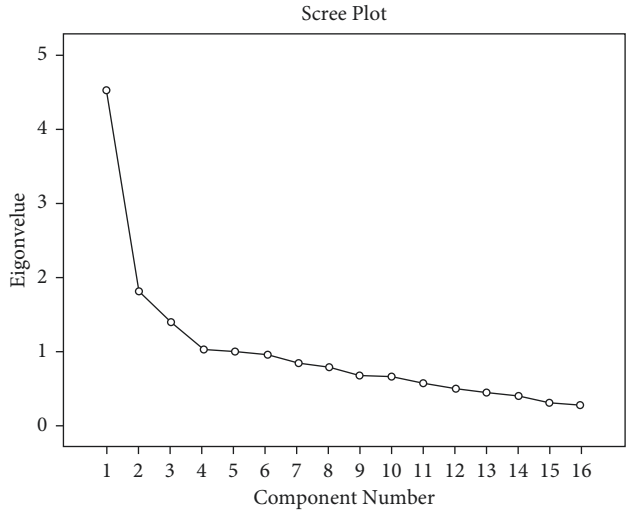


Figure 3: The factors meet the requirements.

actively encourage employees, and let them perceive the positive energy. Employees use this power to guide their own lives in the future.

In order to enable employees to internalize positive core self-evaluation into their own introspective power, they should be appropriately targeted when organizing employee activities, fully consider the existing psychological characteristics and growth characteristics of employees, and organize some activities that are more engaging and interesting, such as carrying out psychological games and trusting seats back to back, so that employees feel trust in the activities. The strength of persistence and the cultivation and improvement of willpower carried out through such activities, on the one hand, mobilize the enthusiasm of employees to participate, and on the other hand, they also convey that the positive energy of trust and persistence develops employees' potential in solidarity and cooperation, improves employees' self-confidence, and thus enhances employees' core self-evaluation.

At present, the combination of work and study is an important talent training mode in various technical colleges and universities, and its benefits in all aspects have been demonstrated and recognized by many parties. For the goal of improving the energy efficiency of entrepreneurship for employees, make full use of the factory environment and consciously cultivate employees to bear hardships and stand hard work. The willpower and ability to deal with problems are also issues that need to be paid attention to by technical colleges today. Among the employees who are about to enter the working environment, a considerable number of them grew up in a superior and coddled environment and cannot well adapt to the company's requirements for their work intensity and efficiency. Faced with a solution, they often choose to withdraw and give up, which is obviously not conducive to the cultivation of their entrepreneurial ability. Therefore, relevant enterprises should do the corresponding publicity work in advance, so that employees have a preliminary understanding of the formal process to be experienced and reduce the psychological gap of employees.

4.2. Improving Employee Self-Efficacy. Studies have found that individuals with higher levels of self-efficacy generally have a stronger motivation for achievement and are willing

to put in more effort and wait to achieve their goals. Individuals with strong achievement motivation tend to be more active when encountering difficulties and setbacks, and they are less likely to take adverse coping methods such as avoidance and withdrawal. Pay attention to cultivating the self-confidence of employees, such as grasping the bright spots of employees, giving them more affirmative evaluation and encouragement, and carrying out community activities to create opportunities for employees to show themselves in various ways; through clever design, let employees learn in the process and experience the joy of success, etc. Improving the self-confidence of employees, stimulating and promoting their formation of high achievement motivation, forming their own sense of responsibility for the future, and having enough confidence to support their actions will ultimately improve employees' self-efficacy.

4.3. Improving Employee Self-Esteem. Individuals with strong self-esteem are more willing to give themselves stronger support in self-action, thereby enhancing self-action. For employees who start a business, everything in the future is new and unfamiliar, and they need to mobilize all their strength and determination to face various difficulties encountered in the process of starting a business. Individuals with low self-esteem tend to give low self-evaluation, easily feel the oppression of difficulties before taking action, and reduce their efforts to start a business. Enterprises can start from the three pillars that constitute self-esteem (self-love, self-confidence, and self-view) and improve employees' self-esteem through caring and respecting employees, communication and trust, and positive guidance, example learning, and action training.

4.4. Helping Employees Learn to Use and Feel Social Support. Humans are creatures of social life, and people grow up with mutual support and help. For entrepreneurial employees, the power of social support cannot be ignored. Social support refers to both the material support given to the individual and the emotional support given to the individual. The existence of these supports enables individuals to obtain both external and psychological satisfaction, which is beneficial for individuals to improve occupational delayed satisfaction in the process of entrepreneurship and ultimately obtain higher entrepreneurial energy efficiency.

Improving employees' sense of social support can start with employees' interpersonal communication and perception. First of all, it is necessary to help employees establish the awareness of active communication and then train their communication behaviors and improve their communication skills by organizing life-like activities. In addition, to improve the perception of employees, when there is social support but the employees do not feel it, the support effect will be greatly reduced, so it is very necessary to cultivate the perception of employees. Through this kind of learning and training, employees can see their social support resources at a glance; they can always maintain the sensitivity of perception and utilization of social support forces, better improve their ability to withstand setbacks in entrepreneurship, and in the future work and entrepreneurship can also achieve a multiplier effect.

In a word, in the process of cultivating entrepreneurial employees, teachers should grasp all aspects that affect employees' entrepreneurial ability, start with improving their self-confidence and self-esteem, and ensure that employees have a positive core self-evaluation ability, so that they can better manage and utilize their own social support resources, improve employees' career delayed gratification, enable employees to make the right choice between short-term and long-term benefits, setbacks, and success, obtain high entrepreneurial energy efficiency, and lay a solid foundation for their future entrepreneurial development creating condition.

5. Conclusion

Based on the literature review and discussion on entry expectations, psychological capital, and organizational socialization, this paper proposes a hypothetical study on the relationship between entry expectations, psychological capital, and organizational socialization and collects data through questionnaires for empirical analysis to obtain entry expectations and analysis results of three dimensions and four dimensions of psychological capital and organizational socialization

The limitations of this paper are mainly shown as follows: First, the samples selected are all new employees who joined the company within two years. In order to ensure the diversity of samples, employees from different industries are selected as subjects as possible. However, due to the limitations of research conditions, the number of people in such industry distribution is not balanced, so the representativeness of research conclusions is limited. Secondly, some values of employees' psychological capital and organizational socialization support dimensions did not get the expected effect. The results of the questionnaire were limited by the fact that the researchers did not adequately monitor the actual situation of the employees filling out the questionnaire data, the limitations of traditional Chinese thinking, and the weak ability of the employees to withstand the environment. There are still some psychological barriers. Therefore, the data obtained from a few questionnaires issued by the company's leaders may be of limited representativeness. Thirdly, the new employee onboarding expectation in this paper is regarded as a value at a static point, and the development cycle of the onboarding expectation level and the degree of development and change in different time series are not studied.

In the future research, we plan to study the new employee management of a specified industry or enterprise type to make the research more pertinent. In addition, the same batch of samples is to be selected for data tracking, to study the interaction effect of adjusted entry expectations and

organizational socialization, as well as the moderating effect of psychological capital on the expectation gap.

Data Availability

The raw data supporting the conclusions of this article will be made available by the corresponding author, without undue reservation.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

The study was supported by Guangdong Philosophy and Social Sciences of the 13th Five-Year Plan Project in 2018, based on the perspective of occupational delay of gratification theory the post-90s generation employee research on organizational socialization outcomes (project no. GD18XGL19); the Scientific Research Project of Department of Education of Guangdong Province in 2018, research on the relationship between post-90s employees' occupational delay of gratification and organizational socialization (project no. 2018WTSCX236); the National Natural Science Foundation of China, cognitive neural mechanism of inequality avoidance and sense of power affecting cooperative behavior in repeated game (project no. 31970990); Guangdong Key Laboratory of Philosophy and Social Sciences, the Neuroeconomics Laboratory of Guangzhou Huashang College (project no. 2021WSYS002); and (Ministry of Education in China) Project of Humanities and Social Sciences, "Research on the Mechanism and Prevention of Cyberloafing in the Context of Digital China" (project no. 21YJC63010).

References

[1] D. G. Allen, "Do organizational socialization tactics influence newcomer embeddedness and turnover Academy of Management Best Conference Paper," *Academy of Management Annual Meeting Proceedings*, vol. 2004, 2004.

[2] N. Anderson, "Towards a theory of socialization impact: selection as pre-entry socialization," *International Journal of Selection and Assessment*, vol. 9, no. 1/2, pp. 84–91, 2001.

[3] E. C. Andrea, A. C. Criffin, and C. Srikanth, "Newcomer and organizational socialization tactics: an interactionist perspective," *Human Resource Management Review*, vol. 10, no. 4, pp. 453–474, 2000.

[4] E. C. Andrea, "Newcomer and organizational socialization tactics: an inter actionist perspective," *Human Resource Management Review*, vol. 10, pp. 453–473, 2000.

[5] B. E. Ashforth and A. M. Saks, "Socialization tacties: longitudinal effects on newcomer adjustment," *Academy of Management Joumal*, vol. 39, no. l, pp. 149–175, 1996.

[6] B. E. Ashforth, A. M. Saks, and R. L. Lee, "Socialization and newcomer adjustment: the role of organizational Context," *Human Relations*, vol. 51, no. 7, pp. 897–926, 1998.

[7] J. B. Avey, F Luthans, R. M. Smith, and N. F. Palmer, "Impact of positive psychological capital on employee well-being over time," *Journal of Occupational Health Psychology*, vol. 15, no. 1, pp. 17–28, 2010.

[8] J. B. Avey, F. Luthans, and C. M. Youssef, "The additive value of positive psychological capital in predicting work attitudes and behaviors," *Journal of Management*, vol. 36, no. 2, pp. 430–452, 2010.

[9] J. B. Avey, J. Patera, and B. West, "The implications of positive psychological capital on employee absenteeism," *Journal of Leadership & Organizational Studies*, vol. 13, no. 2, pp. 42–60, 2006.

[10] B. J. Avolio, L. Gardner, F. O. Walumbwa, E. Luthans, and D. R. May, "Unlocking the mask: a look at the process by which authentic leaders impact follower attitudes and behaviors," *The Leadership Quarterly*, vol. 15, no. 6, pp. 801–823, 2004.

[11] A. Bandura, *Self-efficacy: The Exercise of Control*, Freeman, New York, USA, 1997.

[12] R. M. Baron and D. A. Kenny, "The Moderator-Mediator Variable Distinction in Social Psychological Research," *J Pers Soc Psychol*, vol. 51, no. 6, 1986.

[13] T. N. Bauer, T. Bodner, J. S. Erdogan, D. M. Truxillo, and J. S. Tucker, "Newcomer adjustment during organizational socialization: a meta-analytic review of antecedents, outcomes, and methods," *Journal of Applied Psychology*, vol. 92, no. 3, pp. 707–721, 2007.

[14] T. N. Bauer and S. G. Green, "Testing the combined effects of newcomer information seeking and manager behavior on socialization," *Journal of Applied Psychology*, vol. 83, no. 1, pp. 72–83, 1998.

[15] T. N. Bauer, E. W. Motrison, and R. R. Callister, "Organizational socialization: a review and direetions for future research," *Research in Personnel and Human Resource Managelnent*, vol. 16, pp. 149–214, 1998.

[16] E. Black, "Ashforth and alan M. Saks, socialization tactics: longitudinal effects on newcomer adjustment [J]," *Academy of Management Journal*, vol. 2, pp. 146–178, 1996.

[17] G. W. Bock, R. W. Zmud, Y. G. Kim, and J. N. Lee, "Behavioral intention formation in knowledge sharing: examining the roles of extrinsic motivators, social-psychological forces, and organizational climate," *MIS Quarterly*, vol. 29, no. 1, pp. 87–111, 2005.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2958338, 10 pages
https://doi.org/10.1155/2022/2958338

Research Article

Cartoon-Style Image Rendering Transfer Based on Neural Networks

Lei Wang

College of Arts and Humanities, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China

Correspondence should be addressed to Lei Wang; 4899@cumt.edu.cn

Received 17 May 2022; Accepted 11 June 2022; Published 16 July 2022

Academic Editor: Man Fai Leung

Cartoon rendering of images is a challenging nonphotorealistic image rendering task, which aims to transform real photos into cartoon-style nonphotorealistic images while preserving the semantic content and texture details of the original photos. Based on the understanding of the characteristics of cartoon images and analysis of the defects of established approaches, this paper improves existing methods. The convolutional neural networks' powerful image processing ability and attention mechanism are utilized to get more fine-grained image features, making the result of rendering more realistic. This paper mainly studies how neural networks can better process cartoon pictures, adjust parameters, and choose training methods. Furthermore, the article proposes a new solution of cartoon-style image rendering based on deep learning. The actual test results of real original images have shown that the model is suitable for cartoon-style image rendering.

1. Introduction

Cartoons represent humorous or satirical forms of painting that have a great influence on our lives. Classic cartoon images, such as Mickey Mouse and Donald Duck, have become an irreplaceable part of the post-1990s generation. Disney and Marvel comics are still loved by millions of fans around the world. Animation creates another world that a lot of people yearn for. Reading comics can relax personal mind, improve pupils' imagination, find their own ideas and spiritual sustenance, and see the future belonging to students.

Since computer technology has changed people's daily life, it has significantly impacted on comic creation as well. The research on computer graphics has always been one of the hot fields in the domain of computer science, and its major content includes the representation method of graphics in the computer, the calculation of graphics, and the principle and algorithm of display. In 1962, E. Sutherland from MIT Lincoln Laboratory used the term "Computer Graphics" for the first time in his paper, proving that interactive Computer Graphics is a feasible and useful research field [1, 2]. Thus, Computer Graphics as a new branch of independent science was confirmed. At the beginning of the birth of computer graphics, pioneers were mainly committed to using computer technology to better express the natural world and to truly present the picture seen by bare eyes in the computer system.

At first, scholars used computers to imitate the output of early cameras and gradually began to simulate the reality of images. Computer graphics began to pursue real pictures. With the advancement of technology, the term "Photo-realistic Rendering" has been used to describe a range of techniques that mimic the output of cameras [3]. In pursuit of realism, early work studied the process and characteristics of light reflected off surfaces and then explored the equilibrium solution of light reaching various surfaces when reflected in the environment. Overall, realistic rendering is usually based on the scene and the physical model of objects in the scene. It stimulates colors, textures, and other attributes of objects in the scene in order to keep behaviors and characteristics consistent with colors, light, and shadow of nature as far as possible [4, 5]. Therefore, the computer system presents scenes complying with "real" pictures of human cognition. This pursuit of realistic feelings has greatly promoted the development of computer graphics. Realistic

rendering technology has been widely used in industrial fields such as CAD and postprocessing of film and television. Meanwhile, it has directly drawn forth the industrial field of "CG rendering" [6].

Rendering technology is widely considered to be inseparable from computer models [7, 8], but in fact, it also includes processing two-dimensional images. If the object is a data type of 2D images, then the corresponding technique is called image space nonphotorealistic rendering, whereas if the object is a data type of 3D model, then the technique is called 3D space nonphotorealistic rendering [9]. According to Gooch's classification theory of nonphotorealistic rendering technology, it can be divided into three categories: simulation of artistic media, interactive generation of users' images, and automatic generation of system images.

First and foremost, with the development of social media networks and the increasing popularity of the gaming market, there is a growing demand for personalized, nonrealistic images. For example, more youngsters tend to post photos or art works with personal features in short video platforms. In online games with social functions, many people also desire to have figures with their own characteristics, which require nonphotorealistic rendering. Furthermore, image processing based on deep learning has been investigated by more scholars, such as super-resolution reconstruction, missing image repair, colorization of black-and-white images, and the extremely popular AI face changing recently. Therefore, this kind of deep learning will definitely affect the blossom of comics.

2. Related Studies

In the early stage, nonphotorealistic rendering was primarily used as a supplement to photorealistic rendering [10]. Therefore, nonphotorealistic rendering technology for 3D models has many similarities with photorealistic rendering technology for 3D models in terms of implementation methods [11]. Nonphotorealistic rendering techniques for cartoon-style 3D object space are also rare. Nonphotorealistic rendering is generally defined as a technique that uses computers to simulate the rendering styles of various visual arts [12]. There are great differences between nonphotorealistic rendering and photorealistic rendering in the definition, purposes, and common methods. From the angle of specific rendering methods, photorealistic rendering mainly adopts the way of simulation, while nonphotorealistic rendering usually carries out the stylized process of the material. From the perspective of characteristics, realistic rendering results in the representation of the objective world, while nonrealistic rendering contains subjective expressions. From the point of the effect of the rendering works, realistic rendering will effectively reflect the real physical process, while works of nonrealistic rendering emphasize audiences' emotional cognition rather than visual correctness and usually attach some artistic processing closer to the art works in vision.

With the rapid development of computing technology, more and more researchers have devoted themselves to the study of nonphotorealistic cartoon-rendering algorithms. Lander et al. [13] focused on the relationship between textures and cartoon styles in images and proposed a method to realize cartoon coloring through texture mapping. Claes et al. [14] introduced a fast method to achieve cartoon-style rendering in their paper and smoothed the boundaries in the image during rendering. Cartoon rendering methods based on traditional image processing technology are generally nonparametric methods. Rosa [15] proposed an edge fusion method, which first converts images from RGB color space to Lab color space and performs gradient filtering on the L channel alone to obtain maps based on the edge gradient of images. At the same time, the L channel is smoothed by grayscale quantization in Lab space. Finally, the edge gradient image is fused with a smoothed image in RGB space, and the resulting image has smooth color blocks and prominent edges, with some characteristics of cartoon style.

Raskar et al. [16] offered a method using image depth and edge information in 2004, which takes advantage of multiple flashlights to shoot images to capture geometric information in the images. This approach separates textures from the image to help nonphotorealistic rendering. Winnemoller et al. [17] in 2006 tried to extract the contour of the image by using the bilateral filter combined with Gaussian difference filter, so as to render the image contour. Yan [18] designed a 2D cartoon-style-rendering algorithm based on scan lines in 2009, which can automatically complete cartoon-style rendering in real time. These methods dealing with traditional image processing technology can only process images with simple texture content and uncomplicated background, because they rely on image filtering and edge detection. Their effect is greatly affected by image content while the generalization ability is poor.

Gatys et al. embraced a texture model based on feature space of convolutional neural networks [19]. In this model, textures would be re-presented according to the relationship between feature maps in each layer of the network. When the object information is made clearer, texture extraction will also capture more content features of natural images [20]. Images are input to neural networks, and then, texture analysis is obtained through feature extraction of different convolutional layers and Gram matrix calculation. Next, the image white noise is input to calculate the loss function of textures in different layers to carry out texture synthesis. The document uses the first method of texture synthesis in the paper to transfer image style of oil painting style. The author uses the middle layer of neural networks for the reconstruction of the content, leaving the maximum pixel values of texture image fusion. Finally, it can get the final image containing different artistic styles of oil painting [21].

From the point of endless work in image processing in recent years, people have more research works and ideas in image processing. New breakthroughs have been made in image stylization, image recognition, video stylization, and other aspects [22, 23]. Researchers will introduce image generation and rendering image editing based on the deep learning into the cartoon style and make use of learning ability of deep neural networks to study the mapping relationship between natural images and cartoon-style images. Image transfer networks based on the deep learning can be

used for 2D image style of cartoon rendering tasks, and other methods used for image conversion can also be used for cartoon-style rendering. The former regards cartoon style as a kind of abstract artistic style, while the latter models cartoon-style rendering task as a domain-to-domain conversion task between domain of natural images and domain of cartoon-style images [24].

The existing nonrealistic rendering methods of cartoon style can draw cartoon-style pictures in specific scenes, but they cannot meet the changeable practical application scenes. An excellent cartoon-style nonrealistic rendering algorithm should be able to render natural images with any resolution or any object into a cartoon-style image while keeping the semantic content of the image as far as possible. Nowadays, people's living standard is getting better, and their pursuit of art is also becoming higher. Image stylization algorithm based on deep learning has become more worthy of exploration. In the future, there is a high possibility for deep learning and other related algorithms to achieve results of excellent image stylization. We combine convolutional neural network and attention mechanism to obtain a preliminary improved network and establish the matching and migration algorithm to extract the rendered image features in many aspects.

3. Convolutional Neural Network Image Rendering

3.1. Convolutional Neural Network. The traditional architecture of convolutional neural networks cannot process high-dimensional image input, and the number of neurons in each layer will increase with the square order of image size [25]. Second, the fully connected multilayer network structure does not take into account the spatial information in the image data, so convolutional neural networks are more suitable for processing image data. CNN is a kind of feed-forward neural network, which is stacked by hierarchical structures composed of multiple groups of convolution computation, activation operation, and downsampling. It is one of the most important algorithms for image processing in deep learning consisting of convolution layers and pooling layers. Furthermore, CNN seeks to extract local features, such as emotional word sequences, by limiting the acceptance domain of implicit layers to a local one. The most important advantage of convolutional neural networks is that they have a significant ability of representation learning with a small number of parameters and convenient training. They can automatically learn feature extraction modules for various tasks in a data-driven way. The feature graph extracted by convolutional neural networks has invariability of translation, which is the biggest reason why convolutional neural networks are suitable for processing image data.

The structure of convolutional neural networks is a simulation of the human visual nervous system. Convolutional layers of convolutional neural networks adopt sparse connection and weight sharing strategy, which greatly reduce the number of learnable parameters in each layer of networks, enabling them to train and predict data efficiently. At the same time, the small number of parameters can greatly reduce the risk of network overfitting. Sparse connection and weight sharing strategy are important reasons for the wide application of convolutional neural networks. At first, one-dimensional convolution was the most widely used type. In one-dimensional convolution layers, the convolution vector matrix is calculated through the width q of N filters and the convolution kernel. The weight matrix of filter Fn was expressed as $w \in R^{q\,d}$ to construct local features.

$$c^n = f\left(w^n \otimes X + b^n\right). \tag{1}$$

Here, $\otimes$ represents the convolutional operation of the vector, b is the bias unit, d is the dimension of the word vector, X is the feature extracted by the filter, and f represents nonlinear activation. The most outstanding ability of convolutional neural networks is their ability of feature extraction, which can adapt to learn feature extractors suitable for different tasks according to data and optimization objectives. Therefore, it has been widely used in various image-processing tasks.

Residual Net was proposed by Kaiming from the Microsoft Research Institute [26]. The structure of ResNet can pass down all the previous parameters and greatly improve the accuracy of the network. When it increases the number of network layers to 50, 101, and 152, good results are obtained, but when the network layers are increased to 1202, the results decrease. Because so many network layers are too thick, it will lead to overfitting. The core of Residual Net is to directly transfer each layer of information to the output, which will not cause degradation with the increase of network layers. On the basis of the increasing network layers, the accuracy rate can also be improved to a certain extent to ensure the integrity of extracted features.

With the proposal of further study of VGG network, VGG16 is still used in many problems of image feature extraction. The network uses $3 * 3$ small convolution kernel for feature extraction and $2 * 2$ maximum pooling layer for debasing dimension. The most commonly used ones are VGG16 and VGG19. The network can be roughly divided into five blocks, each of which consists of several convolutional layers in series, followed by a maximum pooling layer, and the last three blocks are fully connected with layers and softmax classifier. The network of VGG is mainly applied in image feature extraction, with $3 * 3$ filters and $2 * 2$ pooling layer. You can improve performance of the network by deepening it. However, the disadvantage of the network is that there are too many parameters needing more storage space accordingly.

3.2. CNN Image Processing. As convolutional neural networks produced good results in image recognition and classification, Gatys et al. applied CNN to the learning of artistic style features, extracted styles based on oil paintings, and combined textures of artistic images with real images, which ultimately obtained artistic painting results with photo content. In the feature extraction process of the style graph, a white noise image is also used, and the Gram matrix

of the target image and the style image is calculated for feature matching. The convolution loss of each layer of the two images can be expressed as follows:

$$E_n = \frac{1}{4\left(N_n^2 M_n^2\right)} \sum_{i,j} \left(G_{ij}^n - A_{ij}^n\right)^2, \tag{2}$$

where E represents the image loss result, A and G represent the feature output in layer n, and N and M represent the filter. The overall style loss function can be expressed as follows:

$$L_s(a, x) = \sum_{i=0}^{n} E_n W_n, \tag{3}$$

where a stands for style image, x represents the final generated image, and W is the weight of the NTH layer loss in the overall loss. Convolutional neural networks use stochastic gradient descent to compare the input content image and style image with the target image. The output result graph is constantly modified through training, and the process is cycled by the gradient descent method. At this time, the derivative of the above formula is obtained:

$$\frac{\partial E_n}{\partial F_{ij}^n} = \frac{1}{N_n^2 M_n^2} \left(G^n - A^n\right)_{ij} F^n. \tag{4}$$

In general, the graph output F of layer n is more than 0. If F is less than 0, the derivative is 0. In order to obtain the characteristic content image with the style image, a white noise image is given. Meanwhile, the style loss and content loss are defined as the total loss function, and the calculation formula is shown as follows:

$$L(p, a, x) = \alpha L(p, x) + \beta L(a, x). \tag{5}$$

Here, α and β are, respectively, the weight ratio of the content and the style of the target image. The result of the loss function is the difference between the final result image and the content and style image. It can be concluded that the loss function gradually decreases with an increasing number of iterations. When the loss function decreases to a stable value, the output target image can be regarded as the result of style transfer. Attention is used to screen out a small amount of important information from a large amount of information and focus on this important information, ignoring most of the unimportant information. The larger the weight is, the more it focuses on its corresponding value, that is, the weight represents the importance of information, and value is its corresponding information. Therefore, we skillfully use the attention mechanism to better obtain the image features and reuse these features in the rendering image, so that the final result is more consistent with the characteristics of human painting.

The strength of CNN lies in that its multilayer structure can automatically learn features, even features of multiple levels. The shallow convolution layer has a small perception domain and can learn features of some local regions, but the deeper layer has a larger perceptual domain and can learn more abstract features. These abstract features are less sensitive to the size, position, and orientation of objects, conducive to improve recognition performance. As Figure 1 shows, the right side is clearly a collection of pixels from many regions. In other words, the image on the left is segmented and the image on the right is aggregated according to different regions. In fact, the network is segmented by features and combined by recognition, and there may be many slow changing processes in the process, as shown in Figure 2.

3.3. Optimization of the Network. After the convolution operation, the pooling layer calculates local statistics to extract the most important features. This process allows the pooling layer to reduce the feature dimension, so as to reduce the calculation time and cost of CNN and prevent overfitting problems of the model. There are two common pooling methods: average pooling and maximum pooling. Average pooling is to only average feature points in the neighborhood. Maximum pooling means maximization of feature points in the neighborhood. The operation of pooling is very similar to convolution but different in the algorithm. The calculation of pooling operation is shown in formula (6). In the process of pooling, the internal values are not considered and only the size of the filter is concerned. In average pooling, images correspond to the position of the filter size and then average pixels that are not 0. This method is more sensitive to the characteristics of the background information and reduces errors of the estimator variance increases. Maximum pooling is the location of the corresponding filter on the image size, applying to a maximum of pixels. This method can get more outstanding characteristics of textures and clearer boundaries:

$$\begin{aligned} h_m &= \frac{1}{\left|N_m\right|} \sum_{i \in N_m} \alpha_i, \\ h_{m,j} &= \max \alpha_{i,j i \in N_m}. \end{aligned} \tag{6}$$

As a result of maximum pooling, the edges are more prominent and the lines are more obvious, and the blurred things in the background will be ignored, making the transfer of styles in the background not obvious. The evenly pooled image is smoother, and the background can transfer features of the original style image without blurring out the details. After the analysis of the cartoon image, the cartoon image often has prominent lines and the background does not need too much color block performance; therefore, this essay chooses maximum pooling in the style transfer of the cartoon image.

Gradient descent is often used in machine learning and artificial intelligence to make the model approximate to the minimum deviation. It is an optimization algorithm. Different optimization algorithms will affect the degree of model convergence and time of model training, so it is very important to choose the most appropriate optimization algorithm. The classical gradient descent method is shown in formula (7). The average loss of all training data is adopted to approximate the objective function:

Figure 1: Image area pixel matching.

Figure 2: An example of an intermediate process for image rendering.

$$L(\theta) = \frac{1}{M}\sum_{i=1}^{M} L\left(f\left(x_i, \theta\right), y_i\right). \tag{7}$$

For the small-batch gradient descent method, parameters are updated according to different blocks, which reduce the randomness of the descent. Since it is carried out in batches, the overall calculation amount will also be reduced and the speed will be accelerated. When parameters are updated, it is often difficult to choose the learning rate; thus, it is necessary to find a balance between accuracy and speed during setting the learning rate. When the learning rate is high, the speed will improve, but it will affect the accuracy. When the learning rate is low, the speed will decrease while the accuracy improves. When the second-order momentum appears, an optimization algorithm of the adaptive learning rate is generated. At the beginning of the training, a high learning rate is selected to speed up the training. After a period of time, when the model gradient descent reaches a certain degree, the learning rate can be reduced to increase the accuracy. Adam combines the first-order momentum with the second-order momentum and accordingly changes the learning rate of different stages by using their estimation of gradient. The learning rate of each stage has a fixed range, and the parameters are relatively stable.

4. Image Style Transfer

4.1. Convolution Image Matching. Rendering that used to mimic the appearance of the light now mimics actual behaviors of the light to make the graphics look more realistic; therefore, style rendering is all about matching relationships. For real-time style transfer based on block matching, iterative algorithm, and forward network can be used without training any model. The substitution of style block and content block can be adapted to parallel calculation. Two convolution layers, one two-dimensional convolution layer and one two-dimensional transpose convolution layer, are used at the same time. Through the average maximum value of channel direction, the convolution layer finally calculates the position of the corresponding replacement style block of each block in the content graph.

The extraction size of two-dimensional convolution is $3 * 3$, the step size is 1, and the normalized cross correlation is calculated. When calculating the average maximum value in the channel direction, the 2-d transpose convolution reconstructs the full activation by placing each patch of the best matching style at its corresponding spatial position. We use the attention mechanism to better find the best match, so that each color is included in the final rendering. The design is shown in Figure 3.

4.2. Adaptive Image Transfer. It can be seen from the result analysis of the traditional matching real-time style transfer, although the model does not need any training and its speed is greatly improved, it loses a lot of style textures and boundary line information [27]. The result is more like the fusion of two images. From the perspective of adaptive algorithm, ResNet is composed of many Residual Blocks and the basic component of each block is the Residual Unit, which is the combined operation of two convolution layers and skip-connection. Based on ResNet-101 architecture, an early cutoff is attempted in the feed-forward calculation of each Residual Block to reduce the redundancy of the model and save consumption of operation, which can support any style conversion. There is a conversion network and loss function calculation network. The generation of style image is faster than the real-time style transfer based on the block matching method.

The main parts of networks are divided into adaptive normalized transformation network and loss function network [28]. The network calculating the loss function will calculate the loss function of the content graph and the style graph during training. The style transformation network consists of three parts: encoding, adaptive normalization, and decoding. VGG network is used for coding, and the fourth convolutional layer is used to extract the features of content graph and style graph. Ada IN layer can normalize the input content graph. Here, it still makes use of the mean and variance of feature graph calculated by each channel matching content graph and style graph. Adaptive Instance Normalization is an improved algorithm based on Conditional Instance Normalization. In the feature space, the algorithm calculates the mean and variance of each channel of the input content image and style image and matches the normalized results. The reconstruction of content images in this paper is also selected from the fourth layer of VGG network. The feature expression based on this algorithm is shown in the following formula:

$$\text{add } IN(x, y) = \sigma(y)\left(\frac{x - u(x)}{\sigma(x)}\right) + u(y). \tag{8}$$

5. Experimental Results and Analysis

In order to verify the authenticity of the model, all the tested pictures in this essay are actually taken or created. Several groups of images have large differences in styles, mainly to verify the processing effect of the model for different styles of images and also to facilitate the adjustment of model parameters. The training data are mainly used in Microsoft Coco database and the standard black-and-white cartoon database for comparison [29]. The Coco data set mainly includes more than 80,000 real images, while the cartoon database primarily contains more than 1,000 black-and-white cartoon images. In the experiment, the size of each training image is adjusted to 256×256, the batch size is set to 5, and the iteration is 50 thousand times. About two cycles are given on Microsoft Coco training data, and about 160 cycles are given on the cartoon data set. It also uses Adam with a learning rate of 3000–3500.

5.1. Basic Image Matching. We need to verify the overall fluency of the model through basic image tests. The style transfer speed of block matching is slower than the speed of the general fast style transfer network. Also, most of the time is spent on the calculation of feature transformation, which accounts for 96% of the total calculation. Its core advantage over the general style transfer is to support images of any style. First, we should ensure that each large area has a corresponding rendering result, as shown in the figure. When the input style image resolution is higher, the conversion time will increase as the image increases, as shown in Figure 4. No matter what the final rendering is, there is some value to be created for the artwork, but the overall layout should not be completely changed.

As shown in Figure 5, the relationship between areas and lines can be clearly seen in Figure D. The boundary of the traditional iterative algorithm is not obvious. At the same time, the final image is more like the fusion of two images, rather than the transfer of styles, and the sense of lines between regions is pretty diluted. For illustration, the traditional method of drawing the body movements of the characters cannot have a perfect style. Although deep learning has solved the referred problems through training, it can also be said that the final rendering result in the figure is relatively stiff. Its content expressions basically prefer to image color segmentation and conversion, and the feeling of emotional style is gradually weakened. With comprehensive consideration, this method has high real-time performance and less loss of style information and boundary information, but lacks expressions of overall artistic emotion.

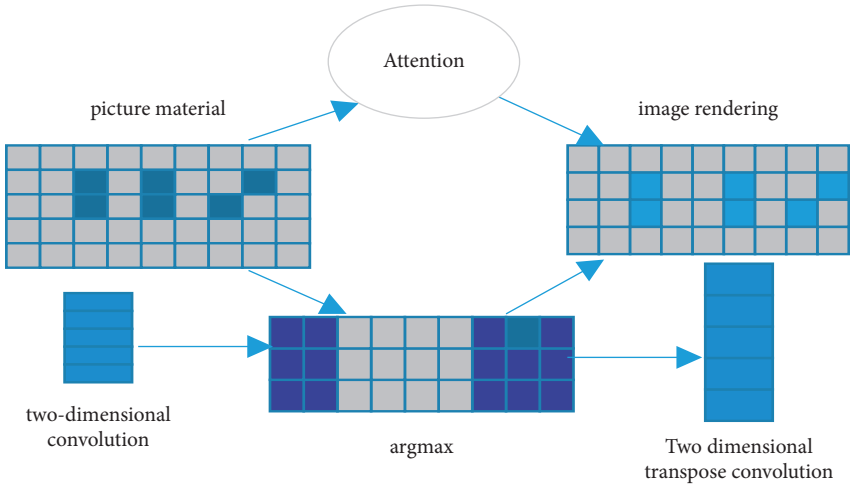


Figure 3: Convolutional image matching schematic diagram.

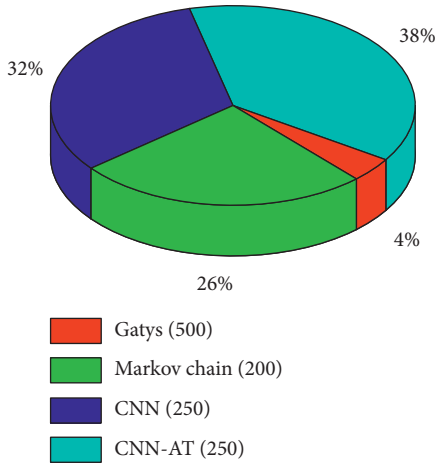
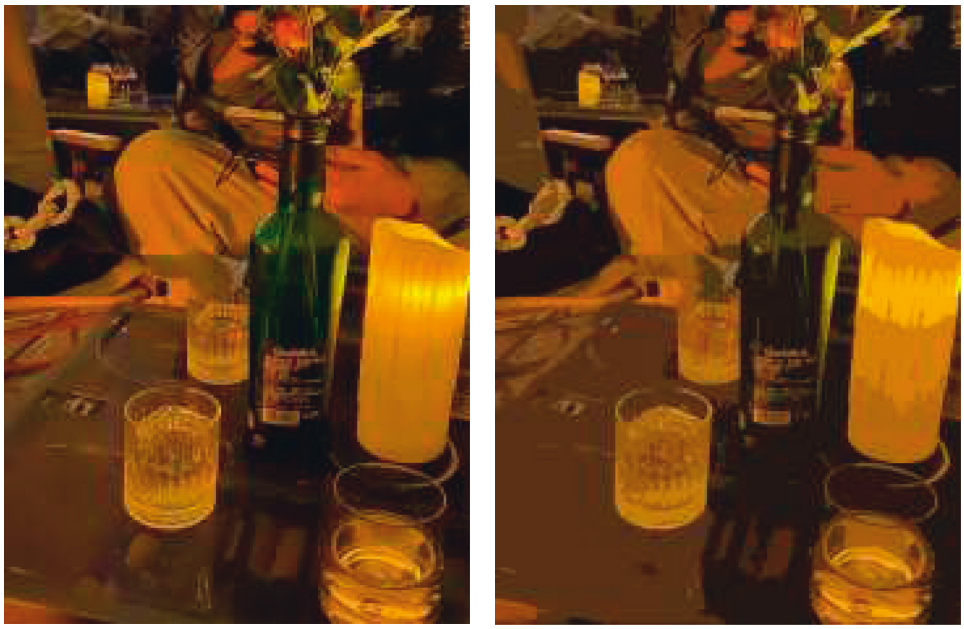


Figure 4: Matching the time scale diagram.

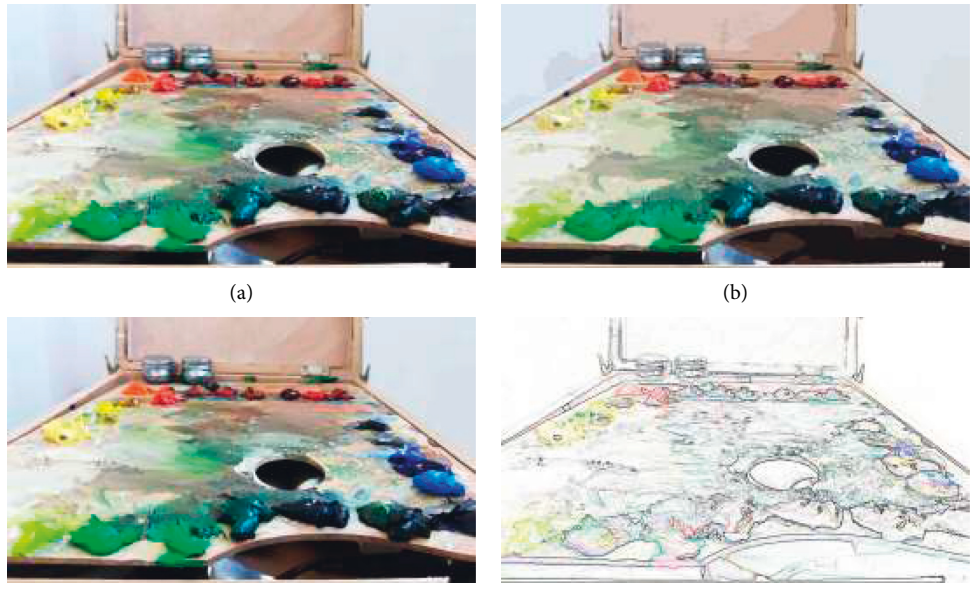


(a) (b)

Figure 5: Continued.

(c) (d)

Figure 5: Area line matching simulation.



(a) (b)

(c) (d)

Figure 6: Oil painting style rendering simulation.

As shown in Figure 4, CNN-AT indicates that convolutional neural networks use attention mechanism for matching. It is obvious that CNN's two methods take longer time. Because they require more features, the total rendering time is relatively long. However, the rendering results of this network algorithm are more diverse and artistic. Combined with Figure 3, CNN needs to extract and process colors, locations, and other features between American and Chinese objects, which contribute to obtain some potential special information, so as to make the final results diverse.

5.2. Adaptive Convolution Transfer Test. Rendering is not only the transfer of color outline but also the detailed description of environment and scenery. These descriptions are described in many ways to highlight the artistic image of a picture. Neural networks are more important to highlight the center and vividly express the connotation of the picture. Therefore, our improved adaptive convolution method is more profound for line extraction and more significant for textures. It has excellent performance in cross style transfer between various types.

As you can see from Figure 6, the lines are clearer. It also displays complex textures, including color textures in complex image areas. The higher the number of convolutional layers is, the better the extraction effect will be; while the lower the number of convolutional layers is, the effect will be like the fusion of two paintings. However, compared with the effect of traditional methods, when more textures exist in high level, there is a lack of style information in low-level extraction. Nevertheless, the background of cartoon image cannot appear too many color blocks and needs to retain the boundary, so it is expected to extract features from low-level convolution. Therefore, this method may create style types that we have not seen before and achieve extremely innovative applications in the practice of comic-style transfer.

6. Conclusion

With the rapid advancement of artificial intelligence and machine learning research, it could be assumed that the application of deep learning in the field of artistic creation has profound research values. To sum up, the paper studies experience and knowledge of utilizing neural networks to create art works and discusses the space of art brought by algorithms. It aims to further help other artists to understand neural networks and present art in fantastic ways. Moreover, the comic-style transfer algorithm based on deep learning can generate comic-style images with clear lines, simple backgrounds, and textures. Compared with results of different data sets with different parameters of different methods, this improved algorithm can generate better comic-style transfer effect in particular. Last but not the least, the method referred in this essay has achieved certain effects in the transfer of cartoon style. However, people cannot forget that cartoons in real life have educational significance, even including emotional expressions. Therefore, we need to constantly strive to expand research fields and achieve more details of neural network rendering results in right place.

Data Availability

The experimental data supporting findings of this study are available from corresponding authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. D. Foley, A. Van Dam, and S. K. Feiner, *Introduction to Computer Graphics*Addison-Wesley, Massachusetts, MA, USA, 1994.

[2] J. D. Foley, F. D. Van, A. Van Dam, S. K. Feiner, and F. J. Hughes, *Computer Graphics: Principles and practice[M]*, Addison-Wesley Professional, Massachusetts, MA, USA, 1996.

[3] K. Agusanto, L. Li, Z. Chuangui, and N. W. Sing, "Photorealistic rendering for augmented reality using environment illumination," in *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 208–216, IEEE, Tokyo, Japan, October 2003.

[4] J. Kronander, F. Banterle, A. Gardner, E. Miandji, and J. Unger, "Photorealistic rendering of mixed reality scenes," *Computer Graphics Forum*, vol. 34, no. 2, pp. 643–665, 2015.

[5] B. A. Barsky, D. R. Horn, S. A. Klein, and J. A. Pang, "Camera models and optical systems used in computer graphics: Part I, object-based techniques," in *Proceedings of the International conference on computational science and its applications*, pp. 246–255, Springer, Montreal, Canada, May 2003.

[6] A. Voloboy and V. Galaktionov, "Realistic rendering technologies in the problems of modern industry," *Scientific Visualization*, vol. 5, no. 4, pp. 31–46, 2013.

[7] M. O'Rourke, *Principles of Three-Dimensional Computer Animation: Modeling, Rendering, and Animating with 3D Computer Graphics*, WW Norton & Company, New York, NY, USA, 2003.

[8] L. Huang, *Research on Real-Time Rendering Technology of Special Effects in Game Engine*, University of Electronic Science and Technology of China, Sichuan, China, 2008.

[9] Y. Zhang, *Research and Implementation of Realistic Real-Time Rendering Technology Based on GPU*, University of Electronic Science and Technology of China, Sichuan, China, 2011.

[10] A. Lu, C. J. Morris, D. S. Ebert, P. Rheingans, and C. Hansen, "Non-photorealistic volume rendering using stippling techniques," in *Proceedings of the IEEE Visualization*, pp. 211–218, IEEE, Boston, MA, USA, October 2002.

[11] M. W. Chang and C. Collins, "Exploring entities in text with descriptive non-photorealistic rendering," in *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 9–16, IEEE, Sydney, NSW, Australia, February 2013.

[12] A. Kolliopoulos, *Image Segmentation for Stylized Non-photorealistic Rendering and Animation*, University of Toronto, Toronto, ON, Canada, 2005.

[13] J. Lander and G. Content, "Shades of disney: opaquing a 3d world," *Game Developer Magazine*, vol. 7, no. 3, pp. 15–20, 2000.

[14] J. Claes, F. Di Fiore, G. Vansichem, F. V. Reeth, and F. D. Fiore, "Fast 3D cartoon rendering with improved quality by exploiting graphics hardware," *Proceedings of Image and Vision Computing New Zealand (IVCNZ)*, pp. 13–18, 2001.

[15] Rosa, "Image non-photorealistic effect rendering algorithm based on edge fusion," *Microelectronics & Computer*, vol. 29, no. 10, p. 5, 2012.

[16] R. Raskar, K. H. Tan, R. Feris, J. Yu, and M. Turk, "Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 679–688, 2004.

[17] H. Winnemöller, S. C. Olsen, and B. Gooch, "Real-time video abstraction," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 1221–1226, 2006.

[18] X. Yan, *Research on 2d Cartoon Rendering Method*, Liaoning Normal University, Liaoning, China, 2011.

[19] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[20] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, https://arxiv.org/abs/1508.06576.

[21] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the*

IEEE conference on computer vision and pattern recognition, pp. 2414–2423, Las Vegas, NV, USA, June 2016.

[22] J. Lei, G. Li, S. Li, D. Tu, and Q. Guo, "Continuous action recognition based on hybrid CNN-LDCRF model," in *Proceedings of the International Conference on Image, Vision and Computing (ICIVC)*, pp. 63–69, IEEE, Portsmouth, UK, August 2016.

[23] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. Lempitsky, "Texture networks: feed-forward synthesis of textures and stylized images," *ICML*, vol. 1, no. 2, p. 4, 2016.

[24] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 1501–1510, Venice, Italy, October 2017.

[25] W. Gao, J. Yang, and C. Zhang, "Research on emotion analysis based on AT-DPCNN model," *Computer Engineering*, vol. 46, no. 11, p. 8, 2020.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[27] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer Science & Business Media, Berlin, Germany, 2012.

[28] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE international conference on computer vision*, pp. 1501–1510, 2017.

[29] T. Y. Lin, M. Maire, S. Belongie, J. Hays, and P. Perona, "Microsoft coco: common objects in context," in *Proceedings of the European conference on computer vision*, pp. 740–755, Springer, Zurich, Switzerland, September, 2014.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3973763, 7 pages
https://doi.org/10.1155/2022/3973763

Research Article

A Comparative Study of Cultural and Traditional Education in Primary and Secondary Schools in Developed Countries Based on the MOPSO-CD-DNN Model

Chunsen Hu

Hubei Business College, Wuhan 430070, Hubei, China

Correspondence should be addressed to Chunsen Hu; huchunsen@hbc.edu.cn

Received 6 June 2022; Accepted 28 June 2022; Published 12 July 2022

Academic Editor: Wenming Cao

In today's globalization, cultural and traditional education in primary and secondary schools has become the core of a country's future development, and how to improve the educational effect of cultural and traditional education in primary and secondary schools and find the development direction of cultural and traditional education has become the top priority. In response to this problem, this study proposes a MOPSO-CD-DNN hybrid prediction model, which introduces an optimization algorithm to optimize the parameters of the deep learning model. In this study, multiple benchmark models and evaluation methods are used for comparative research. The results show that the MOPSO-CD-DNN model has significant advantages in both prediction accuracy and prediction stability. Compared with other models, the prediction accuracy G value (average) is improved by 4.66%, 7.43%, and 9.25%, and the standard deviation (G value) is decreased by 0.001, 0.0502, and 0.0413, indicating its effectiveness and applicability to cultural tradition education. In addition, the introduction of the multiobjective optimization algorithm significantly improves the generalization ability of the model, and the prediction effect is significantly better than the single-objective optimization algorithm.

1. Introduction

Cultural tradition is the cultural spirit that originated from the life and practice of past dynasties and continues to this day in the process of continuous renewal and development [1]. It arises from the past but also exists in the present. As an organized social activity, education can be said to be a cultural phenomenon, the basic form of cultural inheritance and innovation, and the embodiment of the cultural function of education [2]. In the era of globalization, how does education cultivate modern citizens with the spirit of national culture, so that the national cultural tradition can be inherited and the national cultural literacy can be improved. For this issue, we need to broaden our horizons and effectively learn from the educational experience of other countries [3]. Therefore, this paper takes Japan and Singapore as the research objects and deeply analyzes how the two countries cultivate modern citizens who are condensed by cultural traditions through the education of cultural traditions in primary and secondary schools. [4, 5] Therefore, it is extremely necessary to study cultural tradition education.

Some scholars have analyzed [6, 7] the characteristics of school moral education in Japan and Singapore at the level of cultural traditions. The main manifestations are as follows: first, as an oriental cultural tradition, and influenced by Confucianism, moral education is placed in the first place. First place is in school education. Secondly, in the collision and exchange of Eastern and Western cultures, the reasonable elements of Western educational values are drawn to reconstruct the common values of the nation. Moral education not only adapts to national traditions but also integrates Eastern and Western cultural strengths. Third, pay attention to the comprehensive cultivation of people's quality, develop modern content and modern methods, and reflect the coordination in the conflict between traditional

spirit and modern consciousness. Some scholars have done a comparative study on moral education in Japan and Singapore and made a comparison on the basis of analyzing the goals and contents of moral education in Japan and Singapore. The difference is that Japan cultivates perfect personality and emphasizes the formation of the "Yawa soul" [8]; Singapore, because of the historical background of immigrant society and cultural diversity, aims to cultivate "Singaporeans," aiming to achieve the independence and development of a newly established country [9]. The similarities are that they have absorbed the essence of Eastern and Western cultures, and moral education ideas all emphasize loyalty to the country and serve the country [10]. Other scholars have analyzed the characteristics of ideological and political education in Japan and Singapore, which are as follows: advocating common values and enhancing the sense of identity [11]; cherishing and inheriting traditions, highlighting nationality [12]; comprehensive school, family, and society Collaboration; eclectic, integration of East and West; implement ideological and political education with moral courses as the main way, and other courses and activities also tap moral education factors [13]. Some scholars have made a different analysis of moral education in Japanese and Singaporean schools. In terms of guiding ideology, Japan is controlled by neoconservatism and neoliberalism, while Singapore is deeply influenced by Confucian ethics [14]; in terms of mainstream values, Japan cultivates the utilitarian values of "obedient and participatory" citizens, while Singapore pays attention to implementing I am a Singaporean common value [15]; in terms of methods, Japan pays attention to moral reasoning, while Singapore pays attention to moral practice; in the moral education environment, Japan supports the impoverishment of the life and culture of moral education, while Singapore emphasizes family, social moral education environment, and school moral education combined with [16]. In recent years, with the continuous development of deep learning technology, it has been widely used in various fields. The advantage of deep learning is that it learns step by step through multiple networks, extracts complex and effective features, and has higher prediction accuracy and generalization ability. Some researchers have used deep neural networks to predict the development of cultural education. The experimental results show that DNN has good prediction potential for time series [17]. Some researchers have found that the LSTM neural network can effectively capture the changes in time series in the study of the cultural comparison research of different countries [18]. At present, there are few studies on the application of deep learning in cultural and traditional education in primary and secondary schools. At the same time, it should be noted that the choice of parameters will also have an important impact on the prediction effect of the model [19]. At present, the optimization methods for model parameters mainly include cross-validation, grid search, and intelligent optimization algorithms. Some researchers [20] used the CSA algorithm to optimize the weights of the feedforward neural network, and the results showed that the prediction effect of the model was significantly improved, reaching 90.30%.

The above studies generally use a single-objective optimization method, but the prediction of cultural and traditional education is affected by many factors, not only the prediction accuracy but also the model stability should be considered. Zhang et al. [21] introduced the NSGA-III multiobjective algorithm to forecast the power load, and the results showed that the prediction performance of the NSGA-III optimized model was significantly better than that of the single-objective model. Representative multiobjective optimization algorithms mainly include PAES, SPEA2, and NSGA-II. Among them, MOPSO is a multiobjective optimization algorithm based on Pareto domination and external archiving mechanism, which has the advantages of good convergence characteristics and low time complexity. In a comparative study with other multiobjective optimization algorithms, it was found that MOPSO significantly outperformed methods such as NSGA-II, PAES, and micro DE [22]. To sum up, this paper combines MOPSO-CD and DNN to construct a comparative prediction model of cultural and traditional education in primary and secondary schools in developed countries, hoping to learn from the current cultural and traditional education in my country, so as to infiltrate cultural traditions into school education. In goal construction, curriculum construction, and teaching design, let cultural tradition education become an organic component of the school education system and cultivate primary and secondary school students who can inherit and carry forward Chinese cultural traditions and have cultural awareness and cultural spiritual heritage.

2. Building a Predictive Model

2.1. DNN Model. Deep learning has been widely used in the fields of text detection, speech, and face recognition, and natural language processing and achieved excellent results [23]. Compared with the traditional shallow machine learning method, it has the advantages of strong feature extraction ability, simple model structure, low training difficulty, and fast convergence speed. Compared with the traditional neural network, the deep neural network (DNN) inherits the structure of the traditional neural network; that is, it has a three-layer network (see Figure 1), in which the hidden layer reaches more than 6 layers [24].

In Figure 1, the DNN network structure layer can be mainly divided into three layers, the first layer is the input layer, the second layer is the hidden layer, and the third layer is the output layer, in which the layers are fully connected; that is, the first layer of any neuron in the n layer must be connected with any neuron in the $n+1$th layer, and the connection relationship can be expressed by the formula (1), and the formula (2) is the basic model form of the deep neural network.

$$s = \sum_{i=1}^{n} w_i x_i + b, \tag{1}$$

$$y_m = f\left(\sum_{i=1}^{n} w_{\mathrm{im}} x_i + b_m\right), \tag{2}$$

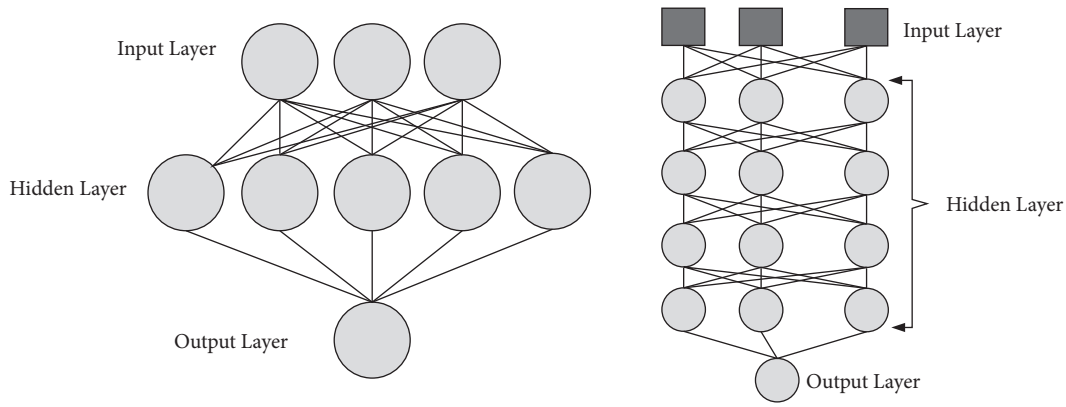


Figure 1: Traditional BP neural network and deep neural network structure diagram.

where w is the weight, b is the bias, x_i is the neural unit, and $f(\cdot)$ is the activation function. Common activation functions include sigmoid function and tanh function. Sigmoid functions are nonlinear functions. When the input data are small, the output is close to 0; as the input increases, the output is close to 1. *tanh* is a symmetric function centered at zero, and its model training process converges faster than the sigmoid function. This paper first selects appropriate samples and input features based on the characteristics and requirements of the comparison of traditional culture and education in primary and secondary schools in different developed countries, then trains the model, determines the relevant parameters of the deep network model, and selects an appropriate prediction model; finally, the test sample data are used. The predictive ability of the model is tested.

2.2. MOPSO Model. All particles fly at a certain speed in the search space and find the global optimal value through the currently searched optimal value. Particle swarm optimization is an iterative-based optimization algorithm. The system is initialized to a set of random solutions, and the optimal value is found through iteration. It is one of the generalized swarm intelligence methods for solving optimization nonlinear problems. The swarm search algorithm and the particle swarm algorithm have few parameters, and it is easy to fall into the problem of local optimality while the multiobjective particle swarm can be used to optimize more than one target parameter, and there are contradictions between the objectives, and there is no uniform measurement standard, and in the case of the unit of measurement, it is an optimization algorithm that can obtain a uniformly distributed noninferior solution set covering the entire search space with as little computing resources as possible.

When using the MOPSO algorithm to achieve target optimization, a set of random solutions (called "particles") will be obtained first, and each particle will have a specific location and initial velocity associated with it V_i. At the same time, according to the fitness determined by the objective function, the optimal position pid and the current position X_i under the current search progress will be determined. At the same time, each particle can obtain the optimal position pgd of the entire population. Therefore, the MOPSO algorithm can be regarded as a self-organizing behavior in which each particle transmits information in a certain space according to a certain rule or method and continuously corrects itself according to the deviation between the actual position and the optimal position.

2.3. MOPSO-CD Model. Multiobjective particle swarm optimization based on crowding distance (MOPSO-CD) is a heuristic evolutionary multiobjective optimization algorithm [25]. As a derivative model of MOPSO, its main idea is to use the strategy of archiving and retaining the elite solution, using the random selection method to select the elite from the external archive set as the global optimal solution of the particle, and at the same time introducing the dynamic nonuniform mutation operator, and finally, the goal of maintaining the diversity of particles, slowing down the convergence speed of the algorithm, and avoiding premature convergence is achieved. Studies have shown that this method is more efficient and practical than the classical method [26].

2.4. MOPSO-CD-DNN Hybrid Prediction Model. This paper uses the multiobjective particle swarm optimization algorithm based on crowding distance (MOPSO-CD) to optimize the parameters of DNN, analyzes the comparison of cultural and traditional education in primary and secondary schools in developed countries, and then predicts the cultural and traditional education of Chinese primary and secondary schools. The specific implementation process of the model is as follows:

Step 1: *Data selection and preprocessing*. This paper selects the experimental sample set according to the target principle of the same industry in the same period and selects the traditional culture education of Japan and Singapore as the main sample, and the rest as the positive sample. Then, the relevant data are standardized (Z-score) to eliminate the influence of the index dimension, and PCA is used to extract factors, and after eliminating redundant characteristic variables, a cultural education prediction model is constructed. In addition, in the data sample division, the

training set (70%) and the test set (30%) are randomly selected.

Step 2: *Model training and parameter optimization*. Using the training set data set, a deep learning model is used to train a cultural and educational development prediction model. At the same time, the MOPSO-CD algorithm is used to optimize the selection of DNN parameters. The optimized parameters are used to predict the cultural and educational situation of primary and secondary schools on the test set. Among them, the objective functions of the MOPSO-CD optimization algorithm are TPR, TNR, EPR, and ENR, with a total of 4 objective functions.

Step 3: *Model testing*. The trained MOPSO-CD-DNN traditional culture education prediction model for primary and secondary schools is tested on the test set, and the corresponding experimental results are obtained to verify the performance of the cultural development prediction model proposed in this paper.

Step 4: *Evaluation of traditional cultural education mixed prediction model*. A variety of model performance evaluation indicators are used to test the MOPSO-CD-DNN mixed traditional culture education prediction model. In addition, this paper uses three comparative studies to further examine the effectiveness and advancement of the MOPSO-CD-DNN model. First, through the comparison and selection of single models such as SVM and BPNN, the superiority of using the DNN model in the hybrid early warning model is demonstrated. Second, the applicability of the MOPSO-CD algorithm is verified by comparing the proposed hybrid early warning model with models such as NSGA-II-DNN. Third, the MOPSO-CD-DNN model proposed in this paper is compared with the SA-DNN and DE-DNN models, and the advantages of the multiobjective optimization algorithm in optimization efficiency and effect are investigated.

2.5. Evaluation of Traditional Culture Education Prediction Model. This paper uses the traditional classification accuracy to analyze the classification accuracy of the traditional cultural education prediction model and uses the evaluation index for unbalanced samples to evaluate the prediction accuracy of the traditional cultural education prediction model. The classification accuracy is the proportion of positive and negative samples that are correctly predicted to the number of positive and negative samples (TPR and TNR) and the overall accuracy (TR) [27]. The evaluation indicators for the classification of unbalanced samples are the geometric mean accuracy rate G and the minority class sample metric F. The geometric average accuracy G comprehensively examines the prediction performance of the model for two types of samples. If G is larger, it means that the model has higher accuracy in predicting both types of samples. Otherwise, the model accuracy is smaller. If F is larger, it means that the prediction performance of the model for traditional cultural education prediction samples is superior; otherwise, the prediction performance is worse [28]. The calculation formula of the evaluation index is as follows:

$$\begin{aligned} \mathrm{TPR} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \\ \mathrm{TNR} &= \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}, \\ \mathrm{TR} &= \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}, \\ G &= \sqrt{\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} \times \frac{\mathrm{TP}}{\mathrm{FP}+\mathrm{TP}}}, \\ F &= \frac{2 \times \mathrm{TN}/\mathrm{TN}+\mathrm{FN} \times \mathrm{TN}/\mathrm{TN}+\mathrm{FP}}{\mathrm{TN}/\mathrm{TN}+\mathrm{FN}+\mathrm{TN}/\mathrm{TN}+\mathrm{FP}}. \end{aligned} \tag{3}$$

Table 1: Comparison of the classification results of each model.

Model	TPR/%	TNR/%	TR/%
BPNN	57.81	57.50	57.69
SVM	65.79	56.06	59.61
DNN	60.00	59.09	59.62
DE-DNN	64.15	60.78	62.50
SA-DNN	66.67	62.26	64.42
NSGA-II-DNN	68.85	72.09	70.19
MOPSO-CD-DNN	72.88	75.56	74.04

2.6. Selection of Characteristic Indicators. Since there is no recognized standard for the characteristic indicators used in the traditional cultural education forecasting model, this paper draws on the relevant research to select the characteristic indicators and initially obtains the traditional cultural education forecasting system. Finally, four key indicators of survival, efficiency, ability, and value are selected as the input features of the MOPSO-CDD-DNN mixed traditional culture education prediction model [29]. In this paper, the above indicators are selected to construct a feature indicator set, and the PCA method is used to reduce the dimensionality of the indicators. Before performing PCA, KMO test and Bartlett sphericity test should be performed on the characteristic indicators to judge whether each standard is suitable for principal component analysis.

3. Comparison of Predictive Performance between Traditional Cultural Education Model and Other Models

3.1. Prediction Accuracy Comparison. In order to evaluate the traditional cultural education prediction model more comprehensively, in addition to the overall accuracy of TR, this paper also gives the prediction accuracy (TPR and TNR) of positive samples and negative samples. Based on the results in Table 1, the MOPSO-CD-DNN model constructed in this paper has the best prediction accuracy, which is about 4 percentage points behind the second best model NSGA-II-DNN in TR. Its evaluation indicators TPR, TNR, and TR are 72.88%, 75.56%, and 74.04%, respectively, which are higher

than the other six prediction models, indicating that the MOPSO-CD-DNN method has better performance in enterprise financial risk prediction. In addition, the prediction performance of NSGA-II-DNN and MOPSO-CD-DNN is better, indicating that the multiobjective optimization algorithm is better than the single-objective method in prediction effect.

The above experimental results only analyze the differences of the models from the numerical value of the prediction accuracy but lack the significance of mathematical statistics. Therefore, in order to strengthen the scientificity and objectivity of the obtained results, on the basis of the results in Table 1, the paired-sample t-test and Friedman test were used to test the significance of the prediction results to judge whether there was a significant difference in the prediction performance of different models. The specific experimental results are shown in Tables 2 and 3.

The results in Table 2 show that, except that the paired-sample t-tests on TNR for MOPSO-CD-DNN and NSGA-II-DNN were significant at the 5% level, the T-tests of the remaining models and MOPSO-CD-DNN on each indicator reject the null hypothesis at the 1% significance level, indicating that the MOPSO-CD-DNN model has significant differences in TPR, TNR, and TR compared with other models, its traditional cultural education prediction effect is better. In addition, it can be found from Table 3 that the F and P values of TPR are 26.657 and 0.001, respectively, while the F values of TNR and TR are 27.771 and 26.225, respectively.

Table 2: Paired-sample t-test for prediction accuracy of each model.

Model	MOPSO-CD-DNN		
	TPR/%	TNR/%	TR/%
NSGA-II-DNN	4.926***	3.336**	7.202***
SA-DNN	12.341***	8.214***	20.396***
DE-DNN	13.046***	8.647***	17.393***
DNN	9.168***	8.810***	10.525***
SVM	6.182***	33.698***	18.779***
BPNN	23.877***	24.342***	32.163***

Note. *means $p < 0.1$, **means $p < 0.05$, ***means $p < 0.01$.

Table 3: F test of prediction accuracy of each model.

	Compare models	TPR/%	TNR/%	TR/%
MOPSO-CD-DNN	NSGA-II-DNN SA-DNN DE-DNN DNN SVM BPNN	$F = 26.657$ $P = 0.001$	$F = 27.771$ $P = 0.001$	$F = 26.225$ $P = 0.001$

Note. *means $p < 0.1$, **means $p < 0.05$, ***means $p < 0.01$.

3.2. Predictive Stability Comparison. Since paired sampling in a 1 : 1 ratio will have a certain impact on the randomness of the sample, and in order to compare the prediction stability of the MOPSO-CD-DNN model with the NSGA-II-DNN, DE-DNN, and SA-DNN models, therefore, this paper examines the robustness of the MOPSO-CD-DNN traditional culture and education prediction model by increasing the number of positive samples paired with negative samples and changing the ratio between positive and negative samples. Specifically, firstly, the ratio between the traditional culture education sample and the normal sample is set to 1 : 2, 1 : 3, 1 : 4, 1 : 5, and 1 : 6, totaling five. Then, the changes in the prediction accuracy of the four types of models under these five proportions are analyzed, so as to judge the prediction stability of the MOPSO-CD-DNN model. Since the samples under the five proportions are no longer balanced samples, it is necessary to introduce metrics specifically for unbalanced samples to evaluate the predictive performance of the model. The results of the model experiment are shown in Figure 2.

As can be seen from the above figure, under different proportions of positive and negative samples, the G value (0.7803) of the MOPSO-CD-DNN model is higher than other models, and the F value is poor, but its F mean (0.5902) is higher than other models with the smallest standard deviation. It shows that the NSGA-II-DNN, DE-DNN, and SA-DNN models are inferior to MOPSO-CD-DNN for both traditional culture education samples and nontraditional culture education samples. From the graph in Figure 2, the prediction accuracy G and F of the MOPSO-CD-DNN model are relatively more stable and less volatile, and the standard deviations of G and F are 0.0097 and 0.0462, respectively, which are also significantly lower than other methods. This shows that the prediction results of the two types of samples are stable under different proportions of traditional cultural education samples.

3.3. Sensitivity Analysis of MOPSO-CD Algorithm. The MOPSO-CD multiobjective optimization algorithm contains three main parameters: the number of particle swarms (popsize), the number of iterations (generations), and the archive size (archivesize), which can significantly affect the effectiveness of the proposed hybrid model for traditional culture and education prediction. Appropriate parameters will improve forecast accuracy and forecast efficiency. Therefore, it is necessary to explore the influence of the parameter changes of the MOPSO-CD model on the prediction performance of traditional culture education by modifying one of the three parameters and keeping the other parameters unchanged (see Table 4).

Table 4 shows that changes in the number of particle swarms, the number of iterations, and the file size have a significant impact on the prediction performance of MOPSO-CD-DNN. For example, with the increase in the number of iterations, TR shows a law of first increasing and then decreasing. When the number of iterations is 8, TR reaches the maximum value (74.04%), and after that, TR gradually decreases. Therefore, changes in the number of iterations will affect the predictive ability of the proposed traditional cultural and educational hybrid prediction

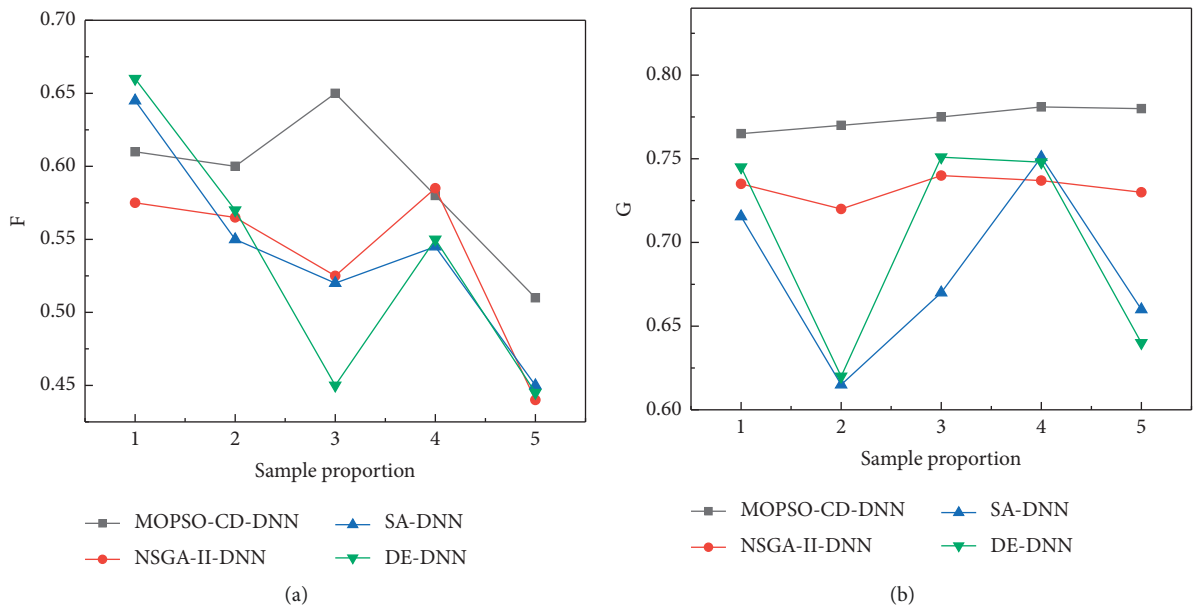


Figure 2: Comparison of the prediction results of the four types of models under different sample proportions. (a) F value, (b) G value.

Table 4: Sensitivity analysis results of MOPSO-CD algorithm.

Parameter	Parameter value	TPR/%	TNR/%	TR/%
Number of particle swarms	50	67.92	64.71	66.35
	100	72.88	75.56	74.04
	150	68.97	69.57	69.23
	200	74.47	66.67	70.19
	250	68.42	68.09	68.27
	300	68.25	73.17	70.19
Number of iterations	58	73.91	65.52	69.23
	15	72.88	75.56	74.04
	23	66.18	75.00	69.23
	35	65.28	78.13	69.23
	50	69.23	65.38	67.31
	130	68.52	66.00	67.31
File size	160	71.11	62.71	66.35
	190	71.69	68.63	70.19
	220	70.83	64.29	67.31
	250	67.24	67.39	67.31
	280	72.88	75.56	74.04
	50	68.42	68.09	68.27

model. The changes in MOPSO-CD parameters will obviously affect the prediction effect of the traditional cultural education mixed prediction model. Inappropriate and poor parameters will lead to the failure to find the optimal initial parameters of the DNN, thus reducing the predictive ability of the MOPSO-CD-DNN hybrid model. According to the experimental results, this paper sets the number of particle swarms to 100, the number of iterations to 8, and the file size to 250.

4. Conclusion

Traditional culture education in primary and secondary schools has always been one of the hotspots in the education sector. This paper fully considers the complexity of traditional culture education forecasting. Singapore and Japan are compared as developed countries. By combining multiobjective optimization algorithm and deep learning technology, a traditional culture education prediction method based on MOPSO-CD is constructed, and it is successfully used in traditional culture education prediction, and compared with other 6 benchmark models. The results show that (1) in terms of prediction accuracy, the MOPSO-CD-DNN method has better prediction ability, and its prediction accuracy is statistically significant compared with other methods; (2) in terms of prediction robustness, MOPSO-CD-DNN can effectively control the fluctuation degree of prediction accuracy and improve the stability of prediction.

Compared with NSGA-II-DNN, DE-DNN, and SA-DNN, the prediction accuracy is improved. The G value (mean value) increased by 4.66%, 7.43%, and 9.25%, and the standard deviation (G value) decreased by 0.001, 0.0502, and 0.0413; (3) the introduction of multiobjective optimization algorithm can effectively enhance the prediction ability and generalization ability of DNN, avoid the shortcomings of single-objective optimization, and improve the learning ability for unbalanced samples.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Y. Kim and K. T. Sung, "Conjugal violence in Korean American families: a residue of the cultural tradition," *Journal of Family Violence*, vol. 15, no. 4, pp. 331–345, 2000.

[2] X. Gao, "Teachers' professional vulnerability and cultural tradition: a Chinese paradox," *Teaching and Teacher Education*, vol. 24, no. 1, pp. 154–165, 2008.

[3] T. L. Cross, "Drawing on cultural tradition in Indian child welfare practice," *Social Casework*, vol. 67, no. 5, pp. 283–289, 1986.

[4] M. Grebska-Kulova, "The balgarchevo II-dolna ribnitsa cultural tradition of southwest Bulgaria," *Ascomycetes*, 2021.

[5] T. I. Lipich and O. S. Gudova, *Philosophical and Anthropological Analysis of the Role of Women in the Russian Cultural Tradition*, 2021.

[6] S. Chai, "Confucianism for the contemporary world: global order, political plurality, and social action, edited by tze-ki ton and kristin stapleton," *Asian Review of World Histories*, vol. 9, no. 2, pp. 303–306, 2021.

[7] S. Y. Kim, "Confucianism and taoism debate in the 18th century sajok(scholar family) novel," *The Studies of Korean Literature*, vol. 69, pp. 41–77, 2021.

[8] C. Michael, "yamato-damashii "Japanese spirit" definitions," *International Journal of Lexicography*, no. 4, pp. 279–306, 1994.

[9] M. D. Be Rg-Yap, S. K. Chew, and P. Deurenberg, "Elevated body fat percentage and cardiovascular risks at low body mass index levels among Singaporean Chinese, Malays and Indians," *Obesity Reviews*, vol. 3, no. 3, pp. 209–215, 2010.

[10] Y. Miao, *Differences and Similarities between Eastern and Western Cultures Reflected in Chinese and English Idioms*, Journal of Suihua University, 2009.

[11] W. Nehru, India's security partnership with Singapore India and Singapore have developed a bilateral security and economic partnership T, https://openresearch-repository.anu.edu.au/bitstream/1885/13061/1/Brewster,%20D.%20India%27s%20security%20partnership%20with%20Singapore%202009.pdf.

[12] X. L. Huang and L. I. Li, *The Enlightenment Pattern of Ideological and Political Education in Foreign Countries to Chinese Universities*, Theory Research, 2013.

[13] Y. L. Sun and S. H. Wang, *The Comparison of College Ideological and Political Education in Japan,Singapore and Korea and the Enlightenment to China*, Journal of Shangrao Normal University, 2009.

[14] J. W. Crimes, "Neoconservatism in Japan," 2015, https://en.wikipedia.org/wiki/Neoconservatism_in_Japan.

[15] T. Mondal and B. Mandal, "A proteolytic functional amyloid digests pathogenic amyloids," *Journal of Materials Chemistry B*, vol. 10, no. 22, pp. 4216–4225, 2022.

[16] G. Scherto, *Ethical Education: Towards an Ecology of Human Development. Cambridge Education Research*, Cambridge University Press, England, 2020.

[17] H. Song, X. Liu, and H. Mao, "Deep compression and EIE: efficient inference engine on compressed deep neural network," *Hot Chips 28 Symposium*, 2017.

[18] F. A. Geres, J. Schmidhuber, and F. Cummins, "learning to forget: continual prediction with LSTM," in *Proceedings of the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99*, 7 September 1999.

[19] H. Che and J. Wang, "A two-timescale duplex neurodynamic approach to mixed-integer optimization," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 36–48, 2020.

[20] A. M. Swartz, S. J. Strath, D. R. Bassett, W. L. O. Brien, G. A. King, and B. E. Ainsworth, "Estimation of energy expenditure using CSA accelerometers at hip and wrist sites," *Medicine & Science in Sports & Exercise*, vol. 32, no. Supplement, pp. S450–S456, 2000.

[21] K. Zhang, Z. Qu, Y. Dong et al., "Research on a combined model based on linear and nonlinear features-a case study of wind speed forecasting," *Renewable Energy*, vol. 130, pp. 814–830, 2019.

[22] C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 256–279, 2004.

[23] S. Chetlur, C. Woolley, P. Vandermersch et al., "cuDNN: efficient primitives for deep learning," *Nural and Evolutionary Computing*, vol. V3, p. 0759, 2014.

[24] J. Li, R. Zhao, J. T. Huang, and Y. Gong, "Learning Small-Size DNN with Output-Distribution-Based Criteria," *interspeech*, 2014.

[25] M. Genard, A. Kadrani, M. Sidi, M. M. Ould Sid, and F. Lescourret, "The relationship between metaheuristics stopping criteria and performances: cases of NSGA-II and MOPSO-CD for sustainable peach fruit design," *International Journal of Geotechnical Earthquake Engineering*, vol. 5, no. 3, pp. 44–70, 2014.

[26] P. Naval, "Mopsocd: MOPSOCD: Multi-Objective Particle Swarm Optimization with Crowding Distance," 2013, https://rdrr.io/cran/mopsocd/.

[27] A. S. Jadhav, "A novel weighted TPR-TNR measure to assess performance of the classifiers," *Expert Systems with Applications*, vol. 152, Article ID 113391, 2020.

[28] C. Fang, Y. Tao, J. Wang et al., "Mapping relation of leakage currents of polluted insulators and discharge arc area," *Frontiers in Energy Research*, vol. 9, Article ID 777230, 2021.

[29] L. Li, Y. Li, Y. Liu et al., "Preparation of a novel activated carbon from cassava sludge for the high-efficiency adsorption of hexavalent chromium in potable water: adsorption performance and mechanism insight," *Water*, vol. 13, no. 24, p. 3602, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4785333, 8 pages
https://doi.org/10.1155/2022/4785333

Research Article

Research on Multimodal Dance Movement Recognition Based on Artificial Intelligence Image Technology

Zhuo Zeng

School of Xihua University, Sichuan, 610000, China

Correspondence should be addressed to Zhuo Zeng; 0120020063@mail.xhu.edu.cn

Received 4 June 2022; Accepted 25 June 2022; Published 12 July 2022

Academic Editor: Man Fai Leung

At present, most robot dances are precompiled. Changing music requires manual adjustment of relevant parameters and metamovements, which greatly reduces the fun and intelligence. In view of the above problems, this paper designed CNN system, studied the multimodal dance movement recognition algorithm of artificial intelligence image technology, and completed the construction of a multimodal dance movement calculation system example. The results show that the CNN algorithm and the Winograd algorithm-based coprocessor-optimized CNN network in multimodal dance movement recognition with image technology reduce from a maximum of 132s to 26s in the runtime criterion, with a maximum reduction of 80%; from a maximum of 73.5% to 16.2% in the memory access criterion, with a maximum reduction of 57.3%; and from a maximum of 93.6% to 25.2% in the power consumption ratio criterion, with a maximum reduction of 68.4%. In the power consumption ratio criterion, the maximum reduction from 93.6% to 25.2% is 68.4%. The maximum accuracy of the proposed optimization method is 95.1%. The solution is proposed to address the problem of insufficient performance of traditional dance movement recognition, which will contribute to the development of artificial intelligence and dance industry.

1. Introduction

Dance is one of the oldest forms of traditional artistic expression, with its unique vitality and creativity. As a special product of human civilization, art has a very important role in promoting the development of the whole society and the prosperity of the country. Among them, dance is a very ancient and unique charming art product, which has its own very important relevance and mechanism as one of the carriers of traditional Chinese culture [1–3]. The revitalization of dance can promote the excellent traditional culture and enhance the soft power of the country culturally. In a diversified world, there are tens of thousands of ways to satisfy spiritual needs, and dance, as an elegant plastic art in motion, is one of the most popular ways and also plays a role in people's socialization. Dance, not only a social and cultural form, is also a dynamic art presented in a rich and varied body language. Dance movements include bending; leaning together, extending; gliding; feet drawing circles on the ground; striking; dissolving; drawing circles in the air; rising; drawing up, pulling up; running of hands; the direction of feet and body. With the rapid development of artificial intelligence, dance also begins to enter the era of information and intelligence. How to identify and optimize multi-modal dance movements through artificial intelligence so as to achieve dance in the intelligence to maintain the reality remains a challenge.

Dance is a special way for people to express their emotions [4, 5], and dance is different from other expressions in that it uses the human body as a tool to express emotions through artistic processing of the human body. Many dances are created in the process of people's work, and they have very strong local characteristics and humanistic feelings. People who express their emotions through dance also carry out the reflection of our socioeconomic conditions, creating many dances rich in local characteristics. Dance relies heavily on the ability of people to synchronize their movements with music in a rhythmic manner. The combination of artificial intelligence and dance has given rise to robotic dance [6–8], which requires

the presentation of a continuous, balanced, and aesthetic movement, in addition to a certain level of intelligence required for the dance. Robotic dance has contributed to the development of dance, which is traditionally limited to the human sphere and uses humans as the main dance vehicle. Due to the rapid development of artificial intelligence, robotic dance has subverted people's perceptions to a certain extent. The related research has also drawn wide attention, so we conducted a research related to dance movements in the relevant database, and the search results are shown in Figure 1, which indicates that the main research work on dance movements in computer artificial intelligence accounts for the largest proportion, 52.07%, and the amount of music and dance research, which is ranked second, is 23.97%, followed by automation-related dance research, accounting for of 12.4%, which indicates that research on artificial intelligence for dance movement recognition is extremely important. Bian [2] et al. proposed to validate the generated model by comparing artificial intelligence (AI) output with dance teachers' assessments, where they first elicited from teachers the dance elements they typically focus on for rhythmic assessment (i.e., tempo, pause, step length, and weight transfer). Then, selected features are extracted from raw motion sensor data related to the rhythmic patterns of the learner's dance, their synchronization with the music beat, and specific features of the song being played. Finally, a machine learning (ML) algorithm is used to create a predictive computational model using these features. To validate quantitative comparisons between ML output and dance teachers' assessments of learners' dance performance, and to provide a qualitative analysis of the potential pedagogical uses of ML model output as envisioned by dance teachers. Sun [3] proposed a new approach to the spatiotemporal dynamics of dance using keypoints integrated with GRU networks, creating a video archive of live recordings from different trained dancers and newspaper clippings from the Internet. The Deep Pose Estimator coupled GRU model uses a deep learning pose estimator to process the spatial aspect and a GRU network to process the temporal perspective. The efficiency of the proposed method is compared with benchmark methods such as 3D convolutional neural network-based models, temporally distributed CNN-LSTM [9–12] models, and hybrid migration learning-LSTM models, and the results show that the proposed method outperforms other method resolutions even under different videos. Hui [13] et al. used artificial intelligence to provide a system of virtual dance partners, providing a new direction for research such as artificially intelligent dance movements. Joo-Wha Hong [5] investigated how people perceive AI music generators and evaluate their songs based on the different characteristics they possess, the independent creativity, and humanization characteristics of AI music generators do not affect music evaluation, relying on the theoretical framework of anthropomorphism and creative machine heuristics, and designed a 2 x 2 experiment in which AI's perceptual anthropomorphism (high vs. low anthropomorphism) and its autonomous creativity (independent vs. dependent creativity) were controlled.

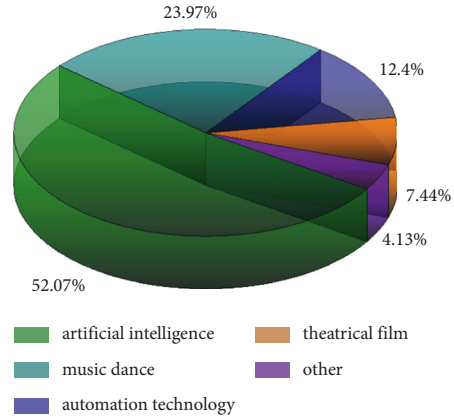


Figure 1: The proportion of dance movement research.

The above literature analysis illustrates the versatility and related excellent stability performance of artificial intelligence in the fields of dance recognition, dance education, etc. Robotic dance is an attractive and highly emerging research field. As an elegant and moving plastic art that perfectly combines auditory and visual arts, it breaks through people's understanding of existing traditional entertainment and makes people's life more colorful. However, in the process of transferring the real dance to the commonly used computer network client, due to the variety of dance movements and the inconsistency of the related dance images, there are often problems such as slow image processing speed, high power consumption, and network bandwidth difficulty in the process of extracting information and calculating related data from the dance images; therefore, this paper addresses the problems related to the dance movement image processing. Artificial intelligence method is used to solve and optimize the related problems in dance movement recognition by combining convolutional neural network with Winograd algorithm.

2. Concepts Related to Artificial Intelligence Dance Movement Recognition

Artificial Intelligence (AI), which refers to the intelligence displayed by programs written by humans, etc., is the technology and science of studying and discovering human intelligence [14–16]. Artificial intelligence technologies are now widely used in various production activities, and the most widely used fields mainly include image processing, natural language processing, and human speech and semantic recognition. Image processing can be combined with a variety of application scenarios and has shown excellent performance in several fields. In the process of dance image processing [17–19], the main system flowchart is shown in Figure 2. First, the initial structure of the image such as the audience is divided, and then the emotions expressed by the characters are divided, while at the same time, the color processing contained inside the image is extremely important, and finally in the input in the information processing module [20, 21].

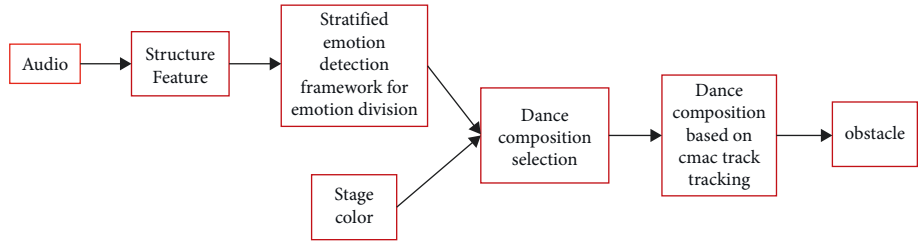


Figure 2: Dance composition flowchart.

In the process of processing image data by artificial intelligence, there are several main problems as follows:

(1) Data transmission is limited by the network. The huge amount of data generated at the edge is uploaded to the cloud for processing, and then the processing results are returned from the cloud, which not only wastes a lot of network resources but also consumes a lot of time and causes a lot of network latency;

(2) Data processing is limited by cloud server computing bottlenecks [22]. It is still unable to meet the demand of the exploding data computing and processing. At the same time computing demand in different time periods, there are peaks and valleys, in the peak of demand edge devices to provide data to wait in line for processing, seriously affecting the normal operation of edge devices; in the trough of demand, a large number of computing resources remains idle, resulting in waste

(3) There are security and privacy issues in data network transmission [23]. The data generated during image processing involves personal privacy, which is likely to be stolen by hackers through loopholes during network transmission, resulting in personal privacy data leakage.

2.1. Feasibility of Artificial Intelligence Technology Applied to Multimodal Dance Movement Recognition. Most of the current robot dances are precompiled, and if you change the music, you need to manually adjust the relevant parameters and metamovements, which is much less interesting and intelligent. If the robot dance is driven not only by the beat of the music but also by the emotion and tension of the music, and it is not just pattern matching, and the interest will be greatly increased. Among them, the demand of dance image processing is one of the main demands of multimodal dance recognition. Image processing usually requires complex operations, and applying artificial intelligence techniques to dance image processing has better results in reducing the amount of operations and improving the performance of image processing. In this paper, an AI-based image recognition multimodal dance movement system is designed to solve the problems of poor real-time, high power consumption, and high network bandwidth requirement of cloud computing system. Convolutional neural network is one of the most widely used and efficient algorithms in the field of artificial intelligence image processing technology. The advantages of convolutional neural networks over other traditional image processing algorithms include the fact that convolutional neural networks can take the original image directly as input without the need for preadjustment of the original image. The CNN training process is mainly divided into forward propagation and backward propagation [24, 25].

Step 1. Firstly, the main purpose of the algorithm processing is to extract a certain feature from the image. The schematic diagram of the feature map and convolution kernel involved in the operation is shown in Figure 3.

In the image information processing, first of all, before the convolutional network starts to work, the basic information of the image is input, and then a preliminary information is extracted, including the basic information of the dance movement, the number of dancers and the stage, etc. Through the preliminary, it is simple that processing the information is input into the convolutional calculation.

$$a_{d,i,j} = f\left(\sum_{d=0}^{D-1}\sum_{m=0}^{F-1}\sum_{n=0}^{F-1} w_{d,m,n}x_{d,j+m,j+n} + b\right), \quad (1)$$

where D, F are the number of filters and the size of the convolution kernel, respectively.

Step 2. In the background conditions of very hot artificial intelligence, the research on hardware acceleration of artificial intelligence algorithms is also getting hotter. Convolutional neural networks are used in many projects as a very effective network structure for image processing. In the convolutional neural network, the convolutional operation occupies more than 80% of the entire convolutional neural network. The data is processed enough to filter out a large number of parameters, improve training accuracy, and reduce errors. Commonly used types of operations are maximum, average, and summation. The basic information of the previous image is weighted, biased, summed, and then outputted by the function, and the basic processing is as follows:

$$y = f(w \cdot x) + b. \quad (2)$$

Step 3. The information of the image has been extracted and processed basically in step 2, and the next step is to reverse update the information in step 2 to further optimize the results and improve the accuracy and stability of the

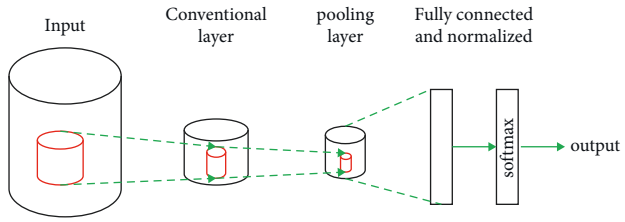


Figure 3: Simplified convolutional neural network structure.

processing process. The process is divided into the following main processing procedures to achieve high accuracy and stability of image action recognition.

(i) By comparing the feature data of the prediction result in the image recognition process with the feature result data of the actual image, the error exists in it so as to facilitate the next update feedback and further processing:

$$\delta^{i,l} = -\left(y_{real}(i) - y_{pre\ di\ ct}(i)\right) * \sigma\left(a^{i,l}\right), \tag{3}$$

(ii) where $y_{predict}$ is the image prediction data result data, and y_{real} is the actual feature result data.

(iii) First, the obtained error is initially input:

(iv) By preliminarily transmitting and processing the error calculated in (i), in the process of intelligently processing image action recognition, the formula for feature recognition extraction is as follows:

$$\delta^{i,l} = \left(w^{l+1}\right)^T \delta^{i,l+1} \odot \delta\left(a^{i,l}\right). \tag{4}$$

(v) Then, through the initial input of the error again, it is determined whether the error is too large to meet the requirements of image feature extraction. If it does not meet the set target, it will be eliminated and the information feature extraction will be performed again; if the requirements are met, the next step refined feature extraction processes the results of the first step again:

$$\delta^{i,l} = \delta^{i,l+1} * rot180\left(w^{l+1}\right) \odot \delta\left(a^{i,l}\right). \tag{5}$$

(vi) After the judgment and processing in the second step, the basic information of the image is well guaranteed and basically meets the set requirements.

(vii) Finally, the processing results are processed in more detail.

$$\delta^{i,l} = \text{upsamples}\left(\delta^{i,l+1}\right) \odot \delta\left(a^{i,l}\right), \tag{6}$$

(viii) where $\delta(a^{i,l})$ module functions that can be handled in more detail. The convolution operation is converted to matrix multiplication, which can rearrange the elements in the feature map, convert the feature matrix into a special matrix, and store it in the memory continuously, which improves the speed of memory reading.

(ix) After the information extraction process described in the previous section, in order to extract and process some more detailed information in the image, we update the weights w and bias b of the processing process twice, in order to better update and track the relevant image information in real time.

After getting the feedback from the previous images, we update them to get some more detailed body movements, such as hand dancing, foot moving, etc.

$$w^l = w^l - \alpha \sum_{i=1}^{m} \delta^{i,l}\left(\alpha^{i,l-1}\right)^T. \tag{7}$$

After the feedback, the deviation update for the image is processed using (8) to obtain the subsequent intermediate processing update values.

$$b^l = b^l - \alpha \sum_{i=1}^{m} \delta^{i,l}. \tag{8}$$

Firstly, the neural network algorithm is deployed to the processing module, which collects images or other information in real time through the data acquisition module. At the same time, the collected image or video information is processed and information analyzed using the neural network algorithm. Finally, in the overall processing process, all the information from the previous period is checked and updated again to ensure the accuracy and authenticity of the images as much as possible.

$$w^l = w^l - \alpha \sum_{i=1}^{m} \alpha^{i,l-1} * \delta^{i,l}. \tag{9}$$

After the whole update process to get the relevant update data, the deviation of the image again follows the processing work, through equation (10) reasoning completed in real-time feedback update to the next step.

$$b^l = b^l - \alpha \sum_{i=1}^{m} \sum_{u,v} \left(\delta^{i,l}\right)_{u,v}, \tag{10}$$

where $\alpha^{i,l-1}$ is expressed as the output of the i-th neuron of the (l-1)-th layer.

After an image action has undergone all the processing methods described above, the extraction of relevant feature information has been completed. In order to further optimize the accuracy of the extracted features, we use the image distortion method in the image processing process. The loss function is used as a feedback indicator.

$$J_c = -\frac{1}{N} \sum_{1}^{N} \sum_{i=1}^{k} \log\left(y_c(i)\right), \tag{11}$$

where $y_c(i)$ is the predicted value, $y(i)$ is the actual value, and N is the number of samples.

2.2. Winograd Algorithm. The Winograd algorithm is a fast convolution algorithm. By finding the common terms in the multiplication of convolution operations and combining

some multiplication operations, the multiplication resources required for calculation can be saved. Multipliers consume a lot of resources in digital processors, so reducing multiplication operations can improve certain performance. The specific calculation process is as follows:

$$\begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_{m-2} & x_{m-1} & \cdots & x_{m+r-2} \\ x_{m-1} & x_m & \cdots & x_{m+r-1} \end{bmatrix} * \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{r-2} \\ w_{r-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{r-2} \\ y_{r-1} \end{bmatrix}, \quad (12)$$

where m is the output size and r is the convolution kernel size.

In addition, the symbol of r input filter is expressed as F, and the minimum number of multiplications required is calculated as follows:

$$\mu(F(m,r)) = m + r - 1, \quad (13)$$

where $\mu(F(m, r))$ is the minimum number of multiplications required for calculation.

For the two-dimensional calculation of the subfilter, the formula can be derived from the one-dimensional formula as follows:

$$\mu(F(m*n, r*s)) = (m + r - 1)(n + s - 1). \quad (14)$$

It can be inferred from the above formula that the minimum number of multiplications is required by CNN calculation after optimization. In addition, in order to further reduce the computational complexity of optimal minimum multiplication, the output length of 2 is selected at this time, and the convolution calculation of the input length of 4 is carried out, and the calculation process is as follows:

$$\begin{aligned} F(2,3) &= \begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix} * \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} \\ &= \begin{bmatrix} x_0 * w_0 + x_1 * w_1 + x_2 * w_2 \\ x_1 * w_0 + x_2 * w_1 + x_3 * w_2 \end{bmatrix}. \end{aligned} \quad (15)$$

It can be obtained that 6 multiplications are needed for direct operation, but only 4 multiplications are needed for Winograd. Specific optimization methods are as follows:

$$m_1 = (x_0 - x_2)w_0, \quad (16)$$

$$m_2 = (x_1 + x_2)\frac{w_0 + w_1 + w_2}{2}, \quad (17)$$

$$m_2 = (x_1 - x_2)\frac{w_0 + w_1 + w_2}{2}. \quad (18)$$

By combining equations (16)–(18), (19) can be obtained as follows:

$$F(2,3) = \begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix} * \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} m_1 + m_2 + m_3 \\ m_2 - m_3 - m_4 \end{bmatrix}. \quad (19)$$

In this way, it takes only four multiplications to complete a calculation that would otherwise require six multiplications, saving two multiplication resources. In fact, it is not difficult to conclude that the larger the base unit used, the more multiplications can be saved. If the dance image you need to identify is large, you can also use a larger base unit. Therefore, customized optimization for different projects can achieve the best optimization effect from the algorithm level.

3. Experimental Verification and Comparative Analysis

3.1. Comparative Analysis of Accuracy of Multimodal Dance Movement Recognition Results. The aesthetic function of dance, including beautiful musical melody, elegant dance posture, and vivid dance image, can accurately identify dance movements through artificial intelligence image processing technology. Through the gradual artificial intelligence image optimization and feature extraction, it can realize the unification of the beautiful dance posture and beautiful melody in the computer client. Therefore, the multimodal dance movement recognition of artificial intelligence can fully extract the format movements in the dance, integrate them into the dance database, and generate relevant videos, which can also provide learning materials for the cultivation and continuous improvement of the aesthetic awareness and ability of dance learners. As a visual art that conveys emotion, dance is a typical embodiment of human imagination and creativity. It not only has the essence of creation and appreciation but also reflects people's appreciation ability. Therefore, the extraction and present of dance features play an important role in activating people's thoughts, cultivating and developing intelligence, and adjusting people's thinking.

In the process of artificial intelligence dance image movement recognition, when the same convolution operation is processed by the CNN network CPU without optimization in the data processing module, a total of 36 times of memory need to be accessed, and 18 times of minimum multiplication and division mathematical operation and 18 times of addition and subtraction mathematical operation need to be carried out. Considering that different compilation options have different optimization effects on the program, the running status of CNN before optimization is analyzed here. Winograd implements the 32 bit single-period multiplier in the execution stage. Considering the zero-cost hardware cycle optimization and ignoring other control signals, it takes at least 4 minutes to complete the convolution operation. As a result, the accelerated CPU achieved more than an eight-fold improvement in performance. The program runtime comparison is shown in Figure 4. The maximum running time is reduced from 132s to 26s, with a

maximum reduction of 80%. This shows that the data processing module after optimization is faster for data processing. In the same time, the processing speed after optimization is about 5 times that of the previous, and the processing capacity of the optimized artificial intelligence image processing module has been greatly improved.

Similarly, in the process of artificial intelligence image processing, the memory ratio of the computer is also a difficulty. On the premise of consistent feature extraction effects for the same image with the processed information extracted, the smaller the memory ratio, the more accurate the feature extraction will be and the more accurate the control of multimode dance movements will be. In our study, the memory access results in the recognition of 105 image actions processed simultaneously are shown in Figure 5. From Figure 5, it can be clearly seen that the ratio of memory in the optimized extraction module decreases from 73.5% to 16.2%. Compared with CNN without Winograd optimization, the number of memory access is reduced by about 57.3%. The optimization used in this paper takes only 2 minutes. However, through ordinary operations (CNN), even excluding the influence of control signals, data activation, and pooling layer computing need a lot of command operations. In this case, the network computing personality can be improved even more by coprocessing.

It is worth mentioning that in the intelligent image action recognition, more and more attention is paid to the energy consumption in the processing process. The lower the energy consumption, the lower the processing cost and the higher the economic benefits. Therefore, in this paper, in order to reduce the power consumption during the network operation, we explore how the optimized network can improve the economic efficiency. This paper further optimizes the network operation cost and conducts an accurate statistical analysis of the energy consumption in the image recognition process before and after optimization. The power consumption ratio comparison of network computing is shown in Figure 6. It can be seen from Figure 6 that the power consumption ratio of the CNN network optimized by the Winograd algorithm can be reduced to 1/4 of the original, from the original maximum of 93.6% to 25.2%, with a maximum reduction of 68.4%. The experimental results show that through this special processing, from the algorithm level, the convolution calculation is optimized, hardware resources are saved, and processor performance is improved. In addition, through the optimization of processing time and memory ratio, and then the energy consumption of the intelligent image processing process is directly optimized, the economic performance is improved, and most of the energy consumption is saved.

In order to further prove the superior performance of our proposed optimization algorithm in multimodal dance action recognition, we selected 2,000 images to implement our algorithm on MATLAB software, of which 1,600 images were used as training set and 400 images were used as test set. As a test set. The training results are shown in Figure 7. After the optimization of the Winograd algorithm is applied, the training set reaches a state of convergence after 60 iterations, and the accuracy can reach 90.2%. In the test set, the accuracy of about 90% can be achieved after 39 iterations. After 60 iterations, the accuracy can be as high as 95.1%.

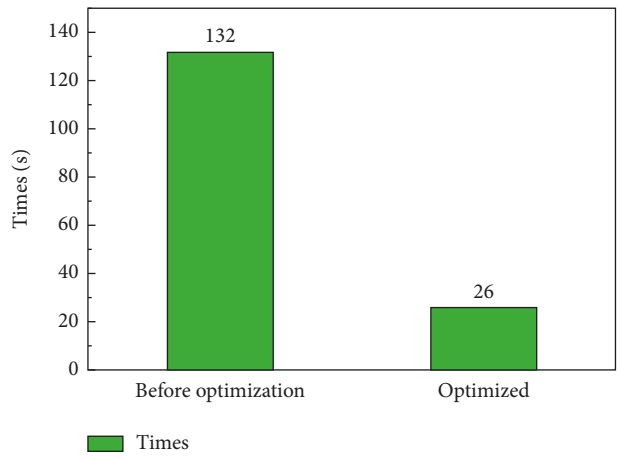


Figure 4: Running time comparison.

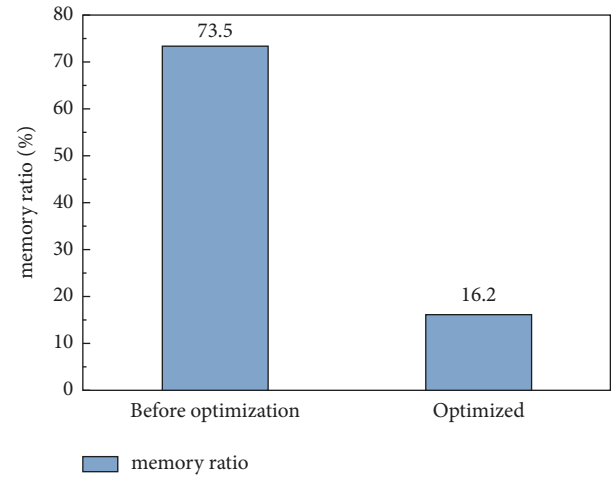


Figure 5: Memory access ratio diagram.

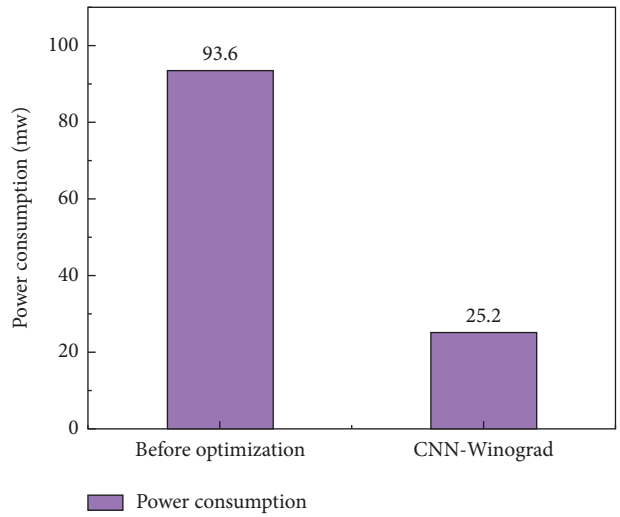


Figure 6: Power consumption performance comparison.

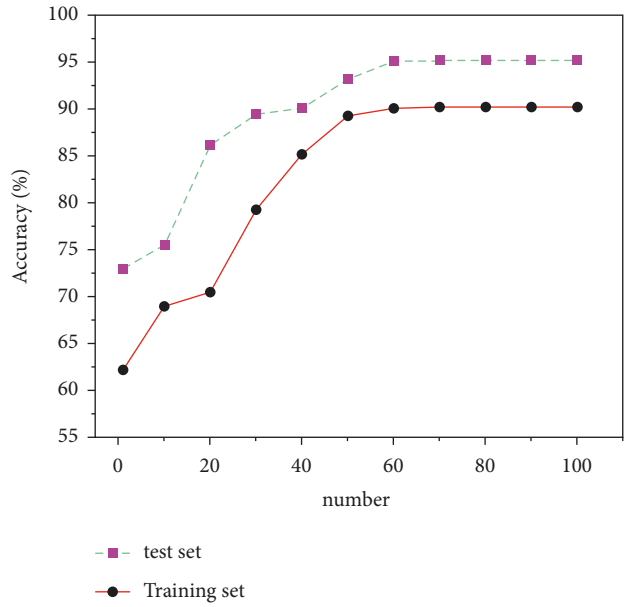


Figure 7: Comparison diagram of algorithm accuracy.

4. Conclusion

In this paper, we give a basic description of the concept of multimodal dance and neural network algorithm, and introduce the CNN algorithm structure and calculation process for multimodal dance recognition. We study the operation process of convolution operation in CPU. Aiming at the situation that the convolution operation requires frequent memory access and a large number of multiplication operations, a coprocessor hardware acceleration module based on the Winograd algorithm is designed to accelerate the convolution operation. And we compared the pros and cons of the CNN and the CNN optimized based on the Winograd algorithm through three evaluation criteria: running time, memory access times, and performance comparison. The result is shown as follows:

(1) The CNN algorithm and the Winograd algorithm-based coprocessor-optimized CNN network in the multimodal dance action recognition of image technology can reduce the running time from a maximum of 132s to 26s, with a maximum decrease of 80%. The judging criteria dropped from a maximum of 73.5% to 16.2%, with a maximum drop of 57.3%. The convolution operation is performed on the feature map by using the operator, and a lot of memory space is saved compared with the operation of converting the convolution operation to matrix multiplication.

(2) Power consumption is one of the economic performances that need to be considered in image processing. In the multimodal dance action recognition in image technology CNN algorithm and coprocessor-optimized CNN network based on Winograd algorithm, the power consumption ratio is reduced from a maximum of 93.6% to 25.2% with a maximum reduction of 68.4% in the judging criteria. The same large performance improvement can be obtained by using the operator to perform convolutional operations on the feature map, which in turn produces a direct optimization of the energy consumption of the intelligent image processing process, improving the economic performance and saving most of the energy consumption.

(3) The CNN algorithm and the Winograd-based algorithm in multimodal dance movement recognition in image technology can reach 90.2% accuracy in the training set and finally stabilize at 95.1% accuracy in the test set after testing. This shows that our proposed optimization method can maintain good accuracy and stability in the recognition of image movements.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declared that there are no conflicts of interest.

References

[1] T. Hui, G. Kumar, Anni, and R. G. Bednarik, "Rock art of heilongjiang province, China," *Journal of Archaeological Science: Report*, vol. 31, Article ID 102348, 2020.

[2] T. Y. Bian, J. Huang, and S. Zhe, "Anchoring effects in the Chinese art market," *Finance Research Letters*, vol. 43, Article ID 102050, 2021.

[3] X. Sun, L. Hu, B. Hu et al., "Remarkable signals of the ancient Chinese civilization since the Early Bronze Age in the marine environment," *Science of the Total Environment*, vol. 804, Article ID 150209, 2022.

[4] M. Grosbras, F. Pollick, and F. Pollick, "Dance and emotion in posterior parietal cortex: a low-frequency rTMS study," *Brain Stimulation*, vol. 5, no. 2, pp. 130–136, 2012.

[5] N. Shikanai and K. Hachimura, "The effects of the presence of an audience on the emotions and movements of dancers," *Procedia Technology*, vol. 18, pp. 32–36, 2014.

[6] X. Meng, T. Zhu, and C. Li, "Construction of perfect dispatch learning model based on adaptive GRU," in *Proceedings of the 2021 The 2nd International Conference on Power Engineering (ICPE 2021)*, vol. 8, pp. 668–677, Nanning, China, December 2021.

[7] Y. Song, M. Zhang, J. Hu, and X. Cao, "Dancing with service robots: the impacts of employee-robot collaboration on hotel employees ' job crafting," *International Journal of Hospitality Management*, vol. 103, Article ID 103220, 2022.

[8] H. Peng, J. Li, H. Hu, L. Zhao, S. Feng, and K. Hu, "Feature fusion based automatic aesthetics evaluation of robotic dance poses," *Robotics and Autonomous Systems*, vol. 111, pp. 99–109, 2019.

[9] K. Zhang, N. Chen, J. Liu, and M. Beer, "A GRU-based ensemble learning method for time-variant uncertain structural response analysis," *Computer Methods in Applied Mechanics and Engineering*, vol. 391, Article ID 114516, 2022.

[10] M. Hu and W. Wang, "Artificial intelligence in dance education: dance for students with special educational needs," *Technology in Society*, vol. 67, Article ID 101784, 2021.
[11] T. H. T. Nguyen and Q. B. Phan, "Hourly day ahead wind speed forecasting based on a hybrid model of EEMD, CNN-Bi-LSTM embedded with GA optimization," *Energy Reports*, vol. 8, pp. 53–60, 2022.
[12] Z. Shen, X. Fan, L. Zhang, and H. Yu, "Wind speed prediction of unmanned sailboat based on CNN and LSTM hybrid neural network," *Ocean Engineering*, vol. 254, Article ID 111352, 2022.
[13] T. Huisheng and G. Kumar, J. Anni and R. G. Bednarik, Rock art of heilongjiang province, China," *Journal of Archaeological Science: Report*, vol. 31, Article ID 102348, 2020.
[14] F. Ma, "Action recognition of dance video learning based on embedded system and computer vision image," *Microprocessors and Microsystems*, vol. 81, Article ID 103779, 2021.
[15] A. Dias Pereira dos Santos, L. Loke, K. Yacef, and R. Martinez-Maldonado, "'Enriching teachers' assessments of rhythmic Forró dance skills by modelling motion sensor data," *International Journal of Human-Computer Studies*, vol. 161, Article ID 102776, 2022.
[16] S. Shailesh and M. V. Judy, "Understanding dance semantics using spatio-temporal features coupled GRU networks," *Entertainment Computing*, vol. 42, Article ID 100484, 2022.
[17] I. Kico and F. Liarokapis, "Enhancing the learning process of folk dances using augmented reality and non-invasive brain stimulation," *Entertainment Computing*, vol. 40, Article ID 100455, 2022.
[18] L. Zhou, A. Xu, and B. Lin, "Image recognition of dance moving based on heterogeneous multiprocessor and dynamic image sampling," *Microprocessors and Microsystems*, vol. 81, Article ID 103658, 2021.
[19] X. Li and H. Zheng, "Target detection algorithm for dance moving images based on sensor and motion capture data," *Microprocessors and Microsystems*, vol. 81, Article ID 103743, 2021.
[20] Y. Zhang, T. Qian, and W. Tang, "Buildings-to-distribution-network integration considering power transformer loading capability and distribution network reconfiguration," *Energy*, vol. 244, 2022.
[21] T. Qian, X. Chen, W. H. Yanli Xin, and L. Wang, "Resilient decentralized optimization of chance constrained electricity-gas systems over lossy communication networks," *Energy*, vol. 239, Article ID 122158, 2022.
[22] S. Gupta and A. D. Dileep, "Relevance feedback based online learning model for resource bottleneck prediction in cloud servers," *Neurocomputing*, vol. 402, pp. 307–322, 2020.
[23] V. Madhav Kuthadi, R. Selvaraj, S. Baskar, and P. M. Shakeel, "Data security tolerance and portable based energy-efficient framework in sensor networks for smart grid environments," *Sustainable Energy Technologies and Assessments*, vol. 52, Article ID 102184, 2022.
[24] K. Feng and Z. Fan, "A novel bidirectional LSTM network based on scale factor for atrial fibrillation signals classification," *Biomedical Signal Processing and Control*, vol. 76, Article ID 103663, 2022.
[25] J. Li, D. Zhu, and C. Li, "Comparative analysis of BPNN, SVR, LSTM, Random Forest, and LSTM-SVR for conditional simulation of non-Gaussian measured fluctuating wind pressures," *Mechanical Systems and Signal Processing*, vol. 178, Article ID 109285, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4286659, 8 pages
https://doi.org/10.1155/2022/4286659

Research Article

Deep Learning Models for Fast Retrieval and Extraction of French Speech Vocabulary Applications

Man Xu

School of Zhejiang International Studies University, Hangzhou 310023, China

Correspondence should be addressed to Man Xu; xuman@zisu.edu.cn

Received 5 June 2022; Accepted 25 June 2022; Published 8 July 2022

Academic Editor: Man Fai Leung

Due to the large French vocabulary, how quickly retrieve and accurately identify the required vocabulary is still a big challenge in French learning. In view of the above problems, we introduce a deep learning algorithm in this study to upgrade and optimize the retrieval system of French words and optimize the acquisition speed of speech words data and the recognition accuracy of speech words, so as to meet the needs of users for word retrieval. The results show that the two training methods of SGD synchronous update network and alternate update network parameters for fast retrieval and extraction of French speech vocabulary reduce from a maximum of 11.65% to 4.25% in the WER criterion, with a maximum reduction of 7.4%; the two training methods of SGD synchronous update network and alternate update network parameters for fast retrieval and extraction of French speech vocabulary reduce from a maximum of 13.52% to 4.4% in the SER criterion. The training methods of fast retrieval and extraction of the SGD synchronous update network and alternate update network parameters in French speech vocabulary reduced from the highest 582 ms to 351 ms in the response time criterion, with a maximum reduction of 8.84%; the maximum reduction of 39.7%. In French speech vocabulary, SGD synchronous updating network and alternating updating network parameter algorithm are used to quickly retrieve and extract French words. When the number of iterations reaches 120, the model fitting accuracy of the training set reaches 90.05%, while the model can reach 94.5% in the test set. The system has a stronger generalization ability and a higher speech vocabulary recognition rate to meet the practical requirements.

1. Introduction

The new era has new requirements for French-language education and a mission to cultivate high-quality and international talents for national strategic development [1, 2]. As a compulsory course in all universities, the teaching of French should keep pace with the times and lay the foundation for cultivating more high-quality talents. With the popularity and use of the Internet, a large number of users have joined the network, and according to the statistics, as of June 2018, the number of Chinese Internet users was 802 million. The huge number of netizens brings a large amount of data. According to IDS estimates, the scale of the global data circle will continue to expand in the next few years. By 2025, the world will create and replicate 163 ZB of data [3, 4]. With the proliferation of text data, it is more and more difficult for users to obtain target data from text databases. Similarly, it is also faced with huge challenges in the fast retrieval and recognition of French vocabulary. A student who speaks more than one language multiplies his or her chances in the national and international job markets. On five continents, 200 million people speak French. French is a major language used for international communication and is one of the most studied foreign languages after English, ranking ninth among the most spoken languages in the world [5, 6]. Mastering French allows one to communicate with French speakers from all continents while expressing another perspective on the world, and to benefit from well-informed international media in French (e.g., TV5, France 24, and Radio France International) [7, 8]. However, French vocabulary is the most crucial basic part in the process of learning French, and if it is difficult to master even the vocabulary, then let alone master the language; at the same time, due to the large vocabulary base of French, it is still a

major challenge in learning French to quickly retrieve and precisely identify the required words [4, 9, 10].

French is an analytical language that has a structure of thought and promotes a critical spirit. French is the language used by great philosophers (such as Descartes, Sartre, and Derrida) and famous scientists (such as the Curies, Pasteur, and Georges Charpak). Students are learning French while learning to make inferences and present different points of view, and learning such methods can be very useful in later discussions or negotiations [11–13]. At present, the first step in learning French is to search for relevant vocabulary, and information retrieval techniques originated from the search and indexing of library materials. The original retrieval system had a single function and could only perform simple searches. Nowadays, it is possible to retrieve information from the whole network, and a lot of new models and methods have been applied to the retrieval system. Aoxiao [14] et al. developed an image retrieval system based on a new deep metric learning algorithm and found that their image retrieval model based on deep metric learning is highly efficient in CXR retrieval, diagnosis, and prognosis with high practical significance through experimental comparison. Da et al. [15] proposed a deep learning-based dual encoder retrieval (DER) model. Pang [16] et al. proposed a new method that uses deep learning techniques to extract high-level and compact features from biomedical images. The deep feature extraction process utilizes multiple hidden layers to capture a large number of feature structures of high-resolution images and represent them at different levels of abstraction, thus improving the indexing and retrieval performance of biomedical images. Beltrán [17] et al. used the VQA model for deep multimodal learning to combine visual and textual representations. Hao [18] investigated the problems related to relevance matching between queries and documents. Gozuacik et al. [19] utilized deep neural networks and natural language processing methods. The above literature analysis shows that deep learning is powerful with superior results in computer vision, image retrieval, speech retrieval, and biological image correlation.

Vocabulary information retrieval refers to the retrieval of vocabulary information from databases to meet user needs and query-related content. The ultimate effect of vocabulary retrieval recognition is to pave the way for French language learning. However, nowadays, vocabulary retrieval has the problems of low retrieval accuracy, slow retrieval speed, and low recognition rate. Therefore, in this study, we introduce the deep learning algorithm to upgrade and optimize the French vocabulary retrieval system, from the speed of acquiring vocabulary data to the accuracy of vocabulary recognition, so as to meet the user's needs for vocabulary retrieval.

2. Concepts Related to Phonetic Word Search

Vocabulary is the most critical and fundamental part of learning French, and mastering it is the first step to success. Vocabulary retrieval and recognition are key techniques in learning vocabulary. The basic task of vocabulary retrieval is to find the answer to the user's given needs from a large database of words. The retrieval query is a literal description of the user's retrieval needs, and the vocabulary is the basic object of retrieval and the basic object returned. To further enhance the effectiveness of the algorithm, the user's behavior on the final feedback results will be used as feedback on the retrieval results, thus further enhancing the retrieval effect [20–22]. The overall framework of vocabulary retrieval and recognition is shown in Figure 1, where each node corresponds to each module of information retrieval.

In the vocabulary retrieval system, the retrieval model and index are the two most critical parts. The retrieval system generally consists of three parts: processing module, retrieval module, and user interface module. The retrieval module can be subdivided into vocabulary index module, user query module, and sorting module. The vocabulary index module first preprocesses all the indexed documents and then performs the inverted index [22–24]; the user query module first preprocesses the user query obtained from the previous interface and quickly retrieves the relevant vocabulary from the inverted index, and then the lexical matching model is used to finally draw a matching conclusion. In order to improve the retrieval accuracy, before the final word feedback, the retrieval system will use the user's result to make further corrections to improve retrieval accuracy [25, 26].

3. Theoretical Foundation of Speech Recognition Based on Deep Learning

In the field of speech signal processing, deep neural networks were first successful in the classification task of speech recognition. After that, many scholars started to apply deep neural networks to speech enhancement tasks. In addition, MMSE objective function-based speech enhancement minimizes the mean square error between the estimated value of the network output and the labeled target value during training, and it treats each time-frequency point as equally important, taking into account the distribution characteristics of the speech spectrum; the trained neural network gets an average optimal result at the time of enhancement, which solves the serious problem of oversmoothing and speech distortion and loss under low signal-to-noise ratio.

3.1. Deep Neural Network (DNN) Architecture. DNN is a deep network structure based on a shallow artificial neural network with stronger nonlinear expression capability by adding hidden layers [27–29]. The network is composed of an input layer, an output layer, and an intermediate L hidden layer, which is characterized by the fact that individual neurons within the same layer are not connected to each other and the neurons in adjacent layers are fully connected to each other, as shown in Figure 2.

Suppose the number of neurons in the lth layer is n_l, the input vector is z^l, the output vector is h^l, and given a training sample x with $h^0 = z^0 = x$, then

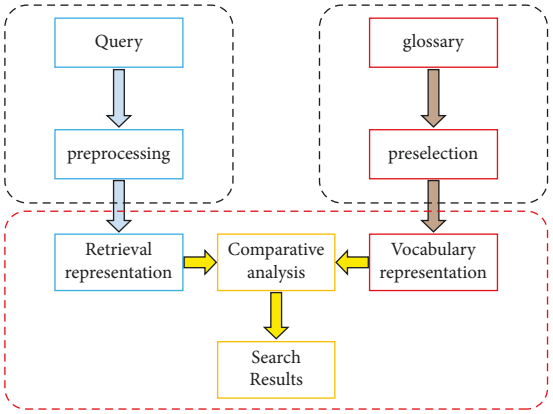


Figure 1: Vocabulary search framework.

$$z^l = W^l z^{l-1} + b^l, \tag{1}$$

where W^l is the weight matrix from the (l-1)th layer to the lth layer and b^l represents the bias matrix of the lth layer. Then, there are

$$h^l = f_1\left(z^l\right), \tag{2}$$

where f_1 represents the activation function of the lth layer and the common form is ReLU, tanh, sigmoid, and softmax.

Speech recognition is a multiclassification problem, so the softmax function is chosen for the activation function of the output layer, and the final output of the DNN is assumed to be $y = h^{L+1}$, which takes the following form:

$$\begin{aligned} y &= \text{softmax}\left(z^l\right) \\ &= \frac{\exp\left(z^L\right)}{\sum_{k=1}^{n^L} \exp\left(z^{Lk}\right)}, \end{aligned} \tag{3}$$

where z^{Lk} denotes the kth component of the vector z^L.

3.2. Derivation of New Objective Function. The DNN inputs the LPS features of the D-dimensional noisy speech signal (2τ+1), and by extending the frames of the input features, the prediction error is defined as follows:

$$e_n = x_n - \hat{x}_n\left(y_{n-r}^{n+r}, W\right), \tag{4}$$

where y_{n-r}^{n+r} is the D-dimensional LPS feature of each extended τ frame on the left and right of the input feature, x_n is the learning target of the network, and W is the neural network parameter.

Assuming that the prediction errors in each dimension are independently distributed and using GGD to model the prediction errors in each dimension, the e_n distribution function is as follows:

$$y\left(e_n\right) = \prod_{d=1}^{D} p_{GGD}\left(e_n, d\right), \tag{5}$$

where e_n represents the prediction error in dimension.

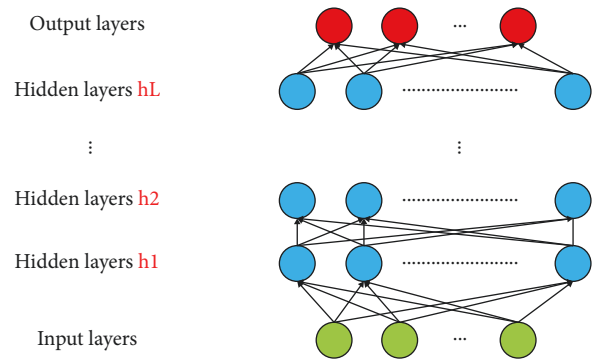


Figure 2: Deep neural network model structure.

Assuming that the prediction errors in each dimension obey the same shape distribution, that is, β has a fixed value for all dimensions, then the expression is as follows:

$$p\left(x_n, W, \alpha\right) = \prod_{d=1}^{D} p_{GGD}\left(x_{n,d} - \hat{x}_{n,d}\left(y_{n-r}^{n+r}, W\right), \alpha_n, \beta\right), \tag{6}$$

where α represents the scale parameter of the prediction error distribution GGD on the dth dimension.

Given N parallel speech data, assuming that the conditional probability distribution is obtained by independent sampling and taking the logarithm of both sides, then the following formula can be obtained:

$$\ln p\left(X, W, \alpha\right) = \sum_{n=1}^{N}\sum_{d=1}^{D}\left(\ln\frac{\beta}{2\alpha_d\Gamma\left(1/\beta\right)} - \frac{\left|x_{n,d} - \hat{x}_{n,d}\left(y_{n-r}^{n+r}, W\right)\right|^{\beta}}{\alpha_d^{\beta}}\right), \tag{7}$$

where $ln\beta/2\alpha_d\left(1/\beta\right)$ is a constant.

In addition, the method of maximum likelihood estimation is introduced in this study to optimize W and α at the same time. Maximizing the log-likelihood function is equivalent to minimizing the formula as follows:

$$E\left(W, \alpha\right) = N\sum_{d=1}^{D}\ln\alpha_d + \sum_{n=1}^{N}\sum_{d=1}^{D}\left(\frac{\left|x_{n,d} - \hat{x}_{n,d}\left(y_{n-r}^{n+r}, W\right)\right|^{\beta}}{\alpha_d^{\beta}}\right). \tag{8}$$

Assuming that the prediction errors in each dimension obey the equal variance distribution, that is, the GGD scale parameters in each dimension are the same, then the following formula can be obtained:

$$E\left(W\right) = \sum_{n=1}^{N}\sum_{d=1}^{D}\left|x_{n,d} - \hat{x}_{n,d}\left(y_{n-r}^{n+r}, W\right)\right|^{\beta}, \tag{9}$$

where, when β is 1, the above formula is the minimum mean absolute error, and when β is 2, the above formula is the minimum mean square error.

3.3. Training Algorithms. The network parameters are generally updated in a minimum batch mode using the stochastic gradient descent method during network training. The new objective function proposed in this section, that is, the objective function in the small-batch sample update mode is obtained as follows:

$$E(W,\alpha) = M\sum_{d=1}^{D}\ln\alpha_d + \sum_{n=1}^{N}\sum_{d=1}^{D}\left(\frac{\left|x_{m,d} - \hat{x}_{m,d}\left(y_{m-r}^{m+r}, W\right)\right|^{\beta}}{\alpha_d^{\beta}}\right), \tag{10}$$

where m is the minimum batch sample size.

This section proposes two training algorithms: one is to use SGD to update network parameters W and α synchronously; the other is to update network parameters W and α alternately. The detailed introduction is as follows:

(1) The parameter α of the network parameter gray-sum distribution function is synchronized by the error back propagation algorithm of SGD, and the updated formula is as follows:

$$W^{(t+1)} = W^{(t)} - \eta\frac{\nabla E_W}{M},$$
$$\alpha^{(t+1)} = \alpha^{(t)} - \zeta\frac{\nabla E_\alpha}{M}, \tag{11}$$

where t represents the number of iterations, η and ξ represent the learning rate of network parameters W and α, and ∇E_W and ∇E_α represent the partial derivative of the objective function with respect to W and α. In addition, the calculation formula of ∇E_α is as follows:

$$\nabla E_\alpha = \frac{M}{\alpha_d} - \frac{\beta}{\left(\alpha_d\right)^{\beta+1}}\sum_{m=1}^{M}\left|x_{m,d} - \hat{x}_{m,d}\left(y_{m-r}^{m+r}, W\right)\right|^{\beta}, \tag{12}$$

where $\hat{x}_{m,d}$ is an abbreviation for $\hat{x}_{m,d}(y_{m-r}^{m+r}, W)$.

(2) Another training algorithm is to alternately update the network parameter W and the distribution function parameter α. When the network parameter W is fixed, the closed-form solution of the distribution function parameter α can be obtained as follows:

$$\alpha_d = \frac{\beta}{M}\sum_{m=1}^{M}\left|x_{m,d} - \hat{x}_{m,d}\left(y_{m-r}^{m+r}, W\right)\right|^{\beta}. \tag{13}$$

In algorithm (2), α is updated by using the closed solution obtained under the maximum likelihood criterion, which saves the trouble of manually adjusting the learning rate ξ in algorithm (1) when using SGD's reverse error propagation algorithm to update, and the algorithm is more robust.

3.4. Speech Lexical Signal Preprocessing. Speech vocabulary signal preprocessing is the basic processing of speech analog signal before feature extraction, and its purpose is to eliminate the impact on the quality of speech signal due to the human articulation organ itself and the equipment used to collect speech signal, such as mixing and high harmonic distortion. It tries to ensure a more uniform and smooth signal for the subsequent speech vocabulary processing and provide a high-quality signal for feature extraction. Preprocessing technology is the premise and foundation of speech recognition, and its key technologies mainly include pre-emphasis, frame-splitting plus windowing, and endpoint detection.

The power spectrum of the voice signal will fall in the frequency range of 6 d/B times in the high-frequency band. Before the voice signal processing, the high-frequency band needs to be increased to make the power spectrum flatter, which is beneficial to the analysis of spectrum or channel parameters. Pre-emphasis technology can remove effects such as lip radiation by boosting the high-frequency band of the speech signal. Pre-emphasis is usually implemented with a first-order high-pass digital filter, and the formula for its transfer function is as follows:

$$H(z) = 1 - \mu Z^{-1}, 0.9 \le \mu \le 1.0, \tag{14}$$

where μ represents the pre-emphasis coefficient, generally 0.94. In the time domain, assuming the input original signal is S(n), the pre-emphasized speech signal is as follows:

$$S'(n) = S(n) - \mu S(n-1). \tag{15}$$

After the speech signal is preprocessed, it needs to be framed and windowed, which is beneficial to the subsequent operations such as feature extraction of the speech signal. The specific formula is as follows:

$$s_w = \sum_{-\infty}^{+\infty} T[s(n)]w(n-m), \tag{16}$$

where $T[\ ]$ is a certain linear or nonlinear operation, $s(n)$ is the speech signal before windowing, w $(n\text{-}m)$ is a certain window function, and s_w (n) is the speech signal after windowing.

4. Experimental Verification and Comparative Analysis

4.1. Comparative Analysis of Accuracy of Speech Word Recognition Results. This study adopts word error rate (WER) and sentence error rate (SER) as the main evaluation criteria for speech recognition. WER represents the relationship between the recognized word sequence and the standard word sequence. The continuous speech recognition results are generally expressed in the form of word sequences. The dynamic programming algorithm is used to align the recognition results with the correct label series and then compare them. There are three types of errors: insertion error I, deletion error D, and substitution error R. Assuming that the total number of word sequences is N, the definition of WER is as follows:

$$WER = \frac{I + D + R}{N} \times 100\%. \tag{17}$$

SER represents the probability of recognizing a correct sentence. Whenever a sentence has a word error, the sentence is considered to be incorrectly recognized, and SER is the number of incorrectly recognized sentences m over the total number of sentences M, which is defined as follows:

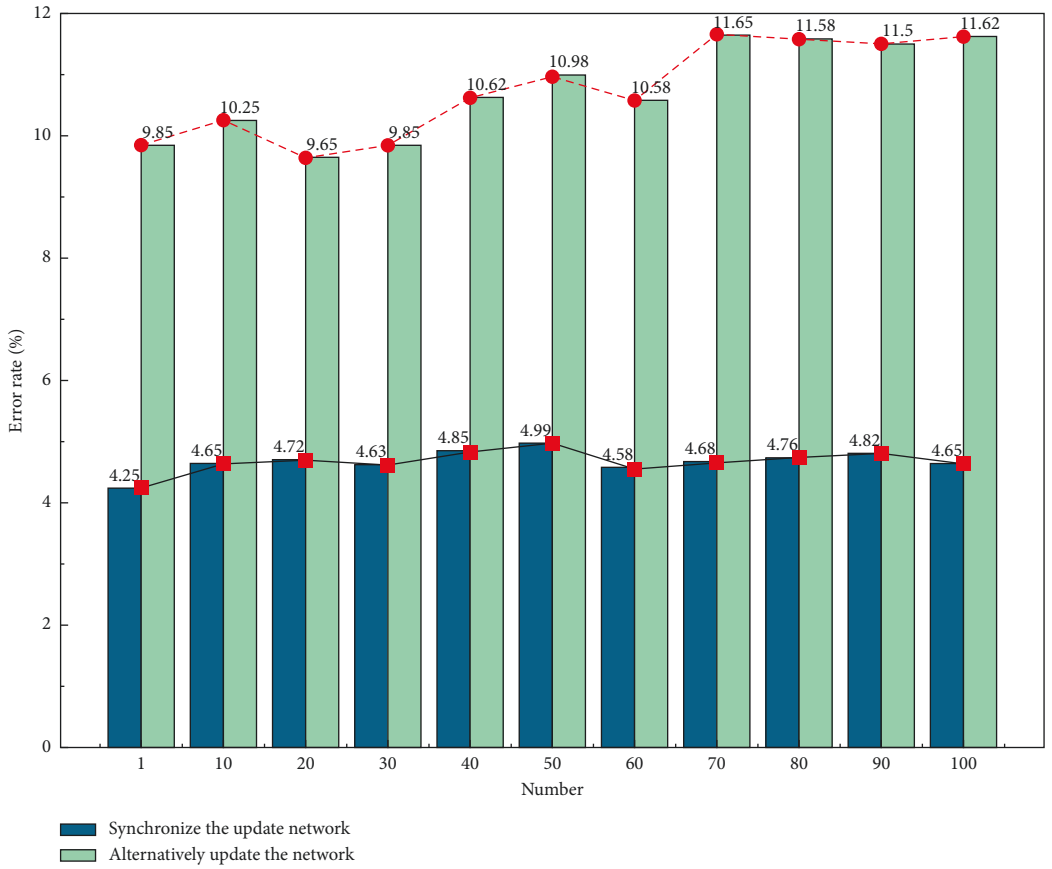


Figure 3: WER analysis graph of two training methods of SGD synchronous update network and alternate update network parameters.

$$SER = \frac{m}{M} \times 100\%. \tag{18}$$

It is worth mentioning that Figure 3 shows the WER analysis for both training methods of SGD synchronous update network and alternate update network parameters. It can be seen that by using alternate update network parameters, it is able to reduce the WER from a maximum of 11.65% to 4.25%, with a maximum reduction of 7.4%. The reason for this is that when increasing N causes more neurons to be inactivated, the sparsity of the entire neural network is increased. It can also be found that as the sparsity increases, the alternate update network training method also has a certain degree of improvement in recognition correctness and computational efficiency with better recognition and lower computational effort.

It is worth mentioning that Figure 4 shows the SER analysis of the two training methods of SGD synchronous update network and alternate update network parameters. It can be seen that by using the alternate update network parameter training method, the SER can be reduced from a maximum of 13.52% to 4.68%, with a maximum reduction of 8.84%. This indicates that the speech bottleneck features extracted based on the alternating update network parameter training method can improve the speech recognition efficiency to a certain extent. The reason for this is that using the sparse regular term as the penalty term of the target function improves the generalization ability of the target function to a certain extent, thus increasing the recognition rate accordingly.

It is worth mentioning that Figure 5 shows the time analysis of the two training methods of SGD synchronous update network and alternate update network parameters. It can be seen that by using the alternate update network parameter training method, the time consumption can be reduced from a maximum of 582 ms to 351 ms, with a maximum reduction of 39.7%. This indicates that the extracted speech based on the alternating update network parameter training method can provide timely feedback to predict the final result and provide faster speech recognition information for the subsequent speech recognition process.

It is worth mentioning that Figure 6 shows the graph of the iteration results of the training set and the test set of the CNN-GRU model. It can be seen from Figure 6 that the network results gradually regionally converge as the number of iterations increases, and when the number of iterations reaches 120, the accuracy of the training set satisfies at 90.05%, while the accuracy of the test set satisfies at 94.5%. This indicates that the CNN-GRU network model is able to retrieve French speech words quickly.

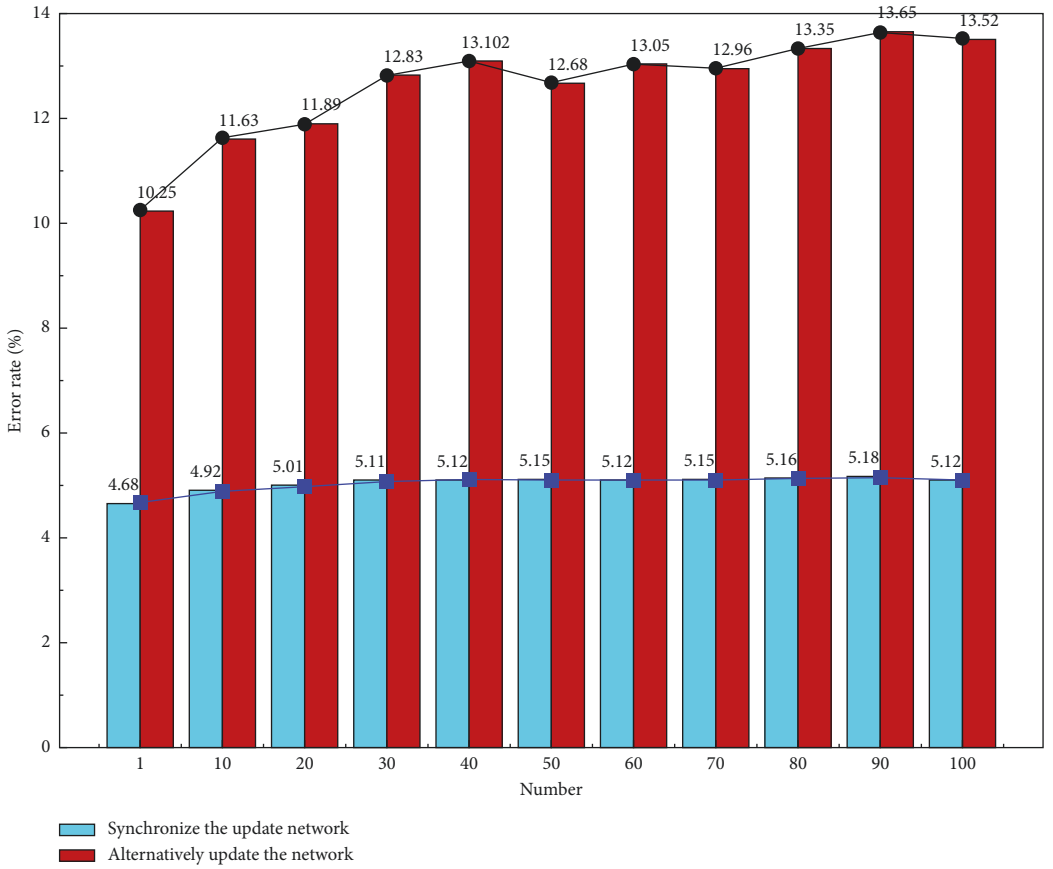


Figure 4: Comparison graph of SER analysis between two training methods of SGD synchronous update network and alternate update network parameters.

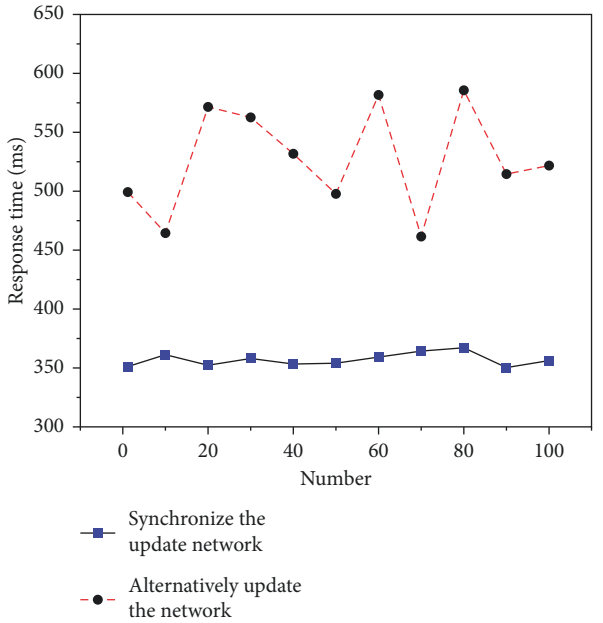


Figure 5: Comparison chart of the response time analysis between two training methods of SGD synchronous update network and alternate update network parameters.

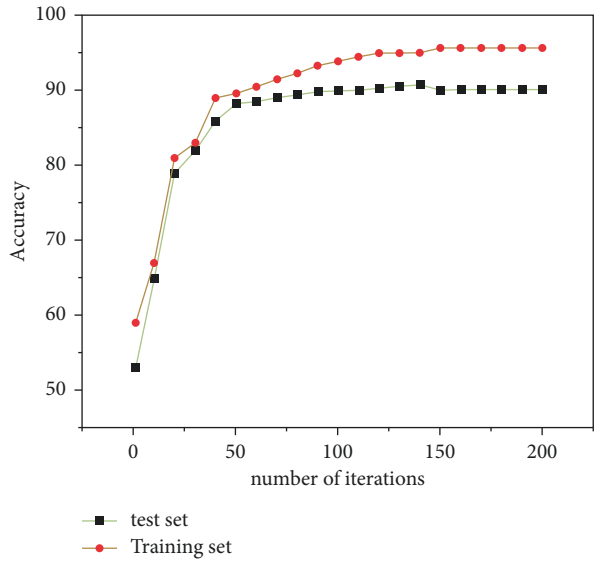


Figure 6: Iteration results of CNN-GRU model training set and test set.

5. Conclusion

In this study, we provide a basic description of the concept of multicommunication framework and neural network algorithm and introduce the DNN algorithm structure and computational flow for French speech retrieval recognition. We also compare the advantages and disadvantages of two training methods, SGD synchronous update network, and alternate update network parameters using three criteria: WER, SER, and response time. The optimization of fast retrieval of French speech vocabulary by deep learning methods improves the system in terms of the accuracy of retrieval of individual words of speech, the accuracy of sentence retrieval, and the response time of retrieval, and comparing the accuracy of the model under a different number of iterations, the conclusions meet the practical needs of retrieval. We hope that our proposed deep learning retrieval optimization can add to the French speech vocabulary learners. The specific results are as follows:

(1) Vocabulary information retrieval is the retrieval of vocabulary information from a database to meet user needs and query-related content. The ultimate effect of lexical retrieval recognition is to pave the way for French language learning. In our proposed model, fast retrieval and extraction of French phonetic words are reduced from a maximum of 11.65% to 4.25% in the WER rubric, with a maximum reduction of 7.4%.

(2) As complete and correct recognition is crucial in the application, it is worth mentioning that our training method based on alternating update network parameters for fast retrieval and extraction of French speech words reduces from a maximum of 13.52% to 4.68% in the SER rubric, with a maximum reduction of 8.84%.

(3) In addition, in the fast retrieval system, the response time for retrieval is a key factor affecting the user experience. In our model, the response time criterion is reduced from a maximum of 582 ms to 351 ms, with a maximum reduction of 39.7%.

(4) In French speech vocabulary, SGD synchronous updating network and alternating updating network parameter algorithm are used to quickly retrieve and extract French words. When the number of iterations reaches 120, the model fitting accuracy of the training set reaches 90.05%, while the model can reach 94.5% in the test set.

Data Availability

The dataset can be obtained from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Science Planning Project of Education of Zhejiang Province, whose title is "Research on the Model of Language Teaching in Universities from an Intercultural Point of View, Taking French as an Example" (project no. 2021SCG339) and the collaborative project between institutional establishments and companies called by the Chinese Ministry of Education in 2021, whose title is "Innovative Reform of the Teaching of French Courses in Universities" (project no. 202101163018).

References

[1] P. Quesnay, M. Gagnayre, and R. Gagnayre, "How does French initial osteopathic training value relational competency, patient education, and the competency-based approach? A cross-sectional survey," *International Journal of Osteopathic Medicine*, vol. 42, pp. 43–50, 2021.

[2] E. Langan, "The normative effects of higher education policy in France," *International Journal of Educational Research*, vol. 53, pp. 32–43, 2012.

[3] Z. Wang, Y. Mao, and X. Mao, "Exploring the value of overseas biodiversity to Chinese netizens based on willingness to pay for the African elephants' protection," *Science of The Total Environment*, vol. 637-638, pp. 600–608, 2018.

[4] R. A. Slaughter, "The IT revolution reassessed part one: literature review and key issues," *Futures*, vol. 96, pp. 115–123, 2018.

[5] G Raffaele, S Stefano, D. P Giuseppe, F Hamido, and E Massimo, "BERT syntactic transfer: a computational experiment on Italian, French and English languages," *Computer Speech & Language*, vol. 71, Article ID 101261, 2022.

[6] C. Cohen, E. Minniear, and J. Minniear, "Exploring how language exposure shapes oral narrative skills in French-English emergent bilingual first graders," *Linguistics and Education*, vol. 63, Article ID 100905, 2021.

[7] B. van Pinxteren, "Language of instruction in education in Africa: how new questions help generate new answers," *International Journal of Educational Development*, vol. 88, Article ID 102524, 2022.

[8] H Sakhaa, S Yuyao, E Farzad, and G Mehdi, "Speech and language abnormalities in myotonic dystrophy: an overview," *Journal of Clinical Neuroscience*, vol. 96, pp. 212–220, 2022.

[9] L Baolong, Z Qi, W Yabing, Z Minsong, D Jianfeng, and W Xun, "FeatInter: exploring fine-grained object features for video-text retrieval," *Neurocomputing*, vol. 496, pp. 178–191, 2022.

[10] Y. Qin, X. Tan, and Z. Tan, "Visual context learning based on textual knowledge for image-text retrieval," *Neural Networks*, vol. 152, pp. 434–449, 2022.

[11] P Sophie, M Henri, P Gilles, G Julien, P Cristian, and D Alain, "Does change in language change the properties of a shortened score previously validated in its complete version? Validation of the French versions of the HOOS-12 and KOOS-12 scores in primary knee and hip arthroplasties," *Orthopaedics and Traumatology: Surgery & Research*, vol. 107, no. 3, Article ID 102824, 2021.

[12] F. Zhang, J. Mairin, and H. Mairin, "Research on language learning motivation in school settings in System," *System*, vol. 107, Article ID 102817, 2022.

[13] E. T. Vold and A. Brkan, "Classroom discourse in lower secondary French-as-a-foreign-language classes in Norway: amounts and contexts of first and target language use," *System*, vol. 93, Article ID 102309, 2020.

[14] Z Aoxiao, L Xiang, W Dufan et al., "Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19," *Medical Image Analysis*, vol. 70, Article ID 101993, 2021.

[15] F. Da, G. Peng, and Y. Peng, "Deep learning based dual encoder retrieval model for citation recommendation," *Technological Forecasting and Social Change*, vol. 177, Article ID 121545, 2022.

[16] S. Pang, M. A. Yu, and Z. Yu, "A novel biomedical image indexing and retrieval system via deep preference learning," *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 53–69, 2018.

[17] L. V. B. Beltrán, C. C Juan, J Nicholas, C Mickaël, L François, and D Antoine, "Deep multimodal learning for cross-modal retrieval: one model for all tasks," *Pattern Recognition Letters*, vol. 146, pp. 38–45, 2021.

[18] H Shufeng, S Chongyang, C Longbing, N Zhendong, and G Ping, "Learning deep relevance couplings for ad-hoc document retrieval," *Expert Systems with Applications*, vol. 183, Article ID 115335, 2021.

[19] N. Gozuacik, C. O. Ozcan, and S. Ozcan, "Social media-based opinion retrieval for product analysis using multi-task deep neural networks," *Expert Systems with Applications*, vol. 183, Article ID 115388, 2021.

[20] P. P. Roy, A. K. Pal, and U. Pal, "Date-field retrieval in scene image and video frames using text enhancement and shape coding," *Neurocomputing*, vol. 274, pp. 37–49, 2018.

[21] L Zhixin, X Xiumin, L Feng, M Huifang, and S Zhiping, "Matching images and texts with multi-head attention network for cross-media hashing retrieval," *Engineering Applications of Artificial Intelligence*, vol. 106, Article ID 104475, 2021.

[22] W Weiwei, S Yuming, Z Haofeng, and L Li, "Semantic-rebased cross-modal hashing for scalable unsupervised text-visual retrieval," *Information Processing & Management*, vol. 57, no. 6, Article ID 102374, 2020.

[23] T. Bogers and V. Petras, "Supporting book search: a comprehensive comparison of tags vs. Controlled vocabulary metadata," *Data and information management*, vol. 1, no. 1, pp. 17–34, 2017.

[24] D. Yang, Y. Furui, and S. Furui, "Vocabulary expansion through automatic abbreviation generation for Chinese voice search," *Computer Speech & Language*, vol. 26, no. 5, pp. 321–335, 2012.

[25] N. Sa and X. J. Yuan, "Examining user perception and usage of voice search," *Data and information management*, vol. 5, no. 1, pp. 40–47, 2020.

[26] L Wei, X Yazhou, T Chao, H Xujing, and X Jianwu, "Multi-user searchable encryption voice in home IoT system," *Internet of Things*, vol. 11, Article ID 100180, 2020.

[27] W. H. Chung, Y. H. Gu, and S. J. Yoo, "District heater load forecasting based on machine learning and parallel CNN-LSTM attention," *Energy*, vol. 246, Article ID 123350, 2022.

[28] X. Fan, X. Feng, Y. Dong, and H. Hou, "COVID-19 CT image recognition algorithm based on transformer and CNN," *Displays*, vol. 72, Article ID 102150, 2022.

[29] X. Li, X. Ma, F. Xiao, C. Xiao, F. Wang, and S. Zhang, "Time-series production forecasting method based on the integration of bidirectional gated recurrent unit (Bi-gru) network and sparrow search algorithm (SSA)," *Journal of Petroleum Science and Engineering*, vol. 208, Article ID 109309, 2022.