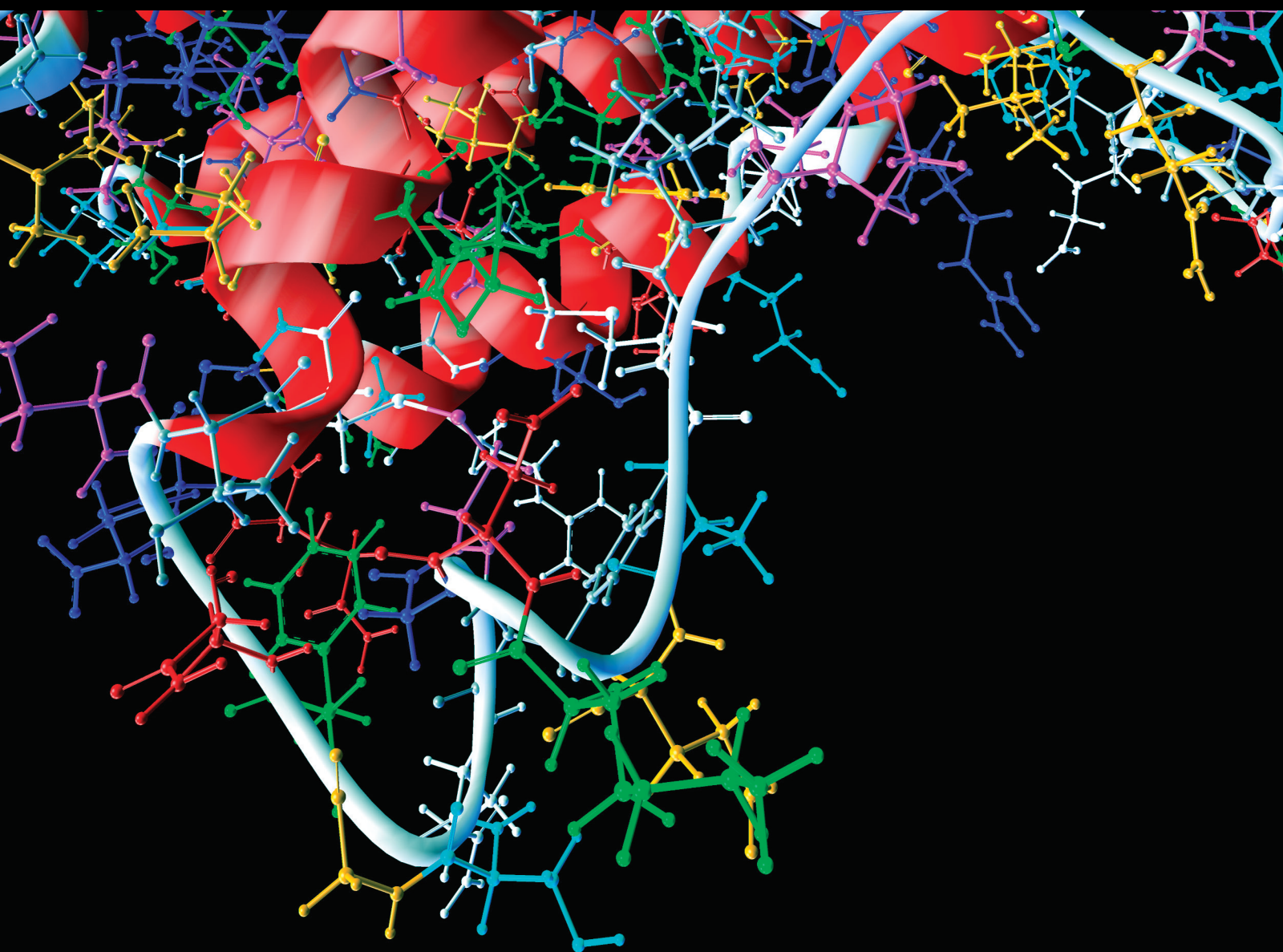


Computational and Mathematical Methods in Medicine

Artificial Intelligence and Cognitive Computing in Medical Image Processing

Lead Guest Editor: Siti Anom Ahmad

Guest Editors: Hamidreza Mohafez and Reza Khosrowabadi





Artificial Intelligence and Cognitive Computing in Medical Image Processing

Computational and Mathematical Methods in Medicine

Artificial Intelligence and Cognitive Computing in Medical Image Processing

Lead Guest Editor: Siti Anom Ahmad




Guest Editors: Hamidreza Mohafez and Reza
Khosrowabadi



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in “Computational and Mathematical Methods in Medicine.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Associate Editors

Ahmed Albahri, Iraq
Konstantin Blyuss , United Kingdom
Chuangyin Dang, Hong Kong
Farai Nyabadza , South Africa
Kathiravan Srinivasan , India

Academic Editors

Laith Abualigah , Jordan
Yaser Ahangari Nanehkaran , China
Mubashir Ahmad, Pakistan
Sultan Ahmad , Saudi Arabia
Akif Akgul , Turkey
Karthick Alagar, India
Shadab Alam, Saudi Arabia
Raul Alcaraz , Spain
Emil Alexov, USA
Enrique Baca-Garcia , Spain
Sweta Bhattacharya , India
Junguo Bian, USA
Elia Biganzoli , Italy
Antonio Boccaccio, Italy
Hans A. Braun , Germany
Zhicheng Cao, China
Guy Carrault, France
Sadaruddin Chachar , Pakistan
Prem Chapagain , USA
Huiling Chen , China
Mengxin Chen , China
Haruna Chiroma, Saudi Arabia
Watcharaporn Cholamjiak , Thailand
Maria N. D.S. Cordeiro , Portugal
Cristiana Corsi , Italy
Qi Dai , China
Nagarajan Deivanayagam Pillai, India
Didier Delignières , France
Thomas Desaive , Belgium
David Diller , USA
Qamar Din, Pakistan
Irina Doytchinova, Bulgaria
Sheng Du , China
D. Easwaramoorthy , India



Esmaeil Ebrahimie , Australia
Issam El Naqa , USA
Ilias Elmouki , Morocco
Angelo Facchiano , Italy
Luca Faes , Italy
Maria E. Fantacci , Italy
Giancarlo Ferrigno , Italy
Marc Thilo Figge , Germany
Giulia Fiscon , Italy
Bapan Ghosh , India
Igor I. Goryanin, Japan
Marko Gosak , Slovenia
Damien Hall, Australia
Abdulsattar Hamad, Iraq
Khalid Hattaf , Morocco
Tingjun Hou , China
Seiya Imoto , Japan
Martti Juhola , Finland
Rajesh Kaluri , India
Karthick Kanagarathinam, India
Rafik Karaman , Palestinian Authority
Chandan Karmakar , Australia
Kwang Gi Kim , Republic of Korea
Andrzej Kloczkowski, USA
Andrei Korobeinikov , China
Sakthidasan Sankaran Krishnan, India
Rajesh Kumar, India
Kuruva Lakshmana , India
Peng Li , USA
Chung-Min Liao , Taiwan
Pinyi Lu , USA
Reinoud Maex, United Kingdom
Valeri Makarov , Spain
Juan Pablo Martínez , Spain
Richard J. Maude, Thailand
Zahid Mehmood , Pakistan
John Mitchell , United Kingdom
Fazal Ijaz Muhammad , Republic of Korea
Vishal Nayak , USA
Tongguang Ni, China
Michele Nichelatti, Italy
Kazuhisa Nishizawa , Japan
Bing Niu , China

Hyuntae Park , Japan
Jovana Paunovic , Serbia
Manuel F. G. Penedo , Spain
Riccardo Pernice , Italy
Kemal Polat , Turkey
Alberto Policriti, Italy
Giuseppe Pontrelli , Italy
Jesús Poza , Spain
Maciej Przybyłek , Poland
Bhanwar Lal Puniya , USA
Mihai V. Putz , Romania
Suresh Rasappan, Oman
Jose Joaquin Rieta , Spain
Fathalla Rihan , United Arab Emirates
Sidheswar Routray, India
Sudipta Roy , India
Jan Rychtar , USA
Mario Sansone , Italy
Murat Sari , Turkey
Shahzad Sarwar, Saudi Arabia
Kamal Shah, Saudi Arabia
Bhisham Sharma , India
Simon A. Sherman, USA
Mingsong Shi, China
Mohammed Shuaib , Malaysia
Prabhishek Singh , India
Neelakandan Subramani, India
Junwei Sun, China
Yung-Shin Sun , Taiwan
Min Tang , China
Hongxun Tao, China
Alireza Tavakkoli , USA
João M. Tavares , Portugal
Jlenia Toppi , Italy
Anna Tsantili-Kakoulidou , Greece
Markos G. Tsipouras, North Macedonia
Po-Hsiang Tsui , Taiwan
Sathishkumar V E , Republic of Korea
Durai Raj Vincent P M , India
Gajendra Kumar Vishwakarma, India
Liangjiang Wang, USA
Ruisheng Wang , USA
Zhouchao Wei, China
Gabriel Wittum, Germany
Xiang Wu, China

KI Yanover , Israel
Xiaojun Yao , China
Kaan Yetilmezsoy, Turkey
Hiro Yoshida, USA
Yuhai Zhao , China








Contents

Tooth CT Image Segmentation Method Based on the U-Net Network and Attention Module

Sha Tao  and Zhenfeng Wang 

Research Article (8 pages), Article ID 3289663, Volume 2022 (2022)

MR Image Classification for Brain Tumor Texture Based on Pseudo-Label Learning and Optimized Feature Extraction

Qianqian Xu , Huachang Xu , Jie Liu , Mingxia Zhou , Min Li , Jinhui Xu , and Hong Zhu 

Research Article (11 pages), Article ID 7746991, Volume 2022 (2022)

A Broad Learning System to Predict the 28-Day Mortality of Patients Hospitalized with Community-Acquired Pneumonia: A Case-Control Study

Jing Yuan , Xin Liu , Wen-Feng Wang , and Jing-Jing Zhang




Research Article (10 pages), Article ID 7003272, Volume 2022 (2022)

A Hemolysis Image Detection Method Based on GAN-CNN-ELM

Xiaonan Shi , Yong Deng , Yige Fang , Yajuan Chen , Ni Zeng , and Limei Fu 






Research Article (12 pages), Article ID 1558607, Volume 2022 (2022)

Development and Validation of Prediction Model for High Ovarian Response in *In Vitro* Fertilization-Embryo Transfer: A Longitudinal Study

Xinsha Tan , Honglin Xi, Jing Yang , and Wenfeng Wang 




Research Article (12 pages), Article ID 7822119, Volume 2021 (2021)

An Approach for Thoracic Syndrome Classification with Convolutional Neural Networks

Sapna Juneja , Abhinav Juneja , Gaurav Dhiman , Sanchit Behl , and Sandeep Kautish 

Research Article (10 pages), Article ID 3900254, Volume 2021 (2021)







Multistage Optimization Using a Modified Gaussian Mixture Model in Sperm Motility Tracking

Mohammed Alameri, Khairunnisa Hasikin , Nahrizul Adib Kadri , Nashrul Fazli Mohd Nasir ,

Prabu Mohandas , Jerline Sheeba Anni , and Muhammad Mokhzaini Azizan 

Research Article (14 pages), Article ID 6953593, Volume 2021 (2021)




Low-Dose CT Image Denoising with Improving WGAN and Hybrid Loss Function

Zhihua Li , Weili Shi , Qiwei Xing , Yu Miao , Wei He , Huamin Yang , and Zhengang

Jiang 



Research Article (14 pages), Article ID 2973108, Volume 2021 (2021)

Algorithm of Pulmonary Vascular Segment and Centerline Extraction

Shi Qiu , Jie Lian, Yan Ding, Tao Zhou , and Ting Liang 







Research Article (10 pages), Article ID 3859386, Volume 2021 (2021)

Comparison of Diagnostic Performance between Perfusion-Related Intravoxel Incoherent Motion DWI and Dynamic Contrast-Enhanced MRI in Rectal Cancer

Ming Li, Xiaodan Xu, Kaijian Xia, Heng Jiang, Jianlong Jiang, Jinbing Sun , and Zhihua Lu 

Research Article (7 pages), Article ID 5095940, Volume 2021 (2021)

CT Image Analysis and Clinical Diagnosis of New Coronary Pneumonia Based on Improved Convolutional Neural Network

Wu Deng , Bo Yang , Wei Liu , Weiwei Song , Yuan Gao , and Jia Xu 
Research Article (12 pages), Article ID 7259414, Volume 2021 (2021)

Research Article

Tooth CT Image Segmentation Method Based on the U-Net Network and Attention Module

Sha Tao  and Zhenfeng Wang 

School of Electrical Engineering, Tongling University, Tongling 244000, China

Correspondence should be addressed to Zhenfeng Wang; wzf860826@163.com

Received 20 August 2021; Revised 5 July 2022; Accepted 19 July 2022; Published 19 August 2022

Academic Editor: Huiling Chen

Copyright © 2022 Sha Tao and Zhenfeng Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional image segmentation methods often encounter problems of low segmentation accuracy and being time-consuming when processing complex tooth Computed Tomography (CT) images. This paper proposes an improved segmentation method for tooth CT images. Firstly, the U-Net network is used to construct a tooth image segmentation model. A large number of feature maps in downsampling are supplemented to downsampling to reduce information loss. At the same time, the problem of inaccurate image segmentation and positioning is solved. Then, the attention module is introduced into the U-Net network to increase the weight of important information and improve the accuracy of network segmentation. Among them, subregion average pooling is used instead of global average pooling to obtain spatial features. Finally, the U-Net network combined with the improved attention module is used to realize the segmentation of tooth CT images. And based on the image collection provided by West China Hospital for experimental demonstration, compared with other algorithms, our method has better segmentation performance and efficiency. The contours of the teeth obtained are clearer, which is helpful to assist the doctor in the diagnosis.

1. Introduction

In recent years, deep learning has developed rapidly and has been applied to many fields of image processing [1]. Among them, image segmentation is one of the key topics in modern computer vision research [2]. The effect of image segmentation in the medical field is getting more and more attention. Whether it is pathological analysis, diagnosis and treatment, or clinical surgical navigation, image segmentation can play a better auxiliary treatment effect. It can quickly identify the lesion and reduce the cost of medical care [3]. Traditional image segmentation is done manually layer by layer by experienced doctors, which takes a long time. The segmentation results of these regions of interest (RoI) are quite different among different doctors [4]. The automatic segmentation of CT images can significantly reduce the workload of doctors and improve the accuracy and consistency of RoI segmentation [5]. At present, modern medical CT mainly includes multiple spiral CT and cone beam CT (CBCT). CBCT is more commonly used in dentistry, which has the

characteristics of short scanning time and low radiation dose [6]. Because the bones of the teeth are closely intertwined with some soft tissues, the CT scan image of the tooth will become very blurred. And there are differences in the grayscale value distribution between teeth. There may also be differences in the grayscale value of the same tooth, the topological structure is more complicated, and it is not easy to distinguish [7]. Therefore, accurate and efficient medical image segmentation algorithms are necessary.

Image segmentation currently does not have a unified general-purpose algorithm. Most algorithms are aimed at specific types of objects, looking for their characteristic points to construct specific energy functions to achieve segmentation [8, 9]. In medical CT image segmentation, the traditional template-based segmentation method has gradually become one of the common methods of RoI delineation in radiotherapy [10]. Kaur et al. [11] proposed a CT image segmentation technique based on a mixture of spatial intuitionistic fuzzy C-means clustering and spatial regularization level set methods. It can integrate the details of the spatial

image and accurately and effectively extract the kidney disease image in the CT image. Li et al. [12] proposed a new validity index after evaluating the tightness and separation of the dataset twice. The effectiveness of the proposed metrics is verified using a large number of synthetic datasets and some typical CT images. This metric is suitable for datasets with arbitrarily shaped clusters that are susceptible to noise corruption. It is beneficial to cluster analysis and computer-aided detection systems. Shi et al. [13] proposed an improved potential well function combined with a level set model for tooth segmentation. The segmentation speed and accuracy are improved. Ji et al. [14] proposed a level set method for tooth shape extraction from head CBCT images, which achieved high accuracy.

However, due to the frequent changes in the anatomy of human organs, it is difficult for traditional methods to establish a universal template based on fixed images. Moreover, even the same RoI may have large differences due to differences in patient size and age [15]. In recent years, deep learning methods have been widely used. For example, stacked autoencoder, deep belief networks (DBN), recurrent neural network (RNN), and convolutional neural network (CNN) have been extensively studied in related fields [16, 17]. Sun [18] proposed the Darknet-53 network for extracting tooth image features and multiscale prediction. The SE module is added to the network to improve the network performance. Li et al. [19] proposed a CNN-based stereo image segmentation framework. The use of coherent parallax propagation solves the inevitable inaccuracy problem of the parallax calculated from the stereo image pair. Then, the pixels propagated through the coherent parallax and the high-confidence pixels of the target probability map generated by the CNN architecture will be used to generate the initial reliable pixels to perform the segmentation based on the energy minimization framework. CNN automatically extracts multilevel visual features by learning a large number of mapping relationships between inputs and outputs [20]. Zhou et al. [21] proposed a three-level network based on 3D U-Net to segment a single tooth. But the relationship between alveolar bone and teeth is very important when orthodontists make orthodontic plans for patients. Kadry et al. [22] used the pretrained VGG-UNet system to extract ISL fragments from brain MRI slice test images and extract COVID-19 infection from lung CT images. Tong et al. [23] propose an extended version of U-Net, which uses the concept of triple attention mechanism to segment skin lesions. The segmentation results are more accurate for complex skin lesions such as shape, size, color, and texture. But the algorithm is more complicated and time-consuming. Zhang et al. [24] proposed a triple crossover U-Nets (TIU-Nets) for glioma segmentation. The proposed TIU-Nets are composed of binary class segmentation U-Net (BU-Net) and multiclass segmentation U-Net (MU-Net) and have achieved better performance. Tooth CT images are relatively simple for skin lesion images and brain MRI images. Miki et al. [25] studied the application of CNN to classify tooth types on dental CT images, which is beneficial to obtain high classification accuracy without

the need for accurate tooth segmentation. Chung et al. [26] proposed a neural network for per-pixel labeling to segment frames using instances that are robust to metal artifacts.

Compared with CNN, U-Net adds upsampling and downsampling processes, generates images of different resolutions, and convolves images of various resolutions [27]. To this end, a tooth CT image segmentation method using the U-Net network is proposed. For this paper, the main contributions are as follows:

- (1) Since the segmentation effect of fully convolutional networks (FCN) is relatively rough, the proposed method uses the U-Net network for image segmentation. And through data enhancement, a more ideal tooth CT image segmentation result can be obtained
- (2) In order to improve the segmentation speed of the network and reduce the loss of precision, the attention module is integrated in the U-Net network and improved. The Subregional Average Pooling (SAP) module is used to obtain the initial channel weight information, and the CE module is used to expand the low-level channel attention

This paper is divided into four sections in total. Section 1 introduces the research significance and technical difficulties of topic selection and classifies the existing image semantic segmentation methods. On this basis, summarize the innovative points of the proposed method. Section 2 elaborates on the proposed method, using the U-Net network combined with the improved attention module to segment the tooth CT image. Section 3 discusses the experimental results and uses relevant standards to evaluate the performance of the proposed method. Section 4 is the conclusion.

2. Materials and Methods

2.1. U-Net. U-Net is a U-Net network structure proposed by Olaf Ronneberger in 2015. It is a semantic segmentation network based on FCN, which matches the form of downsampling during the upsampling process to keep it consistent [28, 29]. On this basis, a large number of feature maps in the downsampling stage are added to the downsampling to fill in the information lost during the operation. Because its structure is symmetrical like U-shaped, it is named the U-Net network.

U-Net has special advantages in processing medical images. In solving the problem of the lack of sample images in medical images, elastic deformation is used to complete data enhancement [30]. Elastic deformation is a relatively common type of deformation in actual cells, so it is very suitable for medical image processing [31]. The algorithm of data enhancement is used to make the neural network model learn the invariance of elastic deformation. In the case of a small dataset, the network has good elastic deformation adaptability. The segmentation can be completed correctly when meeting the elastic deformation of the medical image [32]. The U-Net network has excellent effect on cell segmentation, and its segmentation effect is shown in Figure 1.

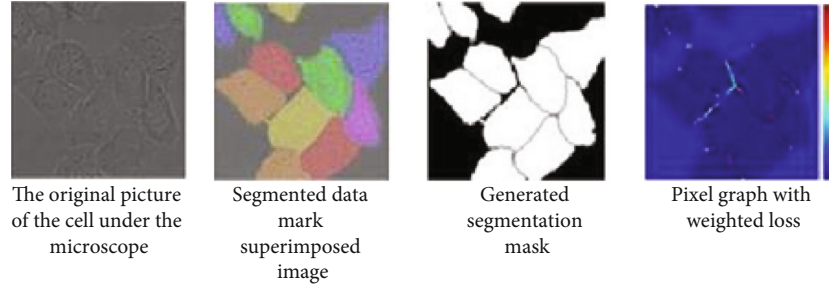


FIGURE 1: U-Net network segmentation cell diagram.

The U-Net network has attracted attention from the medical field since its inception. On the basis of U-Net, many scholars have improved this structure. For example, Milletari's 3D deformation structure V-Net and Drozdal's changes to the skip structure have added many variants to the U-Net network.

2.2. Attention Module. In order to increase the weight of effective features and improve the accuracy of segmentation, we have added an attention mechanism module to the network [33]. The human visual attention mechanism is divided into data-driven and task-driven [34]. This paper uses a data-driven attention mechanism module. The level of attention weight is positively correlated with the importance of the corresponding location information. The input unit with high weight has a decisive effect on the output result. Firstly, global average pooling (GAP) is performed on each channel to obtain the vector of $1 \times 1 \times K$. More than two fully convolutional (FC) layer transformations are done. Sigmoid and ReLU activation functions are used. The extraction network is shown in Figure 2.

In the attention module, first use GAP to shrink the data output from the downsampling layer from $W \times H \times K$ to $1 \times 1 \times K$, which is expressed as follows:

$$z_c = f_{sq}(u_c) \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), \quad (1)$$

where f_{sq} represents the GAP function.

After that, the network performs two FC operations. C represents the dimensionality reduction coefficient, which can be adjusted according to the specific network. The best performance is obtained when $C = 16$ is demonstrated experimentally; the formula is as follows:

$$s_c = f_{ex}(z_c, W) = \varepsilon(g(z_c, W)) = \varepsilon(W_2 \delta(W_1 z_c)),$$

$$W_1, W_2 \in R^{(K/C) \times K}, \quad (2)$$

where ε and δ represent sigmoid and ReLU activation functions, respectively. f_{ex} is the FC function.

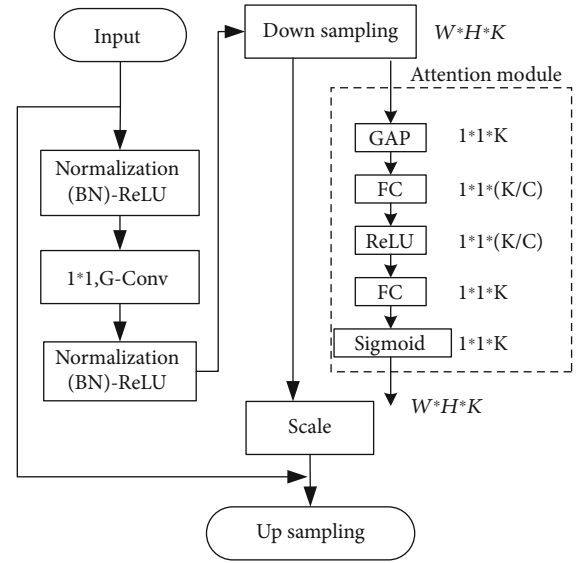


FIGURE 2: Feature extraction network with the attention module.

Then, perform the scale operation. Based on the number of channels unchanged, the data is changed to $W \times H \times K$; the calculation is as follows:

$$X_c = f_{scale}(u_c, s_c) = s_c \cdot u_c, \quad (3)$$

where f_{scale} is the scale operation function.

Before the network joins the attention module, for an input image with a batch of 256 and a size of 224×224 , a forward propagation takes 42 ms. After adding the attention module, it takes 47 ms. After reducing the complexity, there is still time to increase, but it can be ignored compared with the improvement of segmentation accuracy [35].

Because segmentation task scenes are often complex and changeable, using GAP is too simple. The features needed for segmentation cannot be extracted effectively [36]. In response to this problem, the SAP structure was proposed. The structures of GAP and SAP are shown in Figure 3.

As shown in Figure 3, GAP turns a channel feature with a dimension of $H \times W \times K$ into a feature with the size of $1 \times 1 \times K$. In SAP, a feature with a size of $H \times W \times K$ is transformed into a new feature with a size of $m \times n \times k$, so that some spatial information is retained. Then, through a convolution kernel of size $m \times n$, the obtained feature

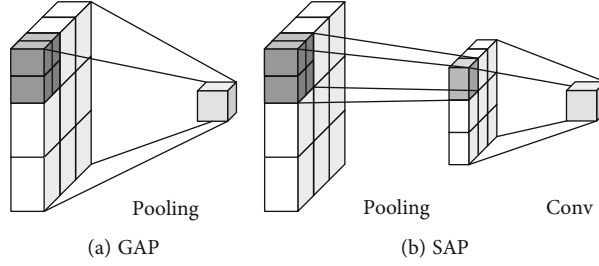


FIGURE 3: Structure comparison between GAP and SAP.

map is transformed into a feature of size $H \times W \times K$. The feature extraction network with the attention module is shown in Figure 4.

First increase the number of channels and use a 1×1 convolution. Because 1×1 convolutional layers are usually used to adjust the number of channels between network layers and control the model complexity, it enables interaction and information integration across channels. At the same time, the nonlinear characteristics can be greatly increased while keeping the size of the feature map unchanged, making the network very deep. Then, add the high-level and low-level channel attention weights. Finally, multiply the new weights with higher-order features [37, 38]. The size range of the updated weight is (0, 2), which can not only reduce but also expand the feature value [39].

The obtained channel feature \hat{F} connects a 1×1 convolution φ_r and sigmoid activation function ε , $r \in [1, K]$. The output ε_{SAP} of the SAP module is

$$\varepsilon_{\text{SAP}} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_r, \dots, \varepsilon_K],$$

$$\varepsilon_r = \varepsilon \left(\sum_{i=1}^K \varphi_r * \hat{F}_i \right). \quad (4)$$

Finally, define the high-order attention distribution obtained by the SAP module as ε_h . The low-level attention distribution after passing the CE module is $\varepsilon_l(\cdot)$. Then, it can be used for the updated attention distribution ε_F :

$$\varepsilon_F = \varepsilon_l(\cdot) + \varepsilon_h(\cdot). \quad (5)$$

The recalibrated characteristic map \hat{U} can be obtained as

$$\hat{U} = U\varepsilon_F. \quad (6)$$

2.3. Network Structure. The attention module is integrated into the U-Net network to obtain a network model for tooth CT image segmentation. The structure is shown in Figure 5.

Using 3 layers of 3×3 convolutional layers, change the ordinary 3×3 convolution to dilated convolution with expansion rate 2, and the resolution of the output feature map is 1/16 of the input. Atrous spatial pyramid pooling (ASPP) is used to provide multiscale information in the U-Net network. That is, on the basis of spatial pyramid pooling, a hole convolution with different expansion rates is added, and global information is added through SAP

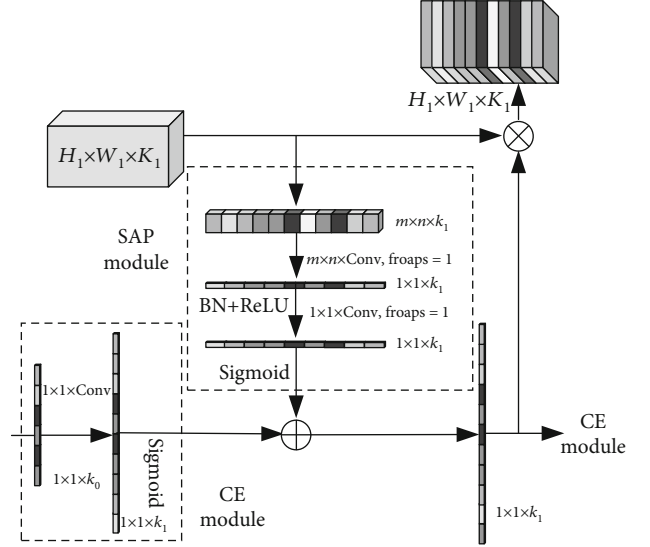


FIGURE 4: Feature extraction network with the attention module.

and image features. The features output by ASPP are upsampled, added to a low-level feature that has undergone a 3×3 convolution, and then merged with a 3×3 convolution. Finally, it is upsampled and restored to the original image size [40].

3. Results and Discussion

The experiment is based on the Keras framework, which is developed by Google. It uses TensorFlow and Theano packages to integrate many basic neural network structures and some mature algorithms. The experimental dataset used is dental CT data provided by West China Hospital. The operating system is Windows 10, the CPU is an 8 GB Intel Core i7-6700, and the GPU is NVIDIA Ge Force GTX 1070.

When training, the optimization method uses random gradient descent (SGD). Its momentum parameter is set to 0.9, and the weight decay rate is 1×10^{-4} . The learning rate uses the initial learning rate learning rate = 8×10^{-3} multiplied by $(1 - \text{current_iter}/\text{max_iter})^{\text{power}}$ to reduce the strategy. Where power = 0.9, current_iter is the current number of iterations, and max_iter is the maximum number of iterations in the training process. The number of experimental training iterations is 800.

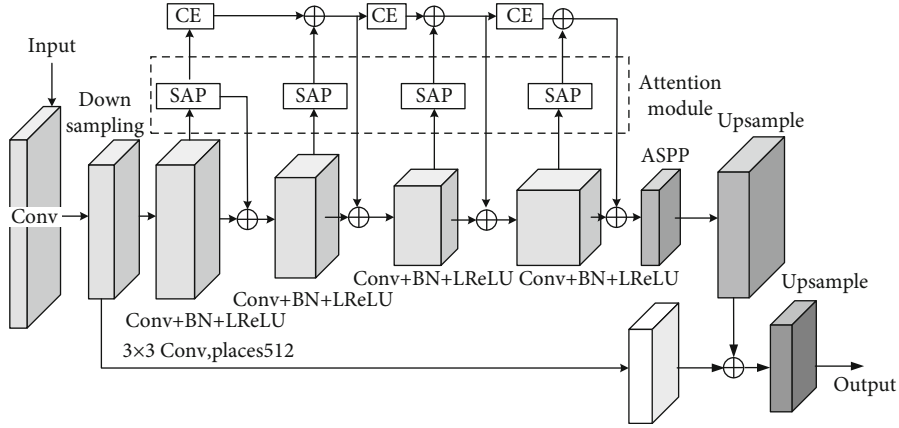


FIGURE 5: U-Net network structure combined with the improved attention module.

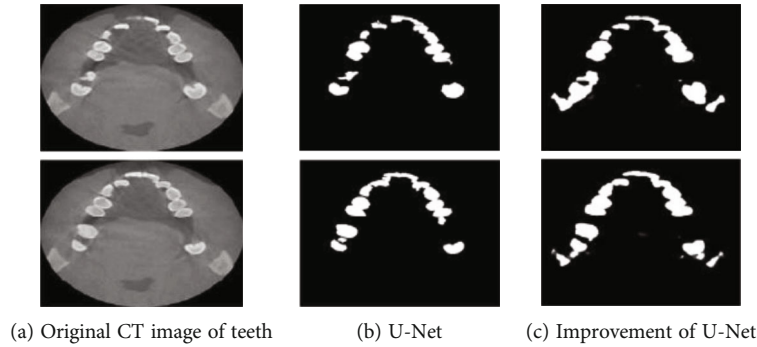


FIGURE 6: Comparison of segmentation results before and after the improvement of the U-Net network.

3.1. Data Preprocessing. Based on the segmentation task of the U-Net network, the final output is closely related to the quality of the input image. In view of the fact that some of the tooth CT images have similar grayscale value, shapes, or textures, image enhancement methods are used to preprocess the input image data to improve image contrast. Histogram equalization (HE) is a basic method of image enhancement. And adaptive histogram equalization (AHE) is improved through HE. AHE improves local contrast, enhances the edge sharpness of each area of the image, and improves the shape, texture, and boundary information of the teeth. Before entering the network, the normalized image grayscale value is (0,1), and the image size is 512×512 .

In order to overcome the overfitting phenomenon of neural networks, the proposed method adopts enhancement techniques such as random cropping, flipping, grayscale perturbation, and shape perturbation on the data image to overcome this problem. The gray level disturbance in the training set can improve the stability of the network, thereby improving the performance of the prediction set network.

The CT image and contour image are deformed by affine transformation to form shape disturbance. In the deformation, the coordinates of the 3 vertices (upper left, upper right, and lower left) are first obtained. Then, each point moves randomly, and the range of random movement is the image length. Finally, affine transformation is performed on the entire image.

TABLE 1: Comparison of segmentation performance before and after the improvement of the U-Net network.

Methods	MPA (%)	MIoU (%)	Execution time (ms)
U-Net	82.37	80.19	98
Improved U-Net	85.91	83.73	126

After image enhancement, the training set (400 images) and the validation set (100 images) are both enlarged 3 times. A total of 1500 image data were obtained for experimental training and parameter adjustment.

3.2. Evaluation Index. In the experiment, mean pixel accuracy (MPA) and mean intersection over union (MIoU) are used as indicators to evaluate the performance of the proposed method.

Assume that n_{ii} represents the number of correct segmentation. n_{ij} represents the number of pixels that originally belonged to class i but were divided into class j . n_{ji} represents the number of pixels that originally belonged to class j but were divided into class i . There are $k + 1$ categories (including k classes and an empty class or background class).

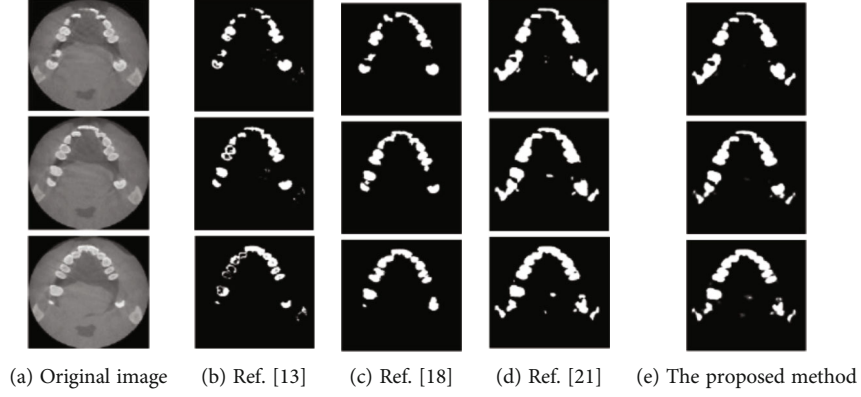


FIGURE 7: Segmentation results of different methods.

MPA calculates the proportion of pixels that are correctly segmented in each class and then finds the average of all classes. The calculation is as follows:

$$\text{MPA} = \frac{1}{k+1} \sum_{i=0}^k \frac{n_{ii}}{\sum_{j=0}^k n_{ij}}. \quad (7)$$

MIOU calculates the ratio of intersection and union of two sets. Calculate the pixel intersection ratio within each pixel category, and then, calculate the average:

$$\text{MIOU} = \frac{1}{k+1} \sum_{i=0}^k \frac{n_{ii}}{\sum_{j=0}^k n_{ij} + \sum_{j=0}^k n_{ji} - n_{ii}}. \quad (8)$$

Due to its strong representativeness, high efficiency, and simplicity, MIOU has become the current general image segmentation evaluation index. Therefore, MIOU is used as the main evaluation index of the experiment.

3.3. Comparison of the Proposed Algorithms. According to the proposed method, the final tooth CT image segmentation result is obtained. The segmentation comparison result of the proposed method and the traditional U-Net network is shown in Figure 6.

It can be seen from Figure 6 that the proposed method can segment the tooth CT image more effectively than the traditional U-Net network. The effect of tooth edge extraction is significantly better than that of U-Net. However, affected by some noise, there is a certain shadow in the center of the segmented image. Therefore, the performance of the U-Net network before and after the improvement is quantitatively analyzed. The results of MPA and MIOU and the time to complete the segmentation are shown in Table 1.

As can be seen from Table 1, the proposed method improves the U-Net network by using the attention mechanism. MPA is 85.91%, and MIOU is 83.73%, which is superior to the U-Net network in terms of segmentation accuracy. However, due to the complexity of the segmentation model, the time to complete the segmentation has also increased. Compared with the U-Net network, the time has increased by 28 ms. However, from an overall point of view,

TABLE 2: Comparison of segmentation performance of different methods.

Methods	MPA (%)	MIOU (%)	Execution time (ms)
Ref. [13]	69.11	68.05	74
Ref. [18]	80.03	79.24	167
Ref. [21]	84.76	83.19	183
The proposed method	85.91	83.73	126

the proposed method still has a large application advantage in tooth CT image segmentation.

In order to further demonstrate the performance of the proposed method, compare it with References [13, 18] and [21]. The segmentation results of tooth CT images obtained by each method are shown in Figure 7.

It can be seen from Figure 7 that the tooth profile obtained by the proposed method is clearer and is closest to the original image. Figure 7(b) adopts the traditional segmentation method combined with the level set model [13], the segmentation effect is not good, and the contour is blurred. Figures 7(c) and 7 (d) use learning algorithms for segmentation [18, 21], but a single algorithm needs to be improved in processing complex tooth images. Therefore, the effectiveness of the proposed method for segmentation of tooth CT images can be demonstrated.

The quantitative analysis results of different methods are shown in Table 2.

It can be seen from Table 2 that the segmentation performance of the proposed method is the best, with MPA of 85.91% and MIOU of 83.73%. It completes the detection speed of 126 ms, but the model is more complicated than that of Reference [13], so the time is slightly longer. The traditional method used in Reference [13] has low segmentation accuracy for complex teeth CT images, with MIOU of only 68.05%. But it takes the shortest time to complete the detection, which is 74 ms. Reference [18] chooses the Darknet-53 network for extracting tooth image features and multiscale prediction. Overall, it has good segmentation performance, but the algorithm is not effective in learning complex features, and the segmentation accuracy needs to

be improved. Reference [21] segmented teeth based on a three-level network of 3D U-Net. But the model is more complicated, so it takes a long time, 183 ms. From a comprehensive point of view, the proposed method has obvious application advantages in tooth CT image segmentation.

4. Conclusions

With the development of today's society, medical standards have put forward higher requirements in people's lives. There are many representative computational intelligence algorithms, such as monarch butterfly optimization (MBO) [41], earthworm optimization algorithm (EWA) [42], elephant herding optimization (EHO) [43], and moth search (MS) algorithm [44, 45]. With the continuous increase in the number of CT images and MRI images, the segmentation of medical images by manpower alone can no longer meet the demand. Therefore, the use of deep learning for automatic segmentation has attracted wide attention from scholars. We propose a tooth CT image segmentation method using the U-Net network. The attention module is integrated into the U-Net network, and the SAP and CE modules are used to improve the performance of the network. Based on the tooth CT image dataset provided by West China Hospital, the improved network was demonstrated experimentally. With comprehensive comparison, the method in this paper has better segmentation performance and segmentation efficiency. The contours of the teeth obtained are clearer, which is helpful to assist the doctor in the diagnosis. The proposed method only performs segmentation for tooth images and has not been used for segmentation of other medical images. In the following research, the proposed method will be applied to other organizations to improve the generalization ability of the proposed method. And in-depth pruning of the network is carried out to improve network efficiency.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Anhui Provincial Natural Science Key Foundation (No. KJ2018A0484).

References

- [1] Y. Chen, L. Liu, J. Tao et al., "The improved image inpainting algorithm via encoder and similarity constraint," *The Visual Computer*, vol. 37, no. 7, pp. 1691–1705, 2021.
- [2] Y. Chen, H. Zhang, L. Liu et al., "Research on image inpainting algorithm of improved total variation minimization method," *Journal of Ambient Intelligence and Humanized Computing*, vol. 6, no. 8, pp. 271–280, 2021.
- [3] F. Guo, C. Shi, X. Li, X. Wu, J. Zhou, and J. Lv, "Image segmentation of nasopharyngeal carcinoma using 3D CNN with long-range skip connection and multi-scale feature pyramid," *Soft Computing*, vol. 24, no. 16, pp. 12671–12680, 2020.
- [4] E. Cuevas, H. Becerra, A. Luque, and M. A. Elaziz, "Fast multi-feature image segmentation," *Applied Mathematical Modelling*, vol. 90, no. 9, pp. 742–757, 2021.
- [5] X. Jiang, H. Yu, and S. Lv, "An image segmentation algorithm based on a local region conditional random field model," *International Journal of Communications, Network and System Sciences*, vol. 13, no. 9, pp. 139–159, 2020.
- [6] H. Tonbul and T. Kavzoglu, "A spectral band based comparison of unsupervised segmentation evaluation methods for image segmentation parameter optimization," *International Journal of Environment and Geoinformatics*, vol. 7, no. 2, pp. 132–139, 2020.
- [7] T. Song, F. Meng, A. Rodriguez-Paton, P. Li, P. Zheng, and X. Wang, "U-next: a novel convolution neural network with an aggregation U-net architecture for gallstone segmentation in CT images," *Access*, vol. 7, no. 6, pp. 166823–166832, 2019.
- [8] Y. Wang, H. Wang, K. Shen, J. Chang, and J. Cui, "Brain CT image segmentation based on 3D slicer," *Journal of Complexity in Health Sciences*, vol. 3, no. 1, pp. 34–42, 2020.
- [9] H. Su, D. Zhao, H. Elmannai et al., "Multilevel threshold image segmentation for COVID-19 chest radiography: a framework using horizontal and vertical multiverse optimization," *Computers in Biology and Medicine*, vol. 146, article 105618, 2022.
- [10] L. Zhou and Z. Zhang, "An image segmentation method based on Mumford-Shah model with mask factor and neighborhood factor," *Pattern Analysis and Applications*, vol. 23, no. 1, pp. 85–94, 2020.
- [11] R. Kaur, M. Juneja, and A. K. Mandal, "A hybrid edge-based technique for segmentation of renal lesions in CT images," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 12917–12937, 2019.
- [12] Q. Li, S. Yue, M. Ding, J. Li, and Z. Wang, "Determining the correct number of clusters in the CT image segmentation," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 11, pp. 2675–2680, 2020.
- [13] Q. Shi, F. Yan, Y. Yang, Y. Chen, X. Lin, and Y. Wang, "Image segmentation of tooth and alveolar bone with the level set model," *Chinese Journal of Magnetic Resonance*, vol. 38, no. 2, pp. 182–193, 2021.
- [14] D. X. Ji, S. H. Ong, and K. Foong, "A level-set based approach for anterior teeth segmentation in cone beam computed tomography images," *Computers in Biology & Medicine*, vol. 50, pp. 116–128, 2014.
- [15] S. Zhou and Z. Xu, "Automatic grayscale image segmentation based on affinity propagation clustering," *Pattern Analysis and Applications*, vol. 23, no. 1, pp. 331–348, 2020.
- [16] S. Wang, Y. Cong, H. Zhu et al., "Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 514–525, 2021.
- [17] B. Lu, "Algorithm improvement of neural network in endoscopic image recognition of upper digestive tract system," *Expert Systems*, vol. 2, no. 6, article e12912, pp. 311–320, 2021.
- [18] S. Z. H. Tooth, *Segmentation Method Based on Deep Learning*, Xi'an University of Science and Technology, 2020.

- [19] X. Li, H. Huang, H. Zhao, Y. Wang, and M. Hu, "Learning a convolutional neural network for propagation-based stereo image segmentation," *The Visual Computer*, vol. 36, no. 1, pp. 39–52, 2020.
- [20] A. P. Bajcsy, S. Feldman, M. Brady, and K. Snyder, "Semantic SEM image segmentation of concrete with contextual labels," *Microscopy and Microanalysis*, vol. 25, no. S2, pp. 144–145, 2019.
- [21] X. R. Zhou, W. Y. Jin, and L. Chai, "Dental CBCT image segmentation," *Modern Computer*, vol. 4, pp. 122–126, 2021.
- [22] S. Kadry, R. Damaševičius, D. Taniar, V. Rajinikanth, and I. A. Lawal, "U-Net supported segmentation of ischemic-stroke-lesion from brain MRI slices," in *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pp. 1–5, Chennai, India, 2021.
- [23] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, and P. Wu, "ASCU-Net: attention gate, spatial and channel attention U-Net for skin lesion segmentation," *Diagnostics*, vol. 11, no. 3, p. 501, 2021.
- [24] J. Zhang, J. Zeng, P. Qin, and L. Zhao, "Brain tumor segmentation of multi-modality MR images via triple intersecting U-Nets," *Neurocomputing*, vol. 421, pp. 195–209, 2021.
- [25] Y. Miki, C. Muramatsu, T. Hayashi et al., "Classification of teeth in cone-beam CT using deep convolutional neural network," *Computers in Biology & Medicine*, vol. 80, pp. 24–29, 2017.
- [26] M. Chung, M. Lee, J. Hong et al., "Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation," *Computers in Biology and Medicine*, vol. 120, no. 11, article 103720, 2020.
- [27] C. Ye, W. Wang, S. Zhang, and K. Wang, "Multi-depth fusion network for whole-heart CT image segmentation," *IEEE Access*, vol. 7, no. 5, pp. 23421–23429, 2019.
- [28] E. A. Audenaert, J. Van Houcke, D. F. Almeida et al., "Cascaded statistical shape model based segmentation of the full lower limb in CT," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 22, no. 6, pp. 644–657, 2019.
- [29] A. Kozei, N. Nikolov, O. Haluzynskyi, and S. Burburska, "Method of threshold CT image segmentation of skeletal bones," *Innovative Biosystems and Bioengineering*, vol. 3, no. 1, pp. 4–11, 2019.
- [30] R. Kulkarni and K. Bhavani, "Analysis of CT DICOM image segmentation for abnormality detection," *International Journal of Engineering and Manufacturing*, vol. 9, no. 5, pp. 46–55, 2019.
- [31] Z. Liu, Y. Q. Song, V. S. Sheng et al., "Liver CT sequence segmentation based with improved U-Net and graph cut," *Expert Systems with Application*, vol. 126, no. 7, pp. 54–63, 2019.
- [32] M. A. Koenrades, E. M. Struijs, A. Klein et al., "Quantitative stent graft motion in ECG gated CT by image registration and segmentation: in vitro validation and preliminary clinical results," *European Journal of Vascular and Endovascular Surgery*, vol. 58, no. 5, pp. 746–755, 2019.
- [33] P. K. Barik and R. C. Barik, "Feed forwarded CT image registration for tumour and cyst detection using rigid transformation with HSV colour segmentation," *International Journal of Computational Systems Engineering*, vol. 5, no. 5/6, pp. 277–286, 2019.
- [34] C. Fan, Z. Wang, G. Li et al., "A novel 3D U-Net deep network with paralleling structure for stroke lesion image segmentation," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 3, pp. 724–730, 2020.
- [35] B. Park, H. Park, S. M. Lee, J. B. Seo, and N. Kim, "Lung segmentation on HRCT and volumetric CT for diffuse interstitial lung disease using deep convolutional neural networks," *Journal of Digital Imaging*, vol. 32, no. 6, pp. 1019–1026, 2019.
- [36] R. Hemke, C. G. Buckless, A. Tsao, B. Wang, and M. Torriani, "Deep learning for automated segmentation of pelvic muscles, fat, and bone from CT studies for body composition assessment," *Skeletal Radiology*, vol. 49, no. 3, pp. 387–395, 2020.
- [37] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.
- [38] X. Dong, Y. Lei, S. Tian et al., "Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network," *Radiotherapy and Oncology*, vol. 141, no. 1, pp. 192–199, 2019.
- [39] M. Sakamoto, Y. Hiasa, Y. Otake et al., "Bayesian segmentation of hip and thigh muscles in metal artifact-contaminated CT using convolutional neural network-enhanced normalized metal artifact reduction," *Journal of Signal Processing Systems*, vol. 92, no. 3, pp. 335–344, 2020.
- [40] J. Fang, Q. B. Zhou, and S. Wang, "Segmentation technology of nucleus image based on U-Net network," *Scientific Programming*, vol. 2021, no. 7, 10 pages, 2021.
- [41] Y. Feng, S. Deb, G. G. Wang, and A. H. Alavi, "Monarch butterfly optimization: a comprehensive review," *Expert Systems with Applications*, vol. 168, article 114418, 2021.
- [42] A. Skrk, B. Ks, and A. NI, "Development of deer hunting linked earthworm optimization algorithm for solving large scale traveling salesman problem," *Knowledge-Based Systems*, vol. 227, article 107199, 2021.
- [43] W. Li and G. G. Wang, "Elephant herding optimization using dynamic topology and biogeography-based optimization based on learning for numerical optimization," *Engineering with Computers*, vol. 1, pp. 1–29, 2021.
- [44] G. G. Wang, "Moth search algorithm: a bio-inspired meta-heuristic algorithm for global optimization problems," *Mematic Computing*, vol. 10, no. 2, pp. 151–164, 2018.
- [45] W. Li, G. Wang, and A. Gandomi, "A survey of learning-based intelligent optimization algorithms," *Archives of Computational Methods in Engineering*, vol. 28, no. 5, pp. 3781–3799, 2021.

Research Article

MR Image Classification for Brain Tumor Texture Based on Pseudo-Label Learning and Optimized Feature Extraction

Qianqian Xu ¹, Huachang Xu ¹, Jie Liu ¹, Mingxia Zhou ¹, Min Li ¹, Jinhui Xu ²,
and Hong Zhu ^{1,2}

¹School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, China

²Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, USA

Correspondence should be addressed to Jinhui Xu; jinhui@buffalo.edu and Hong Zhu; zhuhong@xzhmu.edu.cn

Received 1 July 2021; Revised 5 September 2021; Accepted 21 March 2022; Published 4 April 2022

Academic Editor: Hamidreza Mohafez

Copyright © 2022 Qianqian Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Brain tumors are the deadliest and most difficult to treat of all forms of cancer. Preoperative classification of brain tumors is conducive to the development of corresponding treatment plan. Take pituitary tumors as an example. Precisely judging the image data of pituitary tumor texture before surgery can provide a basis for the selection of surgical plan and prognosis. However, the existing methods require manual intervention, and the efficiency and accuracy are not high. In this paper, we proposed an automatic brain tumor texture diagnosis method for uneven sequence image data. First, for the small sample of pituitary tumor MRI image data, the T1 and T2 sequence data are uneven or missing; we used the CycleGAN model to perform data conversion between different domains to obtain a completely sampled MRI spatial sequence. Then, we used texture analysis+pseudo-label learning to label pituitary tumor data of some unknown labels. After that, we used the improved U-Net model based on CBAM to optimize feature extraction for pituitary tumor image data. Finally, we used the CRNN model to classify the degree of pituitary tumor texture based on the advantages of sequence data. The entire process only needs to provide labels for the entire sequence data, and the efficiency is greatly improved, with an accuracy rate of 94.23%.

1. Introduction

Brain tumors, including various intracranial tumors, are classified into benign and malignant. In recent years, the incidence of intracranial tumors has been on the rise. According to statistics, intracranial tumors account for about 5% of systemic tumors and 70% of childhood tumors. Due to its swelling and infiltrating growth, once occupying a certain space in the skull, whether it is benign or malignant in nature, it will inevitably increase intracranial pressure, compress the brain tissue, cause central nervous system damage, and endanger the lives of patients. Take pituitary tumors as an example. A pituitary tumor is one of the most common intracranial tumors. The pituitary tumor has a serious disease burden, and its treatment involves neurosurgery, biological therapy, and radiation therapy [1]. The softness and density of pituitary tumors are related to many key issues, such as whether surgery is required, what kind of sur-

gery to use, and what is the expected surgical effect. However, due to the nature of cranial cavity, it is often difficult to determine the softness of pituitary tumors before surgery. Therefore, how to accurately determine the softness of pituitary tumors in a noninvasive manner has become an important issue [2]. With the advancement of medical imaging technology, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and other imaging methods have become an important basis for assisting doctors in diagnosis. Therefore, it is very valuable to dig out deep quantitative information (such as the softness of pituitary tumors) that clinicians cannot detect with the naked eye from medical images.

With the popularity of artificial intelligence, deep learning is increasingly used in medical image processing. B. Zhou and S.K. Zhou [3] proposed a dual-domain recurrent network for rapid MRI reconstruction. This kind of network can recover k -space and images at the same time and finally

reconstruct high-quality MRI. Balakrishnan et al. [4] proposed the VoxelMorph framework for deformable pairwise medical image registration, which can speed up the analysis and processing of medical images. Hashimoto et al. [5] developed a new method of cancer subtype classification based on the Convolutional Neural Network (CNN) in combination with multi-instance, domain adversarial, and multiscale learning framework. An autoencoder is an unsupervised model in deep learning which can automatically extract many deep features from the medical image datasets. Gu et al. [6] proposed a context encoder network for 2D medical image segmentation. This network consists of a feature encoder module, a context extractor, and a feature decoder module, which can extract more advanced information and keep the space information. These results all suggest that the autoencoder has good utility and robustness in image feature extraction and model efficiency. Thus, it is quite suitable to use the autoencoder as a feature extractor in medical image processing.

These types of deep learning often require large amounts of labeled data. However, due to privacy concerns, medical image data used in practice are often small samples with incomplete labels. This motivates us to propose in this paper a semisupervised brain tumor image classification method based on CycleGAN, pseudo-label, and optimized feature extraction. Our rationale behind using CycleGAN is that medical images are often incomplete and often have data missing. CycleGAN enables us to generate additional data to make up the missed ones. Also, medical images are often unlabeled, which could make the classification task considerably more challenging. To resolve this issue, our idea is to use the method of texture analysis+pseudo-label learning to label them. Furthermore, we propose a new feature extractor based on improved U-Net and adopt some attention mechanism to extract brain tumor features and perform adaptive optimization. The feature extractor can effectively extract features from medical images. Finally, for sequence data, our method feeds the optimized feature sequences into a Convolutional Recurrent Neural Network- (CRNN-) based classifier to complete the classification of brain tumor texture. In this paper, we use MRI of the pituitary tumor as an example for experiments.

2. Methods

2.1. Theoretical Basis

2.1.1. Pseudo-Label Learning. Pseudo-label learning is a semisupervised learning method. Pseudo-label learning refers to the use of existing labeled data to train a model, and then, use the model to predict and label unlabeled data. Pseudo-label learning generally includes three steps. First, use the labeled data to train a model, then use the model to make label predictions on the unlabeled data, and finally use the pseudo-labeled data and the labeled data to retrain the model.

2.1.2. GRU. Gated recurrent unit (GRU) [7] is a kind of RNN, and its internal structure is similar to LSTM. GRU

uses an update gate and reset gate. Compared with LSTM, it has one less gate, so it is easier to calculate. GRU is widely used because it can achieve the same functions as LSTM while reducing parameters.

2.1.3. CBAM. The Convolutional Block Attention Module (CBAM) [8] is an attention mechanism module. It combines channels and space, so it has a better effect than the attention mechanism that only focuses on channels. The output result of convolutional layers with a CBAM structure will first pass through the channel attention module and then through the spatial attention module to get the final result. CBAM can make the model pay more attention to effective features, thereby improving the efficiency of feature extraction and thus the accuracy of the model.

2.2. Pituitary Tumor Sequence Data Amplification Using CycleGAN. MRI images of patients with pituitary tumors usually consist of different forms of spatial sequences, such as T1WI, T2WI, T1C, and T2FLAIR. The pituitary tumor dataset used in this paper is mainly composed of T1 and T2 spatial sequence images. In actual pituitary tumor MRI image acquisition, due to various practical reasons, such as data loss, data in a certain domain (such as T2) is usually missing. In order to supplement these missing data, we need to generate images of another domain based on the images of one domain and then merge them to generate a more complete dataset.

CycleGAN [9] can convert images of two domains into each other. CycleGAN includes two generators and two discriminators. In view of the characteristics of pituitary tumor data, we chose CycleGAN to complete the supplement of missing data. It can help us to complete the two-way conversion from the T1 domain to the T2 domain and from the T2 domain to the T1 domain. The network structures of the CycleGAN generator and discriminator used in this experiment are shown in Figures 1 and 2, respectively. The convolution layer of the generator consists of Conv2D, leaky ReLU, and instance normalization. The first three deconvolution layers consist of UpSampling2D, Conv2D, ReLU, and instance normalization, and the last deconvolution layer consists of UpSampling2D, Conv2D, and Tanh. The convolution layer of the discriminator consists of Conv2D, leaky ReLU, and instance normalization. Through the mutual conversion of the T1 domain and T2 domain, we expanded the number of slices in each domain to 12, making a total of 24 images of pituitary tumors for each patient to form a complete dataset.

2.3. Data Augmentation Method Based on Texture Analysis and Pseudo-Label Learning. There are 374 cases of pituitary tumor imaging data in this experiment, but only 137 cases have soft and flexible labels. In order to expand pituitary tumor data and improve the data utilization, we used a method of texture analysis+pseudo-label learning to label the remaining 237 unlabeled pituitary tumor data.

We first performed texture analysis on the labeled pituitary tumor images, such as extracting their texture or statistical features, to find the decisive parameters. Then, we used

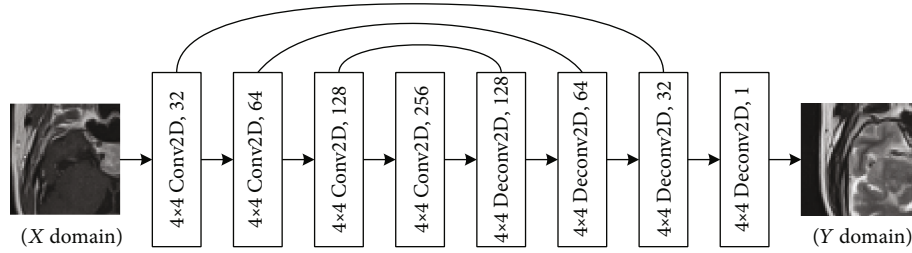


FIGURE 1: Generator architecture.

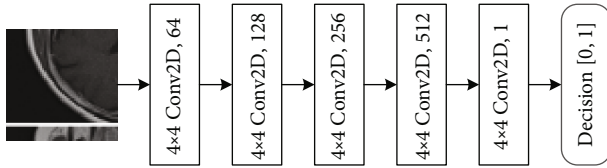


FIGURE 2: Discriminator architecture.

these analyzed parameters as reference standards to label unlabeled data. In order to improve the accuracy of the labeling, we extracted part of the data labeled by texture analysis and part of the data with real labels to synthesize a new labeled training set. Then, we used this new dataset to train a model; here, we chose the GRU model. Finally, we used the trained GRU model to predict the unlabeled data again, generate the final label, and complete the pseudo-label learning. The specific process is shown in Figure 3.

Step 1. Using texture analysis method to label the unlabeled pituitary tumor data.

Step 2. Randomly extracting 50% data from 237 cases of pseudo-label data.

Step 3. Randomly extracting 20% of the 137 cases with labeled data as the test set and synthesizing the new training set with remaining data and 119 cases of pseudo-label data.

Step 4. Using the GRU model to train the new training set.

Step 5. Using the trained model to predict 237 cases of unlabeled data to obtain the final data labels.

2.4. Classifying Pituitary Tumor Texture Based on Adaptively Optimized Texture Extraction. We divided the classification model of pituitary tumor images into two parts: feature extraction and final classification. For the feature extraction of pituitary tumors, we adopted a feature extractor based on improved U-Net and added CBAM between the encoder and the decoder. For the final classification, we used a CRNN-based classifier.

2.4.1. Adaptive Optimization of Feature Extraction. For the feature extractor, an encoder composed of dense blocks, a $1 * 1$ convolutional layer, and a $2 * 2$ AvgPooling layer is used to extract the features of pituitary tumor MRI sequence

data. Each dense block consists of 4 convolutional layers with 64, 64, 128, and 128 kernels, respectively. The kernel size of each convolution layer is $1 * 1$, $3 * 3$, $3 * 3$, and $1 * 1$. Each convolutional layer uses leaky ReLU as the activation function. Dense blocks can enhance the feature propagation ability of pituitary tumor images due to dense connections, while convolutional layers and pooling layers are used to reduce the dimensionality. The entire encoder uses two such dense blocks.

The decoder consists of residual blocks, an upsampling layer, and a convolutional layer and is used to generate a spatial sequence with the same spatial dimensions as the original image. Each residual block consists of 4 convolutional layers with 64, 64, 128, and 128 kernels, respectively. The kernel size of each convolution layer is $1 * 1$, $3 * 3$, $3 * 3$, and $1 * 1$. Each convolutional layer uses leaky ReLU as the activation function. The residual block is used to compress the dimensionality of the feature map, and the upsampling layer and convolutional layer are used to perform the dimensionality upgrade operation. The use of residual structure can help us increase the depth of the network while avoiding the disappearance of gradients, thereby improving the effect of MRI spatial sequence reconstruction. The entire decoder uses two such residual blocks. A skip connection is also used between the encoder and the decoder. The feature maps obtained from each convolutional layer on the left in the U-Net network will be concatenated to the corresponding upsampling layer using skip connections. The use of skip connections can better integrate low-level features and high-level features to avoid information loss.

Finally, in order to further improve the accuracy of feature extraction, we added CBAM to the above-mentioned improved U-Net. We added CBAM between the encoder and the decoder. The feature sequence output by the encoder will be optimized by CBAM and then passed into the decoder as input for decoding.

We input the pituitary tumor image sequence into the encoder to generate the feature sequence and used CBAM to further optimize it. Then, we used the decoder to reconstruct the spatial sequence. Finally, we compared the generated image with the original image at the pixel level. The more similar the two are, the more representative the extracted feature sequence is and the more it shows that the model has learned the features of the pituitary tumor image sequence well. This is the entire feature extraction process, and the extracted feature sequence will be used for further classification.

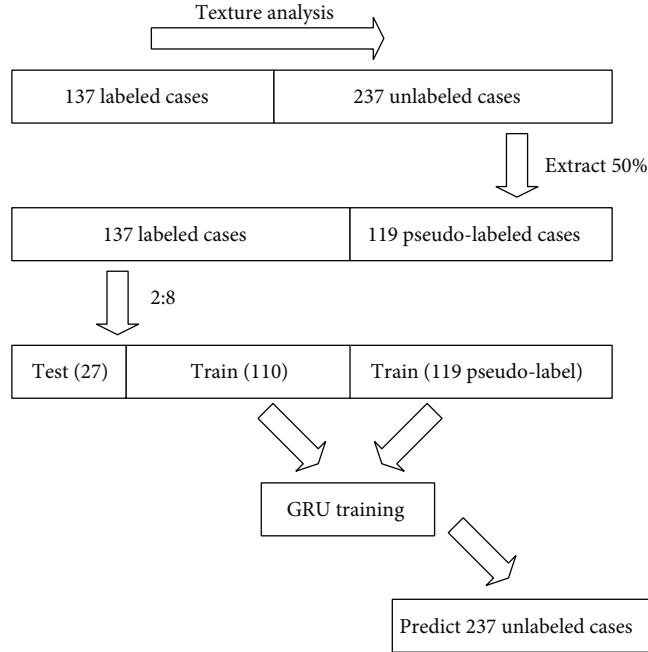


FIGURE 3: Pseudo-label learning.

2.4.2. Multisequence Pituitary Tumor Classification Model. The pituitary tumor image feature sequence extracted by CBAM_U-Net is a low-dimensional data sequence, which is strictly arranged in the order of image slices. Therefore, we used CRNN-based classifiers for the final classification of pituitary tumor texture. CRNN is composed of CNN and RNN and can capture the information of adjacent slices, which is more suitable for sequence data. We input the feature map sequence extracted by the CBAM_U-Net-based feature extractor into the CRNN-based classifier to obtain the classification result of pituitary tumor texture.

The final multisequence pituitary tumor texture classification model is shown in Figure 4.

3. Results

3.1. Experiment Platform and Dataset. The experimental running environment is as follows: the operating system is Windows 10, the processor is 2.10 GHz Intel Xeon (dual core), the memory capacity is 64 GB, the development environment is PyCharm, the deep learning framework is Keras, the programming language is Python, and the 5graphics card is GeForce RTX 2080Ti (three cores).

The pituitary tumor image dataset used in our experiment was provided by the local affiliated hospital. Each patient contains two modalities: T1 and T2. There are a total of 374 patients, of which 137 have labels (90 soft texture, 47 hard texture). The labels of each patient's pituitary tumor image are divided into two levels: soft texture and hard texture.

3.2. CycleGAN-Based Multisequence Data Amplification. The image data of 374 patients used in this experiment included 280 T1 MRI spatial sequences and 94 T2 MRI spatial sequences. The T2 MRI spatial sequence was severely lost.

We trained CycleGAN for 90 epochs using the complete data of the image sequence and used the trained CycleGAN model to augment the image data of pituitary tumors with incomplete T1 or T2 sequences. The domain converter converted the T1/T2 domain to the T2/T1 domain. Finally, we expanded the slices of each patient to 24 (12 T1 slices, 12 T2 slices). Figure 5 shows the test results of the existing T1 and T2 image pairs.

The Figure 5 shows the existing pairs of original T1 and T2 images and their images converted by CycleGAN. We can see that the converted images are very similar to the original images in their corresponding domains.

3.3. Amplification of Pituitary Tumor Data Based on Texture Analysis and Pseudo-Label Learning. Among the 374 cases of pituitary tumor image data amplified by CycleGAN, only 137 cases had labels, and the rest were unlabeled. In order to expand experimental data and improve data utilization, we used the texture analysis+pseudo-label learning method to label the remaining 237 pituitary tumor data.

3.3.1. Texture Analysis. In this experiment, FireVoxel software was used to extract regions of interest (ROI) for pituitary tumors, as shown in Figure 6.

According to the literature [10–12], the feature parameters extracted by FireVoxel software are mainly mean, standard deviation, inhomogeneity, skewness, kurtosis, and entropy. We performed independent t -tests for these parameters. When $P < 0.05$, it indicates that the parameters are statistically significant.

The experiment used SPSS 22.0 software for statistical analysis. Among them, kurtosis (T1), inhomogeneity (T2), and entropy (T2) are statistically significant. The ROC curves of these characteristics are shown in Figure 7.

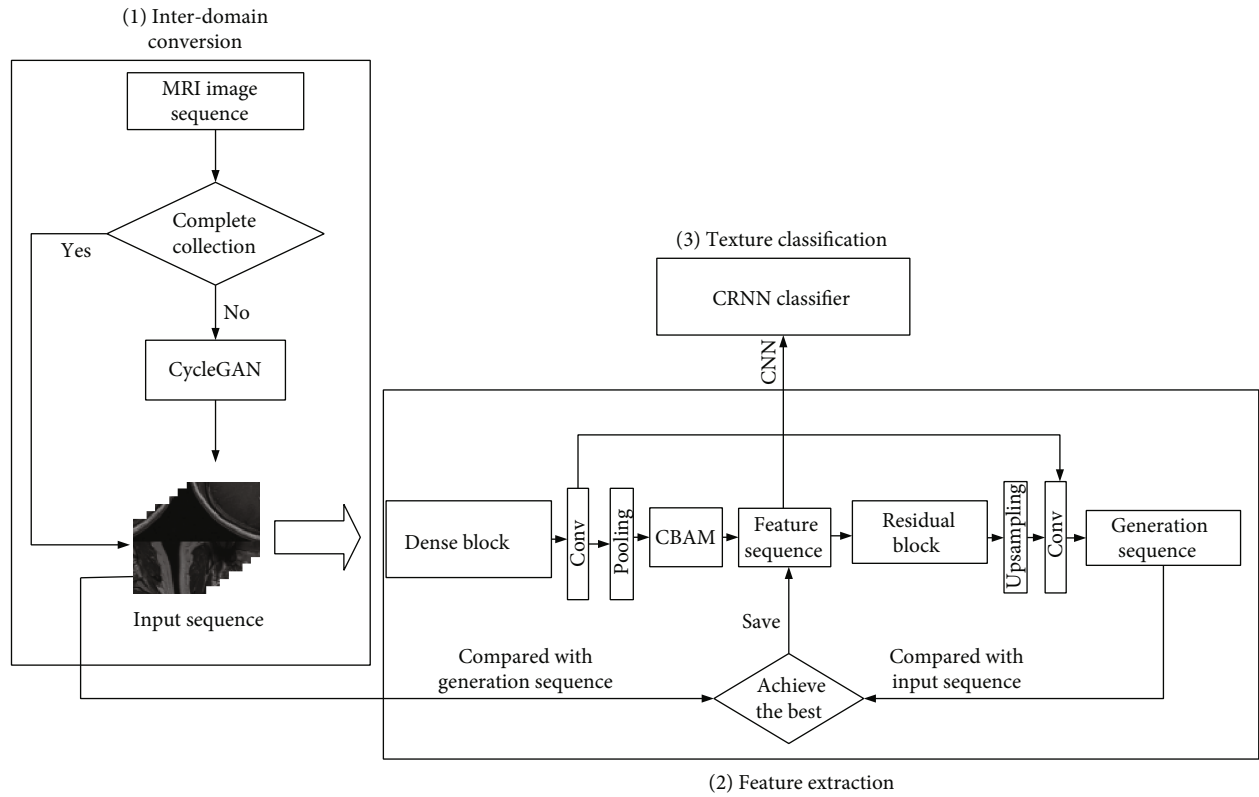


FIGURE 4: Multisequence pituitary tumor classification model.

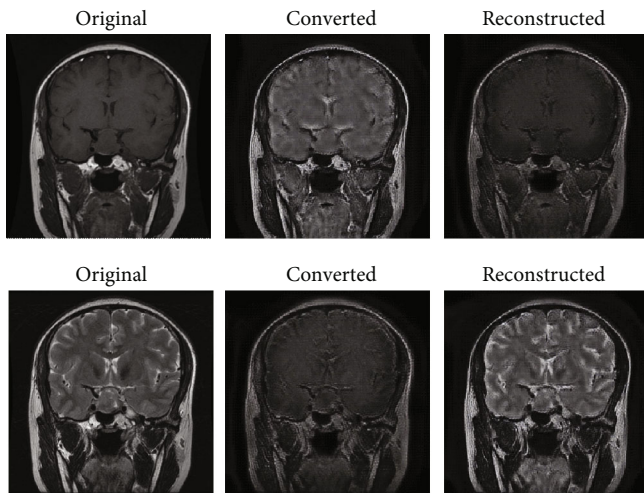


FIGURE 5: Visualization of training results.

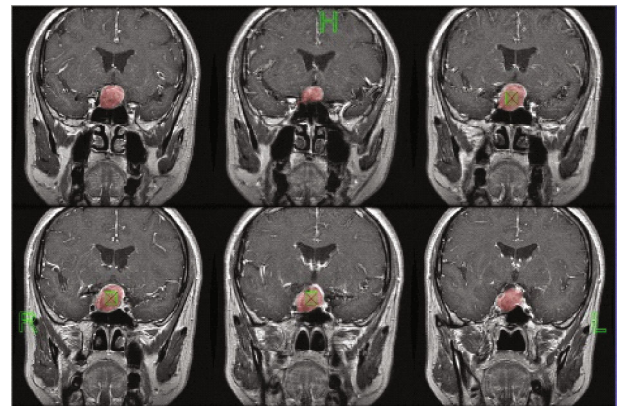


FIGURE 6: ROI.

The specific data is shown in Table 1:

It can be seen from the figures and table that the entropy value of the T2 mode is more effective for judging the soft and flexible texture of pituitary tumors. After analysis, there is a judgment value for entropy statistics, which is 3.91. When the entropy value of the pituitary tumor is less than 3.91, it is judged to be hard. When it is greater than 3.91, it is judged to be soft, and 237 cases of the pituitary tumor without labels were labeled with this criterion.

3.3.2. *Pseudo-Label Learning.* After labeling the pituitary tumors analyzed by texture features, we used the pseudo-label learning method to further evaluate the labels of texture analysis and made the final labels to expand the dataset. The experiment used part of the data labeled by texture analysis and part of the data with real labels to synthesize a new training set to train the GRU model. The training process of the GRU model used in pseudo-label training is shown in Figures 8 and 9. When epoch reached 55, the model converged and reached the optimal.

We used the trained model to predict 237 cases of unlabeled data, and the prediction result is shown in Figure 10.

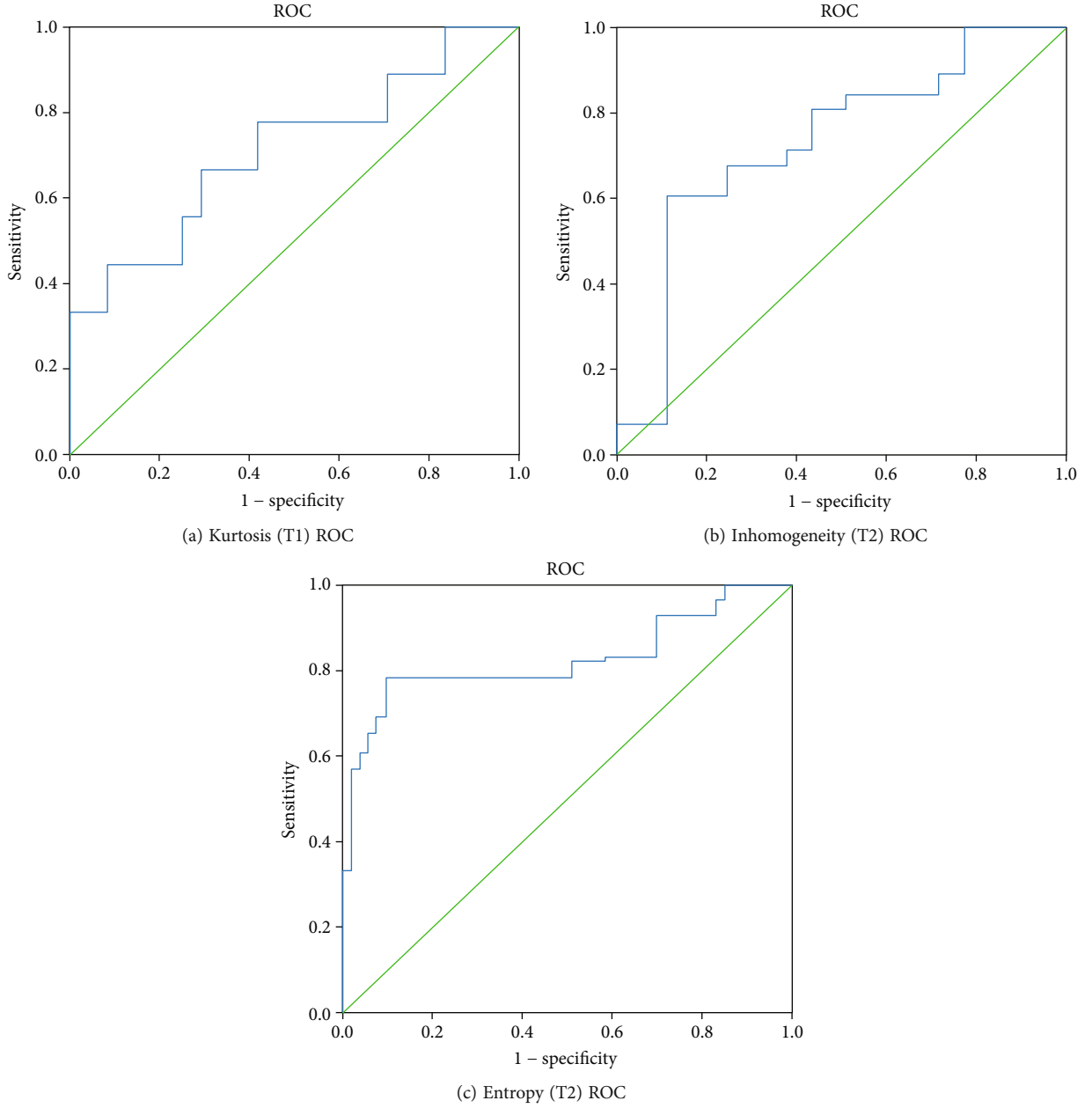


FIGURE 7: Texture feature ROC.

TABLE 1: Characteristic value statistics of pituitary tumors.

	AUC	$P (<0.05)$
Inhomogeneity (T2)	0.732	0.020
Entropy (T2)	0.828	0.010
Kurtosis (T1)	0.713	0.023

There are 200 cases of soft texture label and 37 cases of hard texture label. We integrated pseudo-labeled data with labeled data and finally got a total of 374 cases of pituitary tumor data, of which 290 cases are soft and 84 cases are hard.

3.4. Pituitary Tumor Texture Image Classification Based on Adaptively Optimized Feature Extraction. After data augmentation and pseudo-label labeling, this paper uses a feature extractor based on CBAM_U-Net to extract features from the preprocessed pituitary tumor image data and uses a CRNN-based classifier for texture classification. In order to ensure the effectiveness of the feature extractor, we trained the CBAM_U-Net model and the U-Net model 150 epochs for comparison. Networks were implemented using an adaptive moment estimation optimizer (Adam) and a mean square error loss function (MSE). The initial learning rate was set to 10^{-5} with a batch size of 4. The training process of the models is shown in Figure 11.

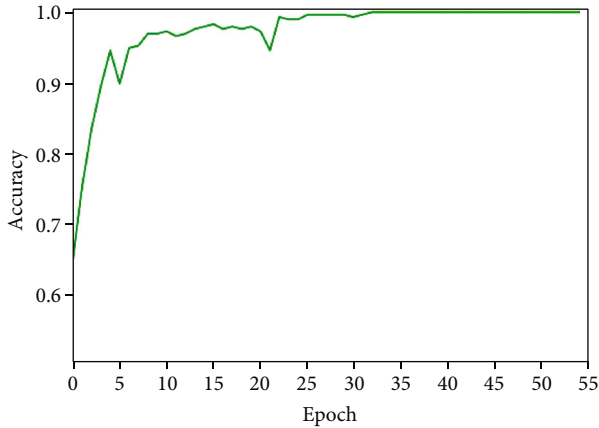


FIGURE 8: Training accuracy of the pseudo-label model.

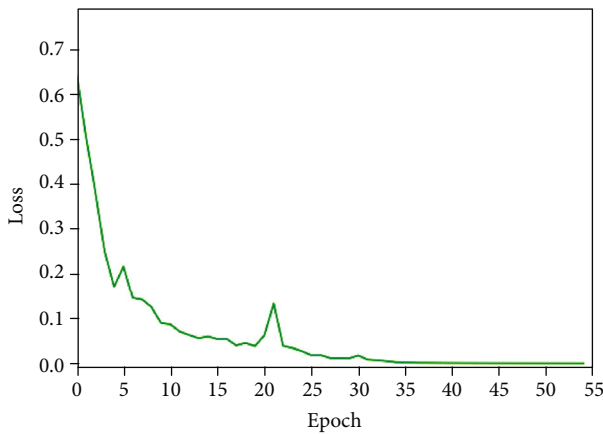


FIGURE 9: Training loss of the pseudo-label model.

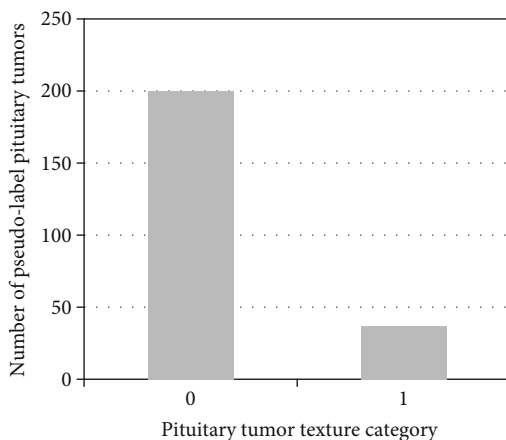


FIGURE 10: Pituitary tumor data distribution.

As can be seen from the figures, when the models were trained for 150 epochs, the accuracy of the CBAM_U-Net feature extraction model was higher than that of the U-Net feature extraction model, and the convergence speed of loss was faster and the minimum value was lower. It can be seen

that adding CBAM makes the improved U-Net model more superior in unsupervised feature extraction.

We put three kinds of pituitary tumor image sequences (multisequence, T1, and T2) into CBAM_U-Net for feature extraction, then put the results into CRNN for training, and divided them into three models, namely, the multisequence model, the T1 domain model, and the T2 domain model. The CRNN classifier was implemented using an adaptive moment estimation optimizer (Adam) and a cross entropy loss function (CE). The batch size was set to 10. We compared them experimentally.

- (1) T1 domain model: only the MRI spatial sequence of the T1 domain was used in the experiment, including the MRI spatial sequence of the T1 domain generated by CycleGAN
- (2) T2 domain model: only the MRI spatial sequence of the T2 domain was used in the experiment, including the MRI spatial sequence of the T2 domain generated by CycleGAN
- (3) Multisequence model: MRI spatial sequences of T1 and T2 domains were used in the experiment, including the MRI spatial sequences of T1 and T2 domains generated by CycleGAN

Figures 12–14 show the training process of the T1 domain, T2 domain, and multisequence model, respectively.

As shown in the figures, we divided the pituitary tumor image data into a training set (80%) and a test set (20%). The test set was randomly selected from the labeled dataset and did not contain pseudo-labeled data. The experiment was repeated 4 times in the T1 domain, T2 domain, and multisequence domain. We recorded each classification accuracy and AUC (95% CI) to calculate the final average and variance. The detailed information is shown in Table 2.

Table 3 shows the classification accuracy and AUC (95% CI) before and after amplification of the pseudo-label data.

In addition, the experiment not only added CBAM between the encoder and the decoder but also added CBAM to different places such as the encoder or the skip connections to compare its different effects. The final results are shown in Table 4. As can be seen from the table, the effect was best when CBAM was added between the encoder and the decoder.

3.5. Compared with Other Models. With the application and development of deep learning, many excellent feature extraction models have emerged. The quality of tumor feature extraction has a great correlation with the final classification result. In order to further illustrate the effectiveness of the pituitary tumor classification method based on CBAM_U-Net, we reimplemented some commonly used classification models on brain tumor image data. Table 5 shows the classification results of the classification model proposed in this paper and some commonly used classification models on brain tumor image data. All results are based on the same cross validation scheme and test set.

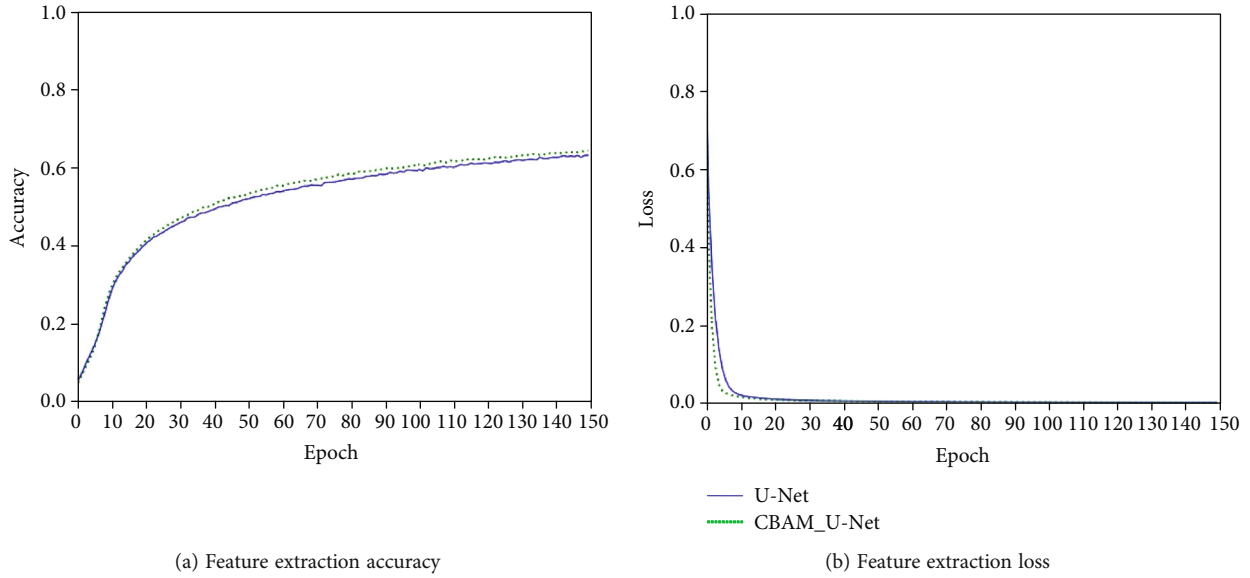


FIGURE 11: Multisequence feature extraction model.

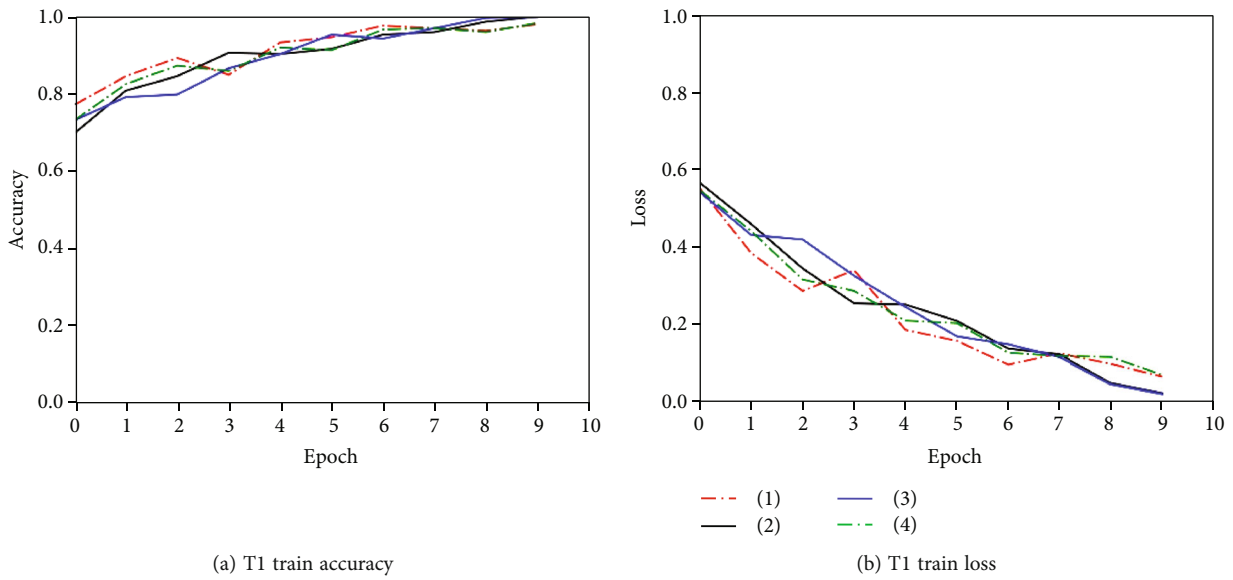


FIGURE 12: T1 domain image classification model training.

As can be seen from the table, the model combining the feature extractor based on CBAM_U-Net and the classifier based on CRNN has the highest classification accuracy. Improved encoders and decoders can help further mine the information of images. Skip connections can combine low-level and high-level features. CBAM can strengthen feature propagation, reduce redundant information that is not related to tumors, and focus on useful information. CRNN can capture the information of adjacent slices of sequence data. The combination of these techniques enables us to improve the classification model accuracy to a certain extent and reduce the amount of network calculations. In addition, the classification model in this paper is more suitable for small-size medical image datasets than those transformer-

based models that often require a large amount of data. This comparative experiment shows the great potential and good application prospects of our proposed method in grading pituitary tumor texture.

4. Discussion

This study designed a series of experiments to classify brain tumor texture and took MRI of the pituitary tumor as an example. First, for incomplete sequence data, this study used CycleGAN to augment the imaging data of pituitary tumors lacking T1 or T2 sequences. CycleGAN converted T1/T2 sequences to T2/T1 sequences. This paper tested it on existing pairs of T1 and T2 images to verify the success of the

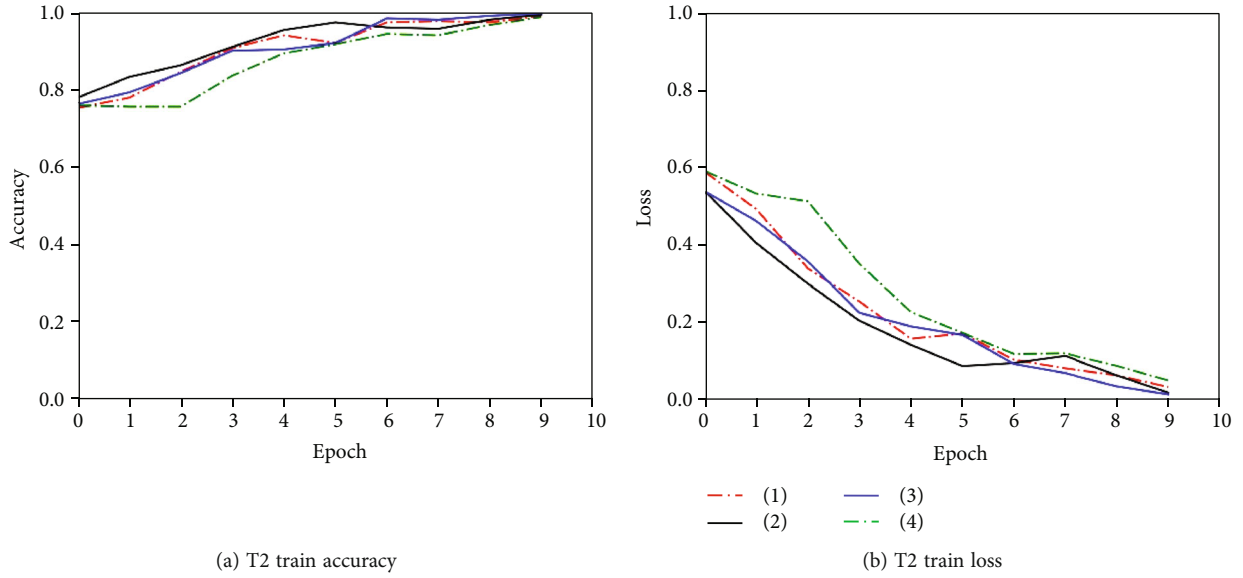


FIGURE 13: T2 domain image classification model training.

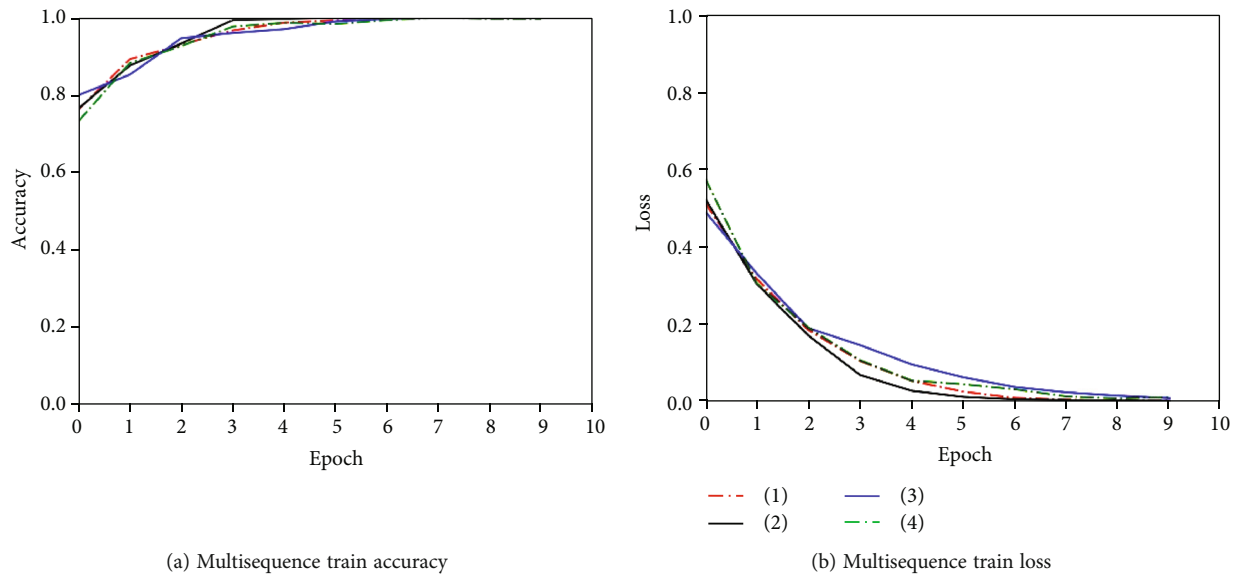


FIGURE 14: Multisequence image classification model training.

TABLE 2: Pituitary tumor classification accuracy.

	Multisequence	T1 domain	T2 domain
Accuracy (%)	94.23 ± 1.19	90.33 ± 1.28	89.00 ± 1.68
AUC (95% CI)	0.976 ± 1.44	0.918 ± 3.60	0.877 ± 3.75

TABLE 3: Comparison of accuracy before and after amplification of pseudo-label data.

	137 cases of data	374 cases of data
Accuracy (%)	91.08 ± 2.06	94.23 ± 1.19
AUC (95% CI)	0.930 ± 3.84	0.976 ± 1.44

method. The texture features extracted from the converted T1/T2 sequence are closer to the T2/T1 sequence. Secondly, we used the method of texture analysis+pseudo-label learning to label the unlabeled pituitary tumor image data. This method first uses texture analysis to initially label unlabeled data and then uses GRU-based pseudo-label learning to complete pseudo-label labeling. This method incorporates statistical features and machine-learned features to improve the reliability of pseudo-label labeling and the utilization of pituitary tumor data. This study used FireVoxel software to extract feature parameters in the texture analysis part of pseudo-label labeling. A total of 12 basic features were extracted from the T1 and T2 images, and the useful features after screening were used for initially labeling of unlabeled

TABLE 4: Comparison of accuracy of CBAM in different positions.

	Between the encoder and decoder	Encoder	Skip connections	Without CBAM
Accuracy (%)	94.23 \pm 1.19	90.67 \pm 1.89	89.33 \pm 2.18	89.00 \pm 2.00
AUC (95% CI)	0.976 \pm 1.44	0.949 \pm 1.07	0.953 \pm 1.76	0.913 \pm 2.77

TABLE 5: Comparisons of classification results of different methods.

Feature extraction	Texture classification	Accuracy (%)	Time (s)
—	VGG	71.89	119
—	ResNet	79.33	108
—	DenseNet	81.39	105
—	ViT	86.67	197
—	CRNN	77.33	72
Autoencoder (DenseNet+ResNet)	CRNN	87.34	46
U-Net (DenseNet+ResNet)	CRNN	89.00	46
U-Net (DenseNet+ResNet)+CBAM	CRNN	94.23	49

data. In future work, we will devote to extracting more texture features and selecting features that are more relevant to brain tumor texture for pseudo-labeling.

This paper proposes an unsupervised feature extractor based on improved U-Net to extract the features of the pituitary tumor image sequence. The feature extractor uses dense blocks and residual blocks to optimize feature extraction to extract more representative features. In the experiment, the pituitary tumor image sequence was input into the encoder of the feature extractor for feature extraction and then restored by the decoder. The generated image sequence was compared with the original input image sequence. The more similar the two, the better the extracted features. This paper also introduces CBAM to the feature extractor. The attention mechanism can help models pay attention to useful information and suppress useless information to improve the reliability and efficiency of model feature extraction. CBAM is an attention mechanism that pays attention to both the spatial and the channel. Compared with the attention mechanisms that only pay attention to a certain aspect, it has a better effect. In future work, we will also try and improve more attention mechanisms to improve the classification model proposed in this paper. Finally, we input the extracted pituitary tumor features into the CRNN-based classifier for the final pituitary tumor texture classification. The entire classification model is end-to-end without the need for tumor presegmentation. In addition, this method is based on image sequences rather than 2D (slice) methods. We repeated the experiment 4 times and tested it on the labeled test set to obtain the classification accuracy and AUC (95% CI). Experimental results and comparative experiments show that our method has advantages in the classification of pituitary tumor texture. This method can also be applied to other types of brain tumors, such as glioma.

For larger-scale multisequence MRI data or multicenter study, the pseudo-label labeling method based on texture analysis and feature extractor proposed in this paper will

also be applicable and can be used to provide direct input to other classification schemes. In conclusion, the method proposed in this paper can be applied to real clinical research to assist doctors in diagnosis and improve the efficiency of doctors' diagnosis. However, the imaging equipment used in different countries and regions is different, so there are still some differences in imaging and subsequent texture analysis. In addition, due to the privacy of medical images, it is difficult to collect a variety of large datasets. Therefore, designing a more versatile model is still a future research direction.

5. Conclusion

This paper proposes a brain tumor texture classification method based on pseudo-label learning and optimized feature extraction and takes pituitary tumors as an example. The experiment finally achieved 94.23% of average accuracy and 0.976 of average AUC (95% CI). Comparative experiments show that compared to some existing methods, our proposed method has more advantages and can improve the classification efficiency while improving the classification accuracy. This study shows great potential and good application prospects of our method in helping clinical diagnosis of brain tumors.

Data Availability

The datasets generated during and analyzed during the current study are not publicly available due to hospital data confidentiality but are available from the corresponding author on reasonable request.

Ethical Approval

We have obtained permission from the Imaging Department of the Affiliated Hospital of Xuzhou Medical University, and the study has been approved by the Ethics Committee of

Xuzhou Medical University. What we need to declare is that the pituitary tumor dataset used in our study has followed all the procedures required by the Chinese government's law. The data has been strictly reviewed by those in charge of such issues, and all sensitive information has been removed. The study is purely for research purpose and does not have any dispute of interest.

Consent

We have obtained verbal informed consent from all participants in the study according to the wishes of the participants. For minors, we have obtained the consent of their parents or legal guardians.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was partially supported by Projects for Research on Medical in Jiangsu Commission of Health (through grant Z2020032), by Projects for 2020 College Student Innovation Program (through grant 202010313014Z), and by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (through grant 19KJB520017).

References

- [1] A. Beckers, L. A. Aaltonen, A. F. Daly, and A. Karhu, "Familial isolated pituitary adenomas (FIPA) and the pituitary adenoma predisposition due to mutations in the aryl hydrocarbon receptor interacting protein (AIP) gene," *Endocrine Reviews*, vol. 34, no. 2, pp. 239–277, 2013.
- [2] H. Zhu, Q. H. Fang, Y. H. Huang, and K. Xu, "Semi-supervised method for image texture classification of pituitary tumors via CycleGAN and optimized feature extraction," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–14, 2020.
- [3] B. Zhou and S. K. Zhou, "DuDoRNet: learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4273–4282, Seattle, WA, USA, 2020.
- [4] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.
- [5] N. Hashimoto, D. Fukushima, R. Koga et al., "Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3852–3861, Seattle, WA, USA, 2020.
- [6] Z. W. Gu, J. Cheng, H. Z. Fu et al., "CE-Net: context encoder network for 2D medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [7] K. Cho, B. V. Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," (2014), <https://arxiv.org/abs/1406.1078>.
- [8] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, Munich, Germany, 2018.
- [9] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2242–2251, Venice, Italy, 2017.
- [10] X. Y. Yang, Y. Zhu, M. Zhang, K. Xu, and Y. Kong, "Comparison between the consistency of pituitary adenoma and MRI texture analysis," *Journal of Clinical Radiology*, vol. 4, pp. 613–617, 2019.
- [11] W. J. Peng, K. Xu, Z. L. Liu, X. W. Gui, X. G. Zhang, and X. Lu, "Mammogram texture analysis in differential diagnosis of benign and malignant breast nodules," *Chinese Journal of Medical Imaging*, vol. 26, no. 12, pp. 19–24, 2018.
- [12] Y. N. Song, S. F. Gao, Y. K. Meng, Z. Sha, G. W. Fan, and R. T. Yu, "Value of texture analysis of peritumoral edema in differential diagnosis between glioblastoma and primary brain lymphoma," *Chinese Journal of Minimally Invasive Neurosurgery*, vol. 24, no. 4, pp. 5–8, 2019.

Research Article

A Broad Learning System to Predict the 28-Day Mortality of Patients Hospitalized with Community-Acquired Pneumonia: A Case-Control Study

Jing Yuan ¹, Xin Liu ², Wen-Feng Wang ³, and Jing-Jing Zhang³

¹Department of Infectious Diseases, Chifeng Municipal Hospital, Chifeng Clinical Medical School of Inner Mongolia Medical University, Chifeng 024000, China

²Department of Neurosurgery, Chifeng Municipal Hospital, Chifeng Clinical Medical School of Inner Mongolia Medical University, Chifeng 024000, China

³School of Science, Shanghai Institute of Technology, Shanghai 201418, China

Correspondence should be addressed to Jing Yuan; yy402293690@126.com

Received 3 August 2021; Revised 20 January 2022; Accepted 31 January 2022; Published 4 March 2022

Academic Editor: Kwang Gi Kim

Copyright © 2022 Jing Yuan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study was to conduct a model based on the broad learning system (BLS) for predicting the 28-day mortality of patients hospitalized with community-acquired pneumonia (CAP). A total of 1,210 eligible CAP cases from Chifeng Municipal Hospital were finally included in this retrospective case-control study. Random forest (RF) and an eXtreme Gradient Boosting (XGB) models were used to develop the prediction models. The data features extracted from BLS are utilized in RF and XGB models to predict the 28-day mortality of CAP patients, which established two integrated models BLS-RF and BLS-XGB. Our results showed the integrated model BLS-XGB as an efficient broad learning system (BLS) for predicting the death risk of patients, which not only performed better than the two basic models but also performed better than the integrated model BLS-RF and two well-known deep learning systems-deep neural network (DNN) and convolutional neural network (CNN). In conclusion, BLS-XGB may be recommended as an efficient model for predicting the 28-day mortality of CAP patients after hospital admission.

1. Introduction

Pneumonia is the most common respiratory disease [1]. Before the advent of antibiotics, pneumonia was one major killer to the human health [2]. With the advances in modern medicine, many pneumonia patients have been cured with antibiotics and adjuvant therapy, but the mortality rate remains high among the very young, the elderly, and those with compromised immune functions [3]. After the initial triage of patients with pneumonia, it is critical for emergency medical staff to assess whether these patients require hospitalization [4]. Unnecessary hospitalizations not only increase the risk of acquired infections but also drain health care resources [5]. Several pneumonia severity scales may be used to assess the severity of a patient's illness, but these scales are mainly used in the inpatients and are not suitable for emer-

gency patients [6]. Community-acquired pneumonia (CAP) is a common infectious disease of respiratory system [7]. A deep insight into the potential factors influencing the quality of antibiotic use is essentially necessary to develop effective and targeted interventions to improve care for patients with CAP [8]. Accurate disease assessment is of great value for the initial treatment, clinical stability, and long-term prognosis [9]. Biomarkers are immune cells and immune proteins that are significantly increased in the process of microbial immunity and have auxiliary diagnostic value in the evaluation of CAP [10].

Nowadays, artificial intelligence is already used to solve emergent problems for medical engineering and particularly, for predicting CAP [11]. In order to avoid the devastating effects of the CAP on the patients' daily lives and healthcare systems and to control the further spread of this virus, we

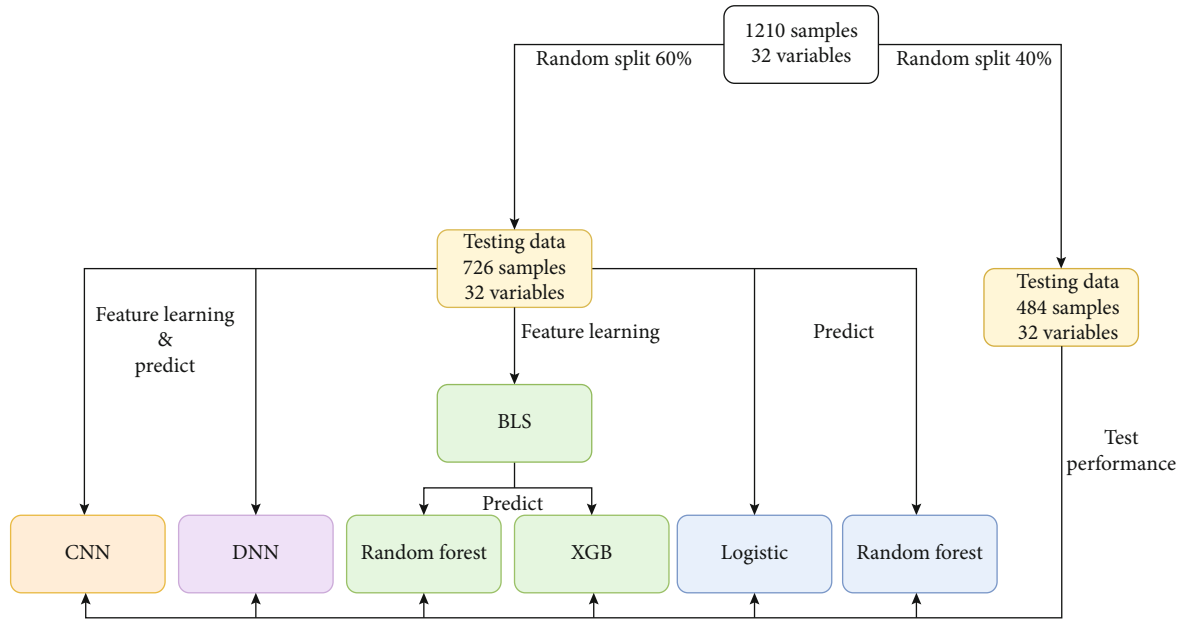


FIGURE 1: Establishment and validation of the prediction models for the 28-day mortality of CAP patients.

not only need to make an effective early diagnosis of infected patients through effective screening but also need to predict the risk of death in CAP patients [12, 13]. A series of models and algorithms were proposed to search for optimal hidden-layer architectures, connectivity, and training parameters for deep learning systems for predicting the CAP risk among patients with respiratory complaints, but the efficiency of these models and algorithms in predicting the death risk of patients hospitalized with CAP needs a further investigation, and meanwhile, novel approaches are quite necessary [14, 15].

Our objectives in the present studies are (1) to develop an efficient model based on the previous models and algorithms for predicting the risk of the 28-day mortality in patients hospitalized with CAP, using the random forest (RF) and eXtreme Gradient Boosting (XGB) models [16]; (2) to utilize the broad learning system (BLS) extract the features and evaluate the importance of BLS features in predicting the 28-day mortality of patients [17]; and (3) to compare the performance of the proposed model with two well-known deep learning systems—deep neural network (DNN) and convolutional neural network (CNN).

2. Materials and Methods

2.1. Study Design and Population. This was a retrospective case-control study. The information of a total of 1,397 CAP patients was collected from the Chifeng Municipal Hospital between August 2019 and December 2020. After excluding cases with age < 18 years ($n = 58$), having recently received chemotherapy ($n = 24$), advanced liver disease ($n = 67$), and the serum creatinine level > 1.5 mg/dl ($n = 38$), 1,210 eligible patients were finally included in this study. This study was approved by the Institutional Review Board (IRB) of Chifeng Municipal Hospital (approval number: no. 2019_24).

The inclusion criteria were as follows: (1) age ≥ 18 years old, (2) patients diagnosed with CAP according to Chinese Guidelines for Diagnosis and Treatment of Adult Community-acquired Pneumonia, and (3) available information of 28-day mortality or survival after hospital admission.

The exclusion criteria were (1) patients who have recently received chemotherapy, corticosteroids, or other immunosuppressants; (2) exposure to antibiotics within 14 days before entering the group; (3) patients with advanced liver disease; (4) being undergoing hemodialysis; (5) patients with serum creatinine level > 1.5 mg/dl; (6) patients with severe infection; and (7) patients with immune dysfunction.

2.2. Data Collection. The demographic and clinical information of CAP patients were collected, including gender, age, nationality, history of diseases (allergy, hypertension, diabetes, lung disease malignant tumor, heart failure (HF)), history of surgery, smoking, drinking, systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate, heart rate (HR), white blood cell (WBC) counts, red blood cell (RBC) counts, hemoglobin (Hb) level, platelet (PLT) counts, aspartate aminotransferase (AST) level, serum albumin (ALB) level, blood urea nitrogen (BUN) level, creatinine (Cr) level, blood glucose (Glu) level, porcine calcitonin (PCT) level, and C-reactive protein (CRP) level. The outcome was the 28-day mortality of patients hospitalized with CAP.

2.3. Establishment and Validation of the Prediction Models. All CAP patients were randomly grouped into the training and testing sets with a ratio of 6 : 4. The balance test was carried out between the two sets. Six prediction models were conducted using the training set (Figure 1). The logistic regression, RF, DNN, and CNN analyses were used to establish four models to predict the risk of 28-day mortality in

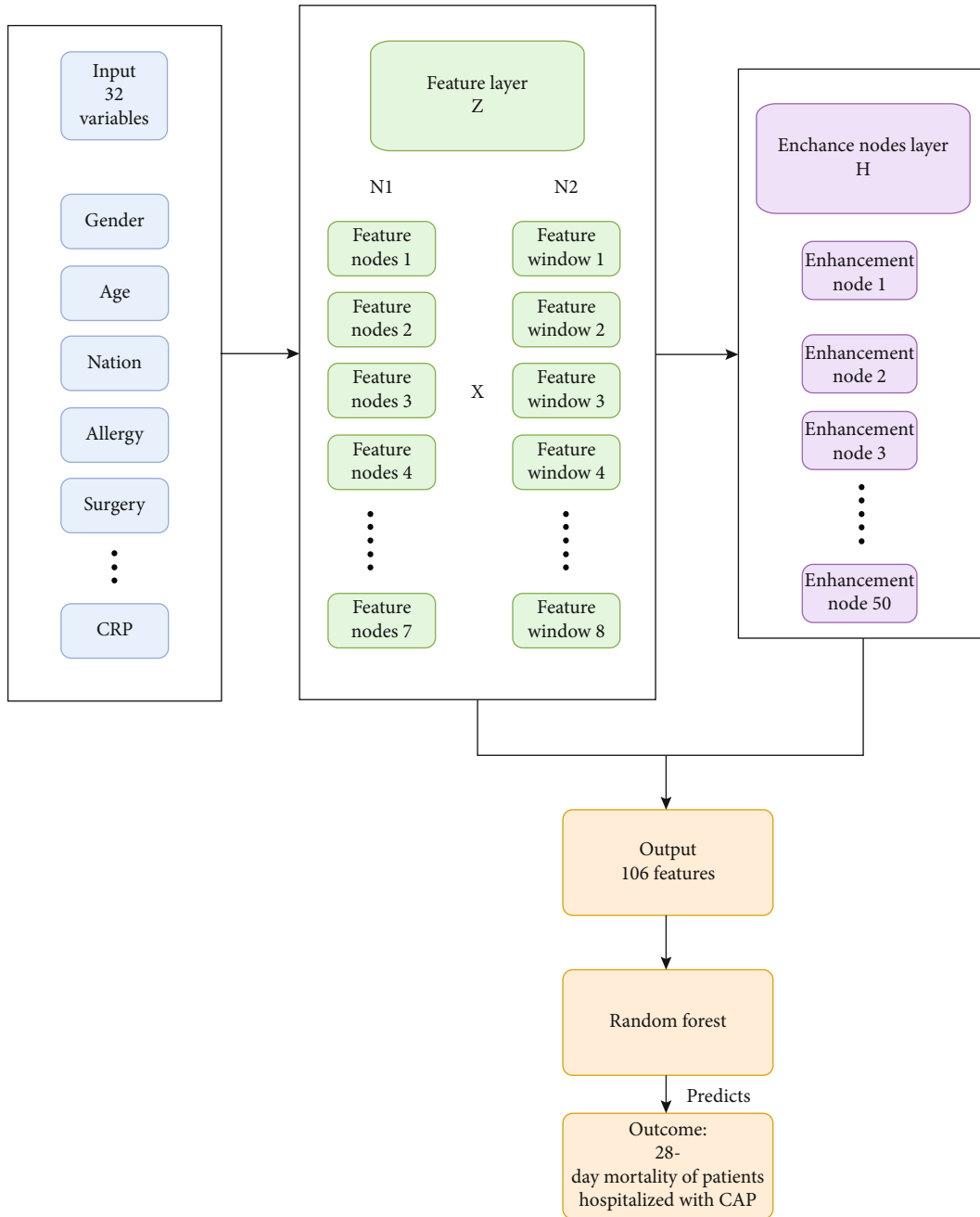
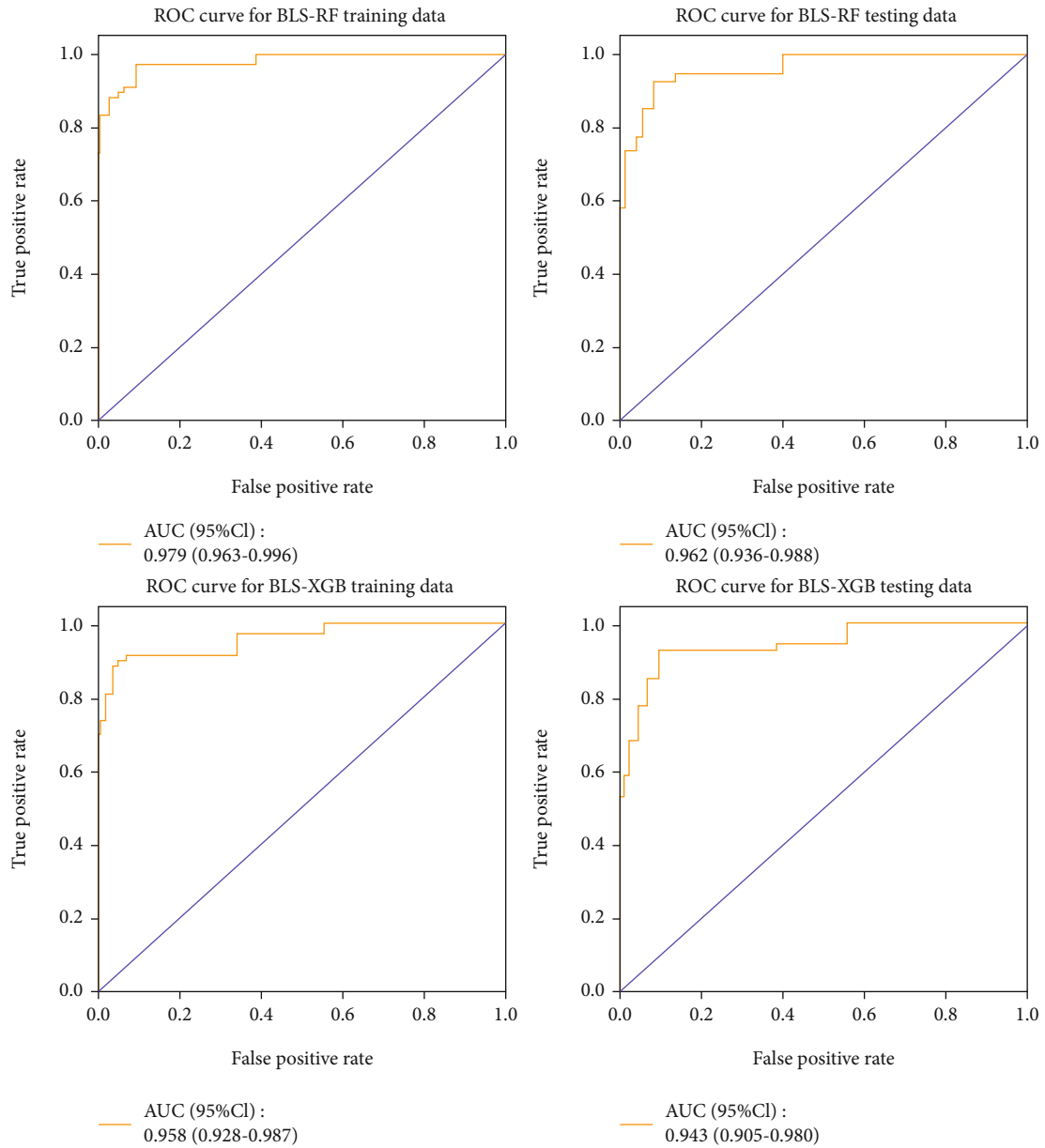


FIGURE 2: Establishment of the BLS-RF model for predicting the 28-day mortality of CAP patients.

patients hospitalized with CAP, respectively. All study variables entered the BLS to generate 106 features. Then, the two models (BLS-RF and BLS-XGB) based on the 106 features were established using RF and XGB analyses, respectively. Figure 2 displayed the establishment of the BLS-RF model. The area under the curve (AUC), accuracy, sensitivity, specificity, positive predict value (PPV), and negative predict value (NPV) evaluated the predictive performance of the six models. Internal validation of the six prediction models was conducted using the testing set. Receiver operating characteristic (ROC) curves of the BLS-RF, BLS-XGB, CNN, and DNN models for predicting the 28-day mortality of CAP patients were shown in Figure 3.

DNN consists of three layers, input layer, hidden layer, and output layer. Each layer is fully connected. Using the original data as the input layer, the sample features are obtained progressively through the hidden layer, and then the features in the output layer are predicted. For deep learning processes, 30 hidden layers are used.

CNN's full name is convolutional neural network, which includes three convolutional layer for feature extraction and max pooling layer for down sampling. And Fully Connected Layer for classification²Features are extracted by the convolutional layer, useless features are excluded by the pooling layer, and finally features in the output layer are classified and predicted by the full connection layer. In this study, four



(a)

FIGURE 3: Continued.

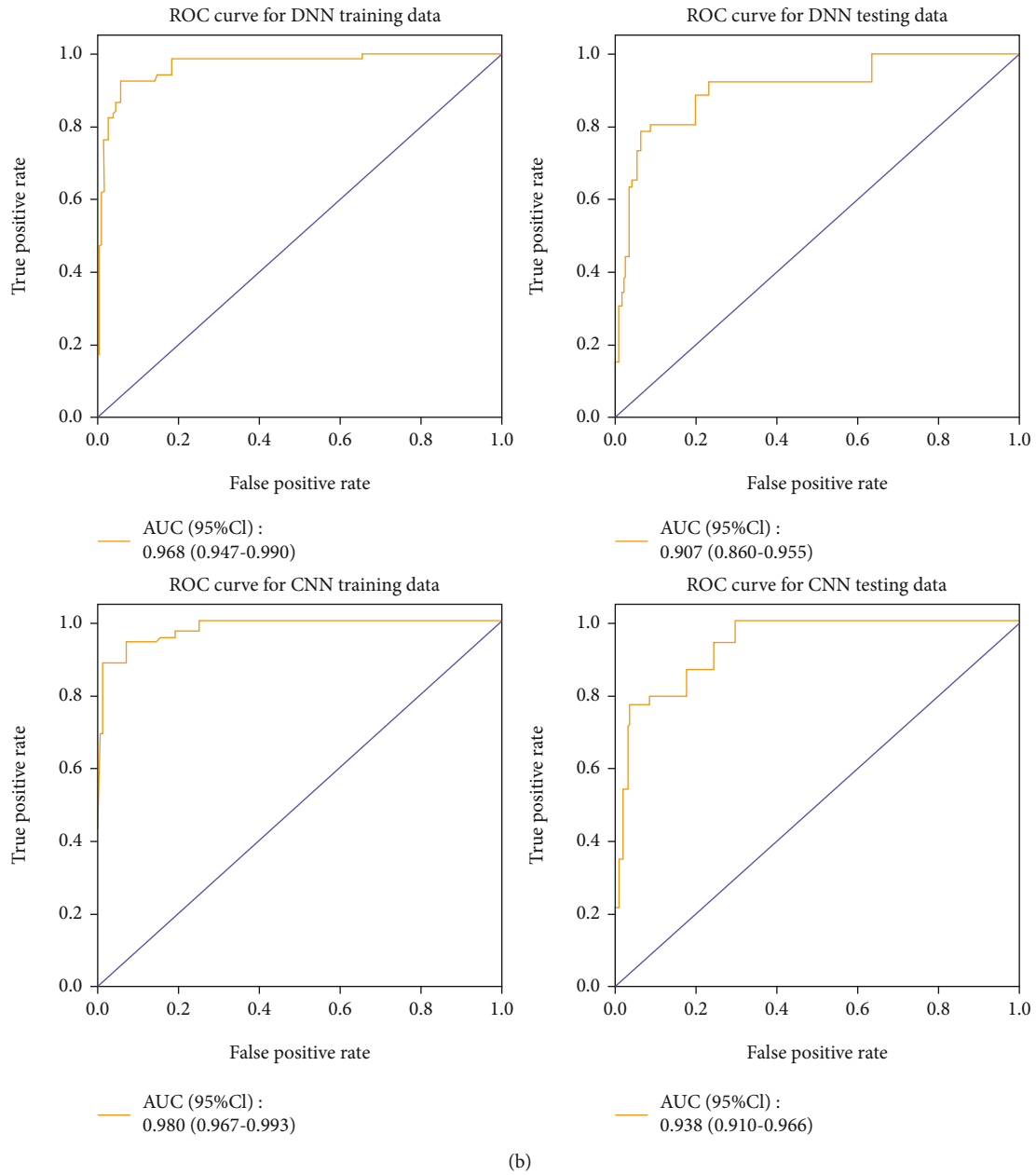


FIGURE 3: (a) ROC curves based on the integrated models. (b) ROC curves based on CNN and DNN.

convolutional layers, one pooling layer, and one full connection layer are adopted.

2.4. Statistical Analysis. The normality test for measurement data was assessed by Shapiro test. The continuous variables with normal distribution were analyzed using T test and expressed by mean \pm standard deviation (Mean \pm SD). Non-normally distributed measurement data were analyzed by Mann-Whitney U test and represented by median and quartile ($M[Q1, Q3]$). Categorical data were evaluated utilizing χ^2 test or Fisher's exact probability method, with the number of cases and the composition ratio ($N(\%)$). All missing data were filled by random forest analysis. The sen-

sitivity analysis was carried out. All statistical analyses were performed using Python software. $P < 0.05$ was considered as a statistical difference.

3. Results and Discussion

3.1. Characteristics of Patients Hospitalized with CAP. A total of 1,210 eligible CAP patients were finally included in this study, with the mean age of 63.58 ± 15.36 years. Of which, 120 cases suffered from death during hospitalization. All patients were randomly grouped into the training ($n = 726$) and testing ($n = 484$) sets according to 6:4. There were no differences in gender, age, nationality, history of diseases (allergy, hypertension, diabetes, lung disease malignant

TABLE 1: The characteristics of patients hospitalized with CAP.

Variables	Total ($n = 1210$)	Testing ($n = 484$)	Training ($n = 726$)	Statistics	P
Gender (female), n (%)	475 (39.26)	185 (38.22)	290 (39.94)	$\chi^2 = 0.361$	0.548
Age, years, mean \pm SD	63.58 \pm 15.36	63.86 \pm 15.20	63.40 \pm 15.48	$t = 0.51$	0.612
Nationality (Han), n (%)	1060 (87.60)	431 (89.05)	629 (86.64)	$\chi^2 = 1.554$	0.213
Allergy (yes), n (%)	110 (9.09)	46 (9.50)	64 (8.82)	$\chi^2 = 0.167$	0.683
Surgery (yes), n (%)	325 (26.86)	140 (28.93)	185 (25.48)	$\chi^2 = 1.753$	0.186
Hypertension (yes), n (%)	480 (39.67)	205 (42.36)	275 (37.88)	$\chi^2 = 2.432$	0.119
Diabetes (yes), n (%)	140 (11.57)	52 (10.74)	88 (12.12)	$\chi^2 = 0.538$	0.463
Smoking (yes), n (%)	395 (32.64)	171 (35.33)	224 (30.85)	$\chi^2 = 2.647$	0.104
Drinking (yes), n (%)	315 (26.03)	129 (26.65)	186 (25.62)	$\chi^2 = 0.161$	0.688
Lung disease (yes), n (%)	160 (13.22)	74 (15.29)	86 (11.85)	$\chi^2 = 3.001$	0.083
Malignant autumn, n (%)	50 (4.13)	23 (4.75)	27 (3.72)	$\chi^2 = 0.782$	0.376
HF (yes), n (%)	35 (2.89)	15 (3.10)	20 (2.75)	$\chi^2 = 0.123$	0.726
SBP, mmHg, mean \pm SD	129.71 \pm 20.07	129.50 \pm 20.01	129.84 \pm 20.11	$t = -0.29$	0.772
DBP, mmHg, mean \pm SD	80.70 \pm 13.24	80.68 \pm 13.12	80.71 \pm 13.33	$t = -0.04$	0.972
Respiratory rate, beats/minute, mean \pm SD	20.81 \pm 2.43	20.85 \pm 2.52	20.77 \pm 2.38	$t = 0.55$	0.579
HR, beats/minute, mean \pm SD	89.56 \pm 18.13	89.59 \pm 18.90	89.54 \pm 17.61	$t = 0.04$	0.966
WBC counts, $10^9/L$, $M(Q_1, Q_3)$	8.23 (6.31, 11.81)	8.46 (6.33, 11.50)	8.14 (6.26, 12.02)	$Z = 0.579$	0.562
RBC counts, $10^{12}/L$, mean \pm SD	4.29 \pm 0.73	4.31 \pm 0.70	4.28 \pm 0.74	$t = 0.74$	0.459
Hb, g/L, mean \pm SD	128.09 \pm 21.78	128.98 \pm 21.11	127.50 \pm 22.22	$t = 1.16$	0.246
PLT counts, $10^9/L$, $M(Q_1, Q_3)$	244.55 (193.00, 320.30)	244.55 (195.40, 328.00)	244.55 (190.00, 314.00)	$Z = 1.013$	0.311
AST, μ/L , $M(Q_1, Q_3)$	21.61 (16.00, 33.00)	21.00 (16.00, 32.00)	22.00 (16.00, 34.00)	$Z = -0.433$	0.665
ALB, μ/L , mean \pm SD	35.48 \pm 5.84	35.67 \pm 5.99	35.35 \pm 5.74	$t = 0.93$	0.350
BUN, mmol/L, $M(Q_1, Q_3)$	5.30 (4.10, 7.30)	5.35 (4.10, 7.30)	5.20 (4.10, 7.26)	$Z = 0.561$	0.575
Cr, $\mu\text{mol}/L$, mean \pm SD	64.16 \pm 20.92	63.28 \pm 20.90	64.75 \pm 20.93	$t = -1.19$	0.233
Glu, mmol/L, $M(Q_1, Q_3)$	5.90 (5.00, 7.26)	5.80 (4.99, 7.17)	5.99 (5.00, 7.30)	$Z = -1.203$	0.229
PCT, $\mu\text{g}/L$, $M(Q_1, Q_3)$	0.10 (0.05, 0.34)	0.10 (0.06, 0.31)	0.10 (0.05, 0.36)	$Z = -0.343$	0.732
CRP, mg/L, $M(Q_1, Q_3)$	50.20 (11.90, 114.00)	43.00 (11.50, 108.00)	54.15 (12.20, 115.00)	$Z = -1.073$	0.283
Survival state, n (%)				$\chi^2 = 0.617$	0.432
Survival	1090 (90.08)	432 (89.26)	658 (90.63)		
Death	120 (9.92)	52 (10.74)	68 (9.37)		

CAP: community-acquired pneumonia; SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; WBC: white blood cell; RBC: red blood cell; Hb: hemoglobin; PLT: platelet; AST: aspartate aminotransferase; ALB: albumin; BUN: blood urea nitrogen; Cr: creatinine; Glu: blood glucose; PCT: porcine calcitonin; CRP: C-reactive protein.

tumor, and HF), history of surgery, smoking, drinking, SBP, DBP, respiratory rate, HR, WBC counts, RBC counts, Hb level, PLT counts, AST level, ALB level, BUN level, Cr level, Glu level, PCT level, and CRP level (all $P > 0.05$). It was indicated that the data was balanced between the two sets. The characteristics of CAP patients in the training and testing sets were shown in Table 1.

3.2. The Predictive Performance of the Models for the 28-Day Mortality of CAP Patients. The AUC values of the BLS-RF model for predicting the 28-day mortality of CAP patients were 0.979 (95% CI: 0.963-0.996) and 0.962 (95% CI:

0.936-0.988) in the training and testing sets, respectively. The AUC values of the BLS-XGB model were 0.958 (95% CI: 0.928-0.988) and 0.943 (95% CI: 0.905-0.980) in the training and testing sets, respectively. The AUC of DNN used in training set is 0.968 (95% CI: 0.947-0.990), and the AUC in test set is 0.907 (95% CI: 0.860-0.955). The AUC of CNN in training set was 0.980 (95% CI: 0.967-0.993), and AUC in testing set was 0.938 (95% CI: 0.910-0.966). Using the basic prediction model, the AUC of RF in the training set is 0.900 (95% CI: 0.861-0.939), and the AUC in the testing set is 0.786 (95% CI: 0.727-0.846). The AUC of logistic model was 0.832 (95% CI: 0.785-0.879) in training

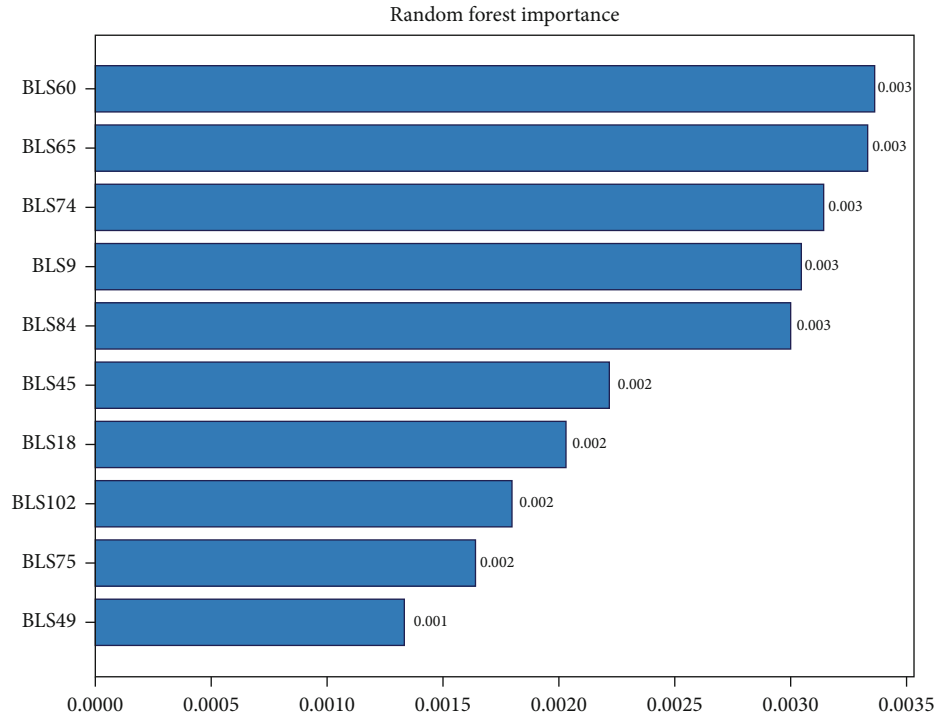


FIGURE 4: Importance of model features.

set and 0.714 (95% CI: 0.649-0.780) in testing set. Finally, BLS is used to learn and output features, and random forest prediction is used, as shown in Figures 3 and 4.

From the AUC, we can find that the AUC of the two training models based on BLS is similar in the testing set ($P = 0.414$). However, there was no significant difference between the AUC of DNN and CNN in the testing set ($P = 0.270$). There is no significant difference between the AUC of the two basic prediction models Logistic and Random Forest in the testing set ($P = 0.371$). The AUC of the testing set of BLS-based stochastic forest model is better than that of DNN ($P = 0.047$). The AUC of integrated models in the testing set not only is better than those of basic model RF and logistic in the testing set.

3.3. Importance Diagram of the BLS-Based Features. As stated in Section 2, BLS is used to learn and output features, and random forest prediction is used. Among the BLS output features, the features with the highest feature importance are the 60th, 65, 74, 9, 84, 45, 18, 102, 75, and 49 among the top 10 features with the highest model importance, BLS60 is the most important, followed by BLS65, BLS49 is the lowest, see details in Figure 4.

Machine learning analysis with text representation has been utilized in some previous studies, such as early detection of readmission risk for decision support based on clinical notes, discovering the predictive value of clinical notes, deep learning approaches in chest radiograph, and deep learning techniques on chest X-ray and CT scan [18, 19]. But a further investigation on applications in predicting the death risk of CAP among hospitalized patients with respiratory com-

plaints is still required, and a novel approach to improve the model performance is also quite necessary [20–23].

3.4. Comparison for the Prediction Models. The accuracy, sensitivity, specificity, PPV, and NPV of the two prediction models are established by using BLS to learn and output features, which is a brain-inspired model [24]. And then using the random forest and XGB to extract features from BLS is the highest among all models. Hence, two integrated models BLS-RF and BLS-XGB are established. The sensitivity and NPV of the BLS – RF model using the training set are 0.970 (95% CI: 0.929-1.000) and 0.997 (95% CI: 0.992-1.000), and those using the testing set are 0.925 (95% CI: 0.853-0.996) and 0.989 (95% CI: 0.979-1.000), respectively. In the training set the accuracy specificity, and PPV of the BLS-XGB model are 0.959 (95% CI: 0.944-0.973), 0.967 (95% CI: 0.953-0.980), and 0.728 (95% CI: 0.632-0.825), respectively. In the testing set, the accuracy, specificity, and PPV of the BLS-XGB model are 0.932 (95% CI: 0.909-0.954), 0.958 (95% CI: 0.939-0.977), and 0.679 (95% CI: 0.565-0.801), as shown in Figure 5.

In Section 3.2, we utilized the BLS to construct better hidden-layer architectures and connectivity to extract the data features, and in this section, we further trained parameters in the integrated broad learning system and compare the efficiency of the integrated models with previous algorithms by performance in predicting the death risk of patients with acquired pneumonia after 28-day hospitalization.

As shown in Table 2, experimental results show that the integrated model BLS-XGB (training accuracy = 95.9%, testing accuracy = 93.2%) as an efficient BLS for predicting

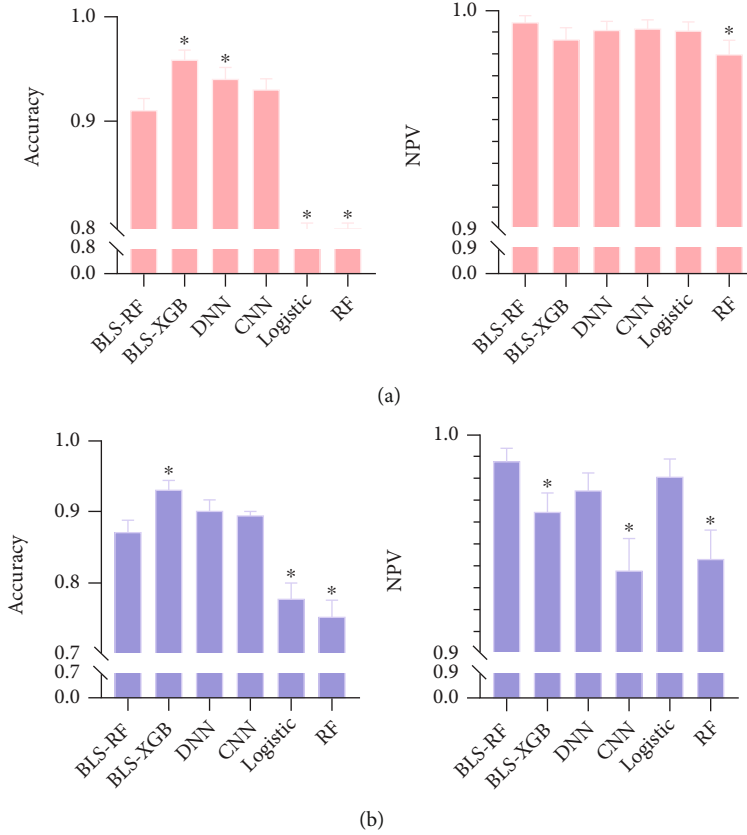


FIGURE 5: Bar graphs of accuracy and NPV of the models on the training (a) and testing (b) set.

TABLE 2: The predictive performance of the models for the 28-day mortality of CAP patients.

Prediction models	Accuracy	Sensitivity	Specificity	PPV	NPV
Training set					
BLS-RF	0.912 (0.891-0.932)	0.970 (0.929-1.000)	0.906 (0.884-0.928)	0.512 (0.425-0.599)	0.997 (0.992-1.000)
BLS-XGB	0.959 (0.944-0.973)	0.881 (0.803-0.958)	0.967 (0.953-0.980)	0.728 (0.632-0.825)	0.988 (0.979-0.996)
DNN	0.941 (0.924-0.958)	0.926 (0.864-0.989)	0.942 (0.924-0.960)	0.624 (0.529-0.718)	0.992 (0.985-0.999)
CNN	0.930 (0.911-0.948)	0.941 (0.885-0.997)	0.929 (0.909-0.948)	0.577 (0.485-0.668)	0.993 (0.987-1.000)
Logistic	0.789 (0.758-0.818)	0.941 (0.885-0.997)	0.774 (0.742-0.806)	0.300 (0.239-0.362)	0.992 (0.985-1.000)
RF	0.791 (0.761-0.820)	0.853 (0.769-0.937)	0.784 (0.753-0.816)	0.290 (0.227-0.353)	0.981 (0.969-0.993)
Testing set					
BLS-RF	0.872 (0.842-0.902)	0.925 (0.853-0.996)	0.865 (0.833-0.898)	0.458 (0.364-0.552)	0.989 (0.979-1.000)
BLS-XGB	0.932 (0.909-0.954)	0.717 (0.596-0.838)	0.958 (0.939-0.977)	0.679 (0.565-0.801)	0.965 (0.948-0.982)
DNN	0.903 (0.877-0.929)	0.808 (0.701-0.915)	0.914 (0.888-0.941)	0.532 (0.422-0.642)	0.975 (0.960-0.990)
CNN	0.897 (0.890-0.924)	0.769 (0.655-0.884)	0.912 (0.885-0.939)	0.513 (0.402-0.624)	0.938 (0.910-0.966)
Logistic	0.779 (0.739-0.815)	0.885 (0.798-0.971)	0.766 (0.726-0.806)	0.313 (0.238-0.388)	0.982 (0.968-0.996)
RF	0.756 (0.718-0.794)	0.615 (0.483-0.748)	0.773 (0.734-0.813)	0.246 (0.172-0.320)	0.944 (0.919-0.968)

CAP: community-acquired pneumonia; BLS: broad learning system; RF: random forest; XGB: eXtreme Gradient Boosting; DNN: deep neural network; CNN: convolutional neural network; PPV: positive predictive value; NPV: negative predictive value.

the death risk of patients, which not only performs better than the two basic models RF (training accuracy = 79.1%, testing accuracy = 75.6%) and the integrated model BLS-RF (training accuracy = 95.9%, testing accuracy = 93.2%) but also performs better than BLS-XGB (training accuracy = 91.2%, testing accuracy = 87.2%) and two well-known deep learning

systems-DNN (training accuracy = 94.1%, testing accuracy = 90.3%) and CNN (training accuracy = 93.0%, testing accuracy = 89.7%), and the competitiveness of the proposed model is further proved. Suggest the integrated model BLS-XGB as an efficient BLS for predicting the death risk of patients.

This study was to develop a prediction model for the risk of the 28-day mortality in patients hospitalized with CAP, which is essentially significant for emergent treating system in intelligent decisions for modern hospitals [24–28]. The potential engineering applications of our proposed model will not be limited to the patients hospitalized with CAP [29–32]. We used RF and XGB methods after learning the sample characteristics of the data by BLS [33–38]. This approach is novel compared to the previous studies on predicting the risk of death among CAP cases [39–43]. Accuracy of the integrated model is more than 90%, indicating a robust prediction. Our model also makes prediction according to various indicators of patients. At the same time, compared with the method of the previous basic models and other competitive models, the integrated model has significantly improved the performance accuracy in practical applications. The unresolved issues are also the main challenges in treating pneumonia is that a patient's condition can deteriorate suddenly, and therefore, the subsequent emergent treatment for saving personal patients needs a further utilization of other methods in medicine and artificial intelligence.

4. Conclusion

BLS offers an alternative way of learning in deep structure and in the present study, after being integrated with XGB, the experiments indicate a robust prediction for control the 28-day mortality risk of CAP patients after hospital admission. The integrated model BLS-XGB was selected as an efficient model to control the 28-day mortality of patients hospitalized with CAP. For subsequent studies, we encourage other researchers to extend potential engineering applications of our proposed model (not be limited to the patients hospitalized with CAP). Another next research priority is to find accompanied methods in medicine and artificial intelligence for the emergent treatment for saving personal patients after the death risk is predicted.

Data Availability

The data utilized to support the findings are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by the Shanghai High-Level Base-Building Project for Industrial Technology Innovation (1021GN204005-A06).

References

- [1] W. Yates, "A review of infectious bovine rhinotracheitis, shipping fever pneumonia and viral-bacterial synergism in respiratory disease of cattle," *Revue canadienne de medecine comparee*, vol. 46, no. 3, pp. 225–263, 1982.
- [2] H. K. Eslamy and B. Newman, "Pneumonia in normal and immunocompromised children: an overview and update," *Radiologic Clinics of North America*, vol. 49, no. 5, pp. 895–920, 2011.
- [3] T. Lancet, "Pleuropneumonia-like organisms in the human vagina," *Annals of Botany*, vol. 109, no. 7, pp. 1263–1275, 2011.
- [4] S. M. Brown, J. P. Jones, D. Aronsky, B. E. Jones, M. J. Lanspa, and N. C. Dean, "Relationships among initial hospital triage, disease progression and mortality in community-acquired pneumonia," *Respirology*, vol. 17, no. 8, pp. 1207–1213, 2012.
- [5] H. Uematsu, K. Yamashita, S. Kunisawa, T. Otsubo, and Y. Imanaka, "Prediction of pneumonia hospitalization in adults using health checkup data," *PLoS One*, vol. 12, no. 6, article e0180159, 2017.
- [6] J. M. Baren, P. L. Henneman, and R. J. Lewis, "Primary varicella in adults: pneumonia, pregnancy, and hospital admission," *Annals of Emergency Medicine*, vol. 28, no. 2, pp. 165–169, 1996.
- [7] J. D. Heffelfinger, S. F. Dowell, J. H. Jorgensen et al., "Management of community-acquired pneumonia in the era of pneumococcal resistance: a report from the drug-resistant Streptococcus pneumoniae therapeutic working group," *Archives of Internal Medicine*, vol. 160, no. 10, pp. 1399–1408, 2000.
- [8] J. A. Schouten, M. E. Hulscher, B. J. Kullberg et al., "Understanding variation in quality of antibiotic use for community-acquired pneumonia: effect of patient, professional and hospital factors," *Journal of Antimicrobial Chemotherapy*, vol. 56, no. 3, pp. 575–582, 2005.
- [9] J. Hedlund, "Community-acquired pneumonia requiring hospitalisation. Factors of importance for the short- and long-term prognosis," *Scandinavian Journal of Infectious Diseases*, vol. 97, no. 97, pp. 1–60, 1995.
- [10] M. Christ-Crain and S. M. Opal, "Clinical review: the role of biomarkers in the diagnosis and management of community-acquired pneumonia," *Critical Care*, vol. 14, no. 1, pp. 203–211, 2010.
- [11] P. S. Heckerling, B. S. Gerber, T. G. Tape, and R. S. Wigton, "Use of genetic algorithms for neural networks to predict community-acquired pneumonia," *Artificial Intelligence in Medicine*, vol. 30, no. 1, pp. 71–84, 2004.
- [12] E. García-Vázquez, S. Soto, J. Gómez, and J. A. Herrero, "Simplificación de los criterios para valorar el riesgo de mortalidad en pacientes con neumonía adquirida en la comunidad," *Medicina Clínica*, vol. 131, no. 6, pp. 201–204, 2008.
- [13] O. Leroy, P. Devos, B. Guery et al., "Simplified prediction rule for prognosis of patients with severe community-acquired pneumonia in ICUs," *Chest*, vol. 116, no. 1, pp. 157–165, 1999.
- [14] S. Singh, "Pneumonia detection using deep learning," in *2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*, NaviMumbai, India, 2021.
- [15] D. Hussein, D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Computers in Biology and Medicine*, vol. 132, article 104348, 2021.
- [16] L. Akter and N. Akhter, "Ovarian cancer prediction from ovarian cysts based on TVUS using machine learning algorithms," *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*, M. S. Arefin, M. S.

- Kaiser, A. Bandyopadhyay, M. A. R. Ahad, and K. Ray, Eds., Springer, Singapore, 2022.
- [17] C. L. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 29, no. 1, pp. 10–24, 2018.
- [18] S. Bellens, P. Vandewalle, and W. Dewulf, "Deep learning based porosity segmentation in X-ray CT measurements of polymer additive manufacturing parts," *Procedia CIRP*, vol. 96, no. 2, pp. 336–341, 2021.
- [19] R. K. Gupta, Y. Sahu, N. Kunhare, A. Gupta, and D. Prakash, "Deep learning-based mathematical model for feature extraction to detect corona virus disease using chest X-ray images," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, no. 6, pp. 921–947, 2021.
- [20] L. Ryan, S. Mataraso, A. Lynn-Palevsky et al., "371: a machine learning approach to predict deep venous thrombosis among hospitalized patients," *Critical Care Medicine*, vol. 49, no. 1, pp. 175–175, 2021.
- [21] P. A. Koul and G. Investigators, "Aspiration risk factors, microbiology, and empiric antibiotics for patients hospitalized with community-acquired pneumonia," *Chest*, vol. 159, no. 1, pp. 58–72, 2021.
- [22] L. Ryan, S. Mataraso, A. Siefkas et al., "A machine learning approach to predict deep venous thrombosis among hospitalized patients," *Clinical and Applied Thrombosis/Hemostasis*, vol. 27, no. 23, 2021.
- [23] C. L. Wu, M. J. Wu, L. C. Chen et al., "AEP-DLA: adverse event prediction in hospitalized adult patients using deep learning algorithms," *IEEE Access*, vol. 9, pp. 55673–55689, 2021.
- [24] W. F. Wang, X. Y. Deng, L. Ding, and L. Zhang, "Brain-inspired intelligence and visual perception: the brain and machine eyes," Springer, 2020.
- [25] I. Eigner and F. Bodendorf, "An intelligent decision support system for readmission prediction in healthcare," *Information Technology*, vol. 60, no. 4, pp. 195–205, 2018.
- [26] R. Summers and E. R. Carson, "Evaluation of intelligent decision aids for application in critical care medicine," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society Volume 13: 1991*, Orlando, FL, USA, 1991.
- [27] R. M. Meneghini, "Resource reallocation during the COVID-19 pandemic in a suburban hospital system: implications for outpatient hip and knee arthroplasty," *The Journal of Arthroplasty*, vol. 35, no. 7, pp. S15–S18, 2020.
- [28] E. Kalantar, A. Kurd, K. Kabir, P. Afrogh, S. Mohammadi, and M. H. Naseh, "Extreme antibiotic resistant acinetobacter baumannii-related pneumonia in a regional hospital," *Infection Epidemiology and Medicine*, vol. 2, no. 4, pp. 29–31, 2016.
- [29] M. A. Dimopoulos, E. Terpos, M. Boccadoro et al., "Daratumumab plus pomalidomide and dexamethasone versus pomalidomide and dexamethasone alone in previously treated multiple myeloma (APOLLO): an open-label, randomised, phase 3 trial," *The Lancet Oncology*, vol. 22, no. 6, pp. 801–812, 2021.
- [30] R. J. Tarpey and M. T. Mullarkey, "Engineering innovative clinical resource management by design: a guided emergent search through a complex adaptive system of systems," *IEEE Transactions on Engineering Management*, pp. 1–15, 2021.
- [31] Y. M. Sakti and R. N. Khadafi, "Emergent spine surgery during COVID-19 pandemic: 10 months experience in Dr. Sardjito general hospital, Indonesia a case series," *Annals of Medicine and Surgery*, vol. 67, no. 18, article 102513, 2021.
- [32] S. Rajendiran, S. Thahir, Y. Veloo et al., "Environmental surface sampling of SARS-CoV-2 in selected hospitals in Malaysia," *Tropical Biomedicine*, vol. 38, no. 3, pp. 462–468, 2021.
- [33] P. V. Opia and R. R. Corpuz, "Suitability mapping of small farm reservoirs using K-nearest neighbors and random forest algorithms," *International Journal of Scientific & Technology Research*, vol. 10, no. 3, pp. 78–80, 2021.
- [34] P. Shanmugam, J. Raja, and R. Pitchai, "An automatic recognition of glaucoma in fundus images using deep learning and random forest classifier," *Applied Soft Computing*, vol. 109, no. 6, article 107512, 2021.
- [35] W. P. Zhao, J. Li, J. Zhao, D. Zhao, J. Lu, and X. Wang, "XGB model: research on evaporation duct height prediction based on XGBoost algorithm," *Radioengineering*, vol. 29, no. 1, pp. 81–93, 2020.
- [36] W. B. Chang, Y. L. Liu, X. Y. Wu, Y. Xiao, S. Zhou, and W. Cao, "A new hybrid XGBSVM model: application for hypertensive heart disease," *IEEE Access*, vol. 7, pp. 175248–175258, 2019.
- [37] C. Chen and Z. Liu, "Broad Learning System: A New Learning Paradigm and System without Going Deep," in *2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Hefei, China, 2017.
- [38] W. F. Wang, H. J. Cai, X. Y. Deng, C. Lu, and L. Zhang, *Interdisciplinary evolution of the machine brain*, Springer, 2021.
- [39] T. T. Bauer, S. Ewig, R. Marre, N. Suttorp, T. Welte, and THE CAPNETZ STUDY GROUP, "CRB-65 predicts death from community-acquired pneumonia," *Journal of Internal Medicine*, vol. 260, no. 1, pp. 93–101, 2006.
- [40] S. Kruger, S. Ewig, R. Marre et al., "Procalcitonin predicts patients at low risk of death from community-acquired pneumonia across all CRB-65 classes," *European Respiratory Journal*, vol. 31, no. 2, pp. 349–355, 2008.
- [41] N. Kokturk, A. Kanbay, N. Bukan, and N. Ekim, "The value of serum procalcitonin in differential diagnosis of pulmonary embolism and community-acquired pneumonia," *Clinical & Applied Thrombosis/Hemostasis*, vol. 17, no. 5, pp. 519–525, 2011.
- [42] N. Daneman, D. E. Low, A. McGeer, K. A. Green, and D. N. Fisman, "At the threshold: defining clinically meaningful resistance thresholds for antibiotic choice in community-acquired pneumonia," *Clinical Infectious Diseases*, vol. 46, no. 8, pp. 1131–1138, 2008.
- [43] J. Gamble, D. T. Eurich, T. J. Marrie, and S. R. Majumdar, "Admission hypoglycaemia portends a substantially increased risk of mortality in patients with community-acquired pneumonia: population-based cohort study," *Canadian Journal of Diabetes*, vol. 33, no. 3, pp. 185–185, 2009.

Research Article

A Hemolysis Image Detection Method Based on GAN-CNN-ELM

Xiaonan Shi , Yong Deng , Yige Fang , Yajuan Chen , Ni Zeng , and Limei Fu 

College of Computer Science, Xi'an University of Science and Technology, Shanxi 710054, China

Correspondence should be addressed to Xiaonan Shi; shi_xn@xust.edu.cn

Received 23 June 2021; Accepted 18 January 2022; Published 22 February 2022

Academic Editor: Luminita Moraru

Copyright © 2022 Xiaonan Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since manual hemolysis test methods are given priority with practical experience and its cost is high, the characteristics of hemolysis images are studied. A hemolysis image detection method based on generative adversarial networks (GANs) and convolutional neural networks (CNNs) with extreme learning machine (ELM) is proposed. First, the image enhancement and data enhancement are performed on a sample set, and GAN is used to expand the sample data volume. Second, CNN is used to extract the feature vectors of the processed images and label eigenvectors with one-hot encoding. Third, the feature matrix is input to the map in the ELM network to minimize the error and obtain the optimal weight by training. Finally, the image to be detected is input to the trained model, and the image with the greatest probability is selected as the final category. Through model comparison experiments, the results show that the hemolysis image detection method based on the GAN-CNN-ELM model is better than GAN-CNN, GAN-ELM, GAN-ELM-L1, GAN-SVM, GAN-CNN-SVM, and CNN-ELM in accuracy and speed, and the accuracy rate is 98.91%.

1. Introduction

Hemolysis is the process by which the red blood cell membrane breaks down, and the hemoglobin in the cell is released into plasma, serum, and other intercellular substances. This abnormal expression can easily lead to many diseases. In particular, neonatal hemolysis has become one of the most important diseases affecting newborn health [1, 2]. In 2018, Zhang [3] analyzed 200 cases of neonatal jaundice. Neonatal hemolysis occurred in 61.61% of the 99 female infants. Among 101 baby boys, neonatal hemolysis accounted for 50.49%. Thus, it can be seen that the probability of jaundice caused by hemolysis of newborns is still relatively high. However, many patients become sicker and even lose their lives since hemolysis is not detected early enough [4].

In clinical hemolysis tests, the specimen character and quality judgment need to be visually measured and recorded by medical staff one by one, and the degree of hemolysis is easily affected by subjective factors. A Vitros 5600 biochemical immune analyzer is used to detect the hemolysis index, realizing automatic quantitative determination of the hemolysis degree [5] and avoiding errors and missing terms

caused by manual operation. The concentration of NSE in neonatal hemolysis samples is calculated by the neuron-specific enolase (NSE) correction formula, which avoided repeating sample collection. It is particularly difficult to collect blood samples from newborns [6]. These methods rely on chemical instruments and the practical experience of medical personnel. The measurement of relevant data is complex and takes a long time.

In recent years, with the rapid development of deep learning technology, computer-aided detection technology has provided better results than traditional methods [7, 8]. In particular, medical image analysis provides medical personnel with more effective diagnostic and therapeutic procedures [9]. However, large-scale tagging data are difficult to obtain. Labeling medical data requires professional knowledge and invests considerable time and manpower. By training the existing data, generative adversarial networks (GANs) can automatically generate a large quantity of data to effectively solve this problem. For example, Qi et al. [10] generated a large number of diverse pulmonary nodule images through GAN, which saved time and cost. Convolutional neural networks (CNNs) also provide great help in medical image analysis. Huynh-the et al. [11] converted

skeletal information into an image representation based on CNN. The experimental results with the highest accuracy were obtained on the most challenging datasets. Bueno et al. [12] achieved higher accuracy by using the sequential CNN segmentation-classification strategy. As a new training framework, the extreme learning machine (ELM) has been widely used in classification, identification, and diagnosis [13]. For example, Wang and Li [14] used ELM to classify pulse data, with an accuracy of 90.3%. Wang et al. [15] used the ELM method to quickly and accurately determine whether the human body has fallen. Cai et al. [16] verified the effectiveness and superiority of ELM in lithology identification through comparison.

To detect hemolysis images using computer-aided techniques instead of manual work and to improve the related algorithm, this paper proposes a hemolysis image detection method based on GAN-CNN-ELM. The prediction accuracy of this method is up to 98.91%, and the recognition accuracy and efficiency are better than GAN-CNN, GAN-ELM, GAN-ELM-L1, GAN-SVM, GAN-CNN-SVM, and CNN-ELM models. Furthermore, it realizes the efficient detection of hemolysis image and helps medical staff diagnose the condition in time and strives for more precious treatment time for patients.

This paper has three main contributions.

- (1) A large number of hemolysis images are generated by GAN to expand the dataset
- (2) The CNN algorithm is used to extract multilayer features from the preprocessed hemolysis experiment images
- (3) The ELM classifier is constructed to classify the eigenvalues and obtain more accurate prediction results

The paper is organized as follows. Section 1 briefly describes the problems existing in hemolysis detection and the solution model. In Section 2, some specific methods adopted are introduced. Section 3 gives the experimental process and results. Finally, in Section 4, several conclusions are presented, and the direction of our future work is mentioned.

2. Methodology

This paper presents an effective hemolysis image detection technique. The algorithm framework is shown in Figure 1. The main method includes three stages as follows:

- (a) Image preprocessing
- (b) CNN extraction features
- (c) ELM classification

2.1. Image Preprocessing. The preprocessing of images mainly includes image segmentation and clipping, image enhancement, graphic transformation data enhancement, and generative adversarial network data enhancement.

Image preprocessing is carried out with X images in M original image sets as the training atlas. $M-X$ images represent the experimental atlas.

2.1.1. Picture Initial Cropping and Labeling. X training pictures are segmented and cropped to obtain pictures of the hemolysis area in the test tube. Then, experts judged and labeled the hemolysis area and divided the images into three categories. The first category is insoluble, and the category number is 0, mainly based on the obvious precipitation in the test tube and obvious stratification. The second category is microsolubility, and the category number is 1. This is mainly based on the fact that the hemolysis solution in the test tube is partially precipitated, but the dissolution is relatively uniform in the upper layer, with obvious particles. The third category is the normal dissolution category, and the category number is 2, mainly based on the solution in the test tube dissolved evenly, excluding extreme cases (completely insoluble or clear water test tube); no larger particles or precipitation appeared. The results of the classification are shown in Figures 2 and 3.

2.1.2. Image Enhancement. The collection of hemolysis images may be affected by uncontrollable factors such as light and photo angle, leading to the phenomenon of extremely concentrated grayscale of hemolysis images or unobvious feature areas [17, 18]. Therefore, image enhancement of the collected hemolysis image can cause different angle features and detail features of the image to appear, reduce the impact of irrelevant information, promote the model to extract more features of the image, and identify hemolysis from multiple features.

This algorithm mainly adopts histogram equalization [19] to enhance the image. For example, for a discrete grayscale image $\{x\}$, the number of occurrences of grayscale i is expressed as n_i , and the probability of occurrence of pixels with grayscale i in the image is expressed as

$$p_x(i) = p(x = i) = \frac{n_i}{n}, \quad 0 \leq i < L, \quad (1)$$

where L is all the grayscale numbers in the image, and the cumulative distribution function corresponding to P_x is defined as

$$\int df_x(i) = \sum_{j=0}^i p_x(j). \quad (2)$$

In this paper, the probability distribution of the pixels of the original image on the grayscale level is changed by the method of histogram equilibrium to achieve the effect of image enhancement.

2.1.3. Data Enhancement. Because of the unique characteristics of medical images, the position and angle of the subject image and other factors directly affect the judgment of the testing personnel. In CNN training, irregular images are likely to make model training unstable, resulting in the model not converging. Therefore, the data after image

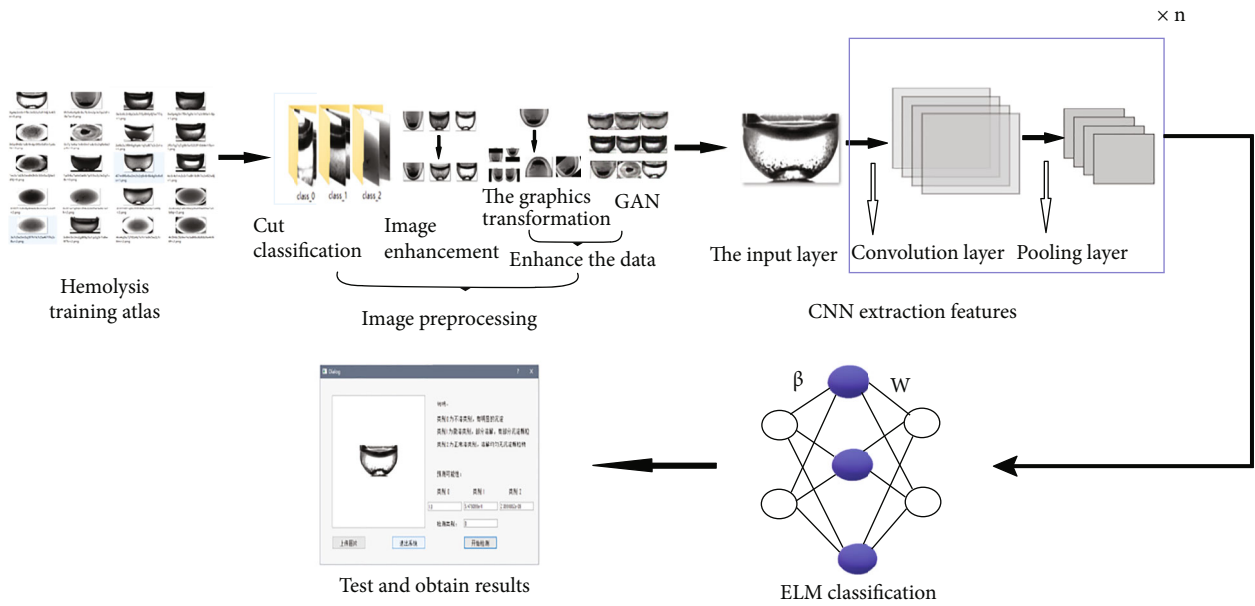


FIGURE 1: Algorithm framework.

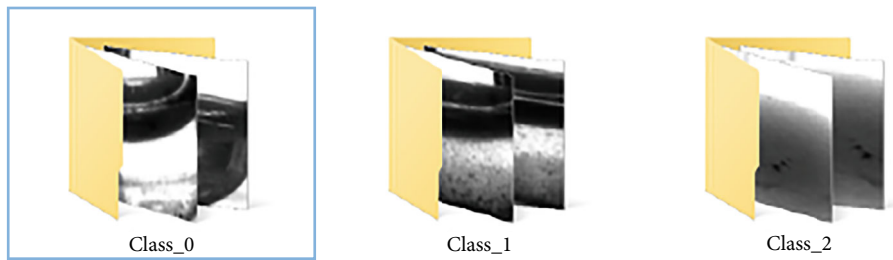


FIGURE 2: Image classification of the hemolysis area.

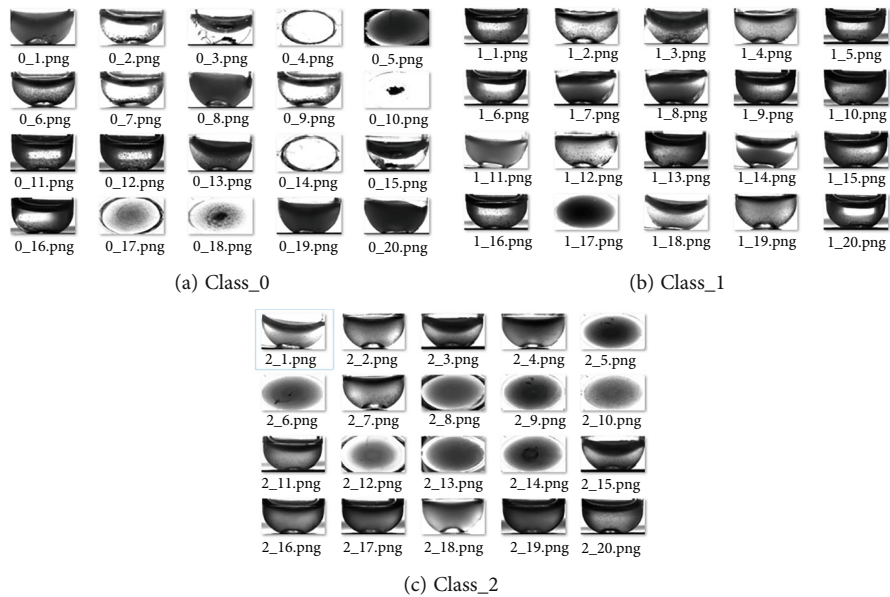


FIGURE 3: Classification of hemolysis images of different categories.

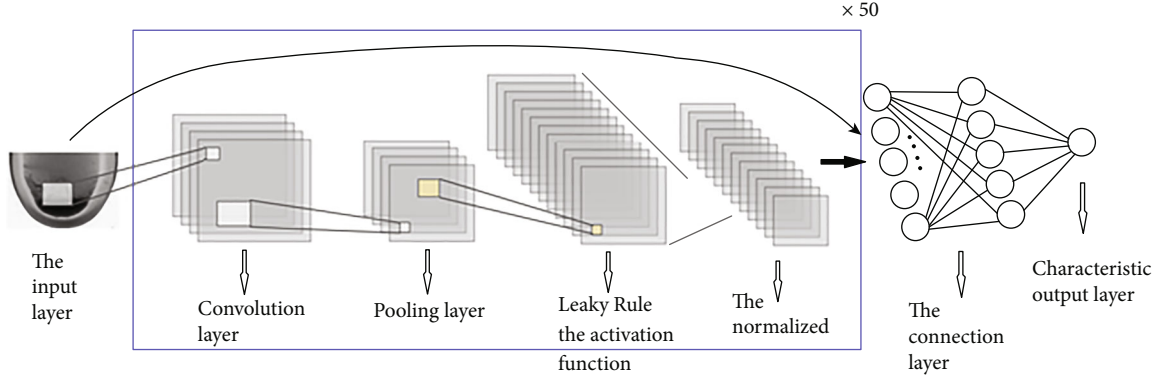


FIGURE 4: Schematic diagram of the ResNet50 algorithm model.

enhancement are enhanced to highlight the main information of the hemolytic image, standardize the dataset, and reduce the noise interference.

(1) *The Graphic Transformation.* The algorithm model adopts three methods of translation transformation, rotation transformation, and horizontal flipping to enhance the data [20–22].

Image translation transformation: assuming that there is a point a in the original image whose coordinates are (x, y) , which is translated by x_0 units along the x -axis and y_0 units along the y -axis and the new coordinate is $(x_1, y_1) = (x + x_0, y + y_0)$, its matrix expression is shown as

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (3)$$

Horizontal flip: setting its original image width to width and height to height. The coordinates of one point in the original image is (x_0, y_0) . After the horizontal mirror flip coordinates become $(x_1, y_1) = (\text{width} - x_0, y_0)$, the matrix expression is represented as

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & \text{width} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}. \quad (4)$$

Rotation transformation: rotating the image clockwise around the origin by angle θ , then the coordinates of x after point (x, y) in the original image is changed to $x \cos \theta - y \sin \theta$ and the coordinates of y is changed to $x \sin \theta + y \cos \theta$. The matrix expression is given as

$$\begin{bmatrix} a(x, y) \\ b(x, y) \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \quad (5)$$

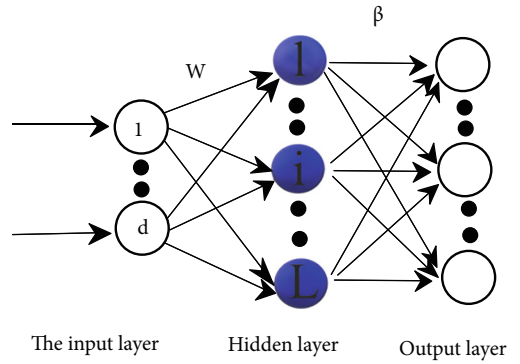


FIGURE 5: ELM model.

TABLE 1: Data distribution tables.

	Total datasets (2250)	Training sets (1800)	Test sets (450)
Low solubility	931	737	194
Medium solubility	668	542	126
High solubility	651	521	130

In this paper, the original image is first translated and transformed and then rotated at random angles. Finally, the image is randomly cropped to discard some irrelevant information in the image. Then, the data are expanded by means of horizontal inversion, and some error samples caused by excessive enhancement are removed.

(2) *Generative Adversarial Networks.* After the graph transformation, the unavailable samples are eliminated, resulting in a further reduction in the available training data. To solve the problem of insufficient training data, the existing data are trained by generating models to obtain a large amount of data to enhance the generalization ability of the network and improve the stability of the network and the universality of new samples.

GANs [23] are composed of two networks: a generator and a discriminator. The generator learns the distribution of real data as much as possible. The discriminator learns

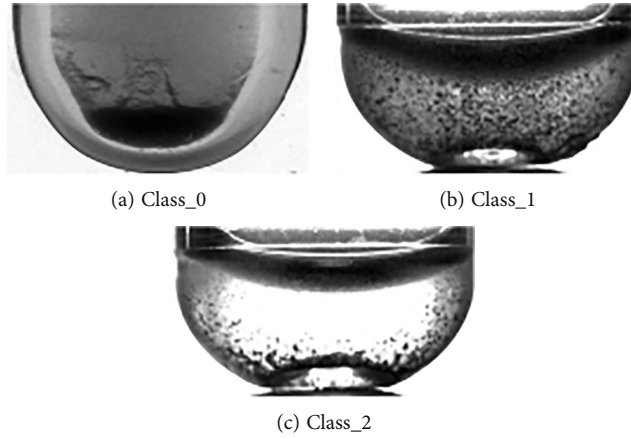


FIGURE 6: Original image.

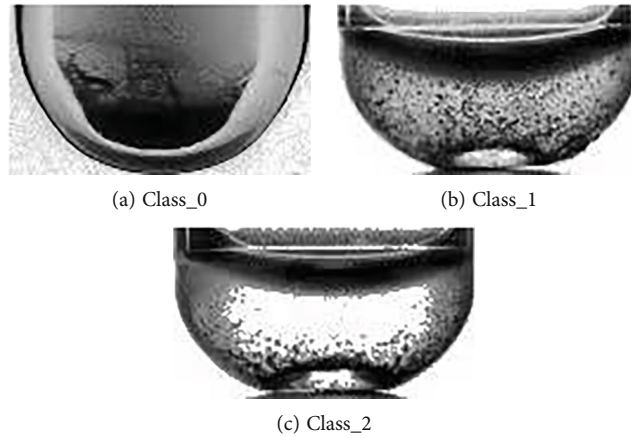


FIGURE 7: Image enhancement process.

to judge whether the input data are from real data or generated data, and they play games and train against each other to finally reach the equilibrium state [24, 25].

Let the formula of sample x generated by a group of random noise z after the generator [17, 18] be as follows:

$$x' = g(z, \theta_g), \tag{6}$$

where $g(x)$ is the final output after several convolution, pooling, and activation operations (see Equations (14) and (15), function with discriminator $D_f(x, \theta_g)$.

The output results of the real sample (x) and the generated sample (x') after passing through the discriminator are shown in Equations (7) and (8).

$$x = D_f(x, \theta_g), \tag{7}$$

$$x' = D_f(x', \theta_g). \tag{8}$$

The discriminant result can be expressed as with the

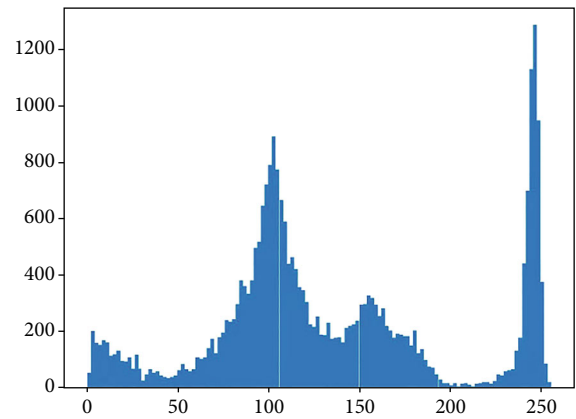


FIGURE 8: Gray histogram of Figure 6(a).

dichotomy result.

$$y = \left(\frac{D(x)}{D(x')} \right) = \left(\frac{D(x)}{D(G(z))} \right), \tag{9}$$

where $D(x)$ represents the probability that x belongs to the real sample distribution, and the loss function [26, 27] is established. It is expressed as

$$\min_G \max_D V(G, D) = E_{x \sim P_{\text{data}}}[\log D(x)] + E_{z \sim P_z}[\log (1 - D(G(z)))]. \quad (10)$$

The GAN is optimized with a backpropagation algorithm.

When optimizing the generator, minimize $V(G, D)$. Maximize $V(G, D)$ when the discriminator is optimized. The gradient descent method can be used to optimize the solution so that the sample distribution generated by the generator is consistent with the actual sample.

2.2. CNN Feature Extraction. CNN is a biologically inspired neural network, which is composed of an input layer, several convolutional layers and pooling layers alternately, a fully connected layer, and an output layer. Among them, the convolution layer is used for convolution calculation, and the pooling layer is used for descending sampling. The main features of the image are extracted through these two layers. This algorithm uses the ResNet50 model [22] to extract features, as shown in Figure 4.

Assume that the input image matrix, the convolution kernel, and the feature map matrix are all in the form of a square. Here, the size of the input image matrix is w , the convolution kernel size is k , convolution kernels are n , convolution stride is s , zero-padding layer number is p , and the size of the n feature graphs generated by the convolution calculation is given as

$$W' = \frac{w + 2p - k}{s} + 1. \quad (11)$$

For an image after several layers of convolution, pooling operation, and the activation function, a number of feature graphs are obtained. Each feature graph is in matrix form, where $f(*)$ is the activation function. The Leaky ReLU function is selected in this paper.

$$y_{ij} = f\left(b + \sum_{i=0}^I \sum_{j=1}^w W_{ij} \times X_{ij}\right). \quad (12)$$

After the convolution calculation of each pixel block in the image, the feature map matrix with the size of is obtained, as shown in the following equation:

$$\begin{bmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nm} \end{bmatrix}. \quad (13)$$

The pooling operation is carried out. During the pooling process, the size of $k * k$ is selected as the pooling kernel. The calculation formulas of the two pooling modes are

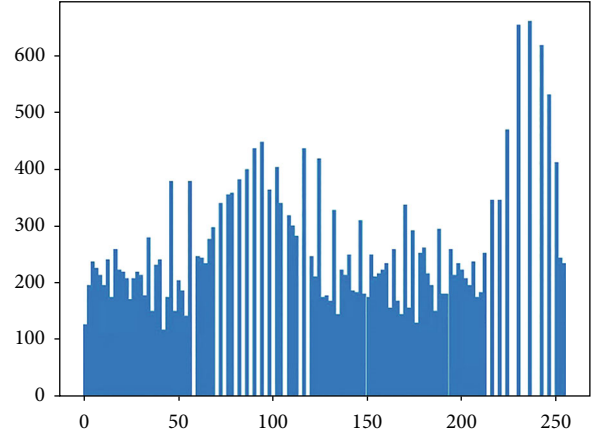


FIGURE 9: Gray histogram of Figure 7(a).

shown in Equations (14) and (15).

$$g(x) = \frac{\sum_{i=1}^{k*k} X_i}{k * k}, \quad (14)$$

$$g(x) = \text{MAX}_{1 \leq i \leq k} (X_i). \quad (15)$$

After several operations, such as layer pooling, convolution, and activation, the last layer feature graph is expanded into the row vector, which is denoted as Y^L , as the training data of ELM, where ϕ is the mapping function and V is the characteristic matrix.

$$Y^L = \phi(V). \quad (16)$$

2.3. ELM Classification. By constructing the ResNet50 model, the pretrained model weight was loaded for feature extraction and preservation. The results of the last layer of the neural network are used for output. After each picture goes through the ResNet50 model, the feature vector $L \times 1$ is extracted, where L is the feature number of each picture and x is the row vector, as shown in the following equation:

$$\vec{f}_i = [x_1, x_2, \dots, x_L], \quad (17)$$

corresponding to the one-hot code labeled 3×1 .

$$\mathbf{I}_i = [I_0, I_1, I_2]. \quad (18)$$

The $N \times L$ feature matrix is extracted from all image features to constitute the dataset for ELM [28] classification, where N represents the number of pictures and L represents the number of features.

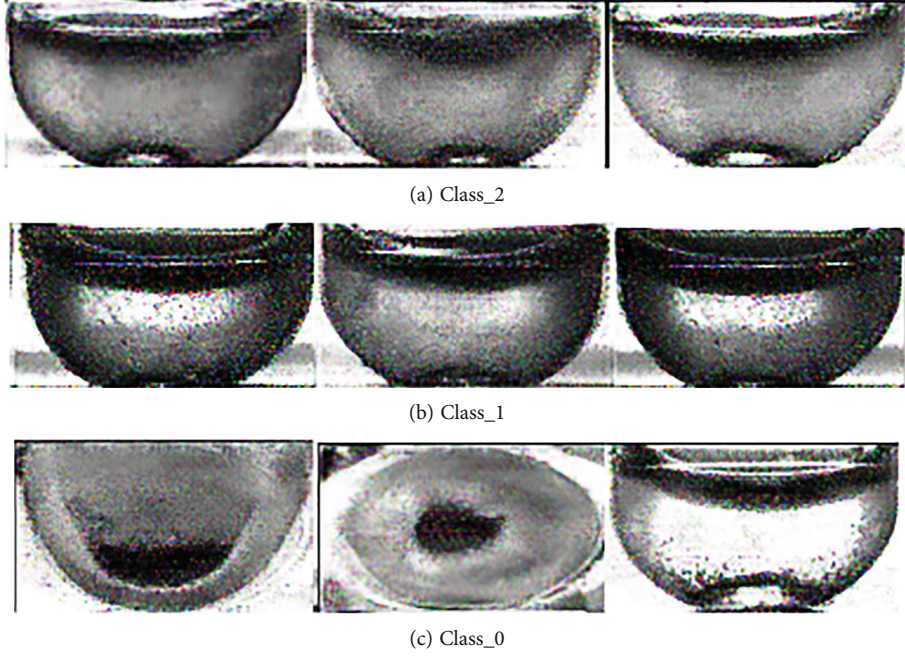


FIGURE 10: Sample instances generated by GAN.

$$\mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} X_{11} & \cdots & X_{1L} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{NL} \end{bmatrix}. \quad (19)$$

The ELM model is built as shown in Figure 5.

The implied nodes in the ELM network do not need to be adjusted through backpropagation (BP) but can be generated randomly based on a set of continuous probability distribution. The ELM contains two stages: a feature mapping stage and a learning stage. In this algorithm, the input data in the input layer is the image feature vector, and the output result of the i th node is shown in the following equation:

$$f(x_i, w_i, b_i) = x_i \cdot w_i + b_i. \quad (20)$$

That is equivalent to mapping a P -dimensional vector to an L -dimensional vector.

$$h(x) = [g(x_0, w_0, b_0), \dots, g(x_L, w_L, b_L)], \quad (21)$$

where w_i is the i th connection between the node of its input layer and the hidden layer, b_i is the bias, g is the activation function, and the sigmoid function is used here.

$$g(x_i, w_i, b_i) = \frac{1}{1 + e^{-(x_i \cdot w_i + b_i)}}. \quad (22)$$

The number of nodes in the output layer is denoted as M , and the weight of the i th hidden layer node and the j

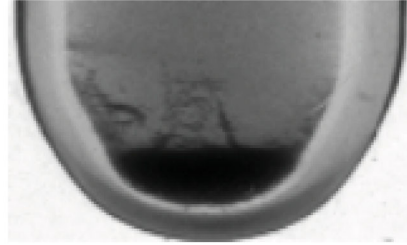


FIGURE 11: Original insoluble image.

th output layer node is β_j . Then, the output (mapping) of node j is

$$f_j(x) = \sum_{i=1}^M \beta_i \cdot g(x, w, b) = h(x)\beta. \quad (23)$$

In the learning stage, the characteristic matrix provided by the data of a given m training samples is denoted as $H(x) = [h_1(x), \dots, h_m(x)]^T$, which contains m mappings generated in the mapping phase. The objective function is represented as

$$\frac{1}{2} \|H\beta - T\|_2^2, \quad (24)$$

where β is solved by the least squares.

$$\beta = \arg \min \left(\frac{1}{2} \|H\beta - T\|_2^2 \right) = H^* T, \quad (25)$$

where H^* is the generalized inverse moment of H .

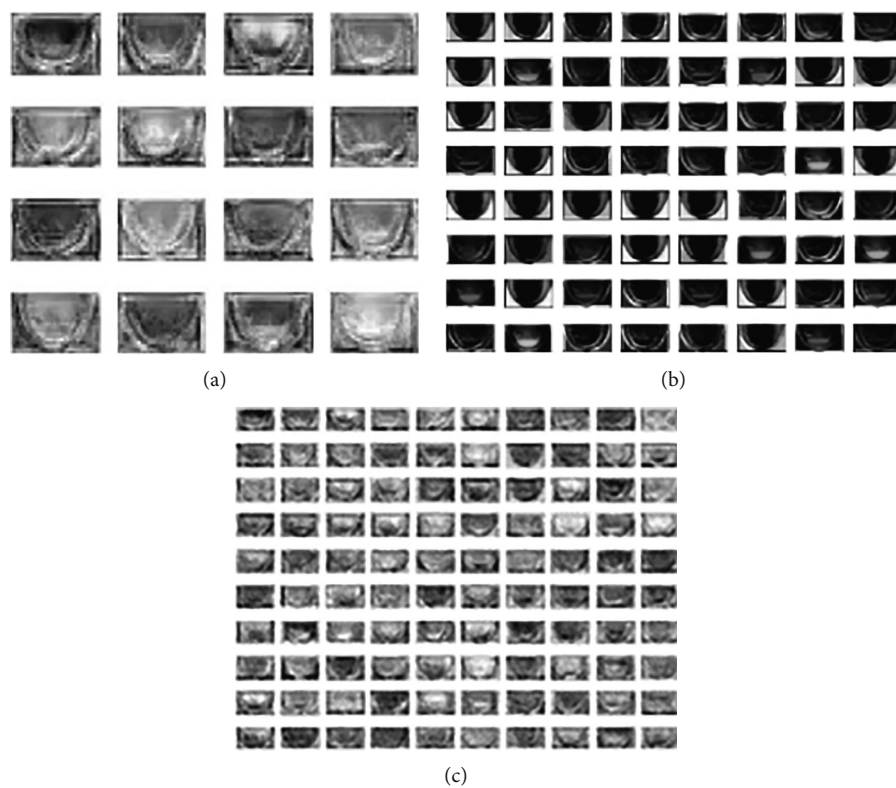


FIGURE 12: Feature plots at different scales.

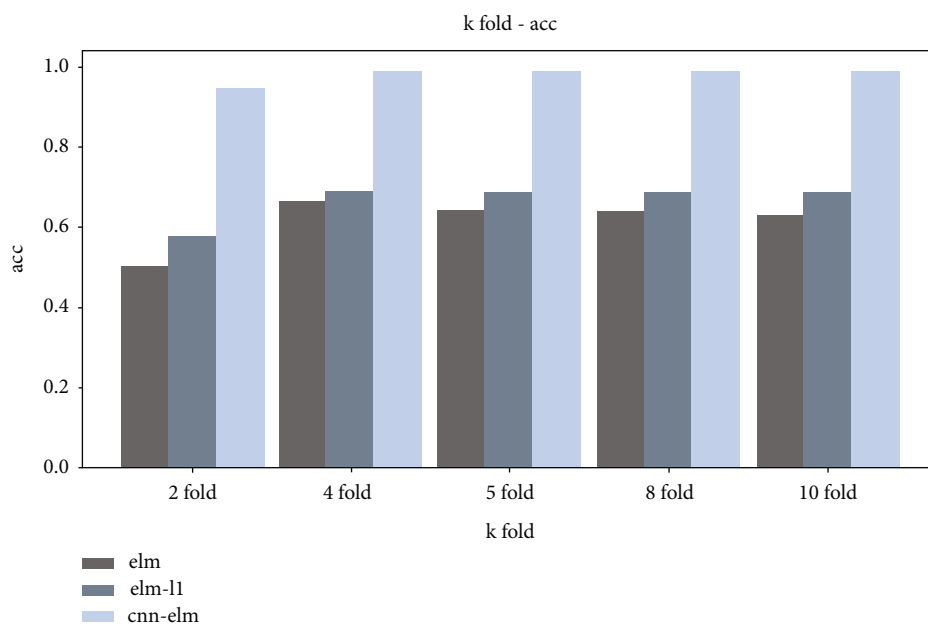


FIGURE 13: Comparison of different fold cross-validation.

To improve the generalization ability of the ELM model and avoid overfitting of the model, the L_1 regular term [29] was added to the error function.

$$\frac{1}{2} \|H\beta - T\|_2^2 + \alpha \|\beta\|. \quad (26)$$

The solution of β is as follows:

$$\beta = (H^T H + \alpha \cdot \mathbf{I})^{-1} H^T. \quad (27)$$

\mathbf{I} is the unit matrix.

In summary, the steps of ELM training can be summarized as follows:

- (1) Initialize the hidden layer parameter $(w_i, b_i) i = 1, \dots, L$ according to the random continuous distribution
- (2) The output matrix $\|H\|$ of the hidden layer is calculated according to Equations (15)–(18)
- (3) The weight matrix $\|\beta\|$ is calculated according to Equation (22)

In the recognition stage, given a set of sample X , it first becomes $H(X)$ after the mapping in the first stage; then, the category of the sample is given as

$$\text{label}(X) = \text{argmax}(H(X) \cdot \beta). \quad (28)$$

The argmax function is the index that returns the largest value in the array; that is, the maximum probability category number is the prediction category.

3. Experiment

This experiment is carried out on a computer with 8 GB of 64-bit memory in Windows10, an i5-6200u CPU, and Python 3.6.

The dataset of this experiment is from the local medical institution, and the hemolytic blood samples are collected and labeled by professionals. The collection standard is the hemolysis index $HI > 15$. The original dataset for this experiment has 2250 hemolysis images used by medical laboratories with irregular resolution. Eighty percent (1800) of the original data are used as training data, and the remaining 20% (450) are used as final test data. The data distribution is shown in Table 1.

3.1. Dataset Processing

- (1) The image is segmented and cropped, and the random cropping resolution of the image is $128 * 72$. As a result, the training scale is expanded, and the clipped area is judged and classified. Histogram equalization is used to enhance the classified image, as shown in Figures 6 and 7

By redistributing the gray value of the original image, the gray value of the image is evenly distributed, so as to increase

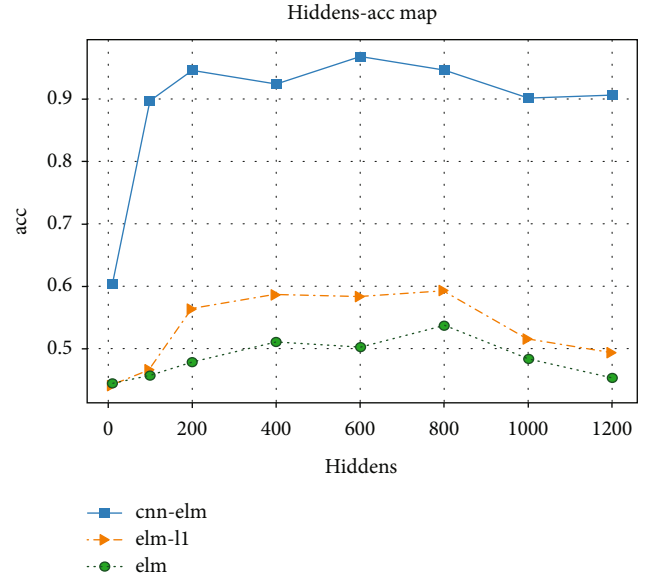


FIGURE 14: Model accuracy comparison under different hidden layers.

the contrast and enhance the details of the image. The gray histogram of Figure 6(a) is shown in Figure 8, and that of Figure 7(a) is shown in Figure 9.

- (2) The quantity of data in the sample is initially expanded through graph transformation, and generative adversarial networks are used to conduct confrontation training on the data after graph transformation, and a large number of available training samples are generated, as shown in Figure 10. The number of training images generated in this experiment is 2350, and the original dataset is extended to 4600

3.2. Extraction of Image Features. Through image enhancement and data enhancement, the original data is utilized more efficiently. The generated samples are processed, and the unavailable samples are rejected. Then, feature extraction is conducted. In the process of feature extraction, three layers of feature maps are extracted for visual observation. The original insoluble hemolysis detection image is shown in Figure 11. After convolution and pooling operations at different scales, a number of feature maps are extracted from the original image to reflect the features of the original image from different angles, as shown in Figure 12.

3.3. Comparison of the ELM Training Effect. ELM has two parameters to determine in advance during training including the input to the model and the number of nodes in the hidden layer. Different parameters affect the training efficiency, fitting degree, and generalization effect of the model. The general principle is that the number of hidden nodes is less than the input dimension, and the number of features should not be too large.

In order to determine the appropriate size of the input layer and hidden layer, K -fold cross-validation is adopted

TABLE 2: Comparison of accuracy on the test set for model cross-validation per fold and with and without cross-validation (%).

Model	Fold-1	Fold-2	Fold-3	Fold-4	With cross-validation	Without cross-validation
GAN-ELM	44.66	45.11	44.00	47.55	64.00	51.66
GAN-ELM-L1	49.55	48.44	53.55	50.44	67.55	58.22
GAN-CNN-ELM	97.66	96.99	97.22	96.33	98.91	97.66

TABLE 3: Model comparison.

Model	Features (number of input features)	Total feature extraction time (unit: s)	Duration of training (unit: s)	Total time (unit: s)	Accuracy (%)
GAN-CNN	—	—	—	5153.56	97.78
GAN-ELM	15360	1250.51	3.96	1254.47	53.78
GAN-ELM-L1	15360	1250.51	3.11	1253.62	59.33
GAN-SVM	15360	1250.51	14.83	1265.34	54.61
GAN-CNN-SVM	15360	1250.51	13.47	1263.98	98.61
CNN-ELM	15360	1085.28	3.41	1088.69	97.97
GAN-CNN-ELM	15360	1250.51	3.44	1253.95	98.77

TABLE 4: Comparison of CNN, CNN-SVM, and CNN-ELM recognition efficiency comparison.

Model	Data size (unit: amplitude)	Identification time (unit: s)	Recognition accuracy (%)
GAN-CNN	450	172.56	94.55
GAN-CNN-SVM	450	54.62	95.89
CNN-ELM	450	39.66	96.33
GAN-CNN-ELM	450	38.28	98.91

in the analysis experiment to reduce overfitting and improve the generalization ability of the model, as shown in Figure 13. Experimental results show that 4-fold cross-validation has the highest accuracy, and 4-fold cross-validation is finally used to train GAN-ELM, GAN-ELM-L1, and GAN-CNN-ELM models. The initialization of hidden nodes is randomly generated based on a continuous probability distribution.

The training parameters are the number of selected features, and the influence of the hidden layer on accuracy is studied. The number of hidden nodes is selected as follows: 10, 100, 200, 400, 800, 1000, and 2000. Then, three models of GAN-ELM, GAN-ELM-L1, and GAN-CNN-ELM are trained separately under 4-fold cross-validation. The training results are shown in Figure 14.

As seen in the experimental results, when the number of hidden layer nodes L increased from 10 to 200, the accuracy is greatly affected and increased rapidly. From the stage of 400 to 800, the accuracy tended to be stable. After the

800th node, the model started to overfit, and the generalization ability is low. Therefore, the number of optimal hidden nodes is approximately 600, which has good performance in this experiment. The experimental results also show that the single ELM classifier is not effective. The accuracy is improved by adding a regularized ELM classifier, and the accuracy is greatly improved by the GAN-CNN-ELM classifier model.

After determining the optimal number of hidden nodes as 600, this paper adopts 4-fold cross-validation to verify the accuracy of each fold, as well as the accuracy of test set with and without cross-validation of the model. The experimental results are shown in Table 2. The experimental results indicate that the accuracy performance of the model in the final test set is improved after using 4-fold cross-validation.

3.4. Comparisons of Experimental Results

- (1) The traditional methods of hemolysis detection mainly include the Vitros 5600 biochemical immune analyzer to detect hemolysis indexes [5], the neuron-specific enolase (NSE) correction formula [6] and microcolumn gel detection card [30]. Because these experimental methods do not rely on human subjective judgment, it avoids the influence of human subjective factors and improves the accuracy of detection. At the same time, it does not rely on chemical instruments, saves chemical reagents, and is simple to operate. Computers operate much faster than physical and chemical analysis instruments. Therefore, it avoids the drawback of time-consuming measurement of relevant data and improves the detection efficiency to a large extent
- (2) This experiment compared seven different algorithms: GAN-CNN, GAN-ELM, GAN-ELM-L1,

GAN-SVM, GAN-CNN-SVM, CNN-ELM, and GAN-CNN-ELM. The results show that the training model of the GAN-CNN-ELM classifier is adopted, since GAN greatly increased the sample data; compared with the CNN-ELM algorithm, the feature extraction time increased, but the accuracy improved. As the full connection layer in CNN network structure requires backpropagation to adjust parameters, ELM is a single hidden layer feedforward neural network, and the weight of the input layer and bias of hidden layer are randomly selected. ELM is used to replace the full connection layer for feature classification, which effectively reduces the computational complexity of the algorithm and improves the efficiency of model detection. The GAN-CNN-ELM model was compared with other models; the training time and accuracy are optimized. In Table 3, the accuracy is increased by at least 0.16%

Four models with relatively high accuracy are selected: GAN-CNN, GAN-CNN-SVM, CNN-ELM, and GAN-CNN-ELM. The remaining 20% of test data are into different models that are trained for prediction. The results are shown in Table 4.

Through the above experimental comparison, the results show that the GAN-CNN-ELM algorithm model is significantly better than the other three models in terms of recognition time and recognition accuracy. The recognition accuracy is increased by at least 2.58%.

4. Conclusion

This paper proposes a GAN-CNN-ELM algorithm model for hemolysis image detection, which uses image enhancement and graphic transformation to enhance image features and expand sample data. Then, we use the GAN to train the existing data to obtain a large quantity of data. The ResNet50 algorithm model is used to extract multilayer features of the pretreated hemolytic experimental atlas. The ELM classifier is constructed to classify the feature values to improve training time efficiency. For the test set, the accuracy of identification is as high as 98.91%. To achieve efficient identification of hemolytic images and accurately determine the hemolytic results.

In the future, multiscale convolution can be used to optimize the generalization degree of the system and the diversity of the recognition. An attempt will be made to improve the accuracy with CNN multiple feature fusion (MFF). The Res2Net framework will be used to extract multilayer features of the image to make the granularity finer and the feature scale more diverse, and the KELM algorithm will be studied to further improve the model training speed.

Data Availability

The image data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work and its data collection and sharing are supported and funded by the research on local stereo matching algorithm based on a vision system (18jk0509).

References

- [1] J. Yang, R. Nuya, and M. Ayegamma, "Correlation between blood type, serum test and ABO neonatal hemolysis in children with hyperbilirubinemia," *Guizhou Medical Journal*, vol. 43, no. 4, pp. 628-629, 2019.
- [2] S. Lin, Q. Wu, and X. Peng, "Analysis of the detection of hemolytic disease of newborn in Xinyu of Jiangxi Province in 2016," *Shanghai Medical & Pharmaceutical Journal*, vol. 40, no. 4, pp. 58-60, 2019.
- [3] Y. Zhang, "Diagnostic value of neonatal hemolytic disease serologic test for neonatal jaundice," *Clinical Laboratory Journal (Electronic Edition)*, vol. 8, no. 1, pp. 63-65, 2019.
- [4] Y. Jia, "Influence of the timing of serological detection of neonatal jaundice on the diagnosis of neonatal hemolytic disease," *Capital Food and Medicine*, vol. 25, no. 17, 2018.
- [5] Q. Yang, L. Zhang, L. Tian, R. Han, and H. Shou, "Clinical application of biochemical immune analyzer in detecting hemolysis index," *Chinese Journal of Clinical Pathologist*, vol. 10, no. 3, pp. 141-142, 2018.
- [6] Q. Wang, X. Lu, G. Wang et al., "Clinical application of biochemical immune analyzer in detecting hemolysis index," *Chinese Journal of Clinical Pathologist*, vol. 10, no. 3, pp. 141-142, 2018.
- [7] H.-a. Li, Q. Zheng, X. Qi et al., "Neural network-based mapping mining of image style transfer in big data systems," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387382, 11 pages, 2021.
- [8] Z. Yu, R. Arif, M. A. Fahmy, and A. Sohail, "Self organizing maps for the parametric analysis of COVID-19 SEIRS delayed model," *Chaos Solitons & Fractals*, vol. 150, article 111202, 2021.
- [9] J. Gao, Q. Jiang, B. Zhou, D. Chen, 1 College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China, and 2 Shanghai University of Medicine & Health Science, Shanghai 201308, China, "Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: an overview," *Mathematical Biosciences and Engineering*, vol. 16, no. 6, pp. 6536-6561, 2019.
- [10] Y. Qi, J. Gu, W. Li, Y. Zhang, and J. Geng, "Image enhancement method for pulmonary nodules based on generating antagonizing network," in *Proceedings of the 23rd Annual Meeting of Network Application Branch of China Computer Users Association*, Hefei, Anhui, China, 2019.
- [11] T. Huynh-The, C. Hua, T. Ngo, and D. Kim, "Image representation of pose-transition feature for 3D skeleton-based action recognition," *Information Sciences*, vol. 513, pp. 112-126, 2020.
- [12] G. Bueno, M. Fernandez-Carrobles, L. Gonzalez-Lopez, and O. Deniz, "Glomerulosclerosis identification in whole slide

- images using semantic segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 184, article 105273, 2020.
- [13] R. Xu, X. Liang, J. Qi, Z. Li, and S. Zhang, “Advances and trends of extreme learning machines,” *Chinese Journal of Computers*, vol. 42, no. 7, pp. 1640–1670, 2019.
- [14] J. Wang and Z. Li, “Wavelet core limit learning machine method for pulse recognition,” *Computer Engineering and Design*, vol. 40, no. 5, pp. 1472–1476, 2019.
- [15] X. Wang, D. Li, X. Zheng et al., “Detection method of human falls based on rapid classification of extreme learning machine,” *Science, technology and engineering*, vol. 19, no. 20, pp. 258–264, 2019.
- [16] L. Cai, G. Cheng, and H. Pan, “Application of limit learning machine in lithology identification,” *Computer Engineering and Design*, vol. 31, no. 9, pp. 128–130, 2010.
- [17] H. Li, Q. Zheng, J. Zhang, Z. Du, Z. Li, and B. Kang, “Pix2Pix-Based grayscale image coloring method,” *Journal of Computer-Aided Design & Computer Graphics*, vol. 33, no. 6, pp. 929–938, 2021.
- [18] H. Li, M. Zhang, K. Yu, X. Qi, and J. Tong, “A displacement estimated method for real time tissue ultrasound elastography,” *Mobile Networks and Applications*, vol. 26, no. 5, pp. 2014–2023, 2021.
- [19] Z. Liu and D. Qi, “Research on tree species identification based on convolutional neural network,” *Forest engineering*, vol. 36, no. 1, pp. 33–38, 2020.
- [20] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Boston, MA, 2015.
- [21] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [22] 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, 2016.
- [23] M. Mardani, E. Gong, J. Cheng et al., “Deep generative adversarial neural networks for compressive sensing MRI,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, 2019.
- [24] J. Wolterink, T. Leiner, M. Viergever, and I. Isgum, “Generative adversarial networks for noise reduction in low-dose CT,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.
- [25] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” 2015, <https://arxiv.org/abs/1511.06434>.
- [26] Y. Wu and Y. Tian, “Coal rock identification method based on maximum pooling sparse coding,” *Journal of Engineering Science*, vol. 39, no. 7, pp. 981–987, 2017.
- [27] Y. Liu, C. Dou, Q. Zhao, Z. Li, and H. Li, “Sketch based image retrieval with conditional generative adversarial network,” *Journal of Computer-Aided Design and Graphics*, vol. 29, no. 12, pp. 2336–2342, 2017.
- [28] G. Huang, Q. Zhu, and C. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.
- [29] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014, <https://arxiv.org/abs/1411.1784>.
- [30] L. Wei, “Analysis of 260 neonates with hyperbilirubinemia hemolysis detection status,” *Smart and Healthy*, vol. 6, no. 9, pp. 178–179, 2020.

Research Article

Development and Validation of Prediction Model for High Ovarian Response in *In Vitro* Fertilization-Embryo Transfer: A Longitudinal Study

Xinsha Tan ¹, Honglin Xi,¹ Jing Yang ², and Wenfeng Wang ^{3,4,5}

¹Department of Reproductive Medical Center, Affiliated Renhe Hospital of China Three Gorges University, Yichang 443002, China

²Department of Reproductive Medical Center, Renmin Hospital of Wuhan University, Wuhan 430060, China

³School of Science, Shanghai Institute of Technology, Shanghai 201418, China

⁴International Academy of Visual Art and Engineering, London E16 1AH, UK

⁵Interscience Institute of Management and Technology, Bhubaneswar 752054, India

Correspondence should be addressed to Jing Yang; dryangjing@whu.edu.cn

Received 22 July 2021; Revised 26 August 2021; Accepted 20 September 2021; Published 16 October 2021

Academic Editor: Siti Anom Ahmad

Copyright © 2021 Xinsha Tan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To develop and validate a prediction model for high ovarian response in *in vitro* fertilization-embryo transfer (IVF-ET) cycles. **Methods.** Totally, 480 eligible outpatients with infertility who underwent IVF-ET were selected and randomly divided into the training set for developing the prediction model and the testing set for validating the model. Univariate and multivariate logistic regressions were carried out to explore the predictive factors of high ovarian response, and then, the prediction model was constructed. Nomogram was plotted for visualizing the model. Area under the receiver-operating characteristic (ROC) curve, Hosmer-Lemeshow test and calibration curve were used to evaluate the performance of the prediction model. **Results.** Antral follicle count (AFC), anti-Müllerian hormone (AMH) at menstrual cycle day 3 (MC3), and progesterone (P) level on human chorionic gonadotropin (HCG) day were identified as the independent predictors of high ovarian response. The value of area under the curve (AUC) for our multivariate model reached 0.958 (95% CI: 0.936-0.981) with the sensitivity of 0.916 (95% CI: 0.863-0.953) and the specificity of 0.911 (95% CI: 0.858-0.949), suggesting the good discrimination of the prediction model. The Hosmer-Lemeshow test and the calibration curve both suggested model's good calibration. **Conclusion.** The developed prediction model had good discrimination and accuracy via internal validation, which could help clinicians efficiently identify patients with high ovarian response, thereby improving the pregnancy rates and clinical outcomes in IVF-ET cycles. However, the conclusion needs to be confirmed by more related studies.

1. Introduction

With the rapid development of assisted reproductive technology (ART), *in vitro* fertilization-embryo transfer (IVF-ET) has become an important treatment for infertility [1]. Controlled ovarian hyperstimulation (COH) is a key step of IVF-ET, where gonadotropin (Gn) stimulates the development of multiple follicles and produces multiple mature oocytes, thereby improving pregnancy rates [2, 3]. However, it cannot be ignored that ovaries' overreaction to Gn could increase the risk of iatrogenic complication-ovarian hyperstimulation syndrome (OHSS) [4], which is characterized

by an increase in ovarian volume and brings more severe and even fatal infertility. Therefore, it is still necessary to identify the risk of OHSS in IVF-ET for patients.

High ovarian response, defined as excessive ovarian response, is reported as an adverse effect of IVF-ET [5]. It is mainly due to changes in the systemic stress state led by the recruitment and development of multiple follicles and abnormally high steroid substances [5]. Hormone level is higher in patients with high ovarian response, which is not conducive to endometrial receptivity and embryo implantation, thereby increasing the incidence of ovarian hyperstimulation syndrome (OHSS) [6]. In view of this, it is of great

clinical significance to find predictors for high ovarian response, which may decrease the risk of OHSS, improve pregnancy rates, and optimize pregnancy outcomes.

Previous studies verified that anti-Müllerian hormone (AMH) and antral follicle count (AFC) were the effective predictors of high ovarian response [7–12]. In the study of Oehninger et al., female age and follicle-stimulating hormone (FSH) were also found to be the predictors of high ovarian response [13]. To our knowledge, most studies focused on the predictive factors of poor ovarian response, and few studies have been done on predictive factors of high ovarian response. Not only that, a complete predictive system on high ovarian response has not been established yet, and there are only some researches of individual indicators, which was not accurate for predicting high ovarian response. Hence, this study was to explore the independent predictors of high ovarian response in patients undergoing IVF-ET and set out to develop a model for predicting the risk of high ovarian response and perform internal validation. We believed that the results will provide better guidance to the clinical use of a reasonable COH protocol, thereby improving pregnancy rates and clinical outcomes in IVF-ET cycles.

2. Methods

2.1. Collection of the Data. A longitudinal study of 1,142 outpatients with infertility who underwent IVF-ET at the Affiliated Renhe Hospital of China Three Gorges University from January 2018 to December 2019 was consecutively selected. The patients were divided into high ovarian response group (>15 oocytes retrieved) and normal ovarian response group (4–15 oocytes retrieved) [9, 14]. The inclusion criteria were as follows: (1) patients aged 20–40 years with normal menstrual cycles (21–35 days); (2) patients undergoing IVF-ET; (3) patients with complete clinical data. Patients were excluded if met any of the following criteria: (1) patients who were diagnosed with polycystic ovary syndrome (PCOS) according to Rotterdam criteria; (2) patients who had primary ovarian insufficiency; (3) patients who had ovarian-related surgery before (such as laparoscopic ovarian drilling, ovarian dissection for endometriosis, unilateral oophorectomy); (4) patients who had hormonal contraceptives before study cycle; (5) patients who had taken other investigational drugs or was participating in other clinical studies within 1 month before study enrolment; (6) physician considered patients who were inappropriate to participate in the clinical investigator.

The study was approved by the Ethics Committee of Affiliated Renhe Hospital of China Three Gorges University with the number of 2020K05. Signed informed consent was obtained from all patients before enrollment.

Gonadotropin-releasing hormone (GnRH) agonist was subcutaneously injected on the 21st day of menstruation to downregulate the function of pituitary gland. After 14 days, COH was initiated with the appropriate amount of Gn (150–300 U/d) according to age, body mass index (BMI), and baseline FSH. The development of follicles was monitored by ultrasound. When 1–2 follicles were ≥ 18 mm in diameter, or 2–3 were ≥ 17 mm in diameter, 10,000 U of human chori-

onic gonadotropin (HCG) was injected. Oocytes were retrieved 36 h after HCG injection, and embryo transfer was performed 3–5 days after retrieval, after which 80 mg of progesterone (P) was intramuscularly injected daily as luteal support. A positive serum HCG pregnancy test after 14 days was defined as biochemical pregnancy, and ultrasound confirmation of a gestational sac or heartbeat (fetal pole) 35 days after transfer was diagnosed as clinical pregnancy.

2.2. Analytic Methods. Baseline characteristics of patients were collected in our study, of which categorical variables included smoking history, type of infertility, and pregnancy history; continuous variables contained age, BMI, age at menarche, mean menstrual cycle, and duration of infertility; AFC, endometrial thickness, luteinizing hormone (LH) level, estradiol (E2) level, P level, FSH level, and AMH level on menstrual cycle day 3 (MC3); dosing days, initial dose, and total dose of Gn; endometrial thickness and hormone levels on the day of HCG injection. AFC was assessed by transvaginal sonography at MC3; endometrial thickness was observed on the day of injection of HCG; LH was defined as a hormone secreted by basophils in the anterior pituitary gland; E2 was a steroidal estrogen with the normal value of follicular stage (94–433 pmol/L), the normal value of luteal phase (499–1580 pmol/L), and the normal value during ovulation (704–1580 pmol/L); P was defined as main progesterone with biological activity secreted by the ovary; FSH was a hormone secreted by basophils in the anterior pituitary gland that promoted follicle maturation; AMH was defined as a hormone secreted by follicles in the predeveloping chambers or small chambers of the ovary. In addition, dosing days, initial dose, and total dose of Gn, endometrial thickness, and hormone level on the day of HCG injection were collected for analyzing the predictive factors of high ovarian response.

The measurement data was tested by the Kolmogorov-Smirnov test for normality. Normally distributed continuous variables were expressed as mean \pm standard deviation (Mean \pm SD), and Student's *t* test was used for comparison between groups; continuous variables with skewed distribution were expressed as median and quartile [$M(Q_1, Q_3)$], and the Mann-Whitney *U* test was applied for comparison. Categorical variables were expressed as number of cases and constituent ratio [n (%)], and the Chi-square test and Fisher's exact test were used for comparison. The two-sided test was performed for all statistical analyses, and $P < 0.05$ was considered statistically significant. All statistical analysis was performed by using SAS 9.4 and R 4.0.2 (model validation and drawing) statistical software. Prediction models were constructed by adopting SAS 9.4 (Logistic model) and Python 3.7 [Broad Learning System (BLS) model] software.

In the present study, the population was randomly divided into training set for developing the prediction logistic model and BLS model, and testing set for validating the models at the ratio of 7 : 3. Univariate and multivariate logistic regressions were carried out to explore the predictive factors of high ovarian response, and then, the prediction

TABLE 1: Baseline characteristics of study population.

Variables, n (%)	Total ($n = 480$)	Group		Statistic	P
		Training set ($n = 336$)	Testing set ($n = 144$)		
Age, years, mean \pm SD	31.36 \pm 3.79	31.26 \pm 3.79	31.59 \pm 3.77	$t = -0.86$	0.389
BMI, kg/m ² , mean \pm SD	22.08 \pm 2.97	21.99 \pm 2.97	22.29 \pm 2.97	$t = -1.00$	0.317
Smoking history, n (%)				$Z = -0.933$	0.351
Active smoking	6 (1.25)	5 (1.49)	1 (0.69)		
Passive smoking	75 (15.63)	48 (14.29)	27 (18.75)		
No	399 (83.13)	283 (84.23)	116 (80.56)		
Type of infertility, n (%)				$\chi^2 = 2.887$	0.089
Primary	174 (36.25)	130 (38.69)	44 (30.56)		
Secondary	306 (63.75)	206 (61.31)	100 (69.44)		
Pregnancy history, n (%)				$\chi^2 = 2.796$	0.094
No	177 (36.88)	132 (39.29)	45 (31.25)		
Yes	303 (63.13)	204 (60.71)	99 (68.75)		
Age at menarche, years, mean \pm SD	13.16 \pm 1.22	13.14 \pm 1.20	13.22 \pm 1.26	$t = -0.60$	0.551
Mean menstrual cycle, days, mean \pm SD	29.23 \pm 2.04	29.28 \pm 2.04	29.10 \pm 2.06	$t = 0.87$	0.385
Duration of infertility, years, $M(Q_1, Q_3)$	3.00 (1.00, 4.00)	3.00 (2.00, 4.00)	3.00 (1.00, 4.50)	$Z = -0.544$	0.586
AFC, $M(Q_1, Q_3)$	15.00 (12.00, 20.00)	15.00 (12.00, 19.00)	15.00 (12.00, 20.00)	$Z = 0.293$	0.770
LH, $\mu\text{u/mL}$, $M(Q_1, Q_3)$	4.79 (3.61, 6.24)	4.79 (3.54, 6.02)	4.75 (3.72, 6.34)	$Z = 1.248$	0.212
MC3					
E2, pmol/L, $M(Q_1, Q_3)$	36.18 (25.95, 46.42)	36.16 (26.26, 46.38)	36.60 (25.57, 46.49)	$Z = -0.521$	0.602
P, nmol/L, $M(Q_1, Q_3)$	0.39 (0.26, 0.59)	0.40 (0.26, 0.59)	0.35 (0.25, 0.56)	$Z = -1.223$	0.221
FSH, U/L, mean \pm SD	6.50 \pm 1.60	6.56 \pm 1.69	6.36 \pm 1.35	$t = 1.33$	0.186
AMH, ng/ml, $M(Q_1, Q_3)$	3.71 (2.46, 5.27)	3.68 (2.46, 5.25)	3.75 (2.46, 5.33)	$Z = -0.108$	0.914
Endometrial thickness, mm, mean \pm SD	10.05 \pm 2.09	10.09 \pm 2.10	9.96 \pm 2.08	$t = 0.63$	0.526
Group				$\chi^2 = 0.004$	0.952
Normal response	241 (50.21)	169 (50.30)	72 (50.00)		
High response	239 (49.79)	167 (49.70)	72 (50.00)		

BMI: body mass index; AFC: antral follicle count; LH: luteinizing hormone; MC3: menstrual cycle day 3; E2: estradiol; P: progesterone; FSH: follicle-stimulating hormone; AMH: anti-Müllerian hormone; Mean \pm SD: mean \pm standard deviation; $M(Q_1, Q_3)$: median and quartile; χ^2 : measured the degree of correlation between two categorical variables; t : compared the difference between the two groups; Z : represented the rank sum test for the comparison of two samples.

model was constructed. Nomogram was plotted for visualizing the model. Area under the receiver-operating characteristic (ROC) curve, the Hosmer-Lemeshow test, and calibration curve were used to evaluate the performance of the prediction model. The Youden index was used to calculate the cutoff point, which was identified as the cutoff point at a high risk for high ovarian response. Then, the comparison of predictive power was carried out between the logistic model and BLS model for high ovarian response.

3. Results

3.1. Patient Characteristics. Among the total 1,142 outpatients, patients who were over 40 years old ($n = 63$) were pretreated with hormonal contraceptives before study cycle ($n = 54$), had ovarian-related surgery before ($n = 47$), had primary ovarian insufficiency ($n = 28$), were diagnosed with

PCOS according to Rotterdam criteria ($n = 196$), had low ovarian response ($n = 264$), or had missing information of age and BMI ($n = 10$) were excluded. 480 eligible outpatients were enrolled with 336 patients in the training set and 144 in the testing set eventually. Then, the patients were divided into the high response group (HR group, $n = 239$) and the normal response group (NR group, $n = 241$). The mean age was 31.36 ± 3.79 years, and the mean age at menarche was 13.16 ± 1.22 years. Only 6 (1.25%) patients reported the history of active smoking, 75 (15.63%) had the history of passive smoking, and the remaining 399 (83.13%) had no smoking history. The mean menstrual cycle was 29.23 ± 2.04 days, and the median duration of infertility was 3.00 (1.00, 4.00) years. Table 1 gives an overview of baseline characteristics of all patients. No significant differences were observed between the training set and the testing set in all variables (all $P > 0.05$).

TABLE 2: Difference analysis for the potential predictors of ovarian response in IVF-ET.

Variables, n (%)	Total ($n = 336$)	HR group ($n = 167$)	NR group ($n = 169$)	Statistic	P
Age, years, mean \pm SD	31.26 \pm 3.79	30.73 \pm 3.67	31.79 \pm 3.85	$t = -2.588$	0.010
Smoking history				—	0.686
Active smoking	5 (1.49)	3 (1.80)	2 (1.18)		
Passive smoking	48 (14.29)	26 (15.57)	22 (13.02)		
No	283 (84.23)	138 (82.63)	145 (85.80)		
Type of infertility				$\chi^2 = 3.530$	0.060
Primary	130 (38.69)	73 (43.71)	57 (33.73)		
Secondary	206 (61.31)	94 (56.29)	112 (66.27)		
Pregnancy history				$\chi^2 = 1.452$	0.228
No	132 (39.29)	71 (42.51)	61 (36.09)		
Yes	204 (60.71)	96 (57.49)	108 (63.91)		
BMI, kg/m ² , mean \pm SD	21.99 \pm 2.97	22.05 \pm 3.06	21.94 \pm 2.88	$t = 0.340$	0.734
Age at menarche, years, mean \pm SD	13.14 \pm 1.20	13.06 \pm 1.11	13.22 \pm 1.29	$t = -1.259$	0.209
Mean menstrual cycle, days, mean \pm SD	29.28 \pm 2.04	29.62 \pm 2.15	28.94 \pm 1.87	$t = 3.091$	0.002
Duration of infertility, years, M (Q_1, Q_3)	3.00 (2.00, 4.00)	3.00 (2.00, 5.00)	3.00 (1.00, 4.00)	$Z = 1.430$	0.153
AFC, M (Q_1, Q_3)	15.00 (12.00, 19.00)	19.00 (16.00, 22.00)	12.00 (10.00, 14.00)	$Z = 13.879$	<0.001
LH, μ u/ml, M (Q_1, Q_3)	4.79 (3.54, 6.02)	4.87 (3.61, 6.01)	4.71 (3.32, 6.04)	$Z = 0.274$	0.784
E2, pmol/L, M (Q_1, Q_3)	36.16 (26.26, 46.38)	35.23 (23.22, 45.70)	37.56 (27.80, 47.44)	$Z = -1.903$	0.057
P, nmol/L, M (Q_1, Q_3)	0.40 (0.26, 0.59)	0.46 (0.31, 0.66)	0.32 (0.22, 0.48)	$Z = 4.626$	<0.001
MC3 FSH, U/L, mean \pm SD	6.56 \pm 1.69	6.34 \pm 1.65	6.77 \pm 1.71	$t = -2.304$	0.022
AMH, ng/ml, M (Q_1, Q_3)	3.68 (2.46, 5.25)	5.20 (4.19, 6.17)	2.69 (2.01, 3.34)	$t = 14.188$	<0.001
Endometrial thickness, mm, mean \pm SD	10.09 \pm 2.10	10.18 \pm 2.07	10.00 \pm 2.13	$t = 0.798$	0.425
Dosing days of Gn, days, mean \pm SD	12.13 \pm 2.63	12.26 \pm 2.55	12.01 \pm 2.71	$t = 0.86$	0.393
Initial dose of Gn, IU, mean \pm SD	217.82 \pm 54.94	208.80 \pm 50.98	226.70 \pm 57.37	$t = -3.02$	0.003
Total dose of Gn, IU, mean \pm SD	2861.99 \pm 919.51	2753.70 \pm 821.50	2969.00 \pm 998.00	$t = -2.16$	0.032
HCG day Endometrial thickness, mm, mean \pm SD	11.10 \pm 2.32	11.31 \pm 2.34	10.89 \pm 2.28	$t = 1.67$	0.095
LH, μ u/mL, M (Q_1, Q_3)	0.95 (0.55, 1.87)	0.84 (0.53, 1.42)	1.15 (0.55, 2.14)	$Z = -2.161$	0.031
E2, pmol/L, M (Q_1, Q_3)	2986.00 (2074.86, 4857.50)	4403.00 (2811.00, 6000.00)	2405.00 (1657.00, 3483.00)	$Z = 8.737$	<0.001
P, nmol/L, M (Q_1, Q_3)	0.95 (0.67, 1.32)	1.16 (0.82, 1.56)	0.81 (0.57, 1.12)	$Z = 6.089$	<0.001

HR group: high response group; NR group: normal response group; BMI: body mass index; AFC: antral follicle count; MC3: menstrual cycle day 3; LH: luteinizing hormone; E2: estradiol; P: progesterone; FSH: follicle-stimulating hormone; AMH: anti-Müllerian hormone; HCG: human chorionic gonadotropin; Gn: gonadotropin; Mean \pm SD: mean \pm standard deviation; M (Q_1, Q_3): median and quartile; χ^2 : measured the degree of correlation between two categorical variables; t : compared the difference between the two groups; Z : represented the rank sum test for the comparison of two samples.

3.2. Risk Factor Selection. The results of Table 2 suggested that age, mean menstrual cycle, AFC, P at MC3, FSH at MC3, AMH at MC3, initial dose of Gn, total dose of Gn, LH on HCG day, E2 on HCG day, and P level on HCG day in the high response group were significantly different from those in the normal response group (all $P < 0.05$), which could be considered as potential predictors and included in the multivariate analysis (Table 3).

3.3. Development and Visualization of the Prediction Model. To identify predictors for high ovarian response in IVF-ET, the multivariate logistic regression was performed. The results indicated that AFC, AMH at MC3, and P level on HCG day were independently associated with high ovarian response. For each additional AFC, the risk of high ovarian response increased by 0.671-fold (95% CI: 1.453-1.921, $P < 0.001$). For every 1 ng/mL increase in AMH at MC3, the risk of high response increased by 0.874-fold (95% CI: 1.404-

TABLE 3: Univariate analysis for the potential predictors of high ovarian response in IVF-ET.

Variables, <i>n</i> (%)	β	S.E	Wald	<i>P</i>	OR	95% CI	
						Lower	Upper
Age	-0.075	0.030	6.489	0.011	0.927	0.875	0.983
BMI	0.013	0.037	0.116	0.734	1.013	0.942	1.088
Smoking history							
No					Ref		
Active smoking	0.455	0.921	0.244	0.621	1.576	0.259	9.576
Passive smoking	0.217	0.313	0.478	0.489	1.242	0.672	2.294
Type of infertility							
Primary					Ref		
Secondary	-0.423	0.225	3.515	0.061	0.655	0.421	1.019
Pregnancy history							
No					Ref		
Yes	-0.270	0.224	1.449	0.229	0.764	0.492	1.185
Age at menarche	-0.115	0.092	1.572	0.210	0.891	0.745	1.067
Mean menstrual cycle	0.172	0.057	8.918	0.003	1.187	1.061	1.329
Duration of infertility	0.056	0.047	1.390	0.238	1.057	0.964	1.159
AFC	0.641	0.070	84.345	<0.001	1.898	1.655	2.176
LH at MC3	0.010	0.025	0.167	0.683	1.010	0.962	1.061
E2 at MC3	-0.002	0.004	0.367	0.545	0.998	0.991	1.005
P at MC3	0.143	0.121	1.389	0.238	1.154	0.909	1.464
FSH at MC3	-0.151	0.067	5.139	0.023	0.859	0.754	0.980
AMH at MC3	1.207	0.128	89.114	<0.001	3.345	2.603	4.298
Endometrial thickness at MC3	0.042	0.052	0.637	0.425	1.043	0.941	1.155
Dosing days of Gn	0.036	0.042	0.729	0.393	1.036	0.955	1.125
Initial dose of Gn	-0.006	0.002	8.675	0.003	0.994	0.990	0.998
Total dose of Gn	-0.000	0.000	4.513	0.034	1.000	0.999	1.000
Endometrial thickness on HCG day	0.079	0.048	2.774	0.096	1.083	0.986	1.189
LH on HCG day	-0.199	0.077	6.743	0.009	0.820	0.705	0.952
E2 on HCG day	0.001	0.000	56.290	<0.001	1.001	1.000	1.001
P on HCG day	0.679	0.197	11.898	<0.001	1.971	1.340	2.898
FSH on HCG day	0.018	0.019	0.911	0.340	1.018	0.981	1.057

BMI: body mass index; AFC: antral follicle count; MC3: menstrual cycle day 3; LH: luteinizing hormone; E2: estradiol; P: progesterone; FSH: follicle-stimulating hormone; AMH: anti-Müllerian hormone; HCG: human chorionic gonadotropin; Gn: gonadotropin; S.E: standard error; OR: odds ratio; CI: confidence interval.

TABLE 4: Multivariate analysis for the predictors of high ovarian response in IVF-ET.

Variables, <i>n</i> (%)	β	S.E	Wald	<i>P</i>	OR	95% CI	
						Lower limit	Upper limit
Constant	-11.094	1.219	82.822	<0.001			
AFC	0.513	0.071	51.996	<0.001	1.671	1.453	1.921
AMH at MC3	0.628	0.148	18.151	<0.001	1.874	1.404	2.502
P on HCG day	0.668	0.273	5.974	0.015	1.950	1.141	3.331

AFC: antral follicle count; AMH: anti-Müllerian hormone; MC3: menstrual cycle day 3; P: progesterone; HCG: human chorionic gonadotropin. S.E: standard error; OR: odds ratio; CI: confidence interval.

2.502, $P < 0.001$). Besides, the risk rose by 0.950-fold for every 1 nmol/L increase in P on the day of HCG injection (95% CI: 1.141-3.331, $P = 0.015$) (Table 4). Then, the predicted risk of high ovarian response was calculated as follows: $\ln(P_{HR}/(1 - P_{HR})) = -11.094 + AFC * 0.513 + AMH$

at MC3 * 0.628 + P on HCG day * 0.668 (P_{HR} represented the probability of high ovarian response).

To visualize our model, the nomogram was plotted (Figure 1). For example, we randomly chose a patient whose P at HCG day was 0.53 nmol/L, AMH at MC3 was

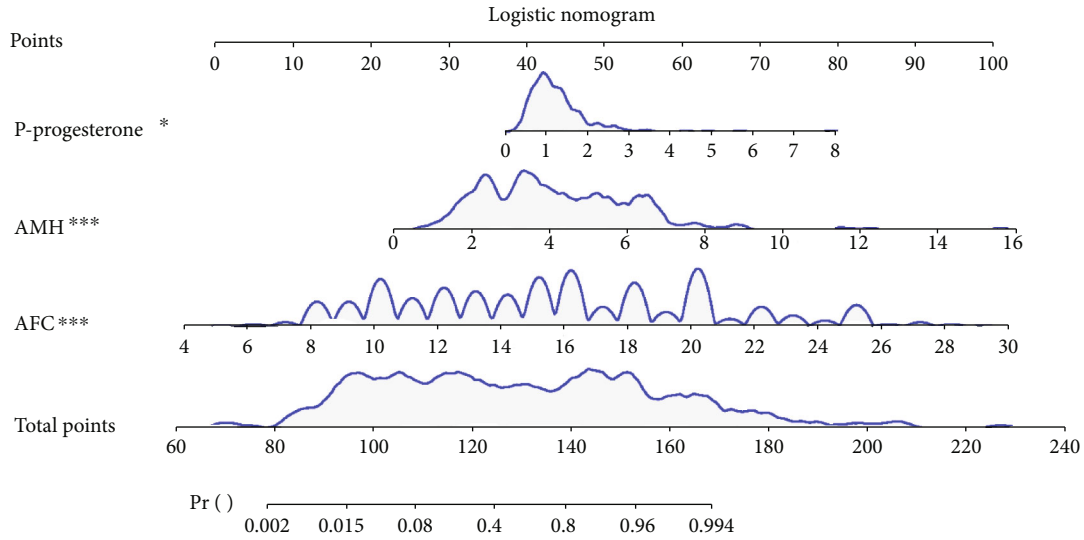


FIGURE 1: Logistic nomogram for predicting high ovarian response.

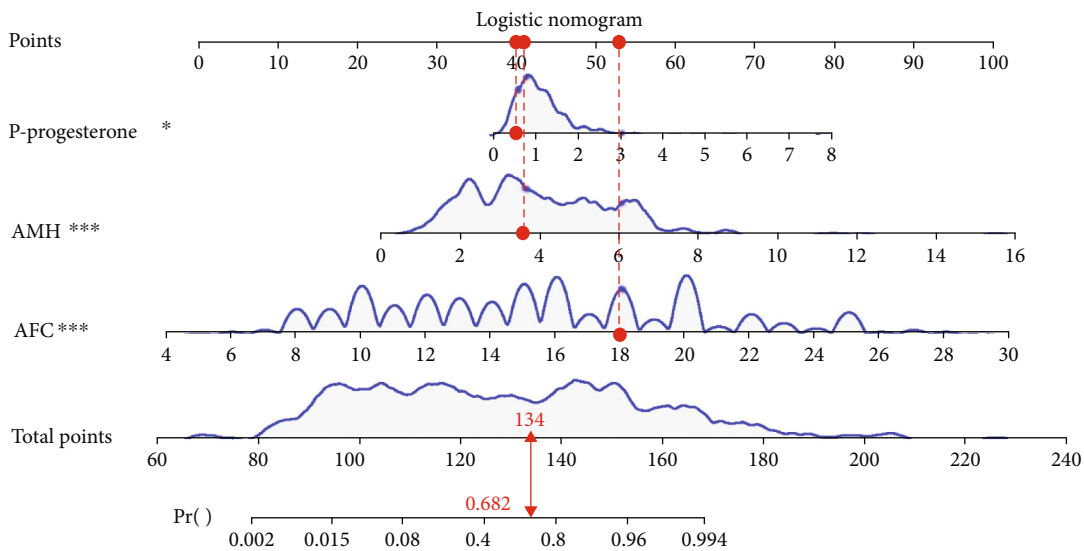


FIGURE 2: Example for the application of the nomogram.

TABLE 5: Performance comparison of predictors for high ovarian response by ROC analysis.

Predictors	AUC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	Z	P
Combination	0.958 (0.936-0.981)	0.491	0.916 (0.874-0.958)	0.911 (0.868-0.954)	—	
AFC	0.937 (0.911-0.963)	0.603	0.868 (0.817-0.920)	0.890 (0.841-0.942)	2.707	0.007
AMH at MC3	0.905 (0.871-0.939)	0.525	0.820 (0.762-0.879)	0.911 (0.868-0.954)	3.905	<0.001
P on HCG day	0.692 (0.636-0.749)	0.484	0.611 (0.537-0.685)	0.662 (0.588-0.767)	8.899	<0.001

AMH: anti-Müllerian hormone; AFC: antral follicle count; MC3: menstrual cycle day 3; P: progesterone; HCG: human chorionic gonadotropin; CI: confidence interval.

3.6 ng/mL, and AFC was 18. The total point was 134, and the predicted probability of high ovarian response was 0.682 (Figure 2), which was higher than optimum cutoff point 0.491 (Table 5) and indicated a higher incidence of high ovarian response.

3.4. Validation of the Prediction Model. According to the ROC analysis, the AUC values for AFC, AMH at MC3, and P level on HCG day were 0.937 (95% CI: 0.911-0.963), 0.905 (95% CI: 0.871-0.939), and 0.692 (95% CI: 0.636-0.749), respectively (Table 5, Figure 3). For the multivariate

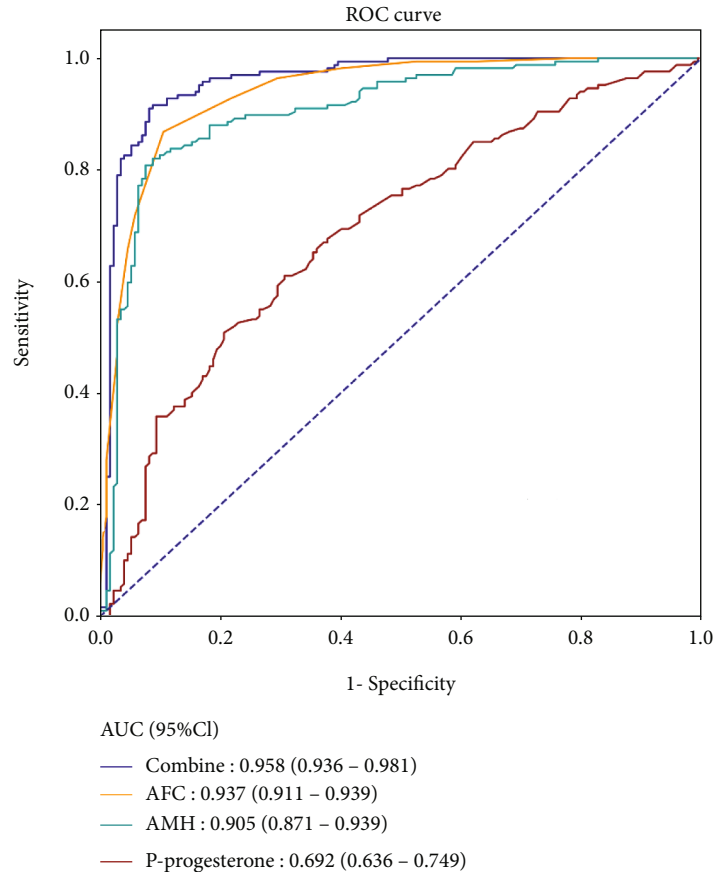


FIGURE 3: ROC curves for single factor and the combined prediction model.

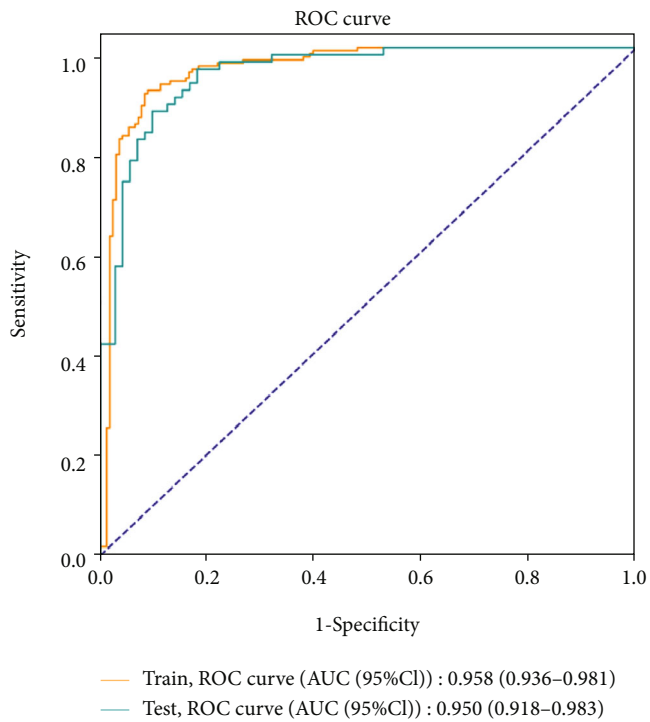
model combining the above predictors, the AUC reached 0.958 (95% CI: 0.936-0.981) with the sensitivity of 0.916 (95% CI: 0.863-0.953) and the specificity of 0.911 (95% CI: 0.858-0.949), revealing a better performance on the prediction of high ovarian response (Figures 3 and 4). Additionally, the AUC of the testing set was 0.950 (0.918-0.983) with the sensitivity of 0.875 (0.799-0.951) and the specificity of 0.875 (0.799-0.951) (Table 6, Figure 4), which confirmed the good discrimination of the model. The Hosmer-Lemeshow test showed the good calibration of the model (training set: $\chi^2 = 0.046$, $P = 0.829$; testing set: $\chi^2 = 7.497$, $P = 0.484$). The calibration curve also confirmed the good calibration of our model (Figure 4).

3.5. Comparison for the Prediction Models. As a novel neural network model based on random vector functional-link neural network (RVFLNN), BLS is suitable for processing relatively simple data and has a faster learning speed [15]. Therefore, we constructed a BLS model with 4 feature nodes per window, 5 feature node windows, 9 enhancement nodes, incremental steps (3), number of reinforcement nodes (50), coefficient of compressibility (0.7), and regularization coefficient (2^{-30}) and compared its predictive power with the logistic model. The predictors of our logistic model were put into the BLS model to assess the predictive power. The results showed that the AUC and accuracy of the BLS model were inferior to the logistic model (Table 7). It was indicated

that the BLS model may be not suitable for simple data which included only three variables. The ROC curves for the training set and the testing set of the BLS model are displayed in Figure 5.

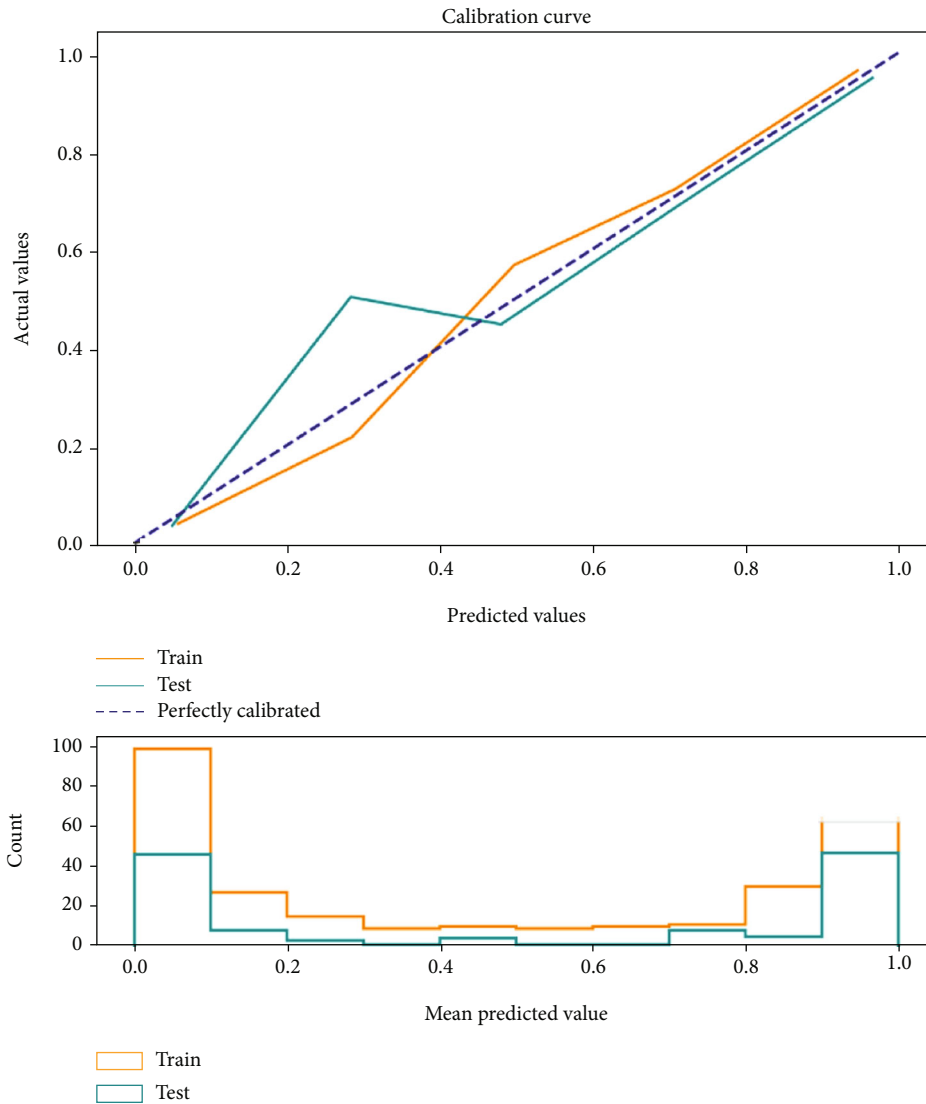
4. Discussion

High ovarian response can induce the risk of OHSS, leading to the increased discomfort in patients and even reducing prospects for pregnancy [16]. Up to 30% of cases with mild or moderate OHSS and 3-8% with severe OHSS were reported in IVF-ET cycles [17]. In this study, we aimed to develop a prediction model to predict the risk of high ovarian response in patients undergoing IVF-ET. Our results suggested that AFC, AMH at MC3, and P level on HCG day were the three effective predictors for high ovarian response in IVF-ET cycles. What is more, a combined prediction model with good performance was developed and validated: $\text{Ln}(P_{HR}/(1 - P_{HR})) = -11.094 + \text{AFC} * 0.513 + \text{AMH at MC3} * 0.628 + \text{P on HCG day} * 0.668$ (P_{HR} represented the probability of high ovarian response). Simultaneously, we plotted a nomogram for visualizing our model; the AUC value of the combined prediction model reached 0.958, which suggested the good discrimination of the model, and the internal validation confirmed the accuracy and feasibility of the model. Further, the Hosmer-Lemeshow test and the calibration curve showed the good



(a)

FIGURE 4: Continued.



(b)

FIGURE 4: ROC (a) and calibration curves (b) for the training set and the testing set of the logistic model.

TABLE 6: Predictive value of the logistic model for high ovarian response.

Variables	Data set	
	Training set	Testing set
AUC (95% CI)	0.958 (0.936-0.981)	0.950 (0.918-0.983)
Accuracy (95% CI)	0.914 (0.884-0.944)	0.875 (0.821-0.929)
Sensitivity (95% CI)	0.916 (0.874-0.958)	0.875 (0.799-0.951)
Specificity (95% CI)	0.911 (0.868-0.954)	0.875 (0.799-0.951)
PPV (95% CI)	0.911 (0.868-0.954)	0.875 (0.799-0.951)
NPV (95% CI)	0.917 (0.875-0.958)	0.875 (0.799-0.951)

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value; CI: confidence interval.

TABLE 7: The predictive value of the BLS model for high ovarian response.

Variables	Training set	Testing set
AUC (95% CI)	0.692 (0.636-0.748)	0.706 (0.622-0.791)
Accuracy (95% CI)	0.613 (0.561-0.666)	0.625 (0.546-0.704)
Sensitivity (95% CI)	0.964 (0.936-0.992)	0.958 (0.912-1.000)
Specificity (95% CI)	0.266 (0.200-0.333)	0.292 (0.187-0.397)
PPV (95% CI)	0.565 (0.507-0.324)	0.575 (0.487-0.380)
NPV (95% CI)	0.882 (0.794-0.971)	0.875 (0.743-1.000)

AUC: area under the curve; PPV: positive predictive value; NPV: negative predictive value; CI: confidence interval.

calibration of our model. We believed that the finding could make it easier for clinicians to predict high ovarian response in IVF-ET cycles and develop individualized treatment strat-

egies for patients, thereby reducing the incidence rates of OHSS and improving pregnancy rates and clinical outcomes.

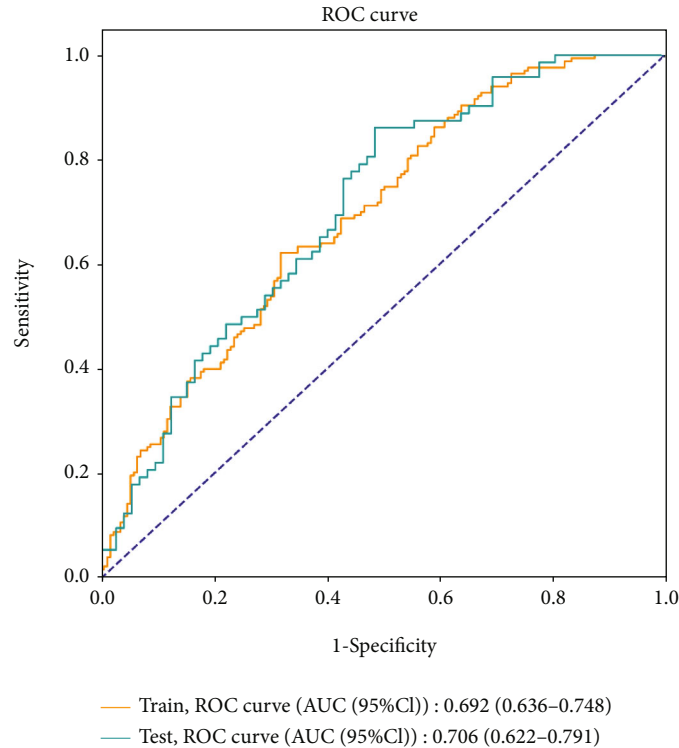


FIGURE 5: ROC curves for the training set and the testing set of the BLS model.

In our study, both AFC and basal AMH were independently associated with high ovarian response, which was consistent with the results of previous researches [9, 18–20]. Aflatoonian et al. reported that AMH and AFC were considered as the accurate and reliable predictors of high ovarian response to COH and could identify the patients who had an increased risk of OHSS before stimulation [18]. The reason may be that AMH concentration correlates significantly with the number of sinusoidal follicles in the ovary before ovulation and the number of oocytes collected after treatment, and patients with high ovarian response have higher AMH concentration compared with patients with normal ovarian response. Not only that, AMH was regarded as an excellent predictor of high ovarian response and could identify the risk of OHSS better than AFC, which may be due to the fact that AMH has more stable periodicity and is less susceptible to exogenous steroid hormones; moreover, AFC requires skilled ultrasound operators to carefully identify, measure, and count ovarian eggs, probably resulting in more interobserver variability in AFC [7, 18, 21–25]. However, few studies have analyzed the association between P level on HCG day and high ovarian response. In a retrospective study, P level at the first day of stimulation was recorded as a potential predictor, but no statistical significance was found [26]. Studies have shown that P level on the day of HCG administration varies among different ovarian responders [27–29]. Whether the P level can affect pregnancy rates and IVF-ET outcomes remains to be verified in further research. More importantly, our study showed that the model including all predictors had a more accurate predictive power for high ovarian response than the one con-

taining independent predictors. The possible explanation may be that considering only one factor to predict the probability of high ovarian response and ignoring the existence of other factors may reduce predictive ability and increase the error brought by independent factor. However, our conclusion needs to be confirmed by more related studies.

The strengths of the study should be noted. We identified predictors for high ovarian response in IVF-ET and developed a prediction model with more accurate predictive power, which could help clinicians efficiently identify patients at a risk of high ovarian response and individualize treatment for these patients. However, there were also some limitations in our study. Firstly, women aged 20–40 years were enrolled in our study, which may be considered as non-representative samples. This is mainly due to the fact that women of this age group have better fertility with more ideal stimulation effect. A wider range of ages could be considered in the future research to improve the universality of the model. In addition, small sample size and lack of external validation may affect the general applicability of our model. A multicenter study with large sample size and external validation is required to improve the accuracy and reliability of the model.

5. Conclusion

The developed prediction model had good discrimination and accuracy through the internal validation, which could help clinicians identify patients with high ovarian response, thereby improving pregnancy rates and clinical outcomes in IVF-ET cycles.

Data Availability

All the data utilized to support the theory and models of the present study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declared no potential conflicts of interests.

Authors' Contributions

Xinsha Tan and Jing Yang designed the study and wrote the manuscript. Xinsha Tan, Honglin Xi and Wenfeng Wang collected and analyzed the data. Xinsha Tan contributed to literature search. Jing Yang critically reviewed and improved the drafts of the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

This research was supported by the Science and Technology Plan Project of Yichang City (No. A19-301-32).

References

- [1] N. Pereira, A. C. Petrini, Z. N. Zhou, J. P. Lekovich, I. Kligman, and Z. Rosenwaks, "Pretreatment of normal responders in fresh in vitro fertilization cycles: a comparison of transdermal estradiol and oral contraceptive pills," *Clinical and Experimental Reproductive Medicine*, vol. 43, no. 4, pp. 228–232, 2016.
- [2] L. Duan, S. Bao, K. Li, X. Teng, L. Hong, and X. Zhao, "Comparing the long-acting and short-acting forms of gonadotropin-releasing hormone agonists in the long protocol of IVF/ICSI cycles: a retrospective study," *The Journal of Obstetrics and Gynaecology Research*, vol. 43, no. 6, pp. 1037–1042, 2017.
- [3] M. Gao, X. Jiang, B. Li et al., "Intrauterine injection of human chorionic gonadotropin before embryo transfer can improve in vitro fertilization-embryo transfer outcomes: a meta-analysis of randomized controlled trials," *Fertility and Sterility*, vol. 112, no. 1, pp. 89–97.e1, 2019.
- [4] S. M. Nelson, "Prevention and management of ovarian hyperstimulation syndrome," *Thrombosis Research*, vol. 151, Suppl. 1, pp. S61–S64, 2017.
- [5] S. K. Sunkara, V. Rittenberg, N. Raine-Fenning, S. Bhattacharya, J. Zamora, and A. Coomarasamy, "Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles," *Human Reproduction*, vol. 26, no. 7, pp. 1768–1774, 2011.
- [6] J. F. Kawwass, D. M. Kissin, A. D. Kulkarni et al., "Safety of assisted reproductive technology in the United States, 2000–2011," *JAMA*, vol. 313, no. 1, pp. 88–90, 2015.
- [7] J. Bedenk, E. Vrtačnik-Bokal, and I. Virant-Klun, "The role of anti-Müllerian hormone (AMH) in ovarian disease and infertility," *Journal of Assisted Reproduction and Genetics*, vol. 37, no. 1, pp. 89–100, 2020.
- [8] F. J. Broekmans, J. Kwee, D. J. Hendriks, B. W. Mol, and C. B. Lambalk, "A systematic review of tests predicting ovarian reserve and IVF outcome," *Human Reproduction Update*, vol. 12, no. 6, pp. 685–718, 2006.
- [9] J. Huang, J. Lin, H. Gao et al., "Anti-Müllerian hormone for the prediction of ovarian response in progestin-primed ovarian stimulation protocol for IVF," *Frontiers in Endocrinology*, vol. 10, p. 325, 2019.
- [10] S. Jang, K. H. Kim, J. H. Jun, and S. You, "Acupuncture for in vitro fertilization in women with poor ovarian response: a systematic review," *Integrative Medicine Research*, vol. 9, no. 2, p. 100395, 2020.
- [11] S. F. Lensen, J. Wilkinson, J. A. Leijdekkers et al., "Individualised gonadotropin dose selection using markers of ovarian reserve for women undergoing in vitro fertilisation plus intracytoplasmic sperm injection (IVF/ICSI)," *Cochrane Database of Systematic Reviews*, vol. 2018, no. 2, article CD012693, 2018.
- [12] N. P. Polyzos, H. Tournaye, L. Guzman, M. Camus, and S. M. Nelson, "Predictors of ovarian response in women treated with corifollitropin alfa for in vitro fertilization/intracytoplasmic sperm injection," *Fertility and Sterility*, vol. 100, no. 2, pp. 430–437, 2013.
- [13] S. Oehninger, S. M. Nelson, P. Verweij, and B. J. Stegmann, "Predictive factors for ovarian response in a corifollitropin alfa/GnRH antagonist protocol for controlled ovarian stimulation in IVF/ICSI cycles," *Reproductive Biology and Endocrinology*, vol. 13, no. 1, p. 117, 2015.
- [14] H. Zheng, S. Chen, H. Du et al., "Ovarian response prediction in controlled ovarian stimulation for IVF using anti-Müllerian hormone in Chinese women: a retrospective cohort study," *Medicine (Baltimore)*, vol. 96, no. 13, article e6495, 2017.
- [15] C. L. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep architecture," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 1, pp. 10–24, 2018.
- [16] S. L. Broer, M. Dölleman, B. C. Opmeer, B. C. Fauser, B. W. Mol, and F. J. M. Broekmans, "AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis," *Human Reproduction Update*, vol. 17, no. 1, pp. 46–54, 2011.
- [17] A. Delvigne and S. Rozenberg, "Epidemiology and prevention of ovarian hyperstimulation syndrome (OHSS): a review," *Human Reproduction Update*, vol. 8, no. 6, pp. 559–577, 2002.
- [18] A. Aflatoonian, H. Oskouian, S. Ahmadi, and L. Oskouian, "Prediction of high ovarian response to controlled ovarian hyperstimulation: anti-Müllerian hormone versus small antral follicle count (2–6 mm)," *Journal of Assisted Reproduction and Genetics*, vol. 26, no. 6, pp. 319–325, 2009.
- [19] L. M. E. Moolhuijsen and J. A. Visser, "Anti-Müllerian hormone and ovarian reserve: update on assessing ovarian function," *The Journal of Clinical Endocrinology and Metabolism*, vol. 105, no. 11, pp. 3361–3373, 2020.
- [20] F. Pilsgaard, A. G.-A. Grynnerup, K. Løssl, L. Bungum, and A. Pinborg, "The use of anti-Müllerian hormone for controlled ovarian stimulation in assisted reproductive technology, fertility assessment and -counseling," *Acta Obstetrica et Gynecologica Scandinavica*, vol. 97, no. 9, pp. 1105–1113, 2018.
- [21] V. L. Baker, C. Gracia, M. J. Glassner et al., "Multicenter evaluation of the access AMH antimüllerian hormone assay for the prediction of antral follicle count and poor ovarian response to controlled ovarian stimulation," *Fertility and Sterility*, vol. 110, no. 3, pp. 506–513.e3, 2018.
- [22] S. Iliodromiti, R. A. Anderson, and S. M. Nelson, "Technical and performance characteristics of anti-Müllerian hormone and antral follicle count as biomarkers of ovarian response,"

- Human Reproduction Update*, vol. 21, no. 6, pp. 698–710, 2015.
- [23] J. Bruno-Gaston, J. Jung, T. Kumar, P. Zarutskie, W. Gibbons, and S. Devaraj, “Association of ovarian response with picoAMH in women undergoing controlled ovarian hyperstimulation,” *Clinical Biochemistry*, vol. 95, pp. 34–40, 2021.
- [24] L. Mohiyiddeen, W. G. Newman, H. McBurney, B. Mulugeta, S. A. Roberts, and L. G. Nardo, “Follicle-stimulating hormone receptor gene polymorphisms are not associated with ovarian reserve markers[J],” *Fertility and sterility*, vol. 97, no. 3, pp. 677–681, 2012.
- [25] B. Sun, Y. Ma, L. Li et al., “Factors associated with ovarian hyperstimulation syndrome (OHSS) severity in women with polycystic ovary syndrome undergoing IVF/ICSI,” *Frontiers in Endocrinology*, vol. 11, p. 615957, 2020.
- [26] F. J. Broekmans, P. J. M. Verweij, M. J. C. Eijkemans, B. M. J. L. Mannaerts, and H. Witjes, “Prognostic models for high and low ovarian responses in controlled ovarian stimulation using a GnRH antagonist protocol,” *Human Reproduction*, vol. 29, no. 8, pp. 1688–1697, 2014.
- [27] A. Mahran, M. Khairy, R. Elkhateeb et al., “The value of serum progesterone level on day of human chorionic gonadotrophin administration/metaphase II oocyte ratio in predicting IVF/ICSI outcome in patients with normal ovarian reserve,” *Journal of Ovarian Research*, vol. 14, no. 1, p. 52, 2021.
- [28] A. Requena, M. Cruz, E. Bosch, M. Meseguer, and J. A. García-Velasco, “High progesterone levels in women with high ovarian response do not affect clinical outcomes: a retrospective cohort study,” *Reproductive Biology and Endocrinology*, vol. 12, no. 1, p. 69, 2014.
- [29] B. Xu, Z. Li, H. Zhang et al., “Serum progesterone level effects on the outcome of in vitro fertilization in patients with different ovarian response: an analysis of more than 10,000 cycles,” *Fertility and Sterility*, vol. 97, no. 6, pp. 1321–1327.e4, 2012.

Research Article

An Approach for Thoracic Syndrome Classification with Convolutional Neural Networks

Sapna Juneja ¹, Abhinav Juneja ², Gaurav Dhiman ³, Sanchit Behl ⁴,
and Sandeep Kautish ⁵

¹IMS Engineering College, Ghaziabad, India

²KIET Group of Institutions, Delhi NCR, Ghaziabad, India

³Govt. Bikram College of Commerce, Patiala, India

⁴BMIET, Sonapat, India

⁵LBEF Campus, Kathmandu, Nepal

Correspondence should be addressed to Sandeep Kautish; dr.skautish@gmail.com

Received 2 July 2021; Revised 13 August 2021; Accepted 25 August 2021; Published 21 September 2021

Academic Editor: Hamidreza Mohafez

Copyright © 2021 Sapna Juneja et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There have been remarkable changes in our lives and the way we perceive the world with advances in computing technology. Healthcare sector is evolving with the intervention of the latest computer-driven technology and has made a remarkable change in the diagnosis and treatment of various diseases. Due to many governing factors including air pollution, there is a rapid rise in chest-related diseases and the number of such patients is rising at an alarming rate. In this research work, we have employed machine learning approach for the detecting various chest-related problems using convolutional neural networks (CNN) on an open dataset of chest X-rays. The method has an edge over the traditional approaches for image segmentation including thresholding, *k*-means clustering, and edge detection. The CNN cannot scan and process the whole image at an instant; it needs to recursively scan small pixel spots until it has scanned the whole image. Spatial transformation layers and VGG19 have been used for the purpose of feature extraction, and ReLU activation function has been employed due to its inherent low complexity and high computation efficiency; finally, stochastic gradient descent has been used as an optimizer. The main advantage of the current method is that it retains the essential features of the image for prediction along with incorporating a considerable dimensional reduction. The model delivered substantial improvement over existing research in terms of precision, *f*-score, and accuracy of prediction. This model if used precisely can be very effective for healthcare practitioners in determining the thoracic or pneumonic symptoms in the patient at an early stage thus guiding the practitioner to start the treatment immediately leading to fast improvement in the health status of the patient.

1. Introduction

The chest carries the vital breath to be dissipated in to the body parts which provides probably nearly all basic survival elements of the body. A huge number of individuals have been detected annually to suffer from chest ailments of various types on the planet. Tuberculosis (TB), chronic obstructive aspiratory disease (COPD), pneumonia, asthma, and lung disease infections are the most significant chest sicknesses, which have been also considered extremely normal diseases on the planet. Tuberculosis (TB) has been emerging as an omnipresent disease across the globe; this may be evi-

denced by the fact that in 2017, tuberculosis was responsible for the highest number of unnecessary deaths worldwide [1].

Radiographs popularly known as X-Rays have been used as one of the reliable sources for tracking the vital body parts over the decades. These radiographs have different sequences of evaluation for different body components. For chest X-rays, the evaluation for the lungs, heart, mediastinum, diaphragm, and bony thorax is performed to read the patients' condition while for the abdomen portion, an evaluation of bowel syndrome, psoas and nephritic issues, liver and spleen, and preperitoneal fat lines and a search for unusual calcination are done in order to find the patients'

parameters. [2]. The interpretation of chest X-rays is a complex issue due to overlapping of the tissues in the chest area [3].

There is an unprecedented growth in the human progression towards increasing the expectancy of life. Medical Sciences and allied fields of research have made a lot of effort to find resolution for many of the life-threatening diseases. A lot of hard work and research is being done worldwide to make the life of human beings better by finding the best of preventive and diagnostic care. In our current work, we have focused on lung illnesses and proposed the detection through machine learning techniques. As per the Global Burden of Disease Study 2015 [4], approximately 3.2 million people lost lives in 2015 because of chronic obstructive aspiratory illness (COPD), essentially invoked due to smoking and tainting, while 4 million population lost their lives due to asthma. The death data of past is quite alarming and indicative of the toll it may create in coming future generations due to rise in the dominance of various factors which are responsible causes for this disease. Fogel [5] mentions that TB is an irresistible ailment, caused much of the time by microorganisms called *Mycobacterium tuberculosis*. The microorganisms generally come to the human being by inward breath across the lungs. They outspread from the underlying area within the lungs to different body parts by means of the circulation system, by the lymphatic framework, by means of the aviation routes, or by direct expansion to different organs [6]. COPD is an inflammatory disease of the lung, which causes an obstruction in the flow of air through the lungs; this phenomena is generally progressive and is not reversible [7]. The restriction in air flow through the lungs is normally both dynamic and related with a strange provocative reaction corresponding to the lungs to harmful molecules or vapors. As per clinical reports, the people suffering from COPD encountered breathlessness and cough and abnormal increase in release of mucus, sometimes even accompanied by wheezing. Ferguson et al. [8] stated that asthma is typically a protracted infection described by intermittent assaults of shortness of breath, tightness of the chest, and wheezing. During an asthma assault [9], the coating of the bronchial cylinders swells, making the aviation routes limited and limiting the progression of oxygen within the lungs. The asthma patients generally experience frequent exhaustion, restlessness, decline in the physical activities, etc. As per the Global Burden of Disease Study 2015 [4], asthma has a moderately low casualty rate that contrasted with other interminable sicknesses. The WHO [10] projected that approximately 300 million individuals experience the ill effects of asthma on an average. Asthma is the most widely recognized constant illness among kids. Cancer of the lung (malignancy) is caused due to uninhibited cell development in tissues of the lung [11]. This development may prompt a problem, i.e., spread of the disease from one body part to other body parts like the lungs. By far, most of lung tumors are cancerous for the lung, obtained from endothelium cells or tissues. Lung disease is identified as the most widely recognized cause for disease-related loss of life of human beings and another one generally common in ladies, as per [12]; the disease is

liable for loss of 1.76 million humans lives worldwide every year.

With a recent advancement in machine leaning-based processes, it is becoming feasible to detect and diagnose the lung diseases more precisely and well in advance. This may further reduce the causalities originated by these diseases and also the expenditure and consultancy on unessential factors. The true and worthy contribution of the researchers is the dedication of their commendable efforts to help the humans by applying AI to the analyze and forecast chest related disease if any. In the present state of practice, there is a lot of data available freely for research and modelling and also the enhanced computing power available to our machines; it has become very convenient and motivating to develop such systems which help the patients by predicting diseases in advance and help those who are not able to even afford the huge medical bills.

Nowadays, there is an increase in the incorporation of artificial neural network (ANN) structures [13] to develop classification systems for medical diagnosis. The multilayer neural network (MLNN) [14], probabilistic neural networks (PNN) [15], learning vector quantization (LVQ) [16] neural networks, generalized regression neural network (GRNN) [17], and radial basis function (RBF) [18] neural network structures have displayed an evidential improvement over the traditional pattern identification techniques for the diagnose system of the diseases including chest diseases. Various classification systems grounded upon neural network have been employed in chest disease diagnosis stream also. In the past, there have been several neural network-based studies that focused on the chest diseases. The taxonomy of the learning vector quantization neural network structure depends upon the nearness of the unknown information and abovementioned models. A learning vector quantization neural network [16] holds two different layers: one is competitive, and another one is linear output layer. The classification of input vectors is done by the competitive layer. Transformation of the classes of the competitive layer in to the classes of target as described by the user is done by the linear output layer.

The datasets generally used for the classification problems using the techniques like machine learning need the various attributes related to the symptoms, age, sex, snapshot data, X-ray data, and few for specific attributes. By inception of this critical data of the patient, it is easier to train a model and use it for predictive analysis of patients by the health workers in practice. As a practice, medical X-ray images are very often employed for diagnosing few typically subtle human body components including bones, chest, tooth, and skull to name a few. Health experts and workers have eventually relied on this process since the past several decades in order to explore and comprehend fractures and anomalies in human body components. It is pertinent to mention that the X-rays are a very result-oriented diagnostic option in enlightening the pathological variations, further complementing its inherent noninvasive operations and economic viabilities. Abiyev and Ma'aitah [19] demonstrated that the chest-related diseases may be projected in the form of CXR images depicting the cavitation, blunted

costophrenic angles, infiltrates, and the consolidations. Upon exploring a typical image of a chest X-ray, it is feasible for a radiologist to evaluate and diagnose certain specific conditions and diseases including effusion, pneumonia, infiltration, bronchitis, cardiomegaly, and fractures [20].

Various researchers have relied on devising methodologies for diagnosis of the chest diseases using the smart and innovative AI-based machine learning methodologies [21]. The detection of prolonged obstructive respiratory and pneumonia infections was devised by inception of the neural systems and an artificial system generating immunity for the environment. For tracking the lung disorders including tuberculosis, pneumonia, and lung disease, chest radiographs are very instrumental. For the purpose of image preprocessing, image segmentation using histogram equalization has been applied. And for the purpose of classification, finally, feed-forward neural networks have been used in the past by the researchers. Though these strategies have been quite helpful in the classification of healthcare-related problems, the deep neural networks outplay these techniques in terms of their efficiency, reliability, and computational time altogether. Many times, to increase the level of accuracy of the classification process, deep neural networks are applied [22]. This is rewarded by deep neural networks with a commendable accuracy outcome. This has led the research community to apply the deep neural networks for classification of medical problems by use of image classification. It has been further evidenced that these networks are able to track and extract classification features which are distinguishing between various classes of possible states of the evidence of disease. Convolutional neural networks are the favorites when deployed in a deep neural network [23] and have been applied for classification of the medical images; these networks are characterized to possess an ability to different layers of features from the sampled images. Deep convolutional neural networks are recognized for upgrading the diagnostic accuracy and mean squared error for the chest diseases.

In the current research work, convolutional neural networks have been employed on the open dataset from the National Institutes of Health (NIH) which has a huge collection of patient X-ray films for the chest, having a high degree of labelling accuracy. The compilation of the similar work done in the past by other researchers has been given in Section 2, literature review. Section 3 of the paper gives the complete process of implementing the convolutional neural networks on the dataset used in this study. It gives a stepwise process flow of the whole set of operations performed to generate the model. Section 4 is the result section, and it shows the finding and achievements of our work and compares them with the related work. Section 5 provides the conclusion of the work done in this current research and highlights the core outcomes of the current work, the limitations, and future scope of the work.

2. Literature Review

Image processing has been an area of interest to the research community for quite a long time, and there are evidences of

motivation behind the success that has been achieved in this domain. Thirumaran and Shylaja [24] showcased a brief overview for medical image processing and the various modalities of the process. Khobragade et al. [25] proposed an automated environment catering to the detection and diagnosis of lung diseases explicitly for thoracic diseases with the incorporation of chest radiographs. From the outcomes, it is evidenced that the image preprocessing approaches like equalization through histograms and image processing yield commendable outcomes using the radiograph of the chest. Specified pattern identification, for example, feed-forward artificial neural network, is presenting acceptable outcomes. Chen et al. [26] explored a methodology to boost the dataset considerably. The researcher used an enhanced dataset for training the CNN model for the identification and diagnosis of thorax disease; the model performance was significantly enhanced. They proposed to create a repository of huge collection of images without labelling from clinical services to improvise the performance of their CNN models. Wang et al. [27] attempted to devise a "machine-human annotated" wide-range chest X-ray database which reveals the practitioner's medical and technical issues related to the handling of several thousands of the patient databases. They executed a comprehensive quantitative performance standard for a set of eight popular thoracic pathology classification and weakly supervised localization with the deployment of chest X-ray database. The prime objective of research was to devise a roadmap for enabling the future efforts to provide aid in promotion of public datasets, which is very critical for this domain of application. Devising a dedicated commercial, robust, reliable, and fully autonomous health diagnosis environment is till date a dream come true. ChestX-ray8 has the potential to equip the data phishing deep neural network models to generate applications which are clinically meaningful which may include instances like pattern mining for some commonly occurring diseases, automation in generating the radiology reports, and analysis correlation of the disease to name a few possibilities. Chan et al. [28] introduced the methodology for segmentation of the lung focused on the abnormal region with the help of numerous overlapping blocks. Texture generated due to computation of multiple overlapping blocks is used to detect the abnormal regions. Concluding this work, the technique efficiently explores lung ailments of the area depicted in chest radiograph image, which further enhances the feasibility of diagnosing the latent issue of the pneumothorax area. Intensity and gradient are the basic fundamentals for texture analysis to detect the pneumothorax. Sharma et al. [29] explored the rib cage area from the lung area using the identification of boundary. To isolate healthy lung area from the cloud of pneumonia Otsu thresholding is incorporated. Despite working on different strategies that may be embraced for thresholding, the pictures of CXR generally yield better outcomes.

After going through the latest research on the feature extraction and use of the same in disease predictions, the current work was motivated. For undertaking the current research work, the objective was to improvise on the existing prediction models and decrease the number of false or

TABLE 1: Brief summary of inspiration from the earlier research in the domain of disease identification.

S. No.	Author and year	Paper title	Technique used	Objective
1	Bharati & Podder; [30]	Disease Detection from Lung X-ray Images Based on Hybrid Deep Learning Subrato	CNN, vanilla NN	The model proposed classification of chest diseases with its metrics as precision, recall, and f -score
2	Rajaraman et al. [31]	Assessment of an Ensemble of Machine Learning Models towards Abnormality Detection in Chest Radiographs	Sequential CNN	Used weighted averaging to in base learners to classify the chest - rays
3	Chan et al. [28]	Effective Pneumothorax Detection for Chest X-Ray Images Using Local Binary Pattern and Support Vector Machine	Support vector machine and local binary pattern	The paper proposed a methodology to detect the lung diseases using the local binary patterns and then further used the SVM technique to classify the type of disease
4	Li et al. [32]	Thoracic Disease Identification and Localization with Limited Supervision	CNN	Identification and localization of abnormalities in the X-rays
5	Sharma et al. [33]	An Analysis Of Convolutional Neural Networks For Image Classification	CNN	The paper focusses on the analysis of real time images of three types of CNN's; these are AlexNets, GoogLeNet, and ResNet50
6	Yao et al. [34]	Learning to diagnose from scratch by exploiting dependencies among labels	LSTM	Used long short-term memory networks for distinction between chest diseases
7	Esteva et al. [35]	Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks	t-SNE-based NN	Analyzed the internal features of the cells by using the CNN with the t-distributed stochastic neighbor embedding
8	Wang et al. [27]	ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases	CNN	The work focusses on how thoracic ailments can be discovered and explicitly located with the help of a combined softly supervised multilabelled image sorting and ailment localization framework; the same is verified with the dataset used in the paper

erroneous predictions using the machine learning models. Table 1 summarizes some of the relevant work done by researchers and provided us a strong foundation to generate the initial framework for the experimental process.

3. Image Classification Using Convolution Neural Networks

Air pollution has the potential to affect human health both with some direct impact or sometimes indirectly, creating discomfort in physical well-being and resulting in disease or maybe death. Research has evidenced that on exposure of the human body to a polluted environment, there is a sudden rise in the mortality rate [36, 37].

3.1. Patient Chest Dataset. Recently, a huge dataset with more than 112120 images of 30805 patients for X-ray lung data was released by the National Institutes of Health [27], for the creation of labels; NLP has been used to convert radiology reports for the classification of the diseases. The approximate accuracy of labelling is more than 90%. The dataset is available for use on data repositories including

Kaggle [38]. This dataset inspired the current experimental work on machine learning. In the current work, the authors have undergone an analysis of this chest dataset and then applied machine learning and deep learning for the prediction of a patient to be suffering for any lung ailment; the type of lung ailment and the degree of accuracy of prediction are determined. There are 11 attributes corresponding to the patient in the dataset which include the image index, patient id, patient age, gender, follow-up, label of disease class, view position, image width, image height, pixel spacing- x , and pixel spacing- y . This project generates a binary classification of the incoming data stream which is basically the patient's individual data pertaining to attributes including patient age, gender, X-ray images, and view position (only the needed attributes have been chosen for modelling), and output is a function that conveys that the patient is suffering from any particular disease or not. This is a relatively new dataset and not much of researchers have explored and presented their work on it.

3.2. Convolutional Neural Networks. Convolutional neural networks (CNN) [39, 40] are typically analogous to the

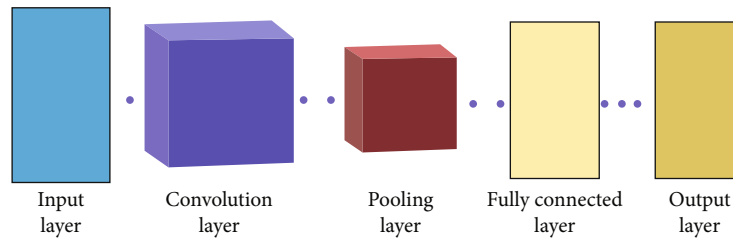


FIGURE 1: A convolutional neural network.

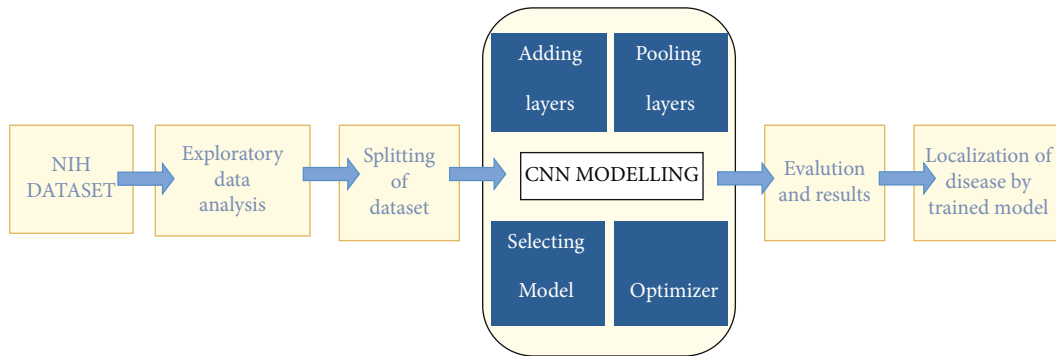


FIGURE 2: Block diagram of Chester—the chest disease predictor.

artificial neural networks (ANN); they comprise of components called neurons which are capable to optimize on their own through a phenomenon that is called as self-learning. Each neuron of the CNN is capable of receiving at its input some typical input and executes an operation (ranging from a scalar product succeeded by a linear function). The input to this network is a set of raw image vectors, and they are processed to a final output format which is the class score for a particular input vector. Perceptive score weight function is present in this entire network structure spanning all its stages and nodes. Loss functions pertaining to different classes are associated with the terminal layer of the network. Also, all the basic functionalities and modalities of the normal artificial neural networks (ANN) still hold for the CNNs. These networks are specialized for image pattern recognition, and this differentiates them from the ANNs [41]. The classical ANNs suffer from a drawback of low computational efficiency for image-related data due to the complexity in calculations. Additionally, the preprocessing required in a CNN is considerably lesser than other comparable algorithms. A typical convolutional neural network is presented in Figure 1.

The CNN [42] consists of 5 major quantifiable stages. The first stage is the input layer, where the input to the network is supplied in the form of an image. The convolution layer, on its part, convolves the image; i.e., it extracts the significant and differentiating features including edges, colours, and corners from the image received from the previous stage. This layer is multiplied into two matrices, wherein one of the matrices is a known learnable parameter matrix and the other one is the portion of image. The dot product generates a reduced matrix in the end with reduced and required features represented by the matrix, also called as the feature matrix. The pooling layer reduces the feature

matrix further in order to generate only dominant image features from the feature matrix. This is done in order to ensure optimized computing efficiency of the system. Average pooling and max pooling are two techniques through which the pooling layer performs this reduction. Till this layer, we reduce the dimensionality of our image. In the fully connected layer [43], the reduced feature matrix is converted to a single vector. The flattened output is sent to a feed-forward neural network [44]; further, back propagation is used during every cycle of training iterations. With maturity of training process, the model is enabled to distinguish certain low-level features and critical features which dominate in images. This facilitates the process of final classification in the output layer. CNNs have been a critical resource in computer vision [29] and image understanding. Generally, the CNNs are realized through a composition of simple linear and nonlinear filters including convolution and rectification, though it becomes a very tedious task to complete them in practice, because the CNNs learn from huge learning datasets, generally, millions of training images, employing intelligent and efficient implementations. Chester explores simple blocks of a CNN in practice, such as convolution, normalisation [45], and pooling [46] which may further be cascaded and drawn out easily to generate CNN architectures. Many of such blocks use optimized CPU and GPU implementations along with CUDA.

3.3. Using CNNs in Image Classification. We have taken a stepwise approach to build our model. For the block diagram of Chester, the proposed model for chest disease prediction has been given in Figure 2, initially, the dataset of National Institutes of Health (NIH).

3.3.1. Playing with Data/Analyzing Data. The dataset was loaded using a standard library. After loading, the data is

TABLE 2: Comparing our proposed work with existing work on considered metrics.

S. No.	Previous work/model	Precision	Recall	f -score	Accuracy
1.	Disease Detection from Lung X-Ray Images Based on Hybrid Deep Learning [30]	0.63	0.69	0.68	0.71
2.	The proposed model for disease prediction	0.77571	0.63098	0.76043	0.80056

divided into two sections, the one with disease and others having no disease. The various attributes of the dataset are analyzed for any missing values or deviations from the standard format of representing them. Some graph plots were done to understand the basic structure of data and have some idea of correlation among the attributes of similar class data members. We used Matplotlib and Seaborn libraries for the analysis of dataset. There are 15 classes in the full dataset which comprise of one no finding class and other 14 disease classes; since this is a drastically reduced version of the full dataset, some of the classes are sparse labelled as “no findings.” The other 14 disease classes are hernia, pneumonia, fibrosis, edema, emphysema, cardiomegaly, pleural thickening, consolidation, pneumothorax, mass, nodule, atelectasis, effusion, infiltration, and no finding. The images belong to any of these identified classes of disease group.

3.3.2. Preprocessing. In the current work, we have used 40000 image samples out of the total dataset, due to limitation of the hardware to process the huge dataset. After the analysis of data, the data is split up into two sets, one for the purpose of training and the other for testing. We have employed 30000 samples for the purpose of training and 10000 samples for the purpose of testing the model.

3.3.3. Chester. As the next process of the Chester Model, we next transfer the model training dataset and transfer it to further layers of the model. The model comprises of three significant layers in the following order.

(1) *Spatial Transformer Layers.* It further comprises three inherent layers. In the first layer, initially, default routing is transferred, which indicates that the λ features of the lung X-ray image correspond to a normal quantum of 0. The next layer is referred to as the batch normalization layer that is responsible for reduction of the amount by which the hidden unit values shift around. Finally, the last layer is the spatial transformer, which corresponds to the removal of maximum significant features for disease classification.

(2) *Extraction of Feature Layers.* For this purpose, the VGG19 model has been pretrained. By default, it loads weight pretrained on ImageNet. There is a group of 19 deep layers where VGG is the feature extraction layer; there are various pretrained classifiers available.

(3) *Classification Layers.* In this case, the first layer defined earlier is used for the purpose of harvesting the VGG19 layers with additional 5 features such as “gender female,” “gender male,” “age,” “view position PA,” and “view position AP.” The purpose of these additional 5 features is to address the issue of sorting. Activation function employed for the

current work is Rectified Linear Units (ReLU) [47]. The ReLU activation function was chosen due to its inherent feature of being simpler computationally. The inherent simplicity of ReLU makes it a favorite; here, forward and backward passes are implemented as simple if statements. There is a considerable reduction in cost for training the network using ReLU. This gives the researcher liberty to train larger networks having substantial parameters though maintaining the same computational cost, thereby providing a better capacity to hold and generally greater test set accuracy as well. The mathematical expression given below in equation (1) describes the ReLU function $f(x)$ for various input values [48].

$$f(x) = \max(0, x_i), \quad (1)$$

$$f(x) = \begin{cases} x_i, & \text{for } x_i \geq 0, \\ 0, & \text{for } x_i < 0. \end{cases} \quad (2)$$

The above equation shows that the ReLU function works well for positive input values and clips the negative values to zero. After creating the model, we define precision, recall, and f -score for our case. To optimize the model, we use stochastic gradient descent as an optimizer and pass binary accuracy, precision, recall, and f -score as its metrics at different threshold values. Execute the model, and further analyse the model performance. For the purpose of evaluation, stochastic gradient descent (SGD) [49] has been used here, which is a very powerful optimizer. And finally, we ran the model in batches of training (32) and validation (256). The model was run till the 5th epoch. Results were visualized, i.e., testing on images and finding out the possible disease and localizing them, and the model delivered considerable performance.

4. Result

The task of analysis of such a huge dataset with X-ray images was very interesting and challenging for us. We could model the system for 40000 images including men and women, selected at random. In the scanned sample of images and from the inferences drawn by our model, it has been analyzed that men are more prone to chest diseases than females. There may be different clinical and behavioral reasons for the same but that is not the scope of our current work. There are various metrics employed for the evaluation of the machine learning models by the researchers. In the current work, we have used the performance evaluation metrics to be precision, recall and F -measure for our Chester classifier. The proposed classifier model fared reasonably well and has been well acquainted to this evaluation.

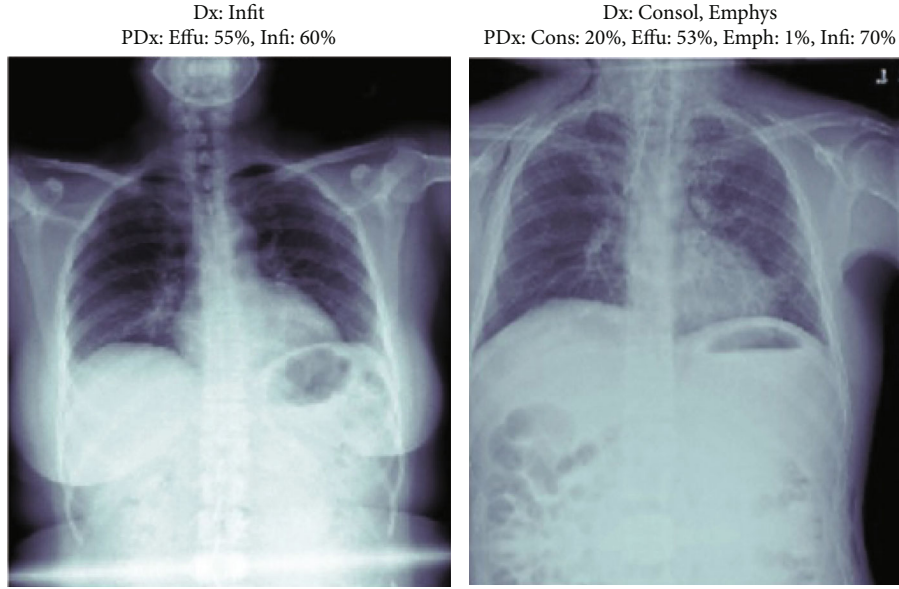


FIGURE 3: Testing on images for the output.

Precision refers to the total quantity of positive class predictions which fit to the positive class actually also. Recall refers to the total number of positive class predictions from the complete set of positive instances from the data bank. F -measure in turn furnishes with a single score that balances both the issues related to precision and recall represented by a number [50, 51]. Equations (3), (4), and (5) below represent the method to calculate precision, recall, and f -score, respectively.

$$\text{Precision} = \text{true} \frac{\text{positive}}{\text{true positive} + \text{false positive}}, \quad (3)$$

$$\text{Recall} = \text{TPR} = \text{true} \frac{\text{positives}}{\text{true positives} + \text{false negatives}}, \quad (4)$$

$$\text{F-score} = (1 + \beta^2) \text{precision} * \frac{\text{recall}}{\beta^2 * \text{precision}} + \text{recall}. \quad (5)$$

The current work considered the above metrics for the model evaluation over AUC_ROC and other available metrics due to some of the inherent properties of these metrics that make them more reliable evaluation metrics. The first reason being the fact that the real world data has a tendency to possess an imbalance among the positive and negative samples. This imbalance has a significant impact on the value of precision but the AUC/ROC do not portray this impact. The ROC/AUC curve has a limitation of not being able to display the performance of the classifier, while in contrast with our used metrics, we can do it with ease.

We have referred to the work of researchers who have previously made their contributions on a similar disease identification with similar dataset or some other datasets. It was a very good foundation for us to get valuable inputs from these already done experimental works. Table 2 below shows a comparative analysis for work explored by us on the

similar data and disease identification problem with same evaluation metrics. Accuracy of our proposed model averages around 80 which is much better than the previous research work being done by the researchers in the same area thus leading to the development of more successful model.

4.1. Testing on Images for the Output. For displaying the result of the patient, we used Matplotlib and NumPy library so that we can calculate the result of the X-ray uploaded by the patient or physician. For this, we are segregating the result in separate classes that are in the chest; for the prediction part of X-ray, we have taken a maximum of 4 classes that have the highest values in the X-ray after prediction. As displayed in the images given below in Figures 3 and 4, it can be seen that the result given in NIH dataset as compared with the result produced by our model are accurate and we can tell what the predicted value of diseases beside them is. In the figures below, 4 random cases have been shown for instance. Here, the user, a physician or a patient, filled in the information about X-ray like age, gender, and view position (PA and AP). Basically judging that the patient is ill, before proceeding with the analysis on more important trials, we have chosen prediction score greater than 0.5 to be taken as a threshold. In the given images below, there are various prediction percentages associated with various potential classes of diseases of a patient may have. But as our chosen prediction score is a threshold more than 50, so it will be needed to consider the patient vulnerable for the class of disease where the percentage is more than 50. In Figures 3 and 4, we have analyzed that our model is capable of identifying and classifying even those X-rays that have multiple chest diseases to what percentile it is confident in what disease that the patient might be suffering from and help them in identification of their diseases. These figures lead to the accuracy of the model by detecting the disease which the other models are not able to diagnose.

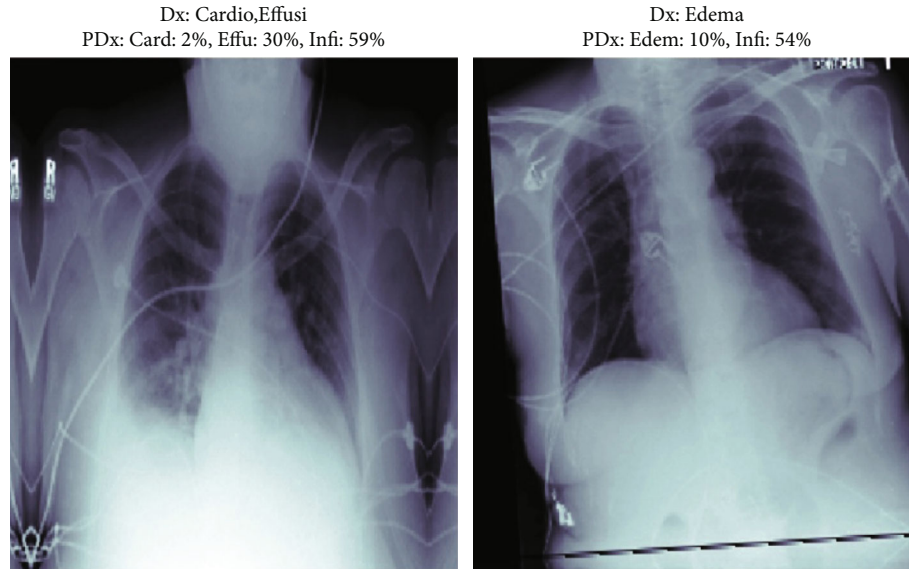


FIGURE 4: Testing on images for the output.

5. Conclusion

Chester is a novel system for analysis and prediction of chest diseases using the profound convolutional systems that is primarily coordinated and permits simple experimentation with clever thoughts. The current work presents a comprehensive mechanism to help the radiologists in diagnosing chest X-rays. This work (Chester) can be utilized as an aide and as an educational tool for students. The framework is intended to process everything locally which guarantees appreciable security. We trust this prompts radiologist to give us criticism which would assist us with improving this instrument and adjust it to their needs. Artificial neural network structures have not been much explored in evaluation of the health issues related to the chest. These examinations have applied distinctive neural system structures to the different chest sicknesses analysis issues utilizing their different datasets. On account of the diverse dataset utilized by the investigations, the immediate correlation of the outcomes was complex. Further, the complexity of the proposed solution is that the current work was very challenging and exciting for the entire team. It is a very complex situation to handle a large volume data, with lakhs of X-ray pictures to scan. The motivation to explore more in the chest disease prediction using the convolutional neural networks for the better disease diagnosis kept the project moving towards its goal. The dataset is not very old, and data is not standardized which makes it difficult to read and map the pictures.

The major outcomes of the research work taken in this paper can be concluded with the following:

- (1) It has been observed that we can use the neural network structures with some confidence bounds to explore the diagnosis of various chest-related ailments

- (2) Reasonable results were obtained for the classification of chest diseases using the CNN model
- (3) The neural networks can self-learn from experience and can be a vital source of help for the practitioners in their diagnostic and treatment efforts

Although we have put our sincere efforts in shaping the current proposed model, with the progression of the work, it was experienced that some more open areas may be addressed to take this work to the next landmark. The modelling may be done with more numbers of epochs, and with parameter tuning, convergence may be achieved conveniently. More degrees of training shall increase the chances of tracking better features to make easier classification. For CNN using VGG, we can experiment on many other pre-trained models along with different tweaks that can be made in the model. For the spatial transformer layer, we can try and implement a more complex and sophisticated localization network. The results of the current have fared well to our initial expectations, but to be able to apply it in hospitals, more improvements are needed to increase the precision of the model and also more training data may be employed. Such CNN-based projects tend to employ systems with good computational resources to be able to provide responsive predictions in the minimum possible time to be implementable in practice.

Data Availability

The data used to support the findings of this study are available from the author upon request (gdhiman0001@gmail.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. Miyahara, S. Piyaworawong, V. Naranbhai et al., "Predicting the risk of pulmonary tuberculosis based on the neutrophil-to-lymphocyte ratio at TB screening in HIV-infected individuals," *BMC Infectious Diseases*, vol. 19, no. 1, pp. 667–669, 2019.
- [2] W. F. Hook, *X-ray Film Reading Made Easy*, CiteSeerX, 2000.
- [3] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," *Biomedical Engineering Online*, vol. 17, no. 1, pp. 113–123, 2018.
- [4] J. B. Soriano, A. A. Abajobir, K. H. Abate et al., "Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015," *The Lancet Respiratory Medicine*, vol. 5, no. 9, pp. 691–706, 2017.
- [5] N. Fogel, "Tuberculosis: a disease without boundaries," *Tuberculosis*, vol. 95, no. 5, pp. 527–531, 2015.
- [6] S. Huddart, T. Bossuroy, V. Pons, S. Baral, M. Pai, and C. Delavallade, "Knowledge about tuberculosis and infection prevention behavior: a nine city longitudinal study from India," *PLoS One*, vol. 13, no. 10, p. e0206245, 2018.
- [7] G. Viegi, F. Pistelli, D. L. Sherrill, S. Maio, S. Baldacci, and L. Carrozzi, "Definition, epidemiology and natural history of COPD," *European Respiratory Journal*, vol. 30, no. 5, pp. 993–1013, 2007.
- [8] J. E. Fergeson, S. S. Patel, and R. F. Lockey, "Acute asthma, prognosis, and treatment," *Journal of Allergy and Clinical Immunology*, vol. 139, no. 2, pp. 438–447, 2017.
- [9] D. Eassey, H. K. Reddel, K. Ryan, and L. Smith, "The impact of severe asthma on patients' autonomy: a qualitative study," *Health Expectations*, vol. 22, no. 3, pp. 528–536, 2019.
- [10] C. Nunes, A. M. Pereira, and M. Morais-Almeida, "Asthma costs and social impact," *Asthma Research and Practice*, vol. 3, no. 1, pp. 1–11, 2017.
- [11] P. E. Bunney, A. N. Zink, A. A. Holm, C. J. Billington, and C. M. Kotz, "Orexin activation counteracts decreases in nonexercise activity thermogenesis (NEAT) caused by high-fat diet," *Physiology & Behavior*, vol. 176, no. 5, pp. 139–148, 2017.
- [12] Forum of International Respiratory Societies, *Fact Sheet-World Lung Day 2019*, 2019.
- [13] O. I. Abiodun, M. U. Kiru, A. Jantan et al., "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019.
- [14] S. T. Lee, S. Lim, N. Y. Choi et al., "Operation scheme of multi-layer neural networks using NAND flash memory as high-density synaptic devices," *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 1085–1093, 2019.
- [15] P. A. Kowalski and M. Kusy, "Determining significance of input neurons for probabilistic neural network by sensitivity analysis procedure," *Computational Intelligence*, vol. 34, no. 3, pp. 895–916, 2018.
- [16] X. Lu, H. Wang, W. Dong, F. Wu, Z. Zheng, and G. Shi, "Learning a deep vector quantization network for image compression," *IEEE Access*, vol. 7, pp. 118815–118825, 2019.
- [17] J. Y. Goulermas, P. Liatsis, Xiao-Jun Zeng, and P. Cook, "Density-driven generalized regression neural networks (DD-GRNN) for function approximation," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1683–1696, 2007.
- [18] N. B. Karayiannis, "New developments in the theory and training of reformulated radial basis neural networks," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, Como, Italy, 2000.
- [19] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," *Journal of Healthcare Engineering*, vol. 2018, Article ID 4168538, 11 pages, 2018.
- [20] A. Dhankhar, S. Juneja, A. Juneja, and V. Bali, "Kernel parameter tuning to tweak the performance of classifiers for identification of heart diseases," *International Journal of E-Health and Medical Communications*, vol. 12, no. 4, pp. 1–16, 2021.
- [21] A. Juneja, S. Juneja, S. Kaur, and V. Kumar, "Predicting diabetes mellitus with machine learning techniques using multi-criteria decision making," *International Journal of Information Retrieval Research*, vol. 11, no. 2, pp. 38–52, 2021.
- [22] R. Geirhos, D. H. J. Janssen, H. H. Schütt, J. Rauber, M. Bethge, and F. A. Wichmann, "Comparing deep neural networks against humans: object recognition when the signal gets weaker," 2017, <https://arxiv.org/abs/1706.06969>.
- [23] B. B. Benuwa, Y. Zhan, B. Ghansah, D. K. Wornyo, and F. Banaseka Kataka, "A review of deep machine learning," *International Journal of Engineering Research in Africa*, vol. 24, no. February 2017, pp. 124–136, 2016.
- [24] J. Thirumaran and S. Shylaja, *Medical Image Processing-An Introduction*, International Journal of Science and Research (IJSR), 2017.
- [25] S. Khobragade, A. Tiwari, C. Y. Patil, and V. Narke, "Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network," in *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, pp. 1–5, Delhi, India, 2016.
- [26] J. Chen, X. Qi, O. Tervonen, O. Silven, G. Zhao, and M. Pietikainen, "Thorax disease diagnosis using deep convolutional neural network," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2287–2290, Orlando, FL, USA, 2016.
- [27] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings -30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 3462–3471, Honolulu, Hawaii, USA, 2017.
- [28] Y. H. Chan, Y. Z. Zeng, H. C. Wu, M. C. Wu, and H. M. Sun, "Effective pneumothorax detection for chest X-ray images using local binary pattern and support vector machine," *Journal of Healthcare Engineering*, vol. 2018, Article ID 2908517, 11 pages, 2018.
- [29] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, no. Iccids, pp. 377–384, 2018.
- [30] S. Bharati and P. Podder, *Disease Detection from Lung X-ray Images Based on Hybrid Deep Learning*, Image and Video Processing, Cornell University, 2020.
- [31] S. Rajaraman, S. Sornapudi, M. Kohli, and S. Antani, "Assessment of an ensemble of machine learning models toward abnormality detection in chest radiographs," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3689–3692, Berlin, Germany, 2019.

- [32] Z. Li, C. Wang, M. Han et al., "Thoracic disease identification and localization with limited supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8290–8299, Salt Lake City, UT, USA, 2018.
- [33] A. Sharma, D. Raju, and S. Ranjan, "Detection of pneumonia clouds in chest X-ray using image processing approach," in *2017 Nirma University International Conference on Engineering (NUICONE)*, pp. 1–4, Ahmedabad, India, 2018.
- [34] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," 2017, <https://arxiv.org/abs/1710.10501>.
- [35] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [36] J. Lelieveld, A. Pozzer, U. Pöschl, M. Fnais, A. Haines, and T. Münzel, "Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective," *Cardiovascular Research*, vol. 116, no. 11, pp. 1910–1917, 2020.
- [37] S. Sanyal, T. Rochereau, C. N. Maesano, L. Com-Ruelle, and I. Annesi-Maesano, "Long-term effect of outdoor air pollution on mortality and morbidity: a 12-year follow-up study for metropolitan France," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, pp. 2487–2488, 2018.
- [38] NIH Chest X-ray Dataset, *National Institutes of Health-Clinical Center*, 2017.
- [39] S. Albawi, T. A. M. Mohammed, and S. Alzawi, *A Data-Driven Approach to Precipitation Parameterizations Using Convolutional Encoder-Decoder Neural Networks Pablo*, Ieee, 2017.
- [40] K. T. O. Shea, "An introduction to convolutional neural networks," 2015, <https://arxiv.org/abs/1511.08458>.
- [41] M. Garg and G. Dhiman, "Deep convolution neural network approach for defect inspection of textured surfaces," *Journal of the Institute of Electronics and Computer*, vol. 2, no. 1, pp. 28–38, 2020.
- [42] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
- [43] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, and B. Shuai, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [44] M. H. Sazli, "A brief review of feed-forward neural networks," *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, vol. 50, 2006.
- [45] Z. Dai and R. Heckel, "Channel normalization in convolutional neural network avoids vanishing gradients," 2019, <https://arxiv.org/abs/1907.09539>.
- [46] Q. Zhao, S. Lyu, B. Zhang, and W. Feng, "Multiactivation pooling method in convolutional neural networks for image recognition," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8196906, 15 pages, 2018.
- [47] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, <https://arxiv.org/abs/1803.08375>.
- [48] Z. Qiumei, T. Dan, and W. Fenghua, "Improved convolutional neural network based on fast exponentially linear unit activation function," *IEEE Access*, vol. 7, pp. 151359–151367, 2019.
- [49] L. Pillaud-Vivien, A. Rudi, and F. Bach, "Statistical optimality of stochastic gradient descent on hard learning problems through multiple passes," pp. 8114–8124, 2018, <https://arxiv.org/abs/1805.10074>.
- [50] C. Goutte and E. Gaussier, *Ch10_Witnesses[8463].Pdf*, *Advances in Intelligent Systems and Computing*, Springer, 2005.
- [51] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "AI 2006: advances in artificial intelligence," in *Australasian joint conference on artificial intelligence*, pp. 1015–1021, Berlin, Heidelberg, December 4–8, 2006.

Research Article

Multistage Optimization Using a Modified Gaussian Mixture Model in Sperm Motility Tracking

Mohammed Alameri,¹ Khairunnisa Hasikin ¹, Nahrizul Adib Kadri ¹,
Nashrul Fazli Mohd Nasir ^{2,3}, Prabu Mohandas ⁴, Jerline Sheeba Anni ⁵,
and Muhammad Mokhzaini Azizan ⁶

¹Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia

²Biomedical Electronic Engineering Program, Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia

³Sport Engineering Research Centre (SERC), Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia

⁴Department of Computer Science and Engineering, National Institute of Technology Calicut, Kerala, India

⁵Department of Computer Science and Engineering, MEA Engineering College, Kerala, India

⁶Department of Electrical and Electronic Engineering, Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

Correspondence should be addressed to Muhammad Mokhzaini Azizan; mokhzainiazizan@usim.edu.my

Received 8 May 2021; Revised 24 June 2021; Accepted 12 August 2021; Published 30 August 2021

Academic Editor: Huiling Chen

Copyright © 2021 Mohammed Alameri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Infertility is a condition whereby pregnancy does not occur despite having unprotected sexual intercourse for at least one year. The main reason could originate from either the male or the female, and sometimes, both contribute to the fertility disorder. For the male, sperm disorder was found to be the most common reason for infertility. In this paper, we proposed male infertility analysis based on automated sperm motility tracking. The proposed method worked in multistages, where the first stage focused on the sperm detection process using an improved Gaussian Mixture Model. A new optimization protocol was proposed to accurately detect the motile sperms prior to the sperm tracking process. Since the optimization protocol was imposed in the proposed system, the sperm tracking and velocity estimation processes are improved. The proposed method attained the highest average accuracy, sensitivity, and specificity of 92.3%, 96.3%, and 72.4%, respectively, when tested on 10 different samples. Our proposed method depicted better sperm detection quality when qualitatively observed as compared to other state-of-the-art techniques.

1. Introduction

Infertility is a medical condition that is represented by the failure of the reproduction system to produce children. Several factors lead to this condition such as genetic disorder, HIV, diabetes, cancer, or overexposure to certain environmental factors [1]. Globally, there are approximately between 48.5 and 52 million reported cases with infertility cases which is estimated to be around 15% of the total couples around the world [2]. In Malaysia, the Government

Department of Statistics reported a fertility rate drop from 4.9 to 1.8 babies per woman between 1970 and 2018. The infertility issue originates from either the man or the woman or in some cases; both couples contribute to the problems [3, 4]. In most male infertility cases, sperm disorder is considered to be the most common cause of the infertility [4]. Males alone are responsible for 20–30% of the infertility disorder, and they contribute around 50% of the overall causes [5].

There are several ways to diagnose the infertility issue among men such as constant measurement of blood and

TABLE 1: WHO standard for normal sperm characteristics.

Sperm property	WHO standard for normal sperm property
Sperm count	10 million sperms per milliliter. Head: oval in shape, 3–5 μm long and 2–3 μm wide
Sperm morphology	Middle piece: less than 1 μm in width, 5–7.5 long, and must be uniform and visible Tail: 45 μm in length, uniform, and visible
Sperm motility	Fast progressive: velocity more than 25 $\mu\text{m}/\text{s}$ Slow progressive: velocity less than 25 $\mu\text{m}/\text{s}$ Immotile: velocity is equal to zero

hormone levels, physical exams, and pH level [6]. However, these approaches rely on manual assessments which are prone to human error, costly, and time-consuming. Another approach is through semen analysis which relies on examining and evaluating the properties of the sperm, and it has been widely explored. Generally, there are three ways to analyze the sperm, namely, (1) sperm count—which concerns about the concentration of the sperms within a sample, (2) sperm morphology—which depends on examining the shape and the size of the sperm’s structure, and (3) sperm motility—which estimates the velocity of the sperm. These three properties provide clues to detect infertility and can indicate the overall quality of the sperm [7, 8]. The World Health Organization (WHO) has established a standard to identify the normal and the abnormal sperm based on these three criteria as tabulated in Table 1 [9].

Sperm analysis is a powerful tool to examine the fertility disorder, and Table 1 tabulates the cut-off values for the sperm parameters. In other words, if the obtained value for the analyzed case is less than the cut-off, then, the case is considered abnormal. Approximately two out of six men in the world are having fertility issues, and around 30% of them encounter problems associated with sperm quality. There are several ways to investigate the sperm properties such as sperms’ DNA fragmentation and sperm chromatin structure assay. However, these methods are time-consuming and costly and need to be handled manually. Therefore, new methods have been introduced based on computer vision which includes the modification of image processing techniques to analyze sperm’s properties. The use of computer vision in the sperm analysis has provided several advantages such as the following: (i) it provides a rapid diagnosis, (ii) it can be modified, (iii) it does not require other chemical condition, and (iv) it is less prone to human error.

This paper is divided into 5 main sections, of which each of them describes the work undertaken for the completion of this research. In Introduction, this paper explores the general understanding of infertility from a global perspective, especially on worldly accepted WHO standard. The cause behind infertility is also looked into, in order to understand the background as well the available diagnostic processes related to the issue. In Research Background, gap analysis was performed through comparisons between currently available methodologies and techniques. Critical discussions on the latest documents on sperm tracking methodologies with their respective ups and downsides are discussed. Find-

ings inside this section provided the motivation behind this manuscript and the way modified Gaussian Mixture Modelling suit the founded gap. Within Methodology, this manuscript discussed the modified Gaussian Mixture Modelling (GMM) which includes the introduction of optimization of sperm detection and the subsequent sperm motility tracking which now uses the input of the optimized detection techniques. Based on the work done in this section, the results are discussed on three main parameters that are accuracy (A), sensitivity (Sn), and specificity (Sp). In Results, the outcome of the proposed modified GMM is benchmarked with a few known, similar techniques to evaluate its performance. Several concluding remarks were performed in Conclusion, where the performance of the proposed algorithm is described to highlight strength and challenges of the work.

2. Research Background

Although sperm count and morphology can give some indications about the statue of the sperms, they are considered nonreliable to draw a decisive conclusion on the fertility on the male in comparison to sperm motility analysis [10]. Since the main concern in fertility analysis is the first sperm that reaches the fallopian tube and fertilizes the egg, therefore, studying the sperm’s velocity properties could provide more accurate results. According to the WHO, sperm’s motility can be categorized in three terms: (i) fast progressive (FP) which means the sperm motility is with velocity $\geq 25 \mu\text{m}/\text{s}$, (ii) slow progressive (SP) which means the sperm motility is with velocity $< 25 \mu\text{m}/\text{s}$, and (iii) immotile which means the sperm does not move at all [4, 9, 11].

There are multiple ways to estimate the velocity of the sperm, which are either through manual visual observation or through using computerised image processing tools. The examples of the manual approaches are photoelectric and multiple exposure phonograph (MEP) [4, 12]. Through manual visual observation, sperm velocity estimation is relatively slow, insufficient, and less accurate. On the other hand, semen analysis with image processing techniques uses algorithms of the computer visions to estimate the sperm’s motility features which are widely acceptable, reliable, fast, and robust [13]. The capability of image processing and computer vision can be seen in various medical field applications [14–18].

Computer-Aided Sperm Analysis (CASA) was the first method and was introduced in 1985. This method is capable to estimate the total number of sperms, detects the sperm’s morphological features, and estimates the velocity of the sperm. The sperm analysis results may vary among different CASA instruments. It is highly dependent on the sperm concentration where high concentration of semen samples may cause motile sperms to be hardly detected. Despite that, this method fails to detect multiple sperms at the same time. Furthermore, it requires big samples for each experiment, and it is not capable of detecting the loss or reborn sperms [19]. The CASA instrument settings should be standardized to ensure the results are comparable where the error trapping rate and the loss trapping rate should be as low as possible.

TABLE 2: Summary of the previous methods in multiple sperm tracking.

Techniques	Advantages	Disadvantages
CASA [28, 29]	(i) Detect the sperm's features	(i) Requires a big sample (ii) Fails in sperm collisions and proximity
JPDAF [20, 21, 22]	(i) Adaptive learning (ii) Works well with sperm collisions	(i) Computational complexity (ii) Slow process termination
HGDT [23]	(i) Combination of object segmentation and tracking (ii) Self-learning capability	(i) Fails in multiple sperm tracking.
CSR-DCF [30]	(i) Implementing of training and testing data for tracking and features extraction (ii) Usage of sperm's features in tracking	(i) Requires large training data (ii) High computational complexity (iii) Difficult to implement
ADT [26, 31]	(i) Fast execution process (ii) Multisperm tracking	(i) Its efficiency decreases in a high dense medium. (ii) It works poorly when many sperms accumulate in a small space
DAT [27]	(i) Easy to implement (ii) Less mathematical complication	(i) Does not predict the dead and newborn sperms (ii) It does not require a testing data (frames)

Joint Probability Data Association Filter (JPDAF) was adapted to track and measure the velocity of hundreds of sperms simultaneously. JPDAF is a mathematical and statistical approach that works based on an approach Global Nearest Neighbour (GNN) which tracks an object by minimizing within distance variable and maximizing the between distance variables. This feature eliminates sperm's collision conditions that intervene the results, and it measures the proximity of close sperms. However, this method has high computational complexity, and its process termination is slower [20–22].

Another modified approach in sperm tracking is the Hybrid Generative-Discriminative Tracker (HGDT). Hybrid Generative system is for object detection, and Discriminative Tracker is for updating the centroid of each sperm for all frames for velocity estimation. This method has features of adaptive learning from frame to another, which enhances the detection and tracking procedure. The main drawback of HGDT that it works properly with single tracker rather than multiple sperm tracking [23, 24].

A recent paper was published earlier this year in which deep learning in sperm detection and tracking was used. A modified method called Channel and Spatial Resolution together with Discriminative Correlation Filter (CSR-DCF) was implemented. It begins with detection in which the RetinaNet method, a deep neural convolutional network approach, was used. This method extracts the main features of the sperm by feeding the Convolutional Neural Network. According to these features, it detects the rest of the sperms for the whole frames. This method requires a higher computational model and large training data set, which makes it difficult to implement and has slow execution time [25].

Khairkhan et al. [26] suggested a method for sperm detection and tracking that is based on Adaptive Distance Tracking (ADT). In this method, the sperm is located by detecting its centroid, and then, its coordinates are determined in each frame. Adaptive Window Average Speed (AWAS) is modified to enhance the accuracy of sperm allocation from frame to another. However, this method works poorly in a highly dense solution or when there are many sperms accumulated in a close region.

Another sperm tracking method was suggested by [27] called Differential Area Trajectory (DAT). In this method, the sperms' location is detected collectively, and at the same time, the footprint is constructed. DAT retains the final results as a velocity function with respect to time in order to estimate the velocity of the sperms. Although this method has a low computational complexity and it is easy to implement, it does not detect the dead and reborn sperms and it does not require a testing data (initial frames) which makes the results less accurate. Table 2 summarizes the previous techniques that have been used in tracking the motile sperms. From the summary presented in this table, we can highlight few research gaps from the existing literatures:

- (1) The updating mechanisms to consider newly emerging sperms reenter the field of views are disregarded in sperm's velocity estimation. Thus, the classification between normal and abnormality of the sperm motility is inaccurate
- (2) The efficiency of sperm motility analysis is often decreased in high density medium where the existing system fails to detect multiple sperms
- (3) Due to this, manual semen analyses are still being used. However, manual analysis has been challenging due to inconsistency of the semen sample. Reproducibility of the analysis is difficult to be achieved which will lead to misclassification

Therefore, in this paper, an automated multistage tracking system is proposed using sperm motility properties. The automated tracking comprises two stages of sperms' head detection and sperm tracking processes. A new modified Gaussian Mixture Model with an automated optimization protocol was proposed to detect multiple sperms head in image frames extracted from sperm motility video. Then, the output from stage 1 will then be utilized to accurately track the motile sperms and eventually determine their velocity.

The main motivation behind this method is to enhance the quality of sperm analysis by providing high accurate

TABLE 3: Summary of the previous methods in multiple sperm tracking.

Specification	Specification value/range
Microscope	Olympus CX31
Video camera	Mounted microscope camera UEye UI-2210C
Magnification	×400
Video format	.AVI
Video resolution	640x480
Video duration	2-7 minutes
Frame rate	50 frames/sec
Participants	85 male participants aged 18 years or older

and reliable results. The main technical contributions and novelty of this work are listed below:

- (1) A modified and optimized GMM function for sperm's head recognition is proposed. With the optimization of determining background threshold, more accurate sperm features can be localized. The proposed modified GMM will be able to localize the sperm's features to be tracked in the second stage of the proposed system
- (2) Multisperm tracking uses similarity and Euclidean distance methods. The detection of the motile sperms will be updated automatically including the new sperms reentering the field of view. This will ensure accurate sperm's velocity can be determined and calculated
- (3) Provide overall sperm analysis system with less computational and mathematical complexity. The proposed modified GMM has provided an improved sperm features that will generate automated threshold to localize the region of interest (i.e., sperm morphology)

3. Methodology

In this paper, we developed the new algorithm based on the available VISEM Dataset [32] where the collected data were used to investigate male reproductive function. The sperm motility videos were recorded and examined under a 400 times magnification using an Olympus CX31 microscope. The sample was placed on a heated microscope stage (37°C), and the videos were captured by a microscope-mounted camera and saved as AVI file. The dataset contains motility videos from 85 male participants aged 18 years or older. Summary of the dataset specification is tabulated in Table 3.

The overall algorithm for multistage automated sperm motility tracking is presented in Figure 1. The proposed algorithm consists of two stages where the first stage focuses on the preprocessing task where debris are removed from the original frame. In order to accurately detect the moving sperms, unwanted debris are eliminated since beginning. This approach will help improving computational efficiency

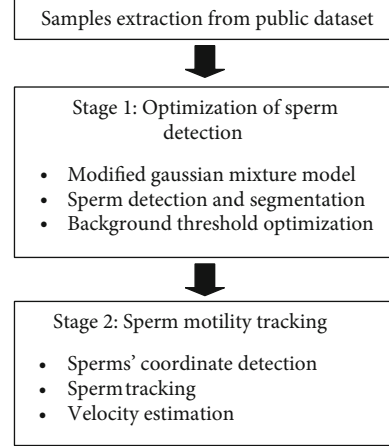


FIGURE 1: Overall process of the proposed method.

of the algorithm in tracing the moving sperms (which will be conducted on the second stage). Then, the cleaned image frame was fed into stage 2 where the center coordinates of the sperms' head were detected. These coordinates were then used for sperm tracking in the consecutive frames. The velocity of the sperm was calculated based on the successful tracked sperm, and thus, classification of normal and abnormal was made according to the recorded average velocity of the motile sperms.

3.1. Stage 1: Optimization of Sperm Detection. This stage started with the sperm detection process with the aim to distinguish object of interest (i.e., the sperm) from the background. In addition, the image frames were cleaned by implementing morphological opening and closing operation to remove dead cells or debris in each image frame. Then, a modified Gaussian Mixture Model (GMM) with a new optimization of Gaussian density function is implemented to detect the sperms and separate them from the background. This technique is clustering-based method that assigns each object to a certain class using the mathematical probability of Gaussian density function, $(\mathcal{N}(x | \mu, \Sigma))$ to segment the objects of interest from the background as shown in

$$\mathcal{N}(x | \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right), \quad (1)$$

where x is the pixel intensity value, μ is the mean, Σ is the covariance matrices, and D is the dimension of the matrix.

The pseudocode of the modified GMM is shown in Figure 2. During this stage, all objects are clustered into two different classes, and every class has a different mean μ to estimate its center, covariance Σ that indicates its length, and weight probability π that describe the how large or small is the class. If the Gaussian density function $(\mathcal{N}(x | \mu, \Sigma))$ is bigger than the set threshold, the pixel is considered the object whereas if the $(\mathcal{N}(x | \mu, \Sigma))$ is smaller than the predetermined background threshold (BG_T), the pixel is considered as the background. To ensure accurate and robust

K : Number of probability distribution
 π : Weight of probability distribution
 μ : Mean of probability distribution
 Σ : covariance of probability distribution
 m : Number of detected sperms

Parameter initialization, π, μ, Σ, m

```

for t= 1:K
    i=1;
    while i ≤ 10
         $BG_T(i) = \frac{i}{10}$ 
         $N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}$   $\exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu))$  %generating gaussian density function
        if  $N \geq BG_T(i)$ 
             $x = 1$ ; % object (moving sperms)
        else
             $x = 0$ ; % background (non-moving sperms)
        end
         $m|x(i) =$  number of detected white pixel
         $i = i + 1$ ;
    end
     $BG_T(\text{final}) =$  maximum ( $m|x(i)$ )
     $N(x|\mu, \Sigma)(\text{final}) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}}$   $\exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu))$ 
end
    
```

Output: $BG_T(\text{final}), x=1, \pi = \{\pi_1, \dots, \pi_K\}, \mu = \{\mu_1, \dots, \mu_K\}, N = \{N_1, \dots, N_K\}$

FIGURE 2: Pseudocode of modified Gaussian Mixture Model.

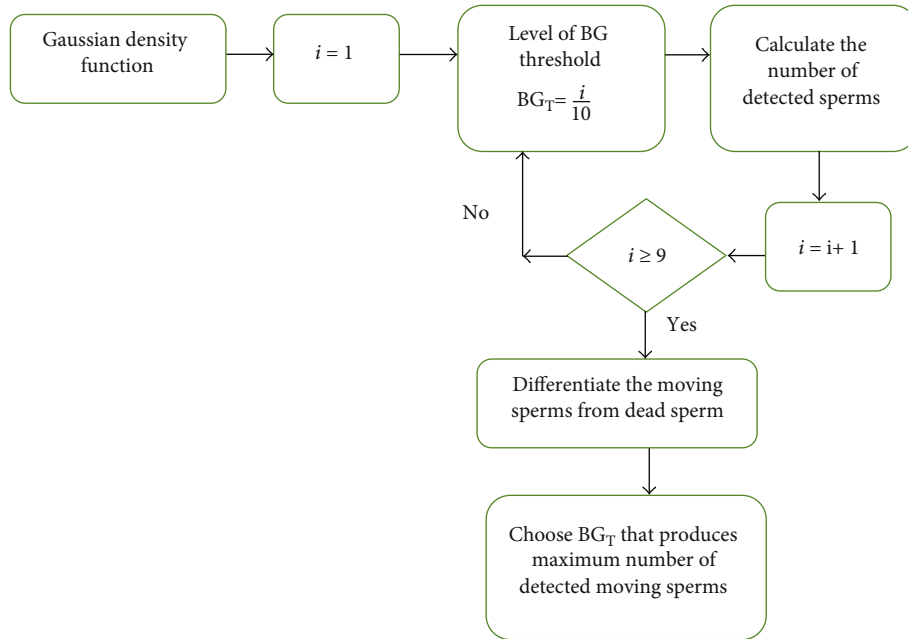


FIGURE 3: Optimization of background threshold.

detection of the sperm, the set threshold was automatically generated based on the proposed optimization algorithm as shown in Figure 3. The BG_T is automatically generated based on the sperm sample. The Gaussian density function is calculated, and the process is repeated to attain the highest detected motile sperms.

To obtain the best value of BG_T , we used the number of detected sperm as the main parameter to check the accuracy of the sperm detection. Since each sperm motility video is recorded in the same environment for each sample, we only computed the optimization procedure on the first frame of every acquired video. This process will ensure efficient

parameter optimization and thus ensuring accurate detection. Therefore, for the highest number of detected sperm attained, the BG_T , the value is chosen as the best value. The optimized sperm detected video based on the automated computation of BG_T is produced by conducting the proposed optimization technique as shown in Figure 3. The BG_T values were set to have ranged from 10% to 100%; therefore, the loop of $i \geq 9$ was applied to ensure automated generation of the threshold values.

The resulted images from GMM are not clear enough to do further analysis. The reason is that there are a lot of dead cell and debris that were mixed or surrounded the edges of the sperm. This will lead to misinterpretation on the results and inaccurate velocity measurement. Therefore, morphological process of open and close operations was implemented to reduce the unwanted noise. This technique is based on either dilating or contracting the shape of the object in the background. In this study, image opening and image closing were applied simultaneously to get a clearer image frame. The size of added or removed pixels is represented by the size of the kernel or structural element (SE). Equations (2) and (3) demonstrate the mathematical expression of the image opening and closing, respectively.

$$A \circ B = (A \ominus B) \oplus B, \quad (2)$$

$$A \circ B = (A \oplus B) \ominus B. \quad (3)$$

where A is the binary image (results of the modified GMM) and B is SE element.

3.2. Stage 2: Sperm Motility Tracking. Once the optimized detected sperms were found in stage 1, the detection and tracking processes were continued during which the centers of the sperms' heads were identified. Objects that are elliptical and have area of 150 pixel² were considered the object of interest (i.e., sperm). Objects that are less than 150 pixel² are considered debris and were eliminated from the tracking process. The elliptical area is calculated using

$$A_{\text{elliptical}} = \pi * \text{major axes} * \text{minor axes}. \quad (4)$$

The coordinates of the sperms' head were recorded for every 10th frame of the sperm motility video. The mathematical procedure of the Mass of the Central Momentum of object was implemented to detect the centers of the sperms' heads. This method is capable of working on multiobject or sperm detection in the same frame. Using the Mass of the where S_{Z_p} the area of the sperm in the previous frame and S_{Z_c} is the area of the sperm in the current frame. The results of this equation were automatically updated and stored in a matrix. t_{size} is the threshold size that is set to 1.2 in this study

Central Momentum technique, the coordinates of the center of the sperms' head are calculated using

$$x_c = \frac{n_{1,0}}{n_{0,0}}, \quad (5)$$

$$y_c = \frac{n_{0,1}}{n_{0,0}}, \quad (6)$$

$$n_{1,0} = \iint_{-\infty}^{\infty} dx dy x b(x, y), \quad (7)$$

$$n_{0,1} = \iint_{-\infty}^{\infty} dx dy y b(x, y), \quad (8)$$

where the coordinate of the sperm's head center is donated as (x_c, y_c) , $n_{0,0}$ is the zero-order moment that describes the area of the object, and $n_{1,0}$ and $n_{0,1}$ are specified in Equations (6) and (7), respectively. $b(x, y)$ is the binary image that resulted from stage 1.

This stage is aimed at finding the sperm in the previous frame that is closest and most similar to the sperm in the current frame. There are two assumptions that we made in this stage: (i) the distance between two sperms (i.e. sperms in current frame and previous frame) must not exceed the distance threshold of 125 pixels to be considered the same motile sperm, and (ii) the shape of the sperms' heads and their deformity of the elliptical must not change much in each frame. The threshold of 125 pixels was set based on observation and simulation test on all samples. In addition, the threshold was set by considering the size ratio between sperm's head and tail. Thus, by observing the behaviour of the sperm's movements in all samples, we concluded that the threshold of 125 pixels is acceptable to be used. The closeness of the sperm was defined by having a minimum Euclidean distance between centers of two sperms of S_c (i.e. sperm in the current frame) and S_p (sperm in the previous frame) given by

$$d(S_c, S_p) = \sqrt{\sum_{i=1}^2 (x_c - x_p)^2 + (y_c - y_p)^2} \leq t_{\text{dis}}. \quad (9)$$

Therefore, we measured the ratio of the sperm's head area from the current (for example, frame-20th) and the previous frames (for example, frame 10th). The ratio should not exceed the threshold which was defined as the positive differences between the area sizes in the current and the previous frames. Equation (10) denotes the mathematical definition of the procedure for tracking process.

with the assumption that the object's size is not expected to change significantly between antecedent and subsequent frames. The threshold was set based on the simulation that was done on 10 different semen samples.

$$\text{Similarity} = \begin{cases} 1, & \text{if } S_{Z_p} > S_{Z_c} \text{ and } \frac{S_{Z_p}}{S_{Z_c}} \leq t_{\text{size}} \text{ or } S_{Z_p} < S_{Z_c} \text{ and } \frac{S_{Z_c}}{S_{Z_p}} \leq t_{\text{size}}, \\ 0, & \end{cases} \quad (10)$$

Based on two conditions as mentioned above, sperms in the current frame were matched with their corresponding frame from the previous frame. If the Euclidean distance and size ratio between sperm in the current and previous frames were less than or equal to the predetermined thresholds, sperm in the previous frame was matched to its corresponding sperm with the sperm in the current frame.

If either closeness or similarity of the object was not achieved, the object was considered the sperm that swims out from the frame. The sperm tracker was reinitialized where the nonmatching sperm was removed from the tracker. If the sperm was successfully matched with its corresponding sperm, the process was repeated until all sperms in the frame were processed. This process was continuously conducted until all image sequence frames were analyzed. Since SE was implemented in the image optimization, the selection of SE does not directly affect the tracking and the detection process. However, the SE was used for morphological operation to improve the clarity of the image and further amplify the sperm's head size, thus indirectly influencing the detection process.

After obtaining the coordinate of the sperms, the sperm's velocity can be measured according to Equation (11). In addition, the unit of the velocity was converted from pixel/second to $\mu\text{m/s}$ (1 pixel = 0.0002 meter) to meet the standard that was stated by the WHO. The centroid was detected once over ten frames so that the variation of the centroid will be large enough for detection to make the system less computational-complex and to increase the processing speed.

$$\text{Velocity} = \frac{\sqrt{(x_c - x_p)^2 + (y_c - y_p)^2}}{\text{frame rates}}. \quad (11)$$

The frame rate depends on the properties of the video, and its unit is frame per second.

The final velocity result is the average velocity of each sperm in the whole video. The results of sperm's velocity in each detection were stored in a matrix. Then, at the end of the video, the average was obtained by dividing the summed velocity values in all frames by the number of the frames. This gave the system the feature of detecting the sperms the exited and entered the study area (i.e., field of view) which resulted in higher accuracy.

3.3. Performance Evaluation. In order to ensure that the system is more accurate as compared with other previous works, three mathematical equations were implemented such as accuracy (A), sensitivity (Sn), and specificity (Sp) as shown in Equations (12), (13), and (14), respectively. There are four main elements integrated in these equations, namely, (i) True Positive (TP) which indicates the number of positive sperm that are correctly recognized by the proposed technique, (ii) True Negative (TN) which indicates the negative sperm samples that are positively recognized by the proposed method, (iii) False Positive (FP) which indicates the number of sperms in which the expert tagged them as negative and the system recognizes them as positive, and

(iv) False Negative (FN) which indicates the number of sperm in which the expert tagged them as positive and the proposed algorithm recognizes them as negative.

$$\text{Accuracy (A)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (12)$$

$$\text{Sensitivity (Sn)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (13)$$

$$\text{Specificity (Sp)} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (14)$$

4. Results

Our results were compared with the other methods such as zero-crossing [33] and classic edge segmentation (Sobel and Prewitt) [34] to evaluate its performance. Figure 4 shows the comparison results to evaluate the differences between the proposed modified GMM and the other two methods. Classical edge segmentation and zero-crossing techniques included the whole sperm structure (i.e., head, midpiece, and tail) and considered motile and immotile sperms. These approaches add unnecessary noises to the background as compared to our approach. Since the head of the sperm is the brightest part of other sperm's structure as shown in Figure 4(a), our current technique focuses on the head area only to ensure accurate and fast detection. The most important capability that has been noticed in our experiment is that the modified GMM managed to detect all moving object or sperms. Since the main concern in the motility measure is the moving sperms, then, the other undetected sperms are considered immotile.

Moreover, the most important capability of the modified GMM is BG_T manipulation. In other words, the background ratio can be modified to get the desired results. The BG_T is uniquely assigned based on the sperm video samples. The proposed algorithm able to automatically generate the BG_T to ensure robust motile sperm detection. Figure 5 shows the results of three different background ratios, and for the tested sample, the 50% ratio shows a more accurate, clear image with smooth edges of the moving sperms. However, this percentage is not fixed for all samples.

In terms of the detection accuracy, we also conducted a comparison analysis (performance evaluation) with other techniques such as adaptive threshold which was reported by [35] and global threshold which was reported by [36]. We evaluated 10 different sperm motility video samples whereby each sample consists of different number of sperms. The evaluation is conducted by calculating the accuracy (A), sensitivity (Sn), and specificity (Sp) as tabulated in Table 4.

In addition, the overall performance of the proposed method in comparison with the adaptive and global thresholds is presented in Table 5. As we can notice from Table 5, the proposed method attained high accuracy, sensitivity values. Although specificity of the proposed method was not consistent for each sample, this drawback could be improved if the immotile sperms were completely removed from the frame before the tracking process. In addition, observing the quality of the resulted image was not as good

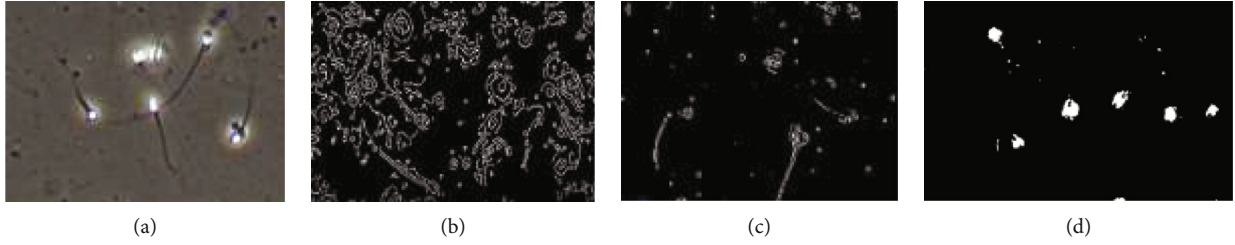


FIGURE 4: Comparison of sperm detection between proposed method and other method in (a) original frame, (b) classical edge segmentation, (c) zero-crossing, and (d) the proposed method.

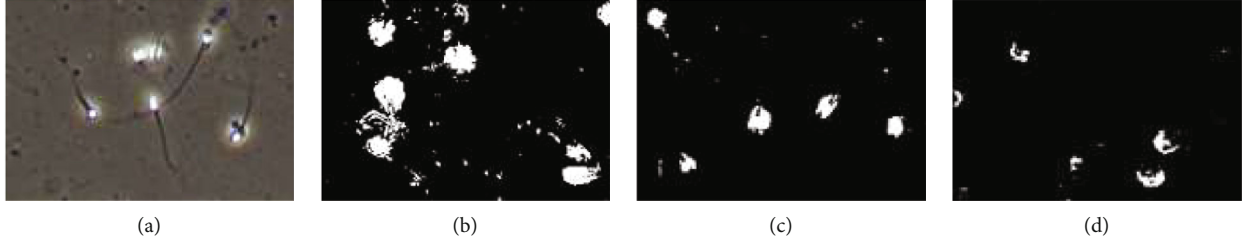


FIGURE 5: Comparison of different BG ratio size in (a) original frame, (b) segmented with 90% BG ratio, (c) segmented with 50% BG ratio, and (d) segmented with 20% BG ratio.

TABLE 4: Sperm detection results of proposed method, adaptive thresholding and global thresholding method.

Sample	No. of motile sperms	No. of immotile sperms	Proposed method			Adaptive threshold			Global threshold		
			A (%)	Sn (%)	Sp (%)	A (%)	Sn (%)	Sp (%)	A (%)	Sn (%)	Sp (%)
1	11	51	91.0	90.0	100	81.3	75.0	81.0	40.0	33.3	46.2
2	9	19	98.0	97.0	97.0	77.8	86.4	79.2	41.2	22.2	62.5
3	21	12	91.3	95.0	98.0	84.6	90.5	80.6	25.8	40.0	40.0
4	19	16	95.0	98.3	66.0	67.5	81.0	72.0	25.0	37.5	20.0
5	16	29	94.4	93.0	96.7	73.8	75.0	72.7	28.1	18.2	33.3
6	9	16	90.0	90.0	50.0	65.7	75.0	60.9	50.0	60.0	45.5
7	6	19	91.2	100	50.0	53.8	42.9	60.0	30.8	33.3	30.0
8	10	15	91.7	100	50.0	65.4	58.8	72.2	37.5	50.0	30.2
9	11	12	92.8	100	66.6	65.7	58.8	72.6	52.2	50.0	53.3
10	7	16	87.5	100	50.0	55.2	58.3	52.9	38.9	25.0	50.0

as it should be, because the structure of the sperms is still nonuniformed, and some debris and dead cells are still in the foreground as shown in Figure 6. This will affect the analysis and will result in inaccurate results. However, as tabulated in Table 5, in comparison with the adaptive threshold and global threshold, both techniques are still detecting the immotile sperms. This has caused inaccuracy in localizing moving sperms and thus can cause miscalculation in determining sperm velocity. The proposed method can discriminate between moving and nonmoving sperms through an optimization procedure in modified GMM. However, due to the additional optimization steps imposed in the proposed method, the processing time is higher as compared to the two nonoptimized approaches of adaptive and global thresholds.

Figure 7 shows the error bar for the standard deviation of the results obtained in Table 4 (performance evaluation). As we can see from the graph, our method has the smallest error bar length (specifically the accuracy and the sensitivity)

which indicates that these values are concentrated, and the plotted average is more likely as compared to the adaptive and global threshold methods. The standard deviation of specificity of the proposed method is higher as compared to other techniques. This could be due to the 10 sperm samples that we used that have different illumination problems, and thus, each sample has different difficulties in determining a localized area. Apart from that, unlike adaptive and global threshold methods, the accuracy and the sensitivity scored the highest points in our method, and they were not overlapped with other methods; therefore, we can tell that our proposed technique is conclusive.

There are two reasons for this to happen: (1) when the number of the detected objects (i.e., sperms) is relatively small; and (2) when we go back to Equation (14), we can tell that the main variable for this equation is the number of detected TP and TN. For instance, samples 6 to 8 have 50% Sp, that is because the number of detected TP is 1 and the number of detected TN is 1 as well, so that makes the

TABLE 5: Comparison of overall performance between proposed method, adaptive threshold, and global threshold.

Criteria	Proposed method	Adaptive threshold	Global threshold
Accuracy (average \pm std)	92.3 \pm 2.78	69.1 \pm 9.76	37.1 \pm 9.04
Sensitivity (average \pm std)	96.3 \pm 3.89	70.2 \pm 14.17	37.2 \pm 12.72
Specificity (average \pm std)	72.4 \pm 21.66	70.4 \pm 9.02	41.1 \pm 12.17
*Processing time (second)	>1 min	<1 min	<0.5 min
Detected sperm features	Head, midpiece, tail	Head, midpiece, tail	Head, midpiece, tail
Immotile sperms	Not detected	Detected	Detected
Advantage	Able to locate only moving sperms due to optimization algorithm in modified GMM	Accurately detect whole sperm body in less than 1 minute	Accurately detect whole sperm body in less than 0.5 min
Disadvantage	Longer processing time	Unnecessary objects such as immotile sperms and debris are considered, thus causing inaccurate tracking	Unnecessary objects such as immotile sperms and debris are considered, thus causing inaccurate tracking

*Processing time is referred to the time required for the method to accurately track the sperms in the consecutive 50 frames. 50 frames were selected for consistency in comparison.



FIGURE 6: Debris and nonuniform sperm structure that resulted from the proposed method.

percentage of Sp low which is reflected on the length of its error bar. To solve the aforementioned issues, morphological operation was employed in which the image opening and closing were simultaneously applied to improve the nonuniform sperms. Since there are several structural elements (SE) that can be applied, some SE were examined to find the best SE that can provide a clear and homogenous structure and images. Figure 8 shows the result of applying different SE to the resulted image frames (from the modified GMM). As we can see from the figure, the results of the morphological images with Line and Cube SE do not provide a uniform sperm's structure as compared with the Disk SE.

The morphological operation in commonly used sperm detection analysis is reported in [37]. They proposed a method that is based on the morphological structure of the sperm using ellipse detection. In this method, gray color is converted to RGB; then, multiple enhanced filters were added to enhance and segment the sperm from the background. However, in our proposed method, the modified GMM is able to convert image class, segmentation, and edge detection and filter the sperm at the same without adding further instructions. This made our proposed method less complex with faster execution. Additionally, the proposed system will detect motile and immotile sperms which could add further noise in velocity measurements. According to Mahdavi et al. [37], the system detects shapes with an ellipse structure only which is not reliable, because they could be

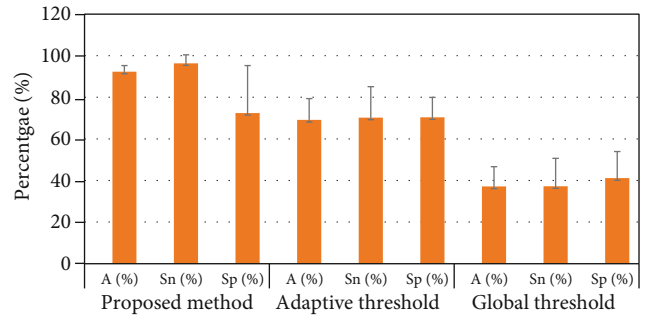


FIGURE 7: Comparison of the performance evaluation between the proposed method and the other common techniques of adaptive and global thresholds. The value presented here is in term of average \pm standard deviation.

other dead cells or debris in the solution that has an ellipse shape.

Sperms' coordinate in the current, and the previous frames are needed to calculate the velocity of the sperm. The similarity index was proposed in this study with a certain threshold. The results of the tracked coordinates are updated and stored automatically in a matrix once each 10 frames. The results of tracking are shown in Figure 9.

In most cases of the sperm tracking, there will be instances of sperm exiting the frame/field of view or sperm entering the frame as shown in Figure 9. These sperms will be automatically updated and do not intervene with the results of the velocity. The sperm exiting could result from several reasons, namely, (i) the sperm losing its kinetic energy and dies, (ii) error in tacking, (i.e.) the similarity index exceeding the threshold, and (iii) the sperms that ran out of the scope of the field of view during the microscopic recording. Figure 10 shows one of the cases of the sperm that left the field of view between the 10th and 20th frames.

Additionally, sperm entering the field of view or the frame is another condition that could occur. This is because either the lost sperm in the previous frame has been updated or it gains a less threshold of similarity index. Figure 11

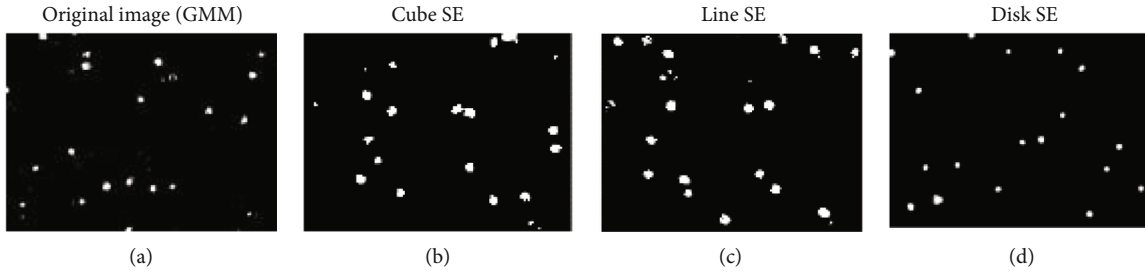


FIGURE 8: Comparison of different SE for image morphology.

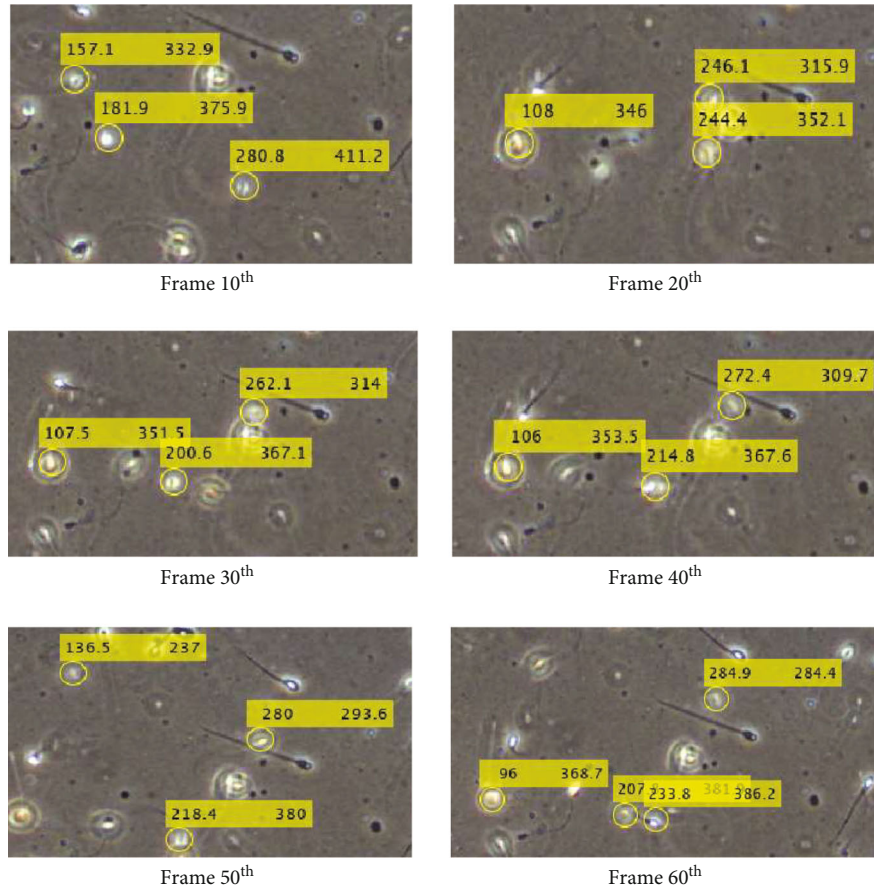


FIGURE 9: Results of sperm coordinate tracking.

shows three additional sperms that were added between the 20th and 30th frames.

Finally, for the sperm motility analysis results, there are ten samples with similar frame rates of 30 frames per second and an average of 800 frames. The number of sperms is different from that of another sample. Table 5 shows the results of motility measurement. The classification of the fast progressive (FP), slow progressive (SP), and immotile sperms was done according to the WHO standard of sperm motility analysis as shown in Table 1. The motility results were expressed in terms of average velocity ($\mu\text{m/s}$). As it can be noticed from Table 6, some samples had smaller deviations and others had larger standard deviations. This condition is anticipated because the standard range of either FP or

SP is quit wide, due to the different number of sperms in each sample. Additionally, in terms of classification, a sample is considered normal if the percentage of the sum of fast progressive (FP) sperm and the slow progressive (SP) sperm exceeds 40% from the total sperm (immotile + FP + SP).

5. Discussions

To understand and verify the results obtained from the previous section, trends of sperm's velocity are plotted in Figure 12 where their velocities have been classified into the WHO standards of either fast progressive, slow progressive, or immotile. Each sample represented a different number of sperms, and from the figure, samples 3, 4, 8, and 9

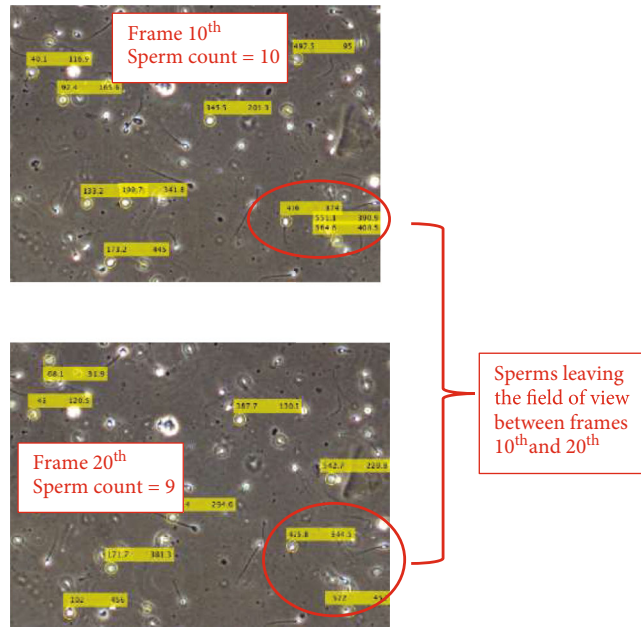


FIGURE 10: Illustration of sperm leaving the field of view condition.

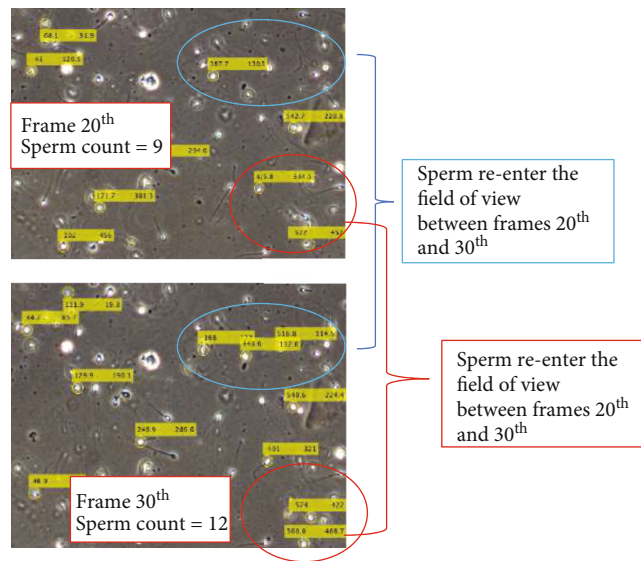


FIGURE 11: An example of sperm reentering the image frame.

are considered a normal sample since the fast progressive and slow progressive sperms exceed 40% from the total sperms. The proposed system is capable of automatically classifying the sperm samples based on the calculated sperm's velocity. The proposed optimization technique will ensure sperms are accurately identified, and the nonmoving sperms as well as debris are excluded from the analysis. This advantage will provide an improved sperm motility classification which will eventually aid the male infertility diagnosis.

The velocity variations can be justified by many factors. For example, the size of the middle piece of the sperm is critical because it contains mitochondria which provides the energy in forms of adenosine triphosphate (ATP) for the

sperm to live and propagate. The larger the size is, the more fuel it has; therefore, it can move faster. There are studies conducted by [38, 39] that proved the higher potential and the size of middle piece of the sperm results in higher sperm's velocity. The same study also concluded that the length of the sperm plays a role in its velocity and that shows a significant linear relationship between the length of the sperm and its swimming velocity.

Moreover, the number of debris and dead cells surrounding the sperms could lower the speed of the sperm. In general, the existence of the debris near or in the area of sperm could result in many collisions between the moving sperm and other debris and dead cells, and that could reduce the sperm's velocity. In addition to that, the high number of

TABLE 6: Classification of sperm motility results.

Sample	Immotile sperms	Fast progressive sperms			Slow progressive sperms			Classification
		No. of sperms	Average velocity ($\mu\text{m/s}$)	Standard deviation	No. of sperms	Average velocity ($\mu\text{m/s}$)	Standard deviation	
1	51	4	34.5	1.3	7	23.2	4.6	Abnormal
2	19	4	49.5	1.2	5	13.6	7.1	Abnormal
3	12	14	36.9	3.9	7	12.8	4.5	Normal
4	16	10	36.6	4.5	9	4.5	3.05	Normal
5	29	4	27.5	2	12	6.34	4.22	Abnormal
6	16	3	25.9	0.77	6	5.4	2.3	Abnormal
7	19	3	26.2	0.7	3	17.4	2.62	Abnormal
8	15	4	26.2	0.78	6	8.8	5.4	Normal
9	12	4	29.5	4.5	7	4.1	2.54	Normal
10	16	5	27.5	2	2	3.85	1.48	Abnormal

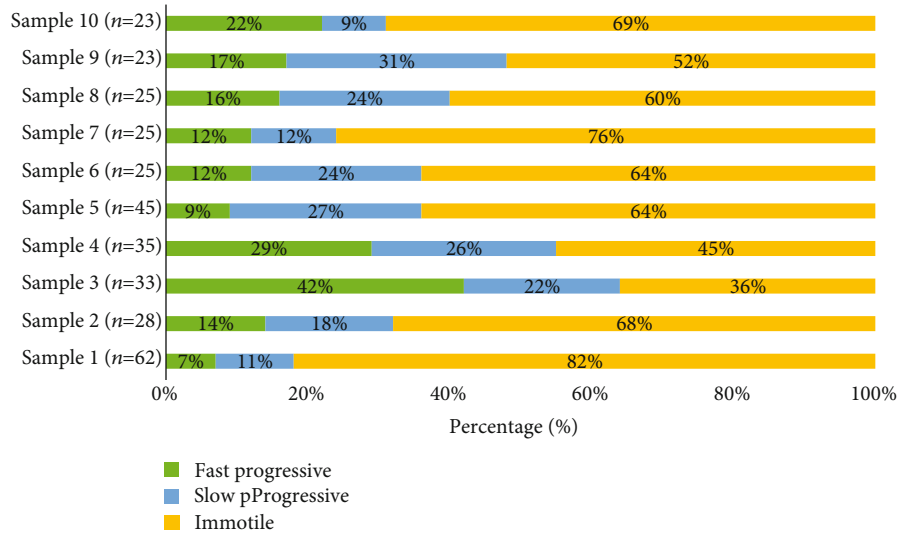


FIGURE 12: Trends of sperm velocity that have been classified for 10 sperm samples.

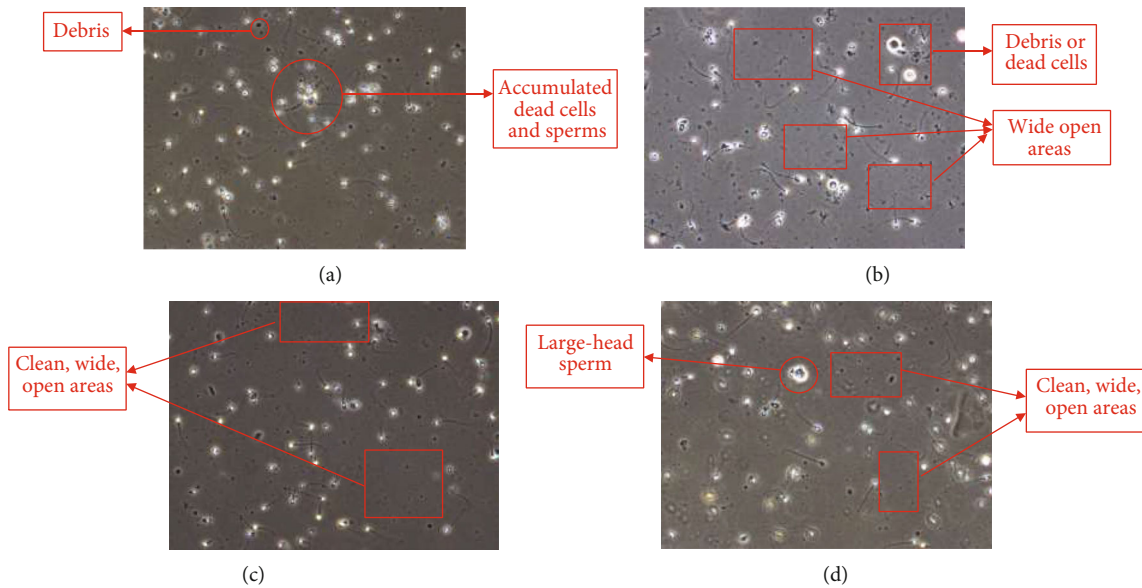


FIGURE 13: Sperm movement scenario for (a) sample 1, (b) sample 2, (c) sample 3, and (d) sample 4.

debris represents high viscosity thus eventually reducing the sperm's velocity. We can observe this incidence by comparing the results obtained in Tables 6 and 4, sample results (frames) as shown in Figure 13, respectively. Sample 1 contains a lot of debris and accumulated dead cell; thus, the sample will have high density which will reduce the speed of the moving objects. In the contrary, sample 3 and sample 4 have many sperms with a high-speed value. The two samples also had less debris and dead cells around the sperms, and that could explain the reason that they have an overall high number of fast progressive sperms.

The number of immotile sperms near the moving sperms may also affect the sperm's velocity. Referring to Table 6, we noticed that the ratio between the numbers of immotile sperms and motile sperms is large in sample 1 and that is an indicator for low velocity value results. On the other hand, sample 3 had less immotile sperm which is why it results in a large number of sperms that have high velocity. This is supported by studies done by [30, 40] where they found out that there is a relationship between the number of dead sperms and overall sperm velocity in a solution; the higher the numbers of dead or immotile sperms, the lesser the velocity of moving sperm. The studies also mentioned that the existence of dead or immotile sperms might result in collisions which slowed down the progression of the sperms. Since the classification of the samples depends on the number of the motile and immotile sperm, the more immotile sperms the sample has, the more likely the sample will be abnormal.

6. Conclusions

In this paper, a modified GMM algorithm with the capability in optimizing sperm detection process was proposed for multisperm tracking for male infertility diagnosis. The proposed method consists of two stages where the first stage focuses on accurate sperm detection while the second stage comprises sperm tracking and velocity measurement. Motility results were evaluated, and 10 sperm samples were presented, and the performance of the proposed method was compared with other state-of-the-art techniques. When tested on 10 sperm motility videos, the proposed method attained 92.3%, 96.3%, and 72.4% in accuracy, sensitivity, and specificity, respectively. These performance indicators ranked the proposed method as the highest as compared to the other methods. Based on the video samples, 4 tested samples were classified as normal and the other 6 turned to be abnormal. The main limitation with the current work is that when two sperms collide and move together in the same line, the system will consider them as one sperm. This problem could be fixed with optimized feature extraction technique where more features of collide sperms will be considered to update the sperm tracker in the tracking process. In addition, despite utilizing Euclidean distance and similarity measures in computing sperm trajectory analysis, feature histogram and foreground information could be added as an indicator to select optimum region for sperm tracking. This will enable object information to be extracted to improve the accuracy of the proposed system. The current state of the

proposed system requires code optimization to improve the computational efficiency especially when dealing with high concentration sample. Furthermore, the research can be continued by considering a mathematical model in estimation sperm motion. The estimation will be able to aid researchers in predicting the location of the corresponding sperms in the next consecutive frames.

Data Availability

The VISEM Multimodal video dataset of human spermatozoa used to support the findings of this study have been deposited in the name of VISEM repository <https://datasets.simula.no/visem/>.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The project was funded by Universiti Malaya research Grant Faculty Program (RF010-2018A) and International Funding of Motorola Solution Foundation (IF014-2019). The authors would like to acknowledge the medical expert's support from Assoc. Prof. Dr. Mahaneem Mohamed, Hospital Universiti Sains Malaysia, Kubang Kerian, Malaysia, in verifying the results.

References

- [1] S. Nawaz, S. Hussain, S. Basit, and W. Ahmad, "First evidence of involvement of *TBC1D25* in causing human male infertility," *European Journal of Medical Genetics*, vol. 64, no. 2, 2021.
- [2] G. Zauner and G. Girardi, "Potential causes of male and female infertility in Qatar," *Journal of Reproductive Immunology*, vol. 141, 2020.
- [3] M. Vander Borgh and C. Wyns, "Fertility and infertility: definition and epidemiology," *Clinical Biochemistry*, vol. 62, pp. 2–10, 2018.
- [4] R. N. Pillai and K. McEleny, "Management of male infertility," *Obstetrics, Gynaecology & Reproductive Medicine*, vol. 31, no. 7, pp. 192–198, 2021.
- [5] M. K. Panner Selvam, R. F. Ambar, A. Agarwal, and R. Henkel, "Etiologies of sperm DNA damage and its impact on male infertility," *Andrologia*, vol. 53, no. 1, 2021.
- [6] K. Hasikin, N. A. Mat Isa, M. Mohamed, S. A. Suandi, and N. A. Kadri, "A new region-based adaptive thresholding for sperm motility segmentation," *Malaysian Journal of Computer Science*, vol. 29, no. 4, pp. 272–286, 2017.
- [7] E. R. S. Roldan, "Assessments of sperm quality integrating morphology, swimming patterns, bioenergetics and cell signaling," *Theriogenology*, vol. 150, pp. 388–395, 2020.
- [8] N. Gatimel, L. Mansoux, J. Moreau, J. Parinaud, and R. D. Léandri, "Continued existence of significant disparities in the technical practices of sperm morphology assessment and the clinical implications: results of a French questionnaire," *Fertility and Sterility*, vol. 107, no. 2, pp. 365–372.e3, 2017.
- [9] W. H. Organization, "WHO laboratory manual for the examination and processing of human semen," 2010.

- [10] C. Wang and R. S. Swerdloff, "Limitations of semen analysis as a test of male fertility and anticipated needs from newer tests," *Fertility and sterility*, vol. 102, no. 6, pp. 1502–1507, 2014.
- [11] A. Fabbrocini, F. Silvestri, V. Vitiello, S. Pelosi, G. Sansone, and R. D'Adamo, "A step towards using the semen of the common pandora *Pagellus erythrinus* for practical applications of cryo-research: the computer-assessed sperm motility pattern in long-term cryostored samples," *Aquaculture*, vol. 528, 2020.
- [12] C. R. Moraes, L. E. Moraes, B. Blawut et al., "Effect of glucose concentration and cryopreservation on mitochondrial functions of bull spermatozoa and relationship with sire conception rate," *Animal Reproduction Science*, vol. 230, 2021.
- [13] S. V. Shojaedini, A. Goldar, and M. Soori, "Correntropy based sperm detection: a novel spatiotemporal processing for analyzing videos of human semen," *Health and Technology*, vol. 8, no. 1-2, pp. 151–158, 2018.
- [14] H. Y. Chai, L. K. Wee, T. T. Swee, S.-H. Salleh, and L. Y. Chea, "An artifacts removal post-processing for epiphyseal region-of-interest (EROI) localization in automated bone age assessment (BAA)," *Bio Medical Engineering OnLine*, vol. 10, no. 1, p. 87, 2011.
- [15] A. Faisal, S. C. Ng, S. L. Goh, and K. W. Lai, "Knee cartilage segmentation and thickness computation from ultrasound images," *Medical & Biological Engineering & Computing*, vol. 56, no. 4, pp. 657–669, 2018.
- [16] M. H. A. Nizar, A. Khalil, C. K. Chan, N. P. Utama, and K. W. Lai, "Pilot study on machine learning for aortic valve detection in echocardiography images," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 1, pp. 9–14, 2019.
- [17] A. Khalil, S.-C. Ng, Y. M. Liew, and K. W. Lai, "An overview on image registration techniques for cardiac diagnosis and treatment," *Cardiology Research and Practice*, vol. 2018, Article ID 1437125, 15 pages, 2018.
- [18] L. K. Meng, A. Khalil, M. H. Ahmad Nizar et al., "Carpal bone segmentation using fully convolutional neural network," *Current Medical Imaging*, vol. 15, no. 10, pp. 983–989, 2019.
- [19] J. Lu, Y. Huang, and N. Lü, "Computer-aided sperm analysis: past, present and future," *Andrologia*, vol. 46, no. 4, pp. 329–338, 2014.
- [20] L. F. Urbano, P. Masson, M. VerMilyea, and M. Kam, "Automatic tracking and motility analysis of human sperm in time-lapse images," *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 792–801, 2017.
- [21] C. Dai, Z. Zhang, J. Huang et al., "Automated non-invasive measurement of single sperm's motility and morphology," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2257–2265, 2018.
- [22] P. Hidayatullah, I. Awaludin, R. D. Kusumo, and M. Nuriyadi, "Automatic sperm motility measurement," in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 1–5, Bandung, Indonesia, 2015.
- [23] X. Zhou, L. Ma, Y. Shang, M. Xu, X. Fu, and H. J. N. Ding, "Hybrid generative-discriminative learning for online tracking of sperm cell," vol. 208, pp. 218–224, 2016.
- [24] A. Arasteh, B. Vosoughi Vahdat, and R. Salman Yazdi, "Multi-target tracking of human spermatozoa in phase-contrast microscopy image sequences using a hybrid dynamic Bayesian network," *Scientific Reports*, vol. 8, no. 1, 2018.
- [25] M. Rahimzadeh and A. Attar, "Sperm detection and tracking in phase-contrast microscopy image sequences using deep learning and modified csr-dcf," 2020, <http://arxiv.org/abs/2002.04034>.
- [26] F. Mostajer Kheirkhah, H. R. Sadegh Mohammadi, and A. Shahverdi, "Modified histogram-based segmentation and adaptive distance tracking of sperm cells image sequences," *Computer Methods and Programs in Biomedicine*, vol. 154, pp. 173–182, 2018.
- [27] S.-Y. Wei, H.-H. Chao, H.-P. Huang, C. F. Hsu, S.-H. Li, and L. Hsu, "A collective tracking method for preliminary sperm analysis," *Biomedical engineering online*, vol. 18, no. 1, 2019.
- [28] C. Schleh and A. L. Leoni, "How to optimize the benefits of computer assisted sperm analysis in experimental toxicology," *Journal of Occupational Medicine and Toxicology*, vol. 8, no. 1, p. 6, 2013.
- [29] J. Talarczyk-Desole, A. Berger, G. Taszarek-Hauke, J. Hauke, L. Pawelczyk, and P. Jedrzejczak, "Manual vs. computer-assisted sperm analysis: can CASA replace manual assessment of human semen in clinical practice?," *Ginekologia Polska*, vol. 88, no. 2, pp. 56–60, 2017.
- [30] M. R. Mohammadi, M. Rahimzadeh, and A. Attar, "Sperm detection and tracking in phase-contrast microscopy image sequences using deep learning and modified CSR-DCF," <http://arxiv.org/abs/2002.04034>.
- [31] Y. Imani, N. Teyfour, M. R. Ahmadzadeh, and M. Golabbakhsh, "A new method for multiple sperm cells tracking," *Journal of Medical Signals & Sensors*, vol. 4, no. 1, pp. 35–42, 2014.
- [32] T. B. Haugen, S. A. Hicks, J. M. Andersen et al., "Visem: A multimodal video dataset of human spermatozoa," in *Proceedings of the 10th ACM Multimedia Systems Conference*, pp. 261–266, Amherst, MA, USA, 2019.
- [33] M. Zhang, X. Li, Z. Yang, and Y. Yang, "A novel zero-crossing edge detection method based on multi-scale space theory," in *IEEE 10th International Conference On Signal Processing Proceedings*, pp. 1036–1039, Beijing, China, 2010.
- [34] R. Muthukrishnan and M. Radha, "Edge detection techniques for image segmentation," *International Journal of Computer Science & Information Technology*, vol. 3, no. 6, pp. 259–267, 2011.
- [35] O. García-Olalla, E. Alegre, L. Fernández-Robles, and M. T. García-Ordás, "Vitality assessment of boar sperm using an adaptive LBP based on oriented deviation," in *Asian Conference on Computer Vision*, pp. 61–72, Springer, 2012.
- [36] I. Susrama, K. Purnama, and M. Purnomo, "Automated analysis of human sperm number and concentration (oligospermia) using otsu threshold method and labelling," *IOP Conference Series: Materials Science and Engineering*, vol. 105, 2016.
- [37] H. S. Mahdavi, A. Monadjemi, and A. Vafae, "Sperm detection in video frames of semen sample using morphology and effective ellipse detection method," *Journal of medical signals & sensors*, vol. 1, no. 3, p. 8, 2011.
- [38] T. Kasai, K. Ogawa, K. Mizuno et al., "Relationship between sperm mitochondrial membrane potential, sperm motility, and fertility potential," *Asian journal of andrology*, vol. 4, no. 2, pp. 97–103, 2002.
- [39] R. C. Firman and L. W. Simmons, "Sperm midpiece length predicts sperm swimming velocity in house mice," *Biology Letters*, vol. 6, no. 4, pp. 513–516, 2010.
- [40] F. Hunter and T. Birkhead, "Sperm viability and sperm competition in insects," *Current Biology*, vol. 12, no. 2, pp. 121–123, 2002.

Research Article

Low-Dose CT Image Denoising with Improving WGAN and Hybrid Loss Function

Zhijia Li , Weili Shi , Qiwei Xing , Yu Miao , Wei He , Huamin Yang ,
and Zhengang Jiang 

Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China

Correspondence should be addressed to Zhengang Jiang; jiangzhengang@cust.edu.cn

Received 22 April 2021; Revised 12 July 2021; Accepted 12 August 2021; Published 27 August 2021

Academic Editor: Reza Khosrowabadi

Copyright © 2021 Zhijia Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The X-ray radiation from computed tomography (CT) brought us the potential risk. Simply decreasing the dose makes the CT images noisy and diagnostic performance compromised. Here, we develop a novel denoising low-dose CT image method. Our framework is based on an improved generative adversarial network coupling with the hybrid loss function, including the adversarial loss, perceptual loss, sharpness loss, and structural similarity loss. Among the loss function terms, perceptual loss and structural similarity loss are made use of to preserve textural details, and sharpness loss can make reconstruction images clear. The adversarial loss can sharp the boundary regions. The results of experiments show the proposed method can effectively remove noise and artifacts better than the state-of-the-art methods in the aspects of the visual effect, the quantitative measurements, and the texture details.

1. Introduction

During recent years, the computed X-ray tomography (CT) has been one of the important practical imaging methods, which has been widely utilized in medical diagnosis. The anatomical structure with high temporal-spatial resolution could be found from CT images, and numerous researchers benefit from CT scans, especially in pathologic diagnosis and treatment domains. However, with the widely use of medical CT, the potential risk of ionizing X-ray radiation to patients has aroused public concern [1, 2].

According to the famous ALARA theory, the minimization of X-ray became one of the research hotspots in CT image fields. Among the many methods, the most popular approach to reduce radiation is reducing X-ray flux by shortening the exposure time and cutting down the operating the X-ray tube current. Unfortunately, the lower the X-ray flux, the noisier the generated CT image. Therefore, one way to address the problem is to reduce the image noise by the algorithm. The common method to reduce noise is filtering. But

it is an ill-posed and challenging problem [3–5]. Recently, deep learning techniques have shown their superiority in denoising the image [6–11]. Various denoising models based on convolutional neural networks (CNNs) have been proposed with different network architecture for LDCT denoising [1, 12–14], which include 2D CNNs [2, 12], 3D CNN [1], residual encoder-decoder CNN [13], and cascaded CNN [14]. Besides, different loss functions, such as the mean squared error (MSE) [1, 12–14], adversarial loss [1, 2], and perceptual loss [2], are presented in the denoising model. Different network architectures and loss function may have a profound impact upon the learning process of the network. According to literature [8], the complexity of the denoising model is determined by the network architecture, and the loss function is related to what the denoising model learns from images and data.

In practice, we found the denoising methods with generative adversarial network could get better results than those with CNNs. However, these methods have difficulties of network training and the gradients disappearance [15]. To solve

this problem, here, we propose an improved GAN with the Wasserstein distance (SSWGAN) to reduce the noise of the low-dose CT images. Specifically, denoising low-dose CT images can be looked as a translation of low-dose CT images into normal-dose CT (NDCT) images. Our proposed GAN could estimate the distance of distribution between low-dose CT and normal-dose CT. In the process, the perceptual loss based on VGG could preserve as many image details as possible when suppressing the noise. The SSIM loss preserves the structural and textural details after the denoising process, and L1 loss keeps the sharpness of the denoised image, especially in the low contrast regions. In summary, our contributions are as follows:

- (i) An improved WGAN network is introduced as the denoising model
- (ii) A novel hybrid loss function is introduced to enhance the denoising model performance
- (iii) Compared with a few latest network models, we found our disadvantages and presented the Q-AE model to improve our generator architecture

2. Related Work

2.1. LDCT Denoising Methods. Generally, the LDCT denoising methods can be divided into three classes:

- (a) *Projection Filtering* [16–18]. Their advantage is the higher computation efficiency. However, they always result in the loss of spatial resolution and edge blur in images
- (b) *Iteration Reconstruction* [19–25]. They outperform in increasing the signal to noise ratio, but they need more computing resources and the accuracy model of the noises
- (c) *Postprocessing* [26–28]. They can be performed on the images directly and have the lower calculating costs so that they have been implied in the CT imaging system and analysis system. There are some residual problems in the processed images yet

With the rapid development of deep learning techniques, associated denoising models have achieved an impressive performance of denoising LDCT images [29, 30]. The learning process includes two major components: network architecture and loss function. The architecture determines the complexity of the denoising model and the loss function controls what the denoising model learn. Recently, lots of methods were proposed. Yi et al [31] summarized these methods and made a comprehensive comparison. Next, we mainly described the approaches with novel network architecture and the ones with improved loss function,

- (1) *Network Architecture.* Chen et al. [32] first proposed the low-dose CT image denoising method based on convolution neural network (CNN), who obtained better effects in visual sense and measurements.

Then, Chen et al. [13] improved the network structure and they developed a residual encoder CNN (RED-CNN). The results were better than the original CNN. However, their network was complex and time-consuming. To overcome the disadvantages of RED-CNN, Zhang et al. [33] proposed a novel network. Compared with RED-CNN, there were less parameters in their network and their results were better

- (2) *Loss Function.* Minimizing the MSE based on the difference between the denoised images and the NDCT easily led to overblurred [1, 2], which was proved to correlate poorly with the human perception of image quality [34, 35]. According to literature [8], the optimal MSE estimator suffered from the regression-to-mean problem, which made denoised LDCT look oversmoothed, unnatural, and implausible. The adversarial loss (AL) could result in a sharp image locally indistinguishable from the NDCT image but it does not exactly correspond to the NDCT image globally [36] since the AL optimizes the distance between distributions of the denoised results and NDCT images. Later, many methods presented the perceptual loss (PL) to make denoised images look more similar to NDCT images in the high-level feature space [2]. However, there are other features to be applied in the images, such as the sharpness and structural similarity index. Here, we extend the wise-used hybrid loss function including AL and PL. Our proposed hybrid loss function includes four terms: AL, PL, sharpness loss, and similarity loss to enhance the denoising performance more effectively

2.2. Wasserstein GAN Framework. Recently, the GAN [34] architecture was developed as a novel way to model the distribution of the given data. But it has the difficulties of network training and the gradient disappearance [8]. To deal with these limitations, the GAN with the Wasserstein distance (WGAN) was widely used [37, 38], which made use of the Wasserstein distance as the measurement of the difference between the distribution loss and perceptual loss [37]. Besides, gradient penalty was employed as a regular accelerated method for training network (WGAN-GP) [39]. It was important that WGAN-VGG [40] was an approach for low-dose CT, which achieved promising denoised CT images [41], and the perceptual loss was utilized by VGG [41] that pretrained on natural images. WGAN-VGG could overcome the problem of image overblur. Also, SMGAN [42] combined the L1 loss and the multiscale structure loss so that it outperformed the WGAN-VGG in convergence accuracy [40]. But sometimes, the reconstruction images were fuzzy. Besides, the gradient penalty term weakened the express ability of GAN [43]. Furthermore, researchers found the denoising model without deconvolutional layers, which is the transpose of convolutional layers [44], implies that the input and the output of the denoising model may have different sizes. To keep the size of denoised CT images equal to that of the input, U-net architecture are used in denoising

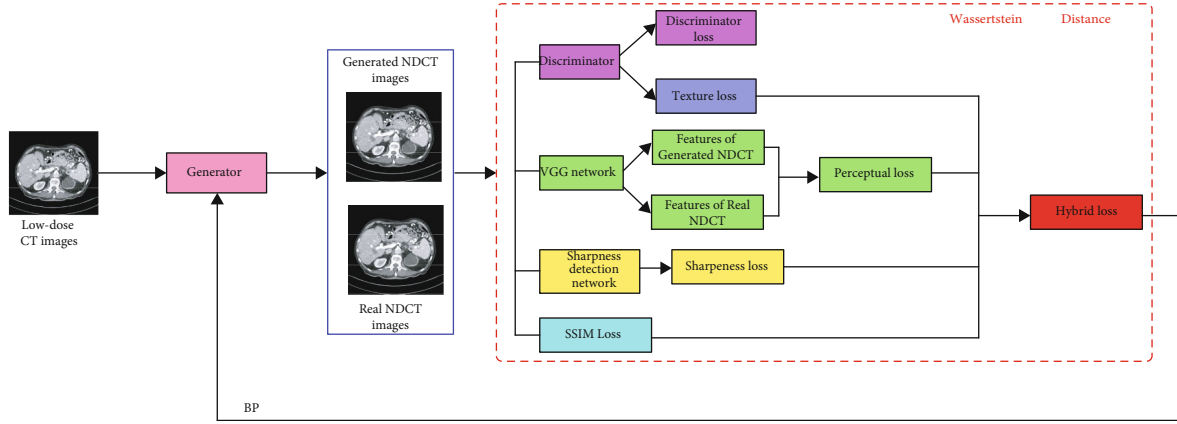


FIGURE 1: The framework of our SSWGAN network for denoising LDCT images.

LDCT images [45–51]. Shan et al. [8] proposed the conveying path-based convolutional U-net denoising model, which is called as CPCE. Fan et al. [15] improved the method and proposed a denoising framework, who replaced the inner product in current artificial neurons with a quadratic operation on input data. Their method is called the Q-AE.

3. Denoising Framework

3.1. Principle and Model of Denoising. Generally, the noise distribution in CT images is treated as the combination of quantum Poisson and electronic Gaussian noise. But the noise in reconstruction images is complex, and its distribution is always nonuniform. Besides, the relationship between NDCT and LDCT cannot be described with an accuracy mathematical model. So only with conventional methods, we could hardly obtain better results of denoising LDCT images. Fortunately, the uncertain noise model can be estimated by deep learning techniques, because of its strong ability of capturing features.

Denoising LDCT images can be represented as the below model. Assume that $\mathbf{x} \in \mathbb{R}^{N \times N}$ represents the NDCT image and $\mathbf{y} \in \mathbb{R}^{N \times N}$ represents the corresponding LDCT, our goal is to confirm a function G which maps \mathbf{y} to \mathbf{x} :

$$\mathbf{G} : \mathbf{y} \longrightarrow \mathbf{x}. \quad (1)$$

The generative and adversarial abilities of GAN can be applied to extract features from deep levels with the spatial information of reconstruction images, so that GAN can identify the noise and effective image details. GAN usually includes a pair of neural networks: a generator G and a discriminator D [52, 53]. The generator G can learn the real distribution of NDCT, and the discriminator D can make the best effort to distinguish between real or fake samples generated by G . This pair of networks is often trained alternately, so the competition encourages the generated samples to be hardly distinguished from real ones. Finally, we could obtain CT images of better quality.

3.2. The Structure of our GAN. Mathematically, G and D could be formed as a two-role minimax game:

$$\min_{GD} \max_{E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{y \sim P_y(y)} [\log (1 - D(G(y)))]}, \quad (2)$$

where E represents the expectation value, P_{data} and P_y represent real and noise distributions, respectively. In the regular GAN, the Jensen-Shannon (JS) divergence is utilized to compute the similarity of two kinds of data distribution [54]. But, as mentioned above, the JS divergence easily results in gradient vanishing. Here, we adopt the Wasserstein distance [38] instead of JS divergence to ensure the training stability of the neural network. The main structure of our network is shown in Figure 1. As shown in the figure, there are four parts in our SSWGAN network, which is the generator, discriminator, sharpness detection network, and hybrid loss function, respectively.

3.2.1. The Architecture of our Generator. As shown in Figure 2, the proposed generator G is different from the traditional noise reduction models. Here, we utilize the ADNet [55] with 17 layers as our generator. There are four parts in our generator network, which represents sparse block (SB), feature enhance block (FEB), attention block (AB), and reconstruction block (RB), respectively. In particular, SB could reduce noise with dilated and common convolution to achieve the optimal balance between performance and efficiency. FEB combines the global and local feature information to improve the representation ability of models. AB is often applied to extract the implied noise in the complex background accurately. Utilizing both FEB and AB could both improve the efficiency and reduce the complexity of training the network model. RB generates the NDCT images of better quality with the obtained noise map and given LDCT images.

3.2.2. The Architecture of our Discriminator. As shown in Figure 3, the input of the discriminator is the NDCT generated by G and the real NDCT. Our discriminator D is designed to distinguish the real one from the two NDCT

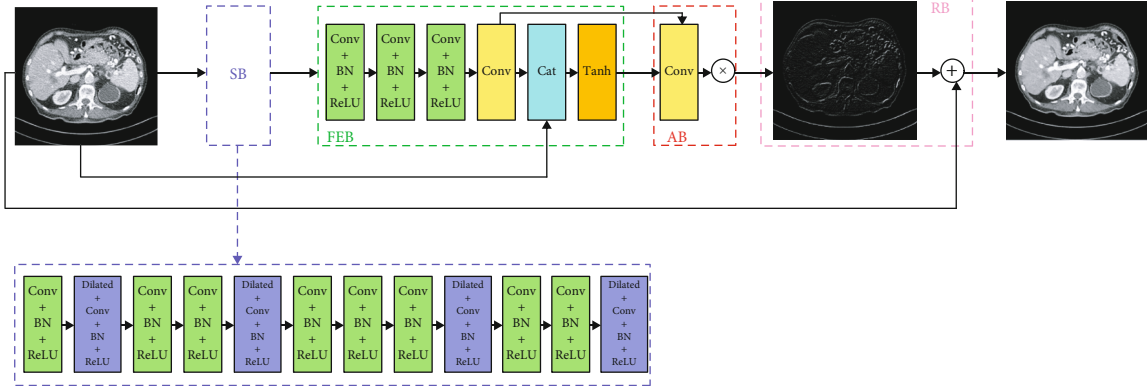


FIGURE 2: The framework of our generator. As shown in the figure, it could learn the noise and then generate the denoised output according to the input LDCT.

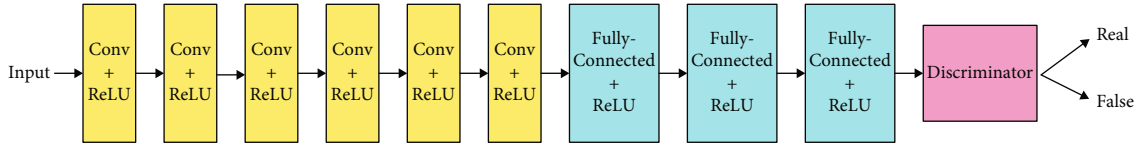


FIGURE 3: The framework of our discriminator. As shown in the figure, it could learn to judge that if the input is generated by G or a real NDCT image.

images. The discriminator includes 6 convolution layers and 3 full-connected layers. Among the convolution layers, there are 64 filters in the first two layers, 128 filters in the middle two layers, and 256 filters in the final two layers. After each convolution operation, there is an activation function ReLU [56]. The step size of convolution is 1 and the filter size is 3×3 . The end of the discriminator is fully-connected layers, and there are 1024 outputs, 512 outputs, and 1 output, respectively. With the discriminator, we could obtain the difference of the generated NDCT and real NDCT.

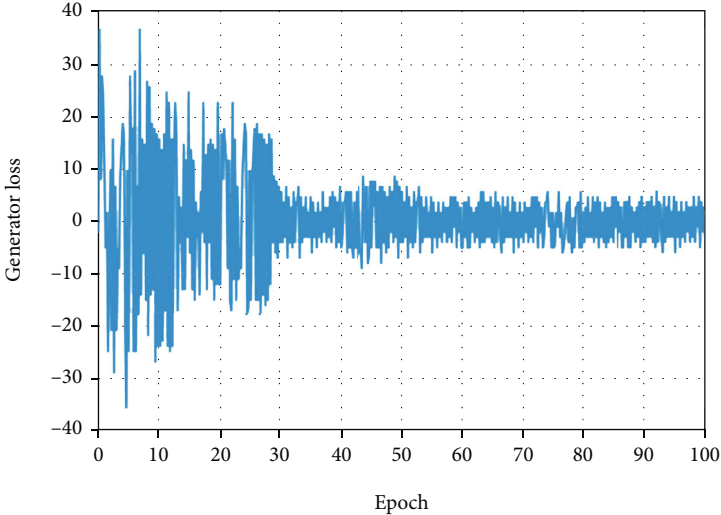
3.2.3. Sharpness Detection Network. Numerous noise reduction methods are poor in fuzzy edge. Traditional nonlinear optimization algorithms are to average adjacent pixels or utilize the self-similar patch. However, when the noise level is high, these optimization algorithms are not efficient because of the high similarity between noise and edge. Although the discriminator of GAN could output more clear images and distinguished images from the candidates, it is not efficient in the low contrast regions because the antagonism loss used in GAN could not ensure that the images are able to be reconstructed accurately.

Recently, a few more flexible and complex methods were proposed, which mainly made use of the statistical differences of the specific properties between the fuzzy regions and the sharp regions, such as the gradient information [57] and discrete cosine coefficient [58]. Other methods utilized the sparse coding way to decompose local paths and obtained sharper images by quantifying the local sharpness. Also, the other methods could generate sharp images, for example, the one based on depth map

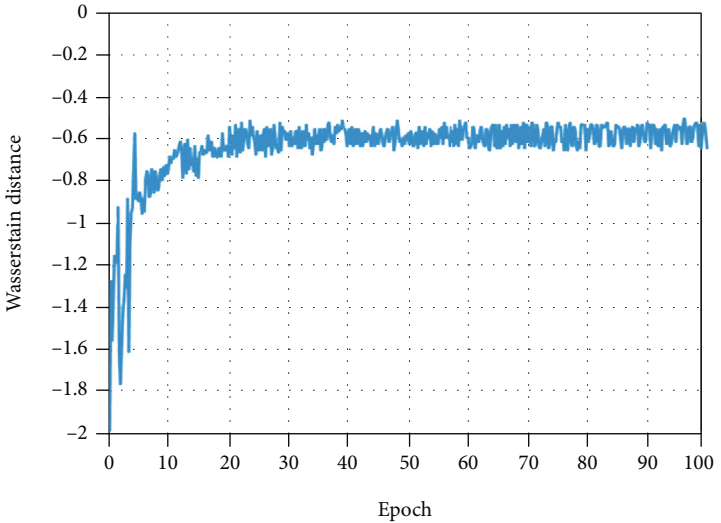
estimation [59]. It is hard to make a mark in the low contrast regions of medical images, so we introduce a sharpness detection network, represented by S , and use the method proposed by Yi and Eramian [60] because of its strong sensitivity in low contrast regions. When implementing SSWGAN, we transfer the NDCT results generated by G to the sharpness detection network S and compare the sharpness images of our generated results with the images of real ones. Because the sharpness images are shown with grayscale, the pixel values represent the local sharpness. With the sharpness images, we can calculate the mean square error between the two sharpness images and update the weight of network according to the calculated results.

3.2.4. Hybrid Loss Function. The main challenge of the training network is to preserve as much texture detail as possible when reducing noise. The hybrid loss function can keep the training process of SSWGAN within bounds. With the hybrid loss function, the differences between the generated NDCT images and real NDCT images can be measured and the weights of generator G can be updated by back propagation (BP). In order to improve the denoising network, our hybrid loss function includes four parts, which is adversarial loss, perceptual loss, sharpness loss, and structural similarity loss, respectively.

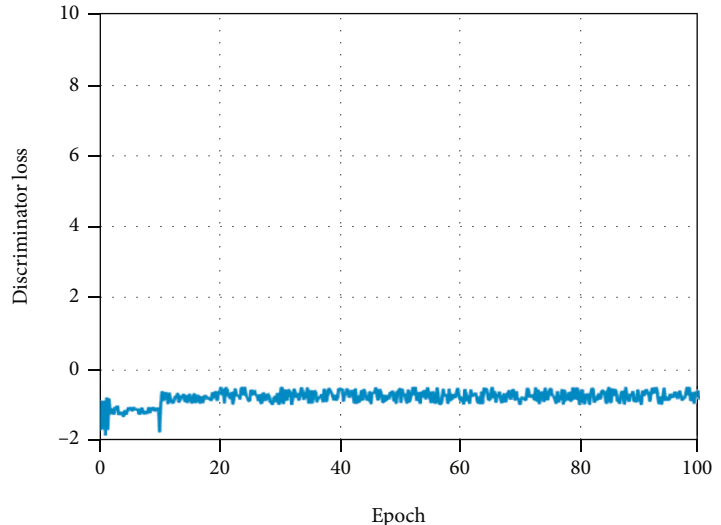
(1) *Adversarial Loss.* As described in Ref [2], minimizing the least-squares loss could approximate the distribution of LDCT according to the NDCT, and finally, we could obtain better denoised images. However, it does not match well the corresponding NDCT in detail. Here, we introduce the adversarial



(a)



(b)



(c)

FIGURE 4: The training curves for different terms of our loss function, which is generator loss (a), Wasserstein distance (b), and discriminator loss (c).

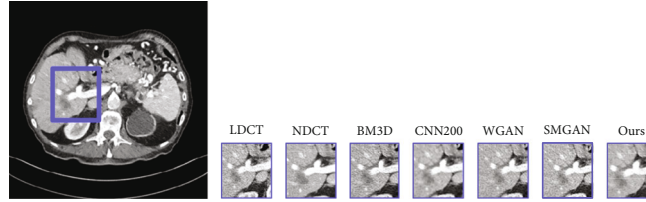


FIGURE 5: Denoising results of the different algorithm on the MDLCT dataset. From left to right, the subgraph indicates the detail of different images: the low-dose CT (LDCT), normal-dose CT (NDCT), BM3D, CNN200, WGAN, SMGAN, and our framework.

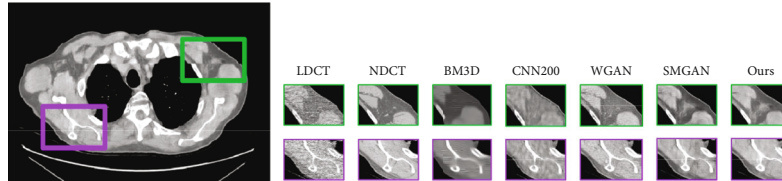


FIGURE 6: Denoising results of the different algorithm on the lung CT images dataset. From left to right, the subgraph indicates the detail of different images: the low-dose CT (LDCT), normal-dose CT (NDCT), BM3D, CNN200, WGAN, SMGAN, and our framework.

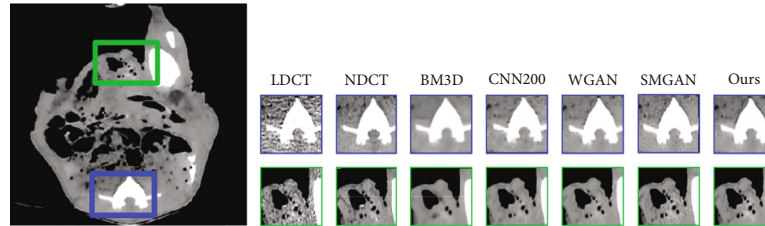


FIGURE 7: Denoising results of the different algorithm on the Piglet CT images dataset. From left to right, the subgraph indicates the detail of different images: the low-dose CT (LDCT), normal-dose CT (NDCT), BM3D, CNN200, WGAN, SMGAN, and our framework.

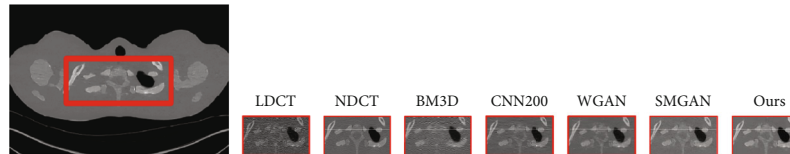


FIGURE 8: Denoising results of the different algorithm on the thoracic CT images dataset. From left to right, the subgraph indicates the detail of different images: the low-dose CT (LDCT), normal-dose CT (NDCT), BM3D, CNN200, WGAN, SMGAN, and our framework.

TABLE 1: Quantitative results for Figure 5 (the results of MDLCT dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	LDCT	BM3D	CNN200	WGAN	SMGAN	Ours
PSNR/dB	24.4411	25.4434	27.8083	28.1414	27.5667	28.8961
SSIM	0.7882	0.7937	0.8362	0.8428	0.8394	0.8596
FSIM	0.912	0.9306	0.9675	0.9484	0.9660	0.9762

TABLE 2: Quantitative results for Figure 6 (the results of lung CT images dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	LDCT	BM3D	CNN200	WGAN	SMGAN	Ours
PSNR/dB	14.5911	24.7603	33.1926	33.7454	34.6880	35.5760
SSIM	0.2008	0.6750	0.8804	0.9281	0.9310	0.9533
FSIM	0.9232	0.9540	0.9683	0.9565	0.9630	0.9769

TABLE 3: Quantitative results for Figure 7 (the results of piglet CT images dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	LDCT	BM3D	CNN200	WGAN	SMGAN	Ours
PSNR/dB	39.93	41.46	44.18	44.31	44.83	45.10
SSIM	0.9705	0.9733	0.9794	0.9802	0.9816	0.9885
FSIM	0.9155	0.9422	0.9614	0.9550	0.9677	0.9765

TABLE 4: Quantitative results for Figure 8 (the results of thoracic CT images dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	LDCT	BM3D	CNN200	WGAN	SMGAN	Ours
PSNR/dB	25.66	30.86	33.57	33.60	33.73	34.03
SSIM	0.4485	0.6552	0.8001	0.8018	0.8049	0.8159
FSIM	0.9103	0.9486	0.9630	0.9565	0.9670	0.9744

loss to let our G generate denoised CT images as real as possible. Adversarial loss could be described as follows:

$$L_{\text{WGAN}}(G, D) = -E_{\mathbf{x}}[D(\mathbf{x})] + E_{\mathbf{z}}[D(G(\mathbf{z}))] + \lambda E_{\tilde{\mathbf{x}}}[(\|\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})\|_2 - 1)^2]. \quad (3)$$

Here, the first two items represent the Wasserstein distance, and the final item represents the gradient vanish one utilized for network normalization. G and D are the generator and discriminator. E is a set of data samples with specific distribution. $\tilde{\mathbf{x}}$ is the generated NDCT image, and λ is the penalty coefficient. Minimize adversarial loss can keep more texture details.

(2) *Perceptual Loss*. The most important for medical images is to keep the necessary features used in pathologic diagnosis [61]. Mean squared error (MSE) is always utilized as the loss function, which can result in images aliasing and details lost. Perceptual loss can calculate the distance between the generated images and the real images in the feature space of human perception instead of the distance in pixel space. With the perceptual loss, the generated denoised NDCT images could preserve the origin feature in real NDCT images, which is not achieved with other loss function. The perceptual loss can be described as follows:

$$L_{\text{Perceptual}}(G) = E_{(x,y)} \left[\frac{1}{whd} \|\phi(G(y)) - \phi(x)\|_F^2 \right], \quad (4)$$

where ϕ is the feature extractor, and $\|\cdot\|_F$ is the Frobenius norm. Here, we adopt the pretrained VGG-19 network [41] as the extractor, w , h , d represent the width, height, and depth, respectively. Because VGG-19 takes the color images as the input and CT images are often in grey scale, we convert the CT images into RGB channel as the input of VGG-19. There are 16 convolution layers and 3 full-connected layers in VGG-19. Among the convolution layers, the output of the 16th layer is the extracted feature of VGG

and is used as the loss function:

$$L_{\text{VGG}}(G) = E_{(x,y)} \left[\frac{1}{whd} \|\text{VGG}(G(y)) - \text{VGG}(x)\|_F^2 \right]. \quad (5)$$

(3) *Sharpness Loss*. Here, we propose a sharpness loss used in sharpness detection network to evaluate the sharpness of images. The generator G is asked to not only generate the image as similar to the real one as possible but also generate the clear image as close to the real image as possible. The sharpness loss is described in mathematical form:

$$L_{\text{Sharp}}(G) = E_{(x,y)} [\|S(G(y)) - S(x)\|_2], \quad (6)$$

where $\|\cdot\|_2$ is L_2 distance.

(4) *Similarity Loss*. In medical CT images of different dose levels, the feature correlation is usually strong. Structural similarity index (SSIM) includes three parts, which is luminance, contrast, and structure. SSIM is a better evaluating indicator than MSE and peak signal-to-noise ratio (PSNR) in visual tasks. To measure the similarity between denoised CT images and normal-dose version, the SSIM can be described:

$$\text{SSIM}(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} * \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} * \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (7)$$

where μ_x , μ_y , σ_x , σ_y , and σ_{xy} represent the means, standard deviations, and the cross-correlation of two images, respectively, and C_1 , C_2 , and C_3 are the constants. Besides, when x and y are more similar, the value of SSIM is closer to 1. Thus, we set the loss function for SSIM as follows:

$$L_{\text{SSIM}}(G) = 1 - \text{SSIM}(x, y). \quad (8)$$

It is worth noting that the SSIM loss can be back-propagated to update the parameters of our network, when giving its property of differentiability. Here, we make use

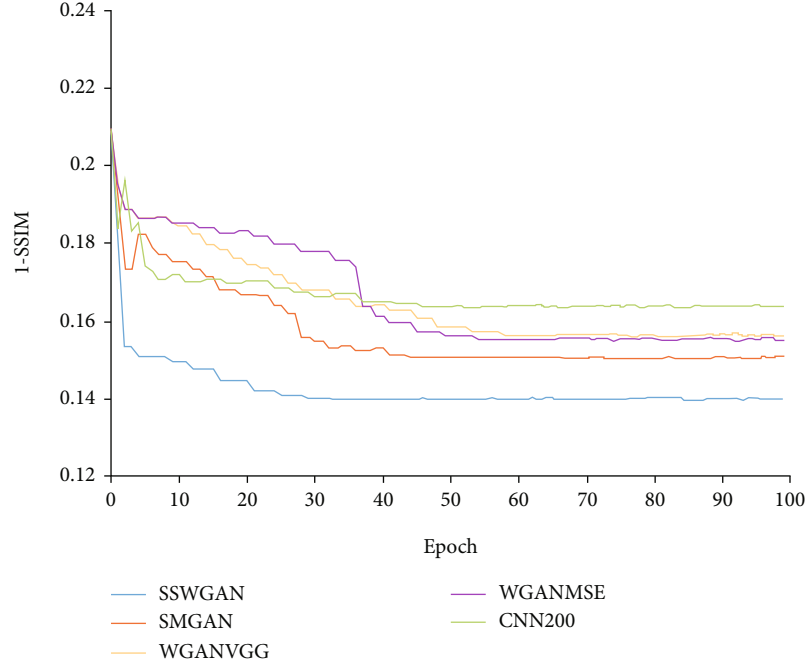


FIGURE 9: Convergence comparison between our framework and different methods.

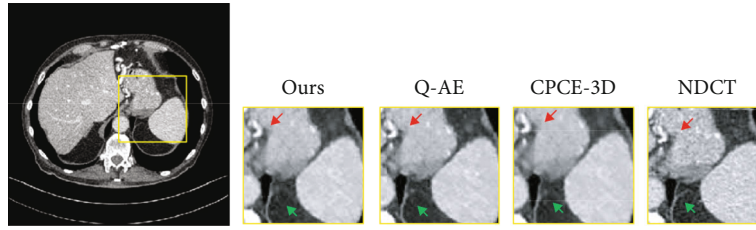


FIGURE 10: Denoising results of the different algorithm on the MDLCT dataset. From left to right, the subgraph indicates the detail of different images: our framework, Q-AE, CPCE-3D, and NDCT images.

of SSIM to calculate the overall similarity between the NDCT images and LDCT images.

In summary, the overall objective function of our adapted SSWGAN is represented as follows:

$$L_{SSWGAN} = \alpha L_{WGAN}(G, D) + \beta L_{\text{perceptual}}(G) + \gamma L_{\text{Sharp}}(G) + \omega L_{SSIM}(G), \quad (9)$$

where α , β , γ , and ω are weight coefficients of the above four terms.

4. Results and Discussion

4.1. Dataset for Experiments. To show the capacity of our proposed denoising SSWGAN for LDCT image, four real clinical CT image datasets were applied in our study in order to avoid overfitting problem. The four datasets were MDLCT dataset authorized by Mayo Clinic for “2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge,”

TABLE 5: Quantitative results for Figure 10 (the results of MDLCT dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	Ours	Q-AE	CPCE-3D
PSNR/dB	28.8961	32.712	30.137
SSIM	0.8596	0.95266	0.905

the lung CT image dataset [62], the real piglet CT image dataset [63], and the thoracic CT image dataset [64].

The MDLCT dataset includes 2378 NDCT images and the corresponding simulated LDCT (quarter dose) from ten anonymous patients [12]. The matrix of each CT images is 512×512 , and the thickness is 3.0 mm. Inspired by Ref. [57], we divided the dataset into two groups. One of the groups includes 2168 paired images from nine patients used in the training process. The other one contains 210 paired images from the last patient utilized as the test dataset. During the training stage, we extracted the patches whose size was 55×55 . Totally, we extracted approximately 106 paired patches used for capturing local details instead of wasting

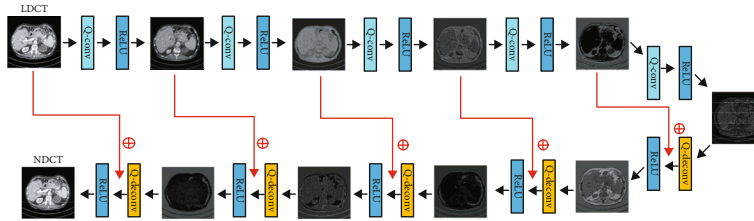


FIGURE 11: Architecture of our improved generator. Q-Conv indicates the “quadratic convolution” [15], and Q-Deconv indicates the “quadratic deconvolution” [15].

huge memories, which improved the efficiency of the training.

The lung CT images dataset is created from a patient with the method proposed in Ref. [64], including 663 slices. The CT scans of the patient are from The Cancer Imaging Archive (TCIA). The piglet CT image dataset contains 900 images with 100 KVp, 0.625 mm thickness. The thoracic CT image dataset includes 407 pairs of CT images from an anthropomorphic thoracic phantom. The current tube for NDCT and LDCT images is 480 m As and 60 m As, respectively, with a peak voltage of 120 KVp and slice thickness of 0.75 mm. We randomly selected 30% of images using in the test stage, and the size of each image is 512×512 .

4.2. Parameter Setting. Our framework is implemented within Python’s platform, Pytorch, and TensorFlow. All experiments run on a personal computer (Intel i5 7400 with 16 G random memory) and accelerated by a NVIDIA RTX 2080 GPU with 16 G memory.

The generator and discriminator of our SSWGAN are both optimized utilizing the adaptive momentum estimation (Adam) proposed in Ref. [65]. The size of our mini-batch is 96. The learning rate is set to 10^{-3} used for training 100 epochs and set to 10^{-4} used for training 100 epochs. The coefficients of our hybrid loss function are set $\alpha = 0.005$, $\beta = 0.0995$, $\gamma = 0.95$, $\omega = 0.95$, and $\lambda = 10$, respectively. As shown in Figure 4, our network can be convergent after training 100 epochs.

4.3. Image Evaluation Criteria. To evaluate the quality of generated images, we adopt three objective evaluation criteria, which are PSNR [12], SSIM [35], and feature similarity index (FSIM) [66]. PSNR calculates the average pixel difference between the generated NDCT images and real NDCT images, which is used for evaluating the denoising ability of different methods. SSIM calculates the structural difference between the generated NDCT images and real NDCT images, which is used for evaluating the similarity of two images. FSIM calculates the feature difference between the two images, which represents the feature-preserving ability of different methods.

5. Experimental Results and Discussion

Note that, we describe the advantages of our algorithm framework in two ways: (1) compared with other widely used traditional LDCT denoising methods and (2) compared with the latest LDCT denoising methods based on GAN.

TABLE 6: Quantitative results for Figure 10 (the results of MDLCT dataset) utilizing different methods. Low dose indicates the noisy low-dose images, and others are the results of different methods.

Quantitative measurements	Our originals	Q-AE	CPCE-3D	Our improvement
PSNR/dB	28.8961	32.712	30.137	32.848
SSIM	0.8596	0.95266	0.905	0.95649

5.1. The Comparison between Ours and the Traditional LDCT Denoising Algorithms. To demonstrate, our proposed method has advantages in denoising LDCT images, and we compare ours with other widely used traditional LDCT denoising methods including BM3D [67], CNN200 [12], WGAN [2], and SMGAN [42]. Among these methods, BM3D is one of the most popular traditional approaches utilized for denoising LDCT images. CNN200, WGAN, and SMGAN are three representative denoising methods based on CNN. CNN200 adopts the encoder-decoder convolutional neural network with MSE loss. WGAN and SMGAN make use of Wasserstein distance and sharing similar network architecture. But their loss function is different between each other.

Figure 5 gives the visual results for the MDLCT dataset. As shown in Figure 5, there are much noise in LDCT images, which results in the blurred images and hard to distinguish the structure and details of images. The corresponding NDCT image is much clearer and of better quality in comparison. The third subgraph is the denoising result of BM3D, where there is a small part of noise. Affected by significant blocky, some edges and small structures are too blurred. The fourth subgraph shows the result of CNN200. From the fourth image, it can be found that this method suppresses noise to some degree; however, there are still some noise and artifacts in the images which are after denoising. From the fifth image and sixth image, the denoising methods based on GAN not only reduce most noise and artifacts but also preserve structural details. Compared with the fourth image, there are less noise in the fifth one. But some edge details are loss. The sixth subgraph is the result of SMGAN. It can be seen that SMGAN smoothens the images excessively, and some crucial structures, like the region of porta, are over blurred. From the right image, it can be seen that our framework outperforms in the content details and textural information than the other methods.

As shown in Figures 6–8, all images are denoised results of the above methods based on lung CT images dataset,

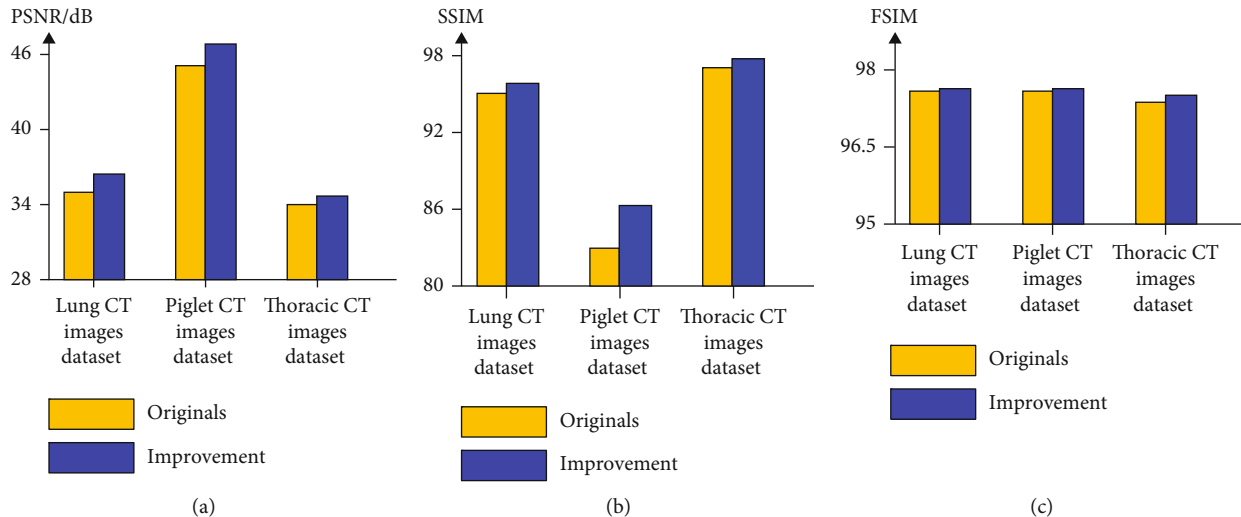


FIGURE 12: Quantitative results of the lung CT images dataset, piglet CT images dataset, and thoracic CT images dataset utilizing our original network and improved network. Seen from the figure, sub-Fig (a) is the PSNR results, sub-Fig (b) is the SSIM results, and sub-Fig (c) is the FSIM results.

piglet CT images dataset, and thoracic CT images dataset. We can obtain the same conclusion from the comparison of all methods in Figures 6–8 as the one in Figure 5. Our method could perform better in reducing artifacts and noise. Our denoised images are closer to the real NDCT images.

We computed the SSIM, FSIM, and PSNR of all denoised images. The results are listed in Tables 1–4, respectively. Here, we evaluated the average values of the dataset.

In Table 1, the results are calculated for the images of the MDLCT dataset. Our framework obtains the best results in aspects of SSIM, PSNR, and FSIM. Our PSNR value is averagely 2.7287 dB higher than other methods, the SSIM of our framework is averagely 0.0385 higher than others, and our FSIM result is averagely 0.0414 higher than others. The result of Table 1 shows that our framework gains the best results in respect of all quantitative measurements. From Tables 2–4, we can get the same conclusion, that is to say, it is important to point out that our statistical value is nearest to that of the NDCT images and obtained the best matching textural statistics to NDCT image than other methods.

Besides, to exhibit that our framework has the advantage in terms of convergence, taking results of the MDLCT dataset as an example, we evaluated the quantitative measurements 1-SSIM (the smaller the values, the better the image is) during the training process of different methods. The results can be seen in Figure 9.

As shown in Figure 9, WGAN-VGG and WGAN-MSE are convergent at the point, where the epoch number is 60. CNN200 and SMGAN could achieve convergence at the point, where the epoch number is 45. Our framework can be convergent at the point where the epoch number equals 30. The efficiency of our method is higher than other methods, and it can be seen in Figure 9 that our images under convergence are of better quality.

5.2. *The Comparison between Ours and the Latest LDCT Denoising Algorithms Based on GAN.* To compare with the

TABLE 7: Quantitative results for Figure 12 utilizing our original network and improved network. “Ori” indicates the results of our original network, and “Imp” indicates the results of our improved network.

Quantitative measurements	Lung CT images dataset	Piglet CT images dataset	Thoracic CT images dataset
PSNR/dB	(Ori) 35.5760	(Ori) 45.10	(Ori) 34.03
	(Imp) 36.3950	(Imp) 45.962	(Imp) 35.2649
SSIM	(Ori) 0.9533	(Ori) 0.9885	(Ori) 0.8159
	(Imp) 0.9645	(Imp) 0.9892	(Imp) 0.8651
FSIM	(Ori) 0.9769	(Ori) 0.9765	(Ori) 0.9744
	(Imp) 0.9774	(Imp) 0.9767	(Imp) 0.9759

latest LDCT denoising algorithms based on GAN including the CPCE algorithm [8] and the Q-AE algorithm [15], we conduct experiments on the MDLCT dataset. Figure 10 shows the results.

Seen from the above figure, we found both of the algorithms perform better than ours. Our image suffers from the oversmoothed details and the loss of texture information (indicated by red arrow and green arrow). The quantitative results can be seen in Table 5. From the results, our proposed denoising framework is not good as the latest algorithms based on GAN.

To analyse the reason why our algorithm is not so good, we compare our network architecture and loss function with the latest algorithms based on GAN. Then, we found the disadvantages of our proposed framework. First, our generator does not involve the deconvolutional layers. As described in literature [15], it easily implies that the input and the output may have different sizes. More seriously, the texture is lost. Then, the convolutional layers do not preserve enough features. To overcome these shortcomings, we improve our

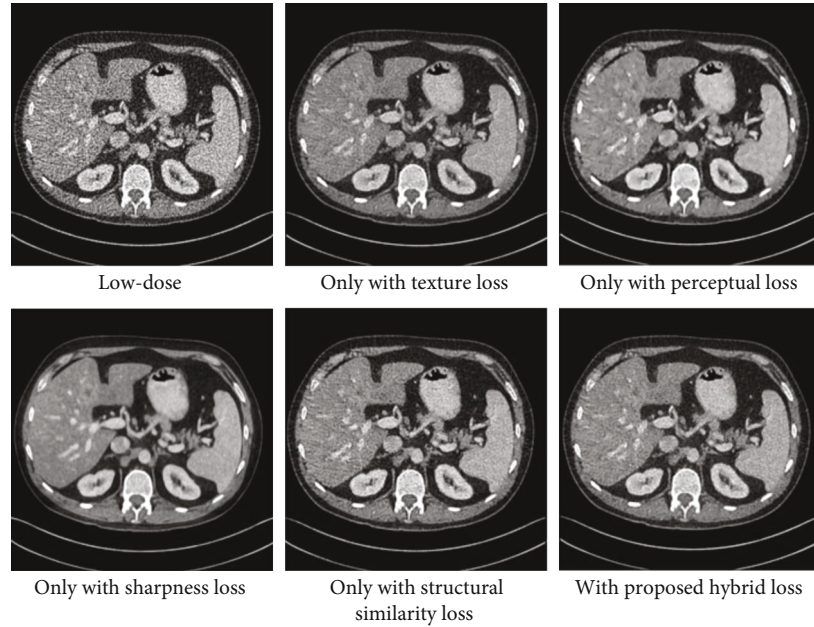


FIGURE 13: Denoising comparison between different loss function.

TABLE 8: Quantitative results for Figure 13 (the results of MDLCT dataset) utilizing different loss function.

Quantitative measurements	Adversarial loss	Perceptual loss	Sharpness loss	Structural similarity loss	Hybrid loss
PSNR/dB	24.66	26.86	26.57	25.60	28.8961
SSIM	0.8485	0.7552	0.8301	0.8058	0.8596
FSIM	0.9485	0.9446	0.9635	0.9665	0.9762

framework. Inspired by the literature [15], we modified the architecture of our generator, and it can be seen in Figure 11.

Seen from the above figure, we replace our original SSWGAN with (network). We keep our loss function unchanged. After improving our network architecture, we compare the new obtained results with the latest denoising methods. The quantitative results are seen in Table 6. From the results, our improved method outperforms better.

Figure 12 and Table 7 show the quality assessment index of the comparison between our improving result and original result on lung CT images dataset and piglet CT images dataset, including the PSNR, SSIM, and FSIM. The results show that our improvement is better than the originals. This is largely due to that the (network) with Q-AE model could give a high-order nonlinear sparse representation with a reasonable model complexity.

5.3. Discussions and Analysis. In our framework, we propose the SSWGAN with hybrid loss function to denoise the LDCT images. Then, inspired by the latest algorithm, we improve our network architecture. The major difference between ours and other methods is the utilization of hybrid loss function except the network architecture. When deep learning approaches are presented in image processing, we can obtain better results than the state-of-the-art LDCT denoising methods because we can capture high-level abstract features from training data. To a large extent, the

loss function of deep learning influences the LDCT image restoration process. Here, we compared different loss function performance on LDCT image restoration: (1) only with adversarial loss, (2) only with perceptual loss, (3) only with the sharpness loss, (4) only with the structural similarity loss, and (5) with the hybrid loss. Also, we took the MDLCT dataset as an example. The results were shown in Figure 13.

Seen from Figure 13, the adversarial loss makes the edge sharper (shown as Figure 13(a)). The perceptual loss makes the edge more obvious (shown in Figure 13(b)), and it easily results in the artifacts. The sharpness loss can generate a clear image (shown in Figure 13(c)), however, it losses part of details. The structural similarity loss can preserve more details and image structures while reducing noise. For evaluating the quality of images, we adapt the PSNR, SSIM, and FSIM. The results are shown in Table 8.

From Table 8, we can find that although any one of the four loss functions has advantages, only with one kind of loss function, the quality of image is lower than the image with hybrid loss function. In addition, with hybrid loss function, we could achieve gradient penalty and acceleration of convergence.

Since (parameter) of our hybrid loss function can make impact on the denoising results. Here, we try to find the relationship between our chose parameters and the quality of the denoised images. In order to determine the optimal

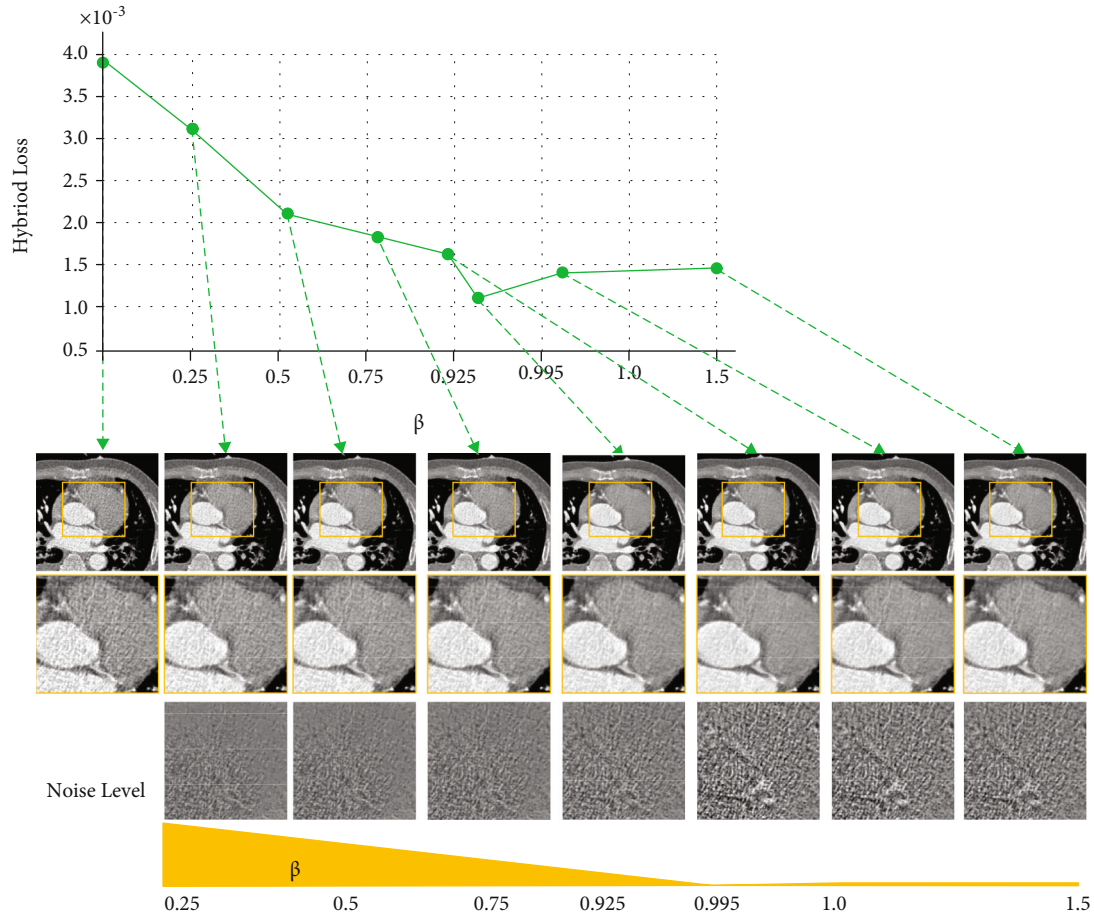


FIGURE 14: The effect of the parameter β in our hybrid loss function on the validation set from the MDLCT dataset. Under the detail images, the noise level of each CT image (the difference between each result and the LDCT image). From the left to the mid column ($[0, 0.995]$), the noise level of CT image gradually decreases as the β increases. From the mid column to the right column ($[0.995, 3.0]$), the noise level of CT image increases as the β increases.

weighting parameter for each loss item in our hybrid loss function, we often rely on our experimental experience. When we need to select the optimal parameters, first, we fix β , γ , and ω and select the optimal α . Then, we fix α , γ , and ω and determine the optimal β . The process of determining the optimal γ is the same as determined optimal β . Finally, we obtain the best value of ω based on optimal α , β , and γ . When choosing the value of parameter, we are used to measuring the denoising performance with different values, as shown in Figure 14. Here, we take the parameter β as an example and use the MSE as the metric.

The results demonstrate that the chosen parameters have influence on the denoising performance.

6. Conclusions

In this paper, we propose a novel framework for denoising low-dose CT images, which utilize noise learning and enhanced a SSWGAN with hybrid loss function, including adversarial loss, perceptual loss, sharpness loss, and structural similarity loss. First, in order to obtain a noise-free CT image, our generator can learn the noise distribution from the LDCT image and then reduce the noise from the

input. After training offline with pairs of the low-dose and normal-dose CT images, our method can reduce the noise of original CT images better than the state-of-the-art methods. In the future, we shall improve our network to obtain noise-free CT images of better quality by denoising the low-dose CT images.

Data Availability

No data were used to support this study.

Conflicts of Interest

There is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the Science and Technology Development Program of Jilin Province, China (nos. 20180201037SF, 20190201196JC, 20190302112GX, 20200404142YY, 20200403127SF, and 20200401078GX).




References

- [1] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.
- [2] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denoising for low-dose CT via wavelet residual network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1358–1369, 2018.
- [3] D. J. Brenner and E. J. Hall, "Computed tomography — an increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.
- [4] R. Smith-Bindman, J. Lipson, R. Marcus et al., "Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer," *Archives of Internal Medicine*, vol. 169, no. 22, pp. 2078–2086, 2009.
- [5] A. Berrington de González, M. Mahesh, K. P. Kim et al., "Projected cancer risks from computed tomographic scans performed in the United States in 2007," *Archives of Internal Medicine*, vol. 169, no. 22, pp. 2071–2077, 2009.
- [6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [7] A. Krizhevsky, I. I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [8] H. Shan, Y. Zhang, Q. Yang et al., "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.
- [9] H. Shan, A. Padole, F. Homayounieh et al., "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 269–276, 2019.
- [10] H. Chen, Y. Zhang, Y. Chen et al., "LEARN: learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333–1347, 2018.
- [11] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [12] H. Chen, Y. Zhang, W. Zhang et al., "Low-dose CT via convolutional neural network," *Biomedical Optics Express*, vol. 8, no. 2, pp. 679–694, 2017.
- [13] H. Chen, Y. Zhang, M. K. Kalra et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
- [14] D. Wu, K. Kim, G. El Fakhri, and Q. Li, "A cascaded convolutional neural network for X-ray low-dose CT image denoising," 2017, <http://arxiv.org/abs/1705.04267>.
- [15] F. Fan, H. Shan, M. K. Kalra et al., "Quadratic autoencoder (Q-AE) for low-dose CT denoising," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2035–2050, 2020.
- [16] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *European Conference on Computer Vision*, 2006, pp. 568–580, Graz, Austria, 2006.
- [17] M. Balda, J. Hornegger, and B. Heismann, "Ray contribution masks for structure adaptive sinogram filtering," *IEEE Transactions on Medical Imaging*, vol. 31, no. 6, pp. 1228–1239, 2012.
- [18] L. Ouyang, T. Solberg, and J. Wang, "Effects of the penalty on the penalized weighted least-squares image reconstruction for low-dose CBCT," *Physics in Medicine & Biology*, vol. 56, no. 17, pp. 5535–5552, 2011.
- [19] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Transactions on Medical Imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.
- [20] Y. Chen, D. Gao, C. Nie et al., "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Computerized Medical Imaging and Graphics*, vol. 33, no. 7, pp. 495–500, 2009.
- [21] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, and G. Wang, "Spectral CT reconstruction with image sparsity and spectral mean," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 510–523, 2016.
- [22] J. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao, "Cine cone beam CT reconstruction using low-rank matrix factorization: algorithm and a proof-of-principle study," *IEEE Transactions on Medical Imaging*, vol. 33, no. 8, pp. 1581–1591, 2014.
- [23] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, pp. 4777–4807, 2008.
- [24] Y. Liu, J. Ma, Y. Fan, and Z. Liang, "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction," *Physics in Medicine & Biology*, vol. 57, no. 23, pp. 7923–7956, 2012.
- [25] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, "Low-dose CT reconstruction via edge-preserving total variation regularization," *Physics in medicine and biology*, vol. 56, no. 18, pp. 5949–5967, 2011.
- [26] Z. Li, L. Yu, J. Trzasko et al., "Adaptive nonlocal means filtering based on local noise level for CT denoising," *Medical Physics*, vol. 41, no. 1, article 011908, 2014.
- [27] P. Fumene Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3D random noise filtering for absorption optical projection tomography," *Physics in Medicine & Biology*, vol. 55, no. 18, pp. 5401–5415, 2010.
- [28] D. Kang, P. Slomka, R. Nakazato, J. Woo, and D. Berman, "Image denoising of low-radiation dose coronary CT angiography by an adaptive block-matching 3D algorithm," in *Medical Imaging 2013: Image Processing*, Lake Buena Vista, Florida, USA, 2013.
- [29] G. Wang, M. Kalra, and C. G. Orton, "Machine learning will transform radiology significantly within the next 5 years," *Medical Physics*, vol. 44, no. 6, pp. 2041–2044, 2017.
- [30] W. Ge, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.
- [31] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: a review," *Medical Image Analysis*, vol. 58, article 101552, 2019.
- [32] H. Chen, Y. Zhang, W. Zhang et al., "Low-dose CT denoising with convolutional neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 143–146, Melbourne, Australia, 2017.

- [33] 章. Zhang Yungang, 易. Yi Benshun, 吴. Wu Chenyue, and 冯. Feng Yu, "Low-dose CT image denoising method based on convolutional neural network," *Acta Optica Sinica*, vol. 38, no. 4, article 0410003, 2018.
- [34] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," in *NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing System*, pp. 2672–2680, Montreal, Canada, 2014.
- [35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [36] M. S. Sajjadi, B. Schölkopf, and M. Hirsch, "EnhanceNet: single image super-resolution through automated texture synthesis," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4501–4510, Shenzhen, China, 2017.
- [37] M. Arjovsky, S. Chinatala, and L. Bottou, "Wasserstein generative adversarial networks," in *International Conference on Machine Learning (PMLR)*, pp. 214–223, Sydney, NSW, Australia, 2017.
- [38] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," 2017, <http://arxiv.org/abs/1704.00028>.
- [39] X. Li, C. Ye, Y. Yan, and Z. Du, "Low-dose CT image denoising method based on WGAN-gp," *Journal of New Media*, vol. 1, no. 2, pp. 75–85, 2019.
- [40] Q. Yang, P. Yan, Y. Zhang et al., "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.
- [41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, <http://arxiv.org/abs/1409.1556>.
- [42] C. You, W. Cong, G. Wang et al., "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, vol. 6, pp. 41839–41855, 2018.
- [43] D. Bau, J. Y. Zhu, J. Wulff et al., "Seeing what a GAN cannot generate," 2019, <http://arxiv.org/abs/1910.11626>.
- [44] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1520–1528, Santiago, Chile, 2015.
- [45] 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.
- [46] S. Rifai, P. Vincent, X. Müller, X. Glorot, and Y. Bengio, "Contractive auto-encoders: explicit invariance during feature extraction," in *ICML'11: Proceedings of the 28th International Conference on Machine Learning*, pp. 833–840, Bellevue, Washington, USA, 2011.
- [47] A. Makhzani and B. Frey, "K-sparse autoencoders," 2013, <http://arxiv.org/abs/1312.5663>.
- [48] D. P. Kingma and M. Welling, "Auto-encoding variational Bayers," 2013, <http://arxiv.org/abs/1312.6114>.
- [49] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, <http://arxiv.org/abs/1511.05644>.
- [50] Y. Chen and M. J. Zaki, "KATE: K-competitive autoencoder for text," in *KDD'17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 85–94, Halifax Ns, Canada, 2017.
- [51] A. Majumdar, "Graph structured autoencoder," *Neural Networks*, vol. 106, pp. 271–280, 2018.
- [52] W. Kunfeng, L. Xuan, Y. Lan, and W. Fei-Yue, "Generative adversarial networks for parallel vision," in *2017 Chinese Automation Congress*, Jinan, Shandong Province, China, 2018.
- [53] E. A. Burlingame, A. A. Margolin, J. W. Gray, and Y. H. Chang, "SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks," in *Medical Imaging 2018: Digital Pathology*, Houston, Texas, USA, 2018.
- [54] I. Goodfellow, "NIPS 2016 tutorial: generative adversarial networks," 2016, <http://arxiv.org/abs/1701.00160>.
- [55] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided CNN for image denoising," *Neural Networks*, vol. 124, pp. 117–129, 2020.
- [56] Y. Ma, B. Wei, P. Feng, P. He, X. Guo, and G. Wang, "Low-dose CT image denoising using a generative adversarial network with a hybrid loss function for noise learning," *IEEE Access*, vol. 8, pp. 67519–67529, 2020.
- [57] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2965–2972, Columbus, USA, 2014.
- [58] S. A. Goldstones and L. J. Karam, "Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 596–605, Hawaii, USA, 2017.
- [59] S. Zhuo and T. Sim, "Defocus map estimation from a single image," *Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011.
- [60] X. Yi and M. Eramian, "LBP-based segmentation of defocus blur," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1626–1638, 2016.
- [61] Z. Shi, J. Li, Q. Cao, H. Li, and Q. Hu, "Low-dose spectral CT denoising method via a generative adversarial network," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 1, pp. 1–10, 2020.
- [62] K. W. Clark, B. A. Vendt, K. E. Smith et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [63] X. Yi and P. Babyn, "Sharpness-aware low-dose CT denoising using conditional generative adversarial network," *Journal of Digital Imaging*, vol. 31, no. 5, pp. 655–669, 2018.
- [64] M. A. Gavrielides, L. M. Kinnard, K. J. Myers et al., "A resource for the assessment of lung nodule size estimation methods: database of thoracic CT scans of an anthropomorphic phantom," *Optics Express*, vol. 18, no. 14, pp. 15244–15255, 2010.
- [65] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, <http://arxiv.org/abs/1412.6980>.
- [66] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [67] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

Research Article

Algorithm of Pulmonary Vascular Segment and Centerline Extraction

Shi Qiu ¹, Jie Lian,² Yan Ding,³ Tao Zhou ⁴, and Ting Liang ^{5,6}

¹Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China

²Department of Pathology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

³Changshu Institute of Technology, Suzhou 215500, China

⁴School of Computer Science and Engineering, North Minzu University, 750021 Yinchuan, China

⁵Department of Radiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

⁶Science and Technology Department, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

Correspondence should be addressed to Ting Liang; liangting31500@126.com

Received 21 June 2021; Revised 22 July 2021; Accepted 12 August 2021; Published 26 August 2021

Academic Editor: Hamidreza Mohafez

Copyright © 2021 Shi Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because pulmonary vascular lesions are harmful to the human body and difficult to detect, computer-assisted diagnosis of pulmonary blood vessels has become the focus and difficulty of the current research. An algorithm of pulmonary vascular segment and centerline extraction which is consistent with the physician's diagnosis process is proposed for the first time. We construct the projection of maximum density, restore the vascular space information, and correct random walk algorithm to satisfy automatic and accurate segmentation of blood vessels. Construct a local 3D model to restrain Hessian matrix when extracting centerline. In order to assist the physician to make a correct diagnosis and verify the effectiveness of the algorithm, we proposed a visual expansion model. According to the 420 high-resolution CT data of lung blood vessels labeled by physicians, the accuracy of segmentation algorithm AOM reached 93%, and the processing speed was 0.05 s/frame, which achieved the clinical application standards.

1. Introduction

In recent years, the air is in poor quality, which seriously affects people's life and health, resulting in an increasing incidence of lung cancer. Because the early symptoms of lung cancer are relatively minor, it is not easy to be discovered, and 80% of lung cancer patients are already in the middle and late clinical stage when they are confirmed, missing the best period of radical operation. If the lung cancer can be detected earlier and get differential diagnosis by category, it can receive standard treatment before the lesion spread, the patient's five-year survival rate can reach more than 60% [1, 2]. If the lung CT data which intuitively reflect the lung condition can be accurately analyzed, lung disease can also be prevented and diagnosed, and more human life can be saved. Pulmonary vascular lesions are one of the early signs of lung cancer, so accurate division and observation of pul-

monary vascular areas are very important for diagnosis and treatment [3].

Due to the small morphological structure of pulmonary blood vessels and complex distribution environment, the process of imaging is easily affected by noise and surrounding tissues. This has brought great difficulty to the complete extraction of lung blood vessels, which is also the focus of the current research. According to the characteristics of the vascular section obeying Gaussian distribution, Foncubierta-Rodríguez et al. [4] constructed an enhancement model of vascular segmentation to determine the direction of vascular growth and extract blood vessels. According to the vascular anatomy, Orkisz et al. [5] calculated the global threshold to divide the lung image, and then, the blood vessels are extracted by discriminating the connectivity. Based on the two-dimensional CT image, Ibrahim et al. [6] selected the initial point manually, calculated the

blood vessel position step by step, and then segmented the blood vessel. Lai et al. [7] enhanced the vascular response by filtering the image from a 3D angle. Gong et al. [8] extracted the blood vessels by constructing the self-adaptive threshold according to the difference between the gray level of the blood vessel and the surrounding area. Charbonnier et al. [9] constructed a matching model in accordance with blood vessels extending throughout the lungs to segment the blood vessels. Based on the local feature model, Gupta et al. [10] predicted the blood vessel direction. According to the samples of axial position, coronal position, and sagittal position, Phellan et al. [11] constructed a CNN model to segment the vascular area. Lidayová et al. [12] established the model of normal blood vessel and abnormal blood vessel to realize the segmentation of blood vessel. Jawaid et al. [13] improved level set to extract the vascular boundary information. Xiao et al. [14] used the grayscale and shape characteristics to constrain and segment blood vessels. Flórez and Orkisz [15] established the cylindrical model to simulate the distribution of blood vessels. Hu et al. [16] improved the crossaccess law function to extract blood vessels.

Because it is difficult for doctors to observe small changes of blood vessels through the naked eye to make accurate diagnosis, computer-assisted diagnosis is required. Based on the above extraction of the vascular centerline, the mainstream diagnostic method is as follows: adjust the window width and window position [17] when blood vessels are observed, the 3D model [18] shows the overall external morphology of blood vessels, a virtual endoscope [19] shows the inside of the blood vessels. The above algorithms cannot give the internal and external characteristics of blood vessels at the same time, which limited the ability of helping physician to diagnose.

The main problems of vascular division, centerline extraction, and auxiliary diagnosis can be summarized as follows: (1) the 3D model of extracting vascular segment is complex, and it needs a large amount calculation, so it cannot be calculated and displayed in real time; (2) it is difficult to divide the vascular area accurately just by gray scale; and (3) there is no accurate and intuitive way to show the internal and external conditions of blood vessels.

For those reasons, we study the process of doctors' diagnosis of vascular lesions and construct a diagnostic model for doctors' vision diagnosis according to the theory of medicine, anatomy, and image graphics: (1) establish a maximum density projection model based on vascular anatomy and imaging principles and significantly reduce the amount of data while retaining the area where the blood vessels are located; (2) adaptive random walk algorithm is constructed by combining gray level with local information; and (3) construct the mechanism of vascular centerline extraction and vascular expansion and display and diagnose vascular from multiangle.

2. Details of Algorithm

When physicians detect pulmonary vascular lesions, removing the interference of the examination bed, muscles, and

soft tissues is the first task. They focus on the lung parenchyma and then determine the starting and ending points of the blood vessels to be observed based on the local images of limited frames, achieving the extraction of blood vessels. Finally, they extract the vascular centerline and give diagnosis based on the anatomical knowledge.

The computer simulates how doctors diagnose lesions as follows: the lung image sequence was firstly pretreated to remove noise and check a bed, and the image was simplified by extracting the lung parenchyma area. Then, it determined the starting and ending points of blood vessels to be observed using global and local models to constrain them and took the characteristics of anatomy and image graphics into consideration to realize the extraction of blood vessels' central line. Figure 1 shows the flow of pulmonary vascular CAD segmentation.

2.1. Pretreatment. CT images can intuitively show the morphological structure of human tissues (blood vessels, musculoskeletal, etc.). According to the physician's diagnosis process, the blood vessels that physician concerned only exist in the lung parenchyma. The ends of the blood vessels are tiny; they are only displayed as a limited number of pixels in the image and are vulnerable to noise. Therefore, the image needs to be preprocessed.

2.1.1. Noise Reduction. CT image always includes pepper and salt noise [20] which presented as a bright discrete pixel; it interferes with the extraction of pulmonary blood vessels' tip and needs to be removed. According to the characteristics that noises are distributed as discrete points, a two-dimensional median filtering algorithm can be used to remove them. The formula is as follows:

$$g(x, y) = \text{med} \{f(x - k, y - m), (k, m \in W)\}, \quad (1)$$

where $f(x, y)$ and $g(x, y)$ are the gray values of the pixels with coordinates (x, y) in the original image and the processed image, respectively; W is a two-dimensional filter template; and med is a median filter function.

2.1.2. Lung Parenchyma Extraction. The area of blood vessels that doctors care about is inside the lung parenchyma. Whether the extraction of the blood vessels is good or bad directly affects the subsequent diagnosis. Therefore, the computer needs to focus on the area where the pulmonary nodules are located. By analyzing the statistical distribution of lung image pixels, Qiu et al. [21] constructed an optimized Gaussian dual-mixing model to calculate the global segmentation threshold quickly, and the complete lung parenchyma region can be extracted. In order to meet the clinical application standards of accuracy and processing speed, the algorithm in literature [21] is applied to extract the lung parenchyma region and confirm the left and right lung regions.

2.1.3. Set the Starting and Ending Points of Blood Vessels. When the doctor identifies the focal point, he will determine the observed vascular area through clinical needs. When the computer simulates this process, according to the

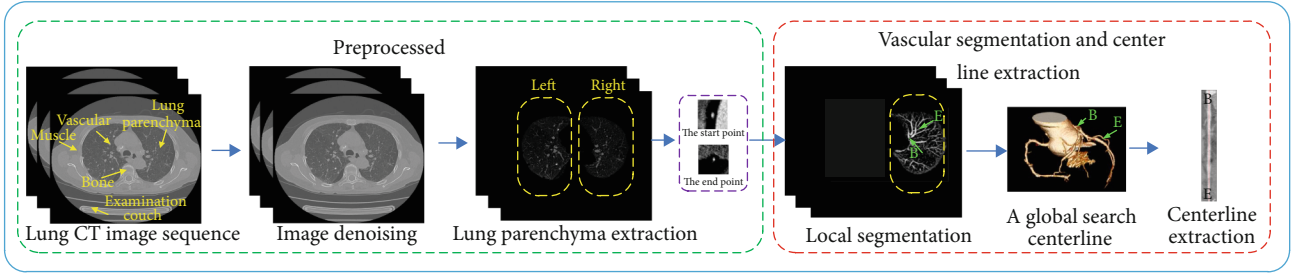


FIGURE 1: Algorithm flow chart.

morphological structure of blood vessels, it sets the blood vessel's starting point P_b and the end point P_e whose corresponding coordinates are (x_b, y_b, z_b) and (x_e, y_e, z_e) on the computer image through human-computer interaction, and $z_b \leq z_e$.

2.2. Rough Extraction by a Local 3D Model. The physician only cares about the blood vessel between point P_b and point P_e , so the blood vessel area needs to be extracted.

Since blood vessels are of connectivity, as shown in Figure 2, only handle $[z_b, z_b + 1, \dots, z_e]$ is not rigorous, where m is the point on the blood vessel.

Anatomically, blood vessels are spatially continuous. From the perspective of image imaging, blood vessels appear as bright and isolated circular areas in two-dimensional CT images.

In order to restore the anatomical characteristics of blood vessels, we introduce the maximum density projection which is integrated with the local 3D information to greatly reduce the amount of data while ensuring the information of blood vessels.

The maximum intensity projection is a method for 3D data that projects in the visualization plane the voxels with maximum intensity that fall in the way of parallel rays traced from the viewpoint to the plane of projection. The equation is as follows:

$$\begin{aligned} \mathbf{MIP}_n(x, y) &= \max (I_{k+1}(x, y) \cdots I_{k+\text{SN}_r}(x, y)), \\ 1 &\leq x \leq H, \\ 1 &\leq y \leq W, \\ k &= (n-1) \times \text{SN}_r, \\ n &= 1, 2 \cdots \frac{\text{Nm}}{\text{SN}_r}, \end{aligned} \quad (2)$$

where $\mathbf{MIP}_n(x, y)$ is the grayscale value at the midpoint (x, y) of the n th frame of the MIP image; H and W are the horizontal and vertical resolutions, respectively; SN_r is the number of projected layers; and Nm is the sum of original CT layers. $I_k(x, y)$ is the grayscale value at the point (x, y) in the k th layer of the original CT sequence images. Figure 3 is a schematic diagram of the MIP projection of the two-dimensional CT data.

The blood vessels in the MIP image are locally highlighted and continuous areas, as shown in Figure 4(a).

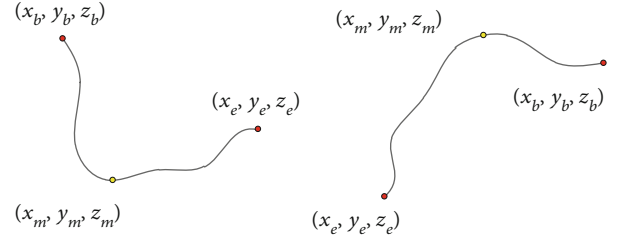


FIGURE 2: Vascular route diagram.

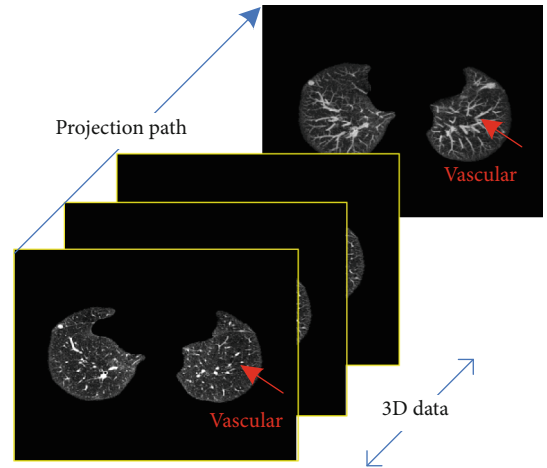


FIGURE 3: MIP schematic.

Subsequent extraction of vascular areas is needed to constrain the vascular centerline method and reduce the amount of data. In the MIP image, P_b and P_e are chosen as seed points, The area of the blood vessel is obtained by using C-means clustering [22, 23], and then, the image of the suspected area I_i as Figure 4(b) is obtained through the 8 connected areas. Follow-up study is based on I_i .

2.3. Segment by Automatic Random Walk Algorithm. Random walk (RW) algorithm [24] regarded the two-dimensional image as a connected undirected weighted graph containing fixed vertices and edges. The unlabeled pixels begin to walk along the edge from the vertices. According to the maximum probability of each pixel arriving at each labeled pixel, the class of each vertex is judged.

The main steps are (1) manually mark the seed point; (2) establish a random walk model $G = (V, E)$, where V is the set of vertices of the image and E is the set of undirected

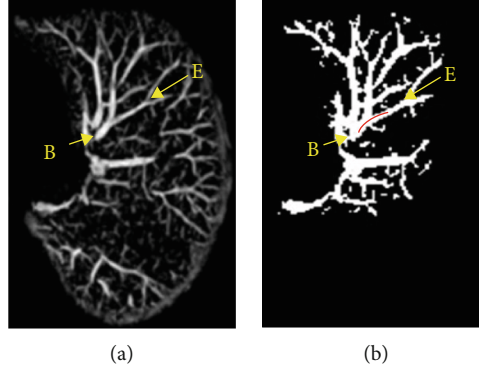


FIGURE 4: MIP image processing: (a) MIP image; (b) I_i .

edges at any two vertices of the image; (3) calculate weights; and (4) compute the distribution of probability and get the segmentation.

In recent years, Dong et.al [25] introduced subMarkov to realize image segmentation. Zheng et al. [26] constrain RW according to target characteristics. Bui et al. [27] established a 3D model to realize the heart division. All of the above methods get good results. As blood vessels run through multiple layers of images, it is inefficient to manually mark seed points on each layer of images. In a 2D image, blood vessels are shown as limited sections of pixels, and the number of pixels is small, so the accuracy of seed selection will directly affect the segmentation effect. For this reason, we study the selection of seed and weight calculation and then propose an algorithm.

2.3.1. Adaptive 3D Seed Selection. The traditional RW algorithm segments the image semiautomatically. The user needs to manually mark the target seed and background seed on the 2D image, and then, the probability of each vertex to these two kinds of seed is used to judge which one they belong to, thus dividing the image into the target area and the background area.

The blood vessel exists on the CT image as a circular or cylindrical section. When using the traditional RW algorithm, the physician needs to spend a lot of time on accurately marking the seed on each 2D image. Besides, the number of blood vessel pixels is limited; the physician may mark the wrong seed because of a slight deviation, which cannot guarantee the completeness of the segmentation.

According to the characteristics of angiology, anatomy, and CT imaging, we optimize the process of selecting seeds as Figure 5. From the medical point of view, the components contained in the blood vessels are uniform, so the pixel values are similar in CT image. Anatomically, it is a continuous region in a 3D space. From the aspect of imaging, blood vessels are locally bright area.

The approximate area of the blood vessel in the 3D space is marked as I_i , with the point P_i ($P_0 = P_b$) as the target seed and the nonvascular region in I_i as the background seed, and the divided blood vessels of the current layer was recorded as C_i . According to the continuity of vascular, $C_i \cap C_{i+1} \neq \emptyset$, according to the uniformity of blood composition, the pixel value of the blood vessel should change a little; then, the coordinate of the central point of $C_i \cap C_{i+1}$ is assigned as P_{i+1} .

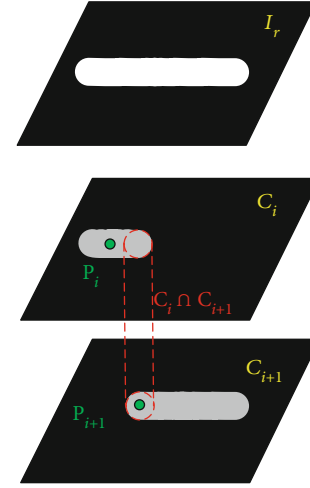


FIGURE 5: Seed selection process.

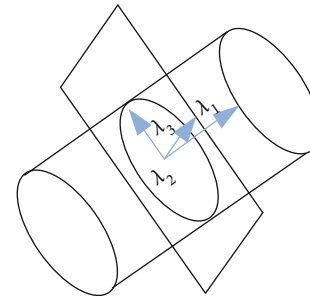


FIGURE 6: Directional diagram of feature vector.

2.3.2. Calculate Weights. The weights determine that the probability of each vertex belongs to different classes. The traditional RW algorithm only considers the grayscale information between pixels and does not consider the geometric information between pixels. For more accurate segmentation, the improved weight function is

$$w_{i,j} = \exp(-\alpha|g(v_i) - g(v_j)| + (1-\alpha)|h(v_i) - h(v_j)|),$$

$$L_{i,j} = \begin{cases} d_i, & i = j, \\ -w_{i,j}, & v_i \text{ is adjacent to } v_j, \\ 0, & \text{others,} \end{cases} \quad (3)$$

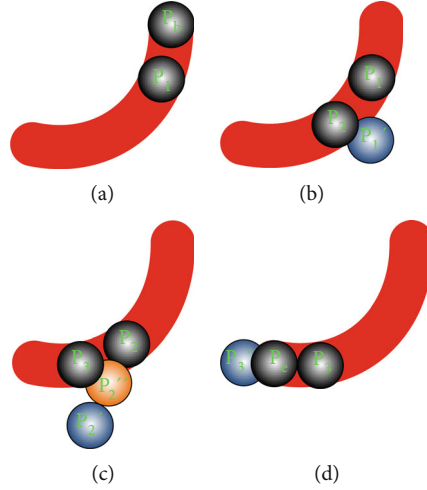


FIGURE 7: Vascular division: (a) Case 1, (b) Case 2, (c) Case 3, and (d) Case 4.

where $g(\cdot)$ represents the grayscale, $h(\cdot)$ represents the coordinate value, α is the weight of grayscale, $(1 - \alpha)$ is the weight of distance, and $L_{i,j}$ is the matrix of distance. RW algorithm is rebuilt to realize the automatic segmentation of blood vessels. The selected seeds are few, the calculation area is small, and the speed is fast. It reduces the workload of doctors and improves the diagnostic efficiency.

2.4. Centerline Extraction and Expansion. After the doctor confirms the vascular area, he reconstructed the 3D characteristics of the blood vessels and set up the center point to observe the pathological conditions of the blood vessels.

Because the spacing of CT data layers and the image resolution are not uniform, in order to restore the true situation of the lungs, the computer needs to interpolate the data to develop the isotropic data; that is, the lung data observed from any direction are uniform. According to MIP image processing, it can be seen that the space of blood vessels is limited, so it needs to be interpolated. The gray scale of the point P_i is used as the seed to cluster on the corresponding layer of the original image sequence to obtain the blood vessel area, whose center and the inscribed circle's radius R are calculated. Then, the point P_i in the image is seen as the center of the ball with a radius of $2R$; the data sequence is interpolated vertically according to the resolution of the axis to form isotropic data.

Calculate the three eigenvalues $\lambda_1, \lambda_2,$ and λ_3 and the eigenvectors $v_1, v_2,$ and v_3 , which represent the trend of movement in different directions, as shown in Figure 6. Take v_1 as the blood vessel direction and $2R$ as the search radius; make some adjustment R to find the position of next blood vessel to obtain a complete blood vessel area.

In the process of the calculation, the following four situations occur, as shown in Figure 7, in which the red strip is a blood vessel and the balls show the adjustment process of centerline.

Case 1. Start off from P_b ; calculate v_1 . In the direction of v_1 , jump with the step $2R_0$. Calculate the coordinates of the center of ball's gravity P_1 and radius R_1 and save it.

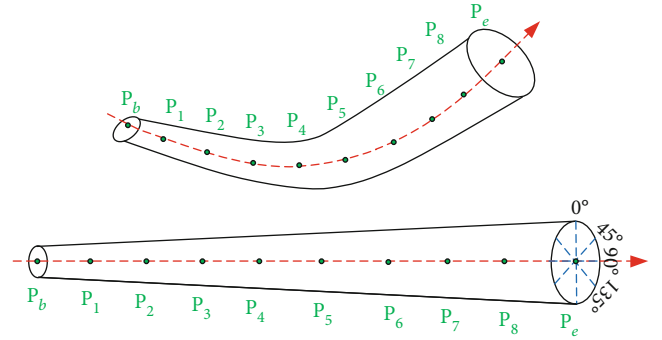


FIGURE 8: Vascular centerline expansion diagram.

Case 2. Start off from P_1 ; calculate v_1 . In the direction of v_1 , jump to P'_1 with the step $2R_1$. Constantly adjust the center of gravity and radius until they do not change; save the center of gravity P_2 and radius R_2 .

Case 3. Start off from P_2 ; calculate v_1 . In the direction of v_1 , jump to P'_2 with the step $2R_2$, and the circle centered on P'_2 with the radius R_2 which is not in the MIP image, indicating that the step length is too large, which needs to be adjusted. When the cross section of circles reaches 10%, the step length is set to R'_2 . The coordinates at this point are P''_2 . Adjust the center of gravity and radius continuously until they do not change; save the center of gravity P_3 and radius R_3 .

Case 4. Start off from P_3 ; calculate v_1 . In the direction of v_1 , jump to P'_3 with the step $2R_3$. If P_e is exceeded, take P_e as the center of the circle and adjust the length of the step, and the center of gravity of the ball is calculated and saved.

When the physician determines the vascular centerline, he needs to observe the internal and external conditions of the blood vessel; this process is simulated by the computer as follows: first, the blood vessels are straightened according to the centerline, and then, according to the anatomical principles, the blood vessels are expanded according to

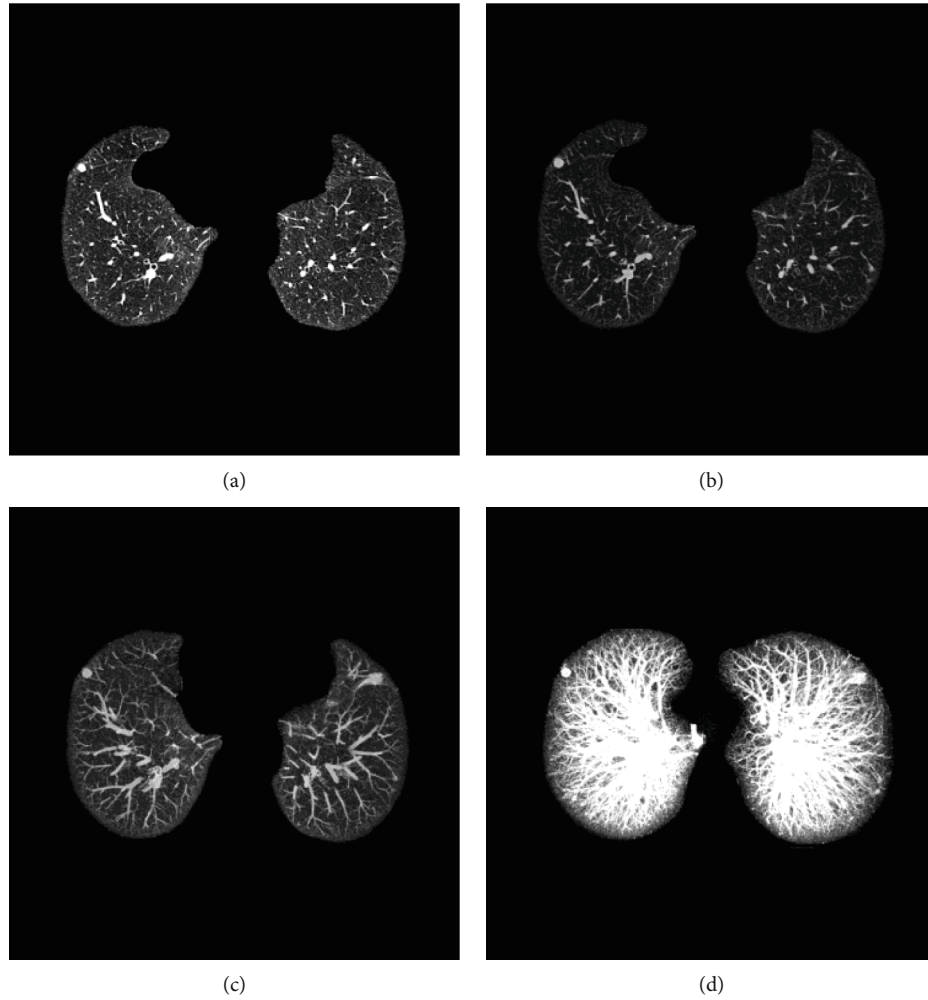


FIGURE 9: The effect of reconstructing MIP diagram: (a) original image, (b) 3 mm, (c) 9 mm, and (d) entire sequence.

different profiles (0° , 45° , 90° , and 135°). The internal and external conditions of different profiles are observed to give a comprehensive judgment. Its schematic diagram is shown in Figure 8.

3. Experiments and Result Analysis

The experiment used 50 sets of CT data with different resolutions and thickness collected in 2020; 420 blood vessel areas were extracted. The programs were compiled in VS2015 with the WIN7 operating system.

3.1. Verify the Effect of MIP Algorithm. The same group of CT sequences was reconstructed as the layer thickness is 3 mm, 9 mm, and the entire sequence, respectively, by the MIP algorithm and compared with the original sequence. The effect is shown in Figure 9. The MIP algorithm can restore the strip shape of vascular and reduce the risk of wrong detection. With the increase of layer thickness, it can display the distribution of peripheral blood vessels more clearly and ensure that the follow-up extraction of vascular centerlines is more complete.

TABLE 1: Quantitative evaluation of vascular segmentation.

Algorithm	AOM%	Average time (s/frame)
CE	73	0.03
3D	79	0.40
MD	83	0.07
LS	86	0.10
RW	85	—
Ours	93	0.05

TABLE 2: Qualitative evaluation of vascular segmentation.

Algorithm	Accurate	General	Poor
CE	223	139	58
3D	258	125	37
MD	265	136	19
LS	280	119	21
RW	225	159	36
Ours	312	93	15

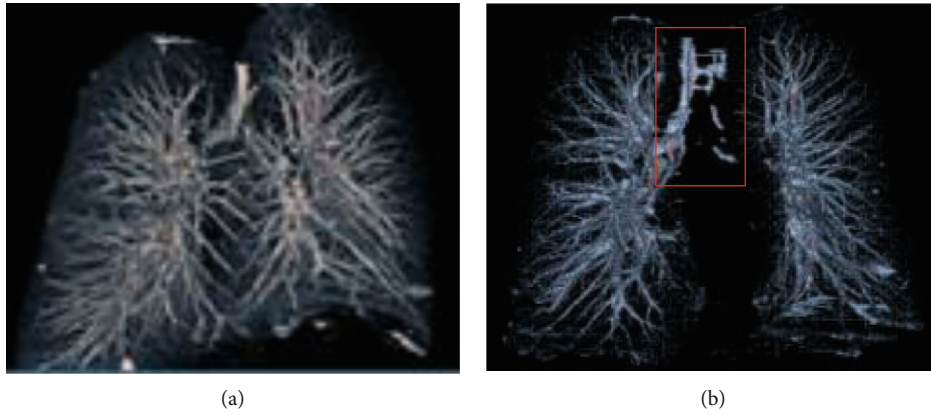


FIGURE 10: Full sequence segmentation results. (a) Segmentation results are “accurate.” (b) Segmentation results are “poor.”

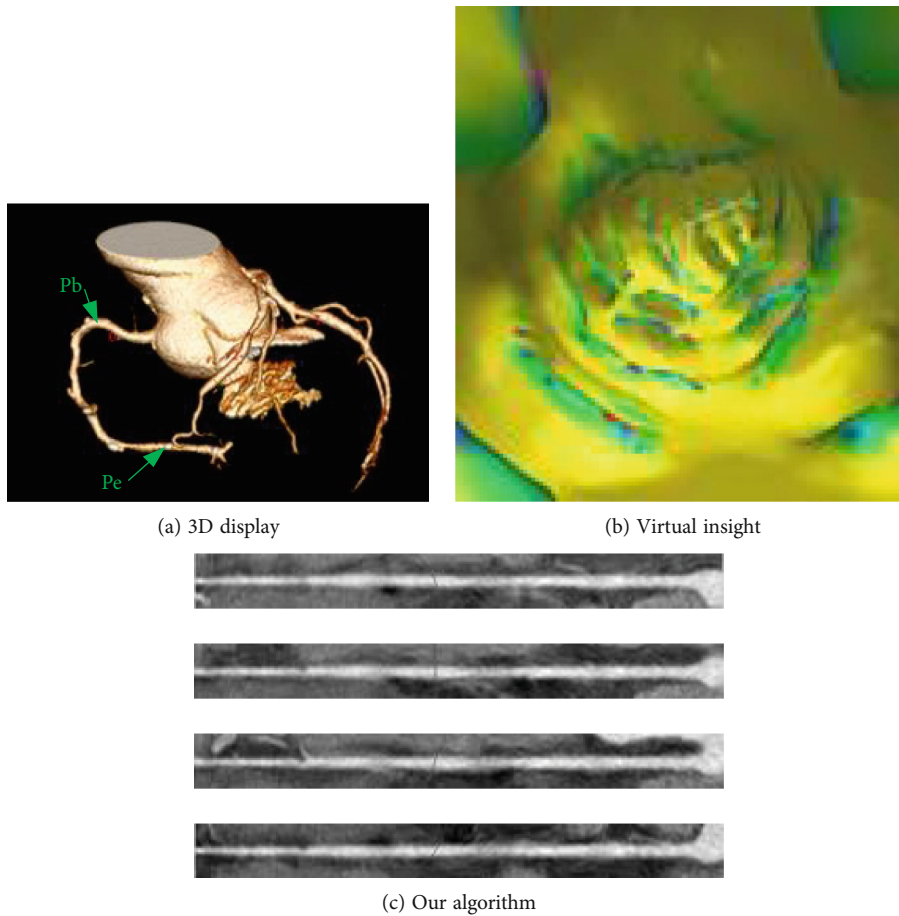


FIGURE 11: Performance of the algorithm.

3.2. *The Effect of Vascular Segmentation.* To verify the effect of blood vessel extraction, three experts who have been engaged in medical imaging for many years were asked to mark blood vessels to verify the effect of the algorithm from the quantitative and qualitative levels.

The quantitative aspects are as follows: take area overlap measure (AOM) [28] as an evaluation indicator of the seg-

mentation effect; it is defined as

$$AOM(A, B) = \frac{S(A \cap B)}{S(A \cup B)} \times 100\%, \quad (4)$$

where AOM is the area overlap measure, A is the image marked by the doctor, B is the image segmented by

computer, and $S(\cdot)$ represents the number of pixels in the corresponding area. The larger the AOM value is, the better effect the segmentation gets. The results are shown in Table 1.

The qualitative aspects are as follows: using medical diagnostic marks: “accurate” means that the vascular division is complete and almost identical with the area marked by experts; “general” indicates that there is a some deviation between the results of the two, but it does not affect the diagnosis; and “poor” indicates that there is a large deviation between the results of the two, which affects the diagnosis. The results are shown in Table 2.

This paper uses the Contrast-Enhanced (CE) algorithm [4] from both qualitative and quantitative aspects to construct a vascular enhancement model and achieve vascular segmentation. 3D algorithm [7] established a 3D model from the inside of the blood vessels. Multidirectional (MD) algorithm [11] observed the blood vessels from the axial, coronal, and sagittal direction, determined the direction of blood vessels, and then segmented them. Level set (LS) algorithm [13] transformed the segmentation problem into the problem of internal and external forces’ balance. RW [24] selected the seed points manually on each layer of images to achieve vascular segmentation, not considering the time spent. Our algorithm takes full account of vascular anatomy and imaging principles, proposes an automatic selection algorithm for seed points, and optimizes the weight function; although a little more time was spent than CE [4] algorithm, the segmentation effect AOM reached 93%, and the effect that achieves the “general” was the best, thus confirming the effectiveness of the proposed algorithm.

To demonstrate the effect of our algorithm, we extract the whole pulmonary blood vessels. It is very difficult. The algorithm proposed in this paper can extract most of the blood vessels, but there are still some incomplete blood vessels extracted. As shown in Figure 10(b), the blood vessels are not completely extracted in the area where they enter the lungs. The main reason is that the left and right lungs of this sequence are relatively close. During the CT imaging process, the left and right lungs are not effectively divided. Further research is needed on this issue.

3.3. The Extraction of Vascular Centerline. In order to verify the effect of the extraction of vascular centerline and show the observation effect of the current major assisted diagnosis method, our algorithm, 3D display, and virtual insight are compared. The 3D display is shown in Figure 11(a); only the external condition of the blood vessel can be observed; the virtual insight is shown in Figure 11(b); only the internal condition of the blood vessel can be observed; and our algorithm is shown in Figure 11(c), which shows the inside and outside of the blood vessel at the same time and intuitively shows the effect of extraction of vascular centerline.

In order to intuitively demonstrate the effect of vascular expansion, we chose unbranched blood vessels, branching blood vessels, and blocked blood vessels, which were displayed at 0° , 45° , 90° , and 135° . Unbranched blood vessels, as shown in Figure 12, are spread in different directions; the centerline is in the center of the blood vessel. For branch-

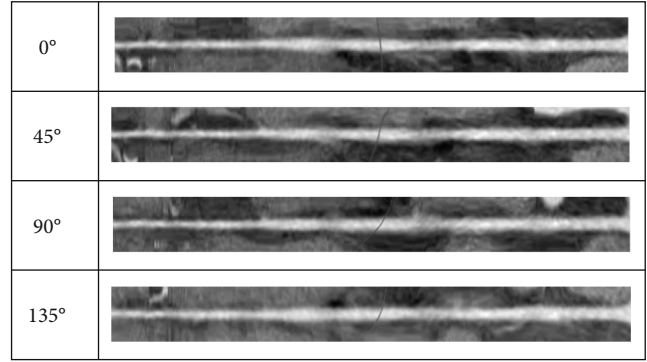


FIGURE 12: Effect of spreading unbranched blood vessel.

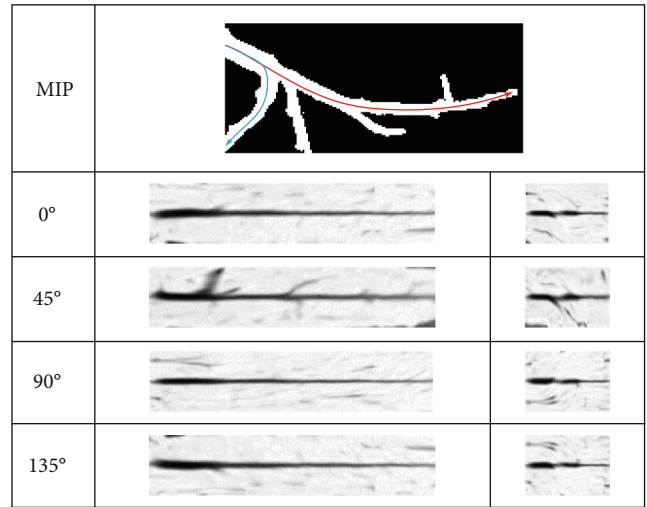


FIGURE 13: Effects of spreading branching blood vessels.

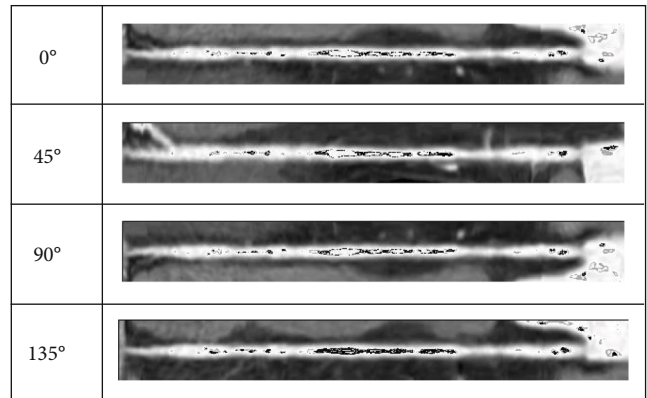


FIGURE 14: Effect of spreading blocked blood vessels.

ing blood vessels, as shown in Figure 13, the longest blood vessel (red) and the shortest blood vessel (blue) are selected for straightening. It can be seen that this algorithm is not affected by the blood vessels next to the branch when extracting the central line. The central line can be extracted accurately. Figures 12 and 13 show the effects of normal vascular centerline’s extraction and spread, and the interior of blood vessel is relatively smooth. Figure 14 shows the

spreading effect of blocked blood vessel. It can be seen that the algorithm can accurately extract the centerline, and a lot of blocked spots can be observed inside the blood vessels. From the above demonstration, the algorithm presented in this paper can accurately extract the vascular centerline and directly display the internal and external conditions of blood vessels, so it can assist the physician to make an accurate diagnosis.

4. Conclusion

Computer-aided detection of pulmonary blood vessels has become a hot topic and difficulty in research. In this paper, the segmentation algorithm of pulmonary vascular CT image is proposed, which combines local 3D information and enhances vascular area. Optimize the random walk algorithm to meet the requirements of separating blood vessel accurately. We applied the knowledge of pathology, anatomy, and image graphics to the vascular centerline extraction and proposed the method of spreading blood vessels which can show the internal and external characteristics of blood vessels more clearly, thus assisting physicians in making accurate diagnoses. It is of great value on clinical application and lays a good foundation for the follow-up detection of vascular lesions. However, the vascular division is incomplete when the left and right lungs are near, and further research is needed.

Data Availability

Data is available at International Early Lung Cancer Action Project (DB/OL) (<http://www.via.cornell.edu/lungdb.html>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the Science and Technology Rising Star of Shaanxi Youth (2021KJXX-61), China Postdoctoral Science Foundation (2020M682144), the Shanxi National Science Foundation (2020JQ-518), and the National Natural Science Foundation of China (62062003).

References

- [1] S. Smith and P. Jennings, "Thoracic intervention and surgery to cure lung cancer: image-guided thermal ablation in primary lung cancer," *Journal of the Royal Society of Medicine*, vol. 112, no. 6, pp. 218–225, 2019.
- [2] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The ensemble deep learning model for novel COVID-19 on CT images," *Applied Soft Computing*, vol. 98, article 106885, 2021.
- [3] S. Qiu, J. Li, M. Cong, C. Wu, Y. Qin, and T. Liang, "Detection of solitary pulmonary nodules based on brain-computer interface," *Computational and Mathematical Methods in Medicine*, vol. 2020, 10 pages, 2020.
- [4] A. Foncubierta-Rodríguez, A. Widmer, A. Depeursinge, and H. Müller, "Enhanced visualization of pulmonary perfusion in 4D dual energy CT images," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6710–6713, Chicago, IL, USA, August 2014.
- [5] M. Orkisz, M. Hernández Hoyos, V. Pérez Romanello, C. Pérez Romanello, J. Prieto, and C. Revol-Muller, "Segmentation of the pulmonary vascular trees in 3D CT images using variational region-growing," *IRBM*, vol. 35, no. 1, pp. 11–19, 2014.
- [6] G. Ibrahim, A. Rona, and S. V. Hainsworth, "Non-uniform central airways ventilation model based on vascular segmentation," *Computers in Biology and Medicine*, vol. 65, pp. 137–145, 2015.
- [7] J. Lai, Y. Huang, Y. Wang, and J. Wang, "Three-dimensions segmentation of pulmonary vascular trees for low dose CT scans," *Sensing and Imaging*, vol. 17, no. 1, pp. 1–15, 2016.
- [8] Z. Gong, W. Tan, Y. Song et al., "Segmentation of pulmonary blood vessels based on fractional differentiation in 3D chest CT images," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 7, pp. 1746–1749, 2016.
- [9] J. Charbonnier, M. Brink, F. Ciompi, E. Scholten, C. Schaefer-Prokop, and E. M. van Rikxoort, "Automatic pulmonary artery-vein separation and classification in computed tomography using tree partitioning and peripheral vessel matching," *IEEE Transactions on Medical Imaging*, vol. 35, no. 3, pp. 882–892, 2016.
- [10] G. Gupta, S. Kulasekaran, K. Ram, N. Joshi, M. Sivaprakasam, and R. Gandhi, "Local characterization of neovascularization and identification of proliferative diabetic retinopathy in retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 55, pp. 124–132, 2017.
- [11] R. Phellan, A. Peixinho, A. Falcão, and N. Forkert, "Vascular segmentation in TOF MRA images of the brain using a deep convolutional neural network," in *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pp. 39–46, Springer, Cham, 2017.
- [12] K. Lidayová, H. Frimmel, E. Bengtsson, and Ö. Smedby, "Improved centerline tree detection of diseased peripheral arteries with a cascading algorithm for vascular segmentation," *Journal of Medical Imaging*, vol. 4, no. 2, article 024004, 2017.
- [13] M. Jawaid, S. Zai, and M. Ansari, "Improving vascular segmentation by geometric fitting in level set evolution," in *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)*, pp. 1–6, Karachi, Pakistan, April 2017.
- [14] R. Xiao, H. Ding, F. Zhai, T. Zhao, W. Zhou, and G. Wang, "Vascular segmentation of head phase-contrast magnetic resonance angiograms using grayscale and shape features," *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 157–166, 2017.
- [15] L. Flórez-Valencia and M. Orkisz, "Right generalized cylinder model for vascular segmentation," in *Computing and Visualization for Intravascular Imaging and Computer-Assisted Stenting*, pp. 131–156, Academic Press, 2017.
- [16] K. Hu, Z. Zhang, X. Niu et al., "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function," *Neurocomputing*, vol. 309, pp. 179–191, 2018.
- [17] H. Shmilovich, V. Cheng, D. Dey et al., "Optimizing image contrast display improves quantitative stenosis measurement in heavily calcified coronary arterial segments on coronary

- CT angiography: a proof-of-concept and comparison to quantitative invasive coronary angiography,” *Academic Radiology*, vol. 21, no. 6, pp. 797–804, 2014.
- [18] J. Touati, M. Bologna, A. Schwein, F. Migliavacca, and M. Garbey, “A robust construction algorithm of the centerline skeleton for complex aortic vascular structure using computational fluid dynamics,” *Computers in Biology and Medicine*, vol. 86, pp. 6–17, 2017.
- [19] W. Li, R. Germain, and M. Gerner, “Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D),” *Proceedings of the National Academy of Sciences*, vol. 114, no. 35, pp. E7321–E7330, 2017.
- [20] K. Dimililer, B. Ugur, and Y. Ever, “Tumor detection on CT lung images using image enhancement,” *The Online Journal of Science and Technology*, vol. 7, no. 1, pp. 133–138, 2017.
- [21] S. Qiu, D. Wen, Y. Cui, and J. Feng, “Lung nodules detection in CT images using Gestalt-based algorithm,” *Chinese Journal of Electronics*, vol. 25, no. 4, pp. 711–718, 2016.
- [22] J. Gu, L. Jiao, S. Yang, and F. Liu, “Fuzzy double c-means clustering based on sparse self-representation,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 612–626, 2018.
- [23] M. Yang and Y. Nataliani, “Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters,” *Pattern Recognition*, vol. 71, pp. 45–59, 2017.
- [24] L. Grady, “Random walks for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.
- [25] X. Dong, J. Shen, and L. Van, “Segmentation using subMarkov random walk,” in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 237–248, Springer, Cham, 2015.
- [26] Y. Zheng, D. Ai, P. Zhang et al., “Feature learning based random walk for liver segmentation,” *PloS One*, vol. 11, no. 11, article e0164098, 2016.
- [27] V. Bui, L. Hsu, L. Chang, and M. Chen, “An automatic random walk based method for 3D segmentation of the heart in cardiac computed tomography images,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1352–1355, Washington, DC, USA, 2018.
- [28] S. Qiu, K. Cheng, L. Cui, D. Zhou, and Q. Guo, “A moving vehicle tracking algorithm based on deep learning,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–7, 2020.

Research Article

Comparison of Diagnostic Performance between Perfusion-Related Intravoxel Incoherent Motion DWI and Dynamic Contrast-Enhanced MRI in Rectal Cancer

Ming Li,¹ Xiaodan Xu,² Kaijian Xia,³ Heng Jiang,⁴ Jianlong Jiang,¹ Jinbing Sun ¹ and Zhihua Lu ⁴

¹Department of General Surgery, Changshu Hospital Affiliated to Soochow University, Changshu No.1 People's Hospital, Changshu, 215500 Jiangsu Province, China

²Department of Gastroenterology, Changshu Hospital Affiliated to Soochow University, Changshu No.1 People's Hospital, Changshu, 215500 Jiangsu Province, China

³Department of Central Laboratory, Changshu Hospital Affiliated to Soochow University, Changshu No.1 People's Hospital, Changshu, 215500 Jiangsu Province, China

⁴Department of Radiology, Changshu Hospital Affiliated to Soochow University, Changshu No.1 People's Hospital, Changshu, 215500 Jiangsu Province, China

Correspondence should be addressed to Jinbing Sun; [sunjb515@163.com](mailto:sunj515@163.com) and Zhihua Lu; luzhihua4100@suda.edu.cn

Received 11 May 2021; Accepted 9 July 2021; Published 30 July 2021

Academic Editor: Reza Khosrowabadi

Copyright © 2021 Ming Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study was aimed to determine the diagnostic performance of perfusion-related parameters derived from intravoxel incoherent motion diffusion-weighted imaging (IVIM-DWI) by comparing them with quantitative parameters from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) based on differentiation grades of rectal cancer. We retrospectively analyzed 98 patients with rectal cancer. Perfusion-related IVIM parameters (D^* , f , and $f \cdot D^*$) and quantitative DCE parameters (K^{trans} , K_{ep} , V_e , and V_p) were obtained by plotting the volume-of-interest on in-house software. Furthermore, we compared the difference and diagnostic performance of all well-moderately and poorly differentiated rectal cancer parameters. Finally, we analyzed the correlation between those DCE and IVIM parameters and pathological differentiation grade. The values of f , K^{trans} , and K_{ep} significantly differentiated poor and well-moderate rectal cancers. K^{trans} achieved the highest area under the curve (AUC) value compared to perfusion-related IVIM and DCE parameters. Furthermore, K^{trans} showed a better correlation with pathological differentiation grade than f . The diagnostic efficiency of DCE-MRI was greater than perfusion-related IVIM parameters. The f value derived from perfusion-related IVIM offered a diagnostic performance similar to DCE-MRI for patients with renal insufficiency.

1. Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a quantitative technique that measures the capillary blood perfusion parameters of the tissue using intravenous contrast media and evaluates tissue properties such as capillary permeability and extracellular volume fraction that indirectly reflect the microcirculation and hemodynamics in the tissue [1]. Previous studies have demonstrated

the efficiency of DCE-MRI in preoperative diagnosis of rectal cancer [2, 3], determining malignant inner wall irregularity in cystic-cavitary lung lesions [4] and evaluating disease activity of juvenile idiopathic arthritis [5]. However, DCE-MRI is not suitable for patients with contraindication to contrast media.

Intravoxel incoherent motion diffusion-weighted imaging (IVIM-DWI) collects multiple images containing low b values without using intravenous contrast media and

analyzes them using a biexponential model. Quantitative parameters derived from IVIM-DWI can separate true molecular diffusion from the motion of water molecules [6], including pure diffusion coefficient (D), which represents the true tissue cellularity and diffusion; pseudo diffusion coefficient (D^*), which represents incoherent microcirculation; and perfusion fraction (f), which represents the number of protons linked to microcirculation [7]. D^* and f are parameters related to perfusion. The clinical utility of IVIM-DWI has been established in a variety of tumors [7–10]. D values derived from IVIM-DWI demonstrated superior performance compared with ADC value, and perfusion-related IVIM parameters showed a correlation with quantitative parameters from DCE-MRI. Therefore, IVIM-DWI is suitable for patients with renal impairment. However, further research is still needed to compare the diagnostic performance of perfusion-related IVIM parameters and quantitative parameters from DCE-MRI.

Therefore, this study compared the diagnostic performance of perfusion-related parameters derived from IVIM-DWI with quantitative parameters derived from DCE-MRI based on differentiation grades of rectal cancer.

2. Materials and Methods

2.1. Patients. Our institutional review board approved this retrospective study and waived the requirement to obtain written informed consent. Selection criteria were as follows: (1) MRI examination including high-resolution T2WI, multiple b value DWI, and DCE-MRI and (2) the availability of pathology reports of surgical specimens. A total of 128 consecutive patients were enrolled between October 2016 and February 2021. Among these patients, 30 patients were excluded because of (1) poor image quality on T2WI, multiple b value DWI, and DCE-MRI ($n = 19$), (2) preoperative chemoradiotherapy ($n = 6$), and (3) other especial pathological conditions, such as mucinous adenocarcinoma ($n = 5$). Finally, 98 patients were included in the study.

2.2. MRI Technique. All patients were scanned with a 3.0-T MRI system (Intera Achieva 3.0 T TX, Philips Medical System, Best, the Netherlands) with a 16-channel phased-array surface coil. The patients underwent enema before MR imaging to reduce gas-induced artifacts in the rectum. Each patient received 10 mg anisodamine (Hangzhou Minsheng Pharmaceuticals Co. Ltd., Zhejiang, China) intramuscularly 10 minutes prior to the scan to minimize bowel peristalsis. The MRI protocol included the following parameters: (1) axial T1-weighted turbo spin-echo (TSE) (TR/TE = 550/10 ms, slice thickness = 6 mm, slice gap = 1.5 mm, field of view (FOV) = 300 × 300 mm, matrix = 376 × 336, number of excitations (NEX) = 2) (2) axial T2-weighted (T2W) fat suppression TSE (TR/TE = 3554/80 ms, slice thickness = 6 mm, slice gap = 1.8 mm, FOV = 300 × 379 mm, matrix = 376 × 374, NEX = 1), (3) high-resolution sagittal T2W TSE (TR/TE = 3577/70 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 240 × 240 mm, matrix = 300 × 266, NEX = 3), (4) high-resolution oblique coronal T2W (TR/TE = 3000/75 ms, slice thickness = 2 mm, slice gap = 0

mm, FOV = 180 × 180 mm, matrix = 300 × 218, NEX = 3), and (5) high-resolution oblique axial T2W (TR/TE = 3000/75 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 180 × 180 mm, matrix = 368 × 273, NEX = 3). The oblique coronal T2WI was parallel to the tumor axis. Moreover, as identified on sagittal T2WI, the oblique axial T2WI was perpendicular to the tumor axis.

We performed multiple b value DWI in a single-shot spin-echo echo-planar-imaging sequence in the orientation similar to axial oblique T2WI with the following parameters: TR/TE = 2750/76 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 220 × 220 mm, matrix = 112 × 108, NEX = 2, and nine b values (0, 10, 20, 50, 100, 200, 500, 800, and 1000 s/mm²).

The DCE T1W THRIVE sequence was performed with the same orientation as axial oblique T2WI with the following parameters: TR/TE = 6.9/3.5 ms, slice thickness = 3 mm, slice gap = 0 mm, FOV = 220 × 220 mm, matrix = 184 × 182, NEX = 1, flip angle 10°, number of slices 64, and temporal resolution 5.3 seconds. The baseline T1 map was calculated using three precontrast T1-weighted fast field echo sequences with different flip angles (5°, 10°, and 15°, respectively) as references. These three precontrast reference series were similar to the DCE sequences. Each patient received a gadodiamide injection (Omniscan; Ge Healthcare, Ireland) intravenously at a flow rate of 2.5 ml/s (a dose of 0.2 ml/kg of body weight) during the scan.

2.3. Image Analysis. A focal mass or irregular wall thickening with intermediate-signal intensity on T2WI, hyperintensity on DWI with b value of 1000 s/mm², and corresponding hypointensity on grey-scale ADC map were the criteria for rectal cancer. Two experts (Ming Li and Heng Jiang) manually drew the regions of interest (ROIs) slice by slice to cover as much of the entire tumor area as possible while excluding necrosis, cysts, and gas areas and keeping the ROIs on DWI and DCE consistent. When both of these experts disagreed, the third expert (Zhihua Lu) made the decision. An example of the placement of ROIs is shown in Figure 1.

IVIM-DWI raw data were postprocessed using an image analysis software FireVoxel that is based on Le Bihan's biexponential model [11]:

$$S_b/S_0 = (1 - f) \exp(-bD) + f \exp(-bD^*), \quad (1)$$

where S_b is the signal intensity in the pixel with diffusion gradient b value, S_0 is the signal intensity in the pixel with b value of 0 s/mm², D is the true diffusion representing pure molecular diffusion, f is the fractional perfusion representing microcirculation, and D^* is the pseudodiffusion coefficient. The MATLAB software was used to calculate the average D^* and f values in each pixel.

We analyzed DCE-MRI using the pharmacokinetic Extended Tofts Linear model. Furthermore, using the Omni Kinetics software (GE Healthcare, Shanghai, China), we quantitatively analyzed DCE. After importing the dynamic enhancement sequence, three precontrast T1-fast field echo sequences with different flip angles were imported. We drew the internal iliac artery vessel to obtain the time-

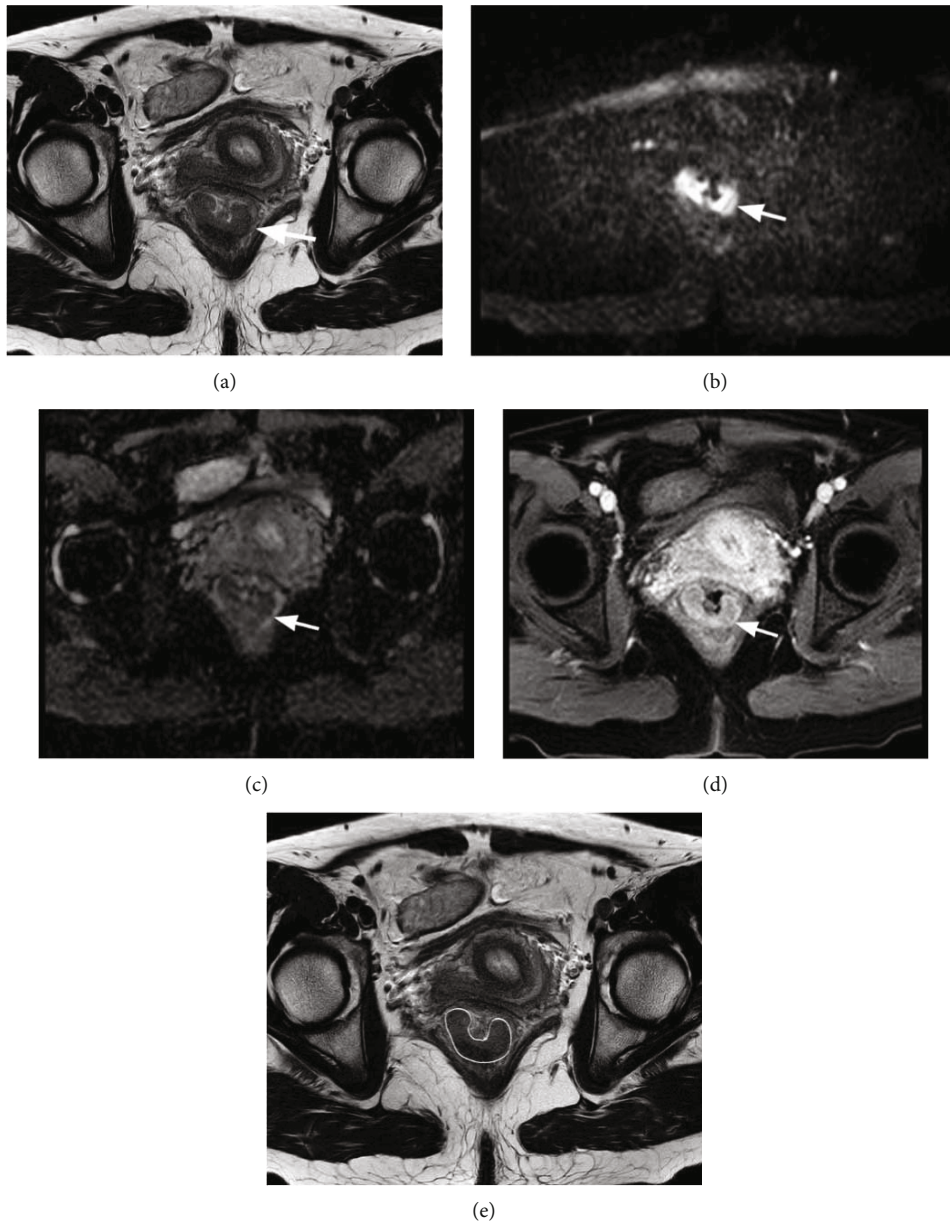


FIGURE 1: A 48-year-old woman with rectal adenocarcinoma was performed by IVIM and DCE. (a) High-resolution oblique axial T2W showed intermediate-signal intensity lesion (white arrow). (b) DWI ($b = 1000 \text{ s/mm}^2$) showed a hyperintensity lesion (white arrow). (c) ADC map showed a hypointensity lesion (white arrow). (d) DCE-MRI showed an abnormal enhanced lesion (white arrow). (e) ROI was manually drawn to cover as much of the entire tumor area as possible (white circle).

concentration curve of the blood supply vessel as the arterial input function and used this curve as a reference. Moreover, the pseudocolor images of K^{trans} , K_{ep} , V_e , and V_p and the average value of each parameter were analyzed.

2.4. Statistical Analysis. We analyzed our data statistically using the SPSS and MedCalc software. We estimated the distribution of all quantitative parameters using the Kolmogorov–Smirnov test. Normally distributed data were presented as mean \pm standard deviation, whereas abnormally distributed data were presented as median value and interquartile range. To compare well-moderately differentiated and poorly

differentiated tumors on all quantitative parameters, we used the independent sample t -test or Mann–Whitney U -test. Furthermore, we analyzed the diagnostic performance for well-moderately differentiated and poorly differentiated tumors using the receiver operating characteristic (ROC) curve. The correlations between those DCE and IVIM parameters and pathological differentiation grade, as well as between perfusion-related parameters of IVIM and DCE parameters, were evaluated using Spearman correlation analysis. We calculated the intraclass correlation coefficients (ICC) to determine the interobserver agreement between the measurement values of two experts. The ICC value of

TABLE 1: Comparison of perfusion-related IVIM and DCE parameters between well-moderate differentiation and poor differentiation tumor.

Parameters	Well-moderate differentiation ($n = 57$)	Poor differentiation ($n = 41$)	P value	ICC
D^* ($\times 10^{-3}$ mm ² /s)	87.15 \pm 14.46	82.95 \pm 19.71	0.252	0.824
f (%)	0.45 \pm 0.09	0.52 \pm 0.08	<0.001	0.882
$f \cdot D^*$	39.24 \pm 10.87	42.05 \pm 11.03	0.213	0.813
K^{trans} (min ⁻¹)	0.53 \pm 0.29	0.92 \pm 0.51	<0.001	0.914
K_{ep} (min ⁻¹)	0.64 \pm 0.22	0.78 \pm 0.19	0.002	0.894
V_e	0.10 (0.05, 0.33)	0.15 (0.08, 0.28)	0.204	0.826
V_p	0.08 (0.04, 0.20)	0.10 (0.06, 0.21)	0.351	0.863

TABLE 2: Diagnostic performance of perfusion-related IVIM and DCE parameters to discriminate poor differentiation tumor.

Parameters	AUC	P value	95% CI	Sensitivity (%)	Specificity (%)	Cutoff value
f	0.728	<0.0001	0.629, 0.813	92.7	43.9	>0.43
K^{trans}	0.775	<0.0001	0.680, 0.854	85.4	63.2	>0.53
K_{ep}	0.710	0.0001	0.609, 0.797	92.7	49.1	>0.60

0.00–0.20 demonstrated a poor agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 good agreement, and 0.81–1.00 excellent agreement.

3. Results

3.1. Clinical and Pathological Findings. This study enrolled 98 patients, of which 60 were male and 38 female. The median age of patients was 66 years (range, 41–89 years). According to the histopathological analysis, 8 tumors were pT1, 32 tumors were pT2, 50 tumors were pT3, and 8 tumors were pT4. 56 patients were staged as N0, 26 patients were staged as N1, and the remaining 16 patients as N2. The differentiation grades of rectal cancer according to pathological results were as follows: well-differentiated in 6 patients, moderately differentiated in 51 patients, and poorly differentiated in 41 patients. Patients were divided into two groups. The first group included 57 patients with well-moderately differentiated grades, and the second group 41 patients with poorly differentiated grades.

3.2. Correlations between Perfusion-Related IVIM and DCE Parameters. There was a weak correlation between f and K_{ep} ($r = 0.279$, $p = 0.005$). However, no correlation was found among other perfusion-related IVIM and DCE parameters.

3.3. Perfusion-Related IVIM and DCE Parameters. Table 1 summarizes the comparison between well-moderately differentiated and poorly differentiated tumors using perfusion-related IVIM and DCE parameters. f , K^{trans} , and K_{ep} were significantly higher in the group with poor differentiation than in the group with well-moderate differentiation ($p < 0.001$, $p < 0.001$, $p = 0.002$, respectively). There were no significant differences in D^* , $f \cdot D^*$, V_e , and V_p between the two groups.

3.4. Comparison of Diagnostic Performance of Perfusion-Related IVIM and DCE Parameters. The diagnostic efficiency of statistically significant perfusion-related IVIM and DCE parameters to discriminate poorly differentiated tumors is presented in Table 2. Among DCE parameters, K^{trans} had a higher AUC (77.5%) with a sensitivity of 85.4% and a specificity of 63.2% (Figure 2), whereas K_{ep} had a lower AUC of 71% with sensitivity and specificity of 92.7% and 49.1%, respectively. Among perfusion-related IVIM parameters, f had an AUC of 72.8% with a sensitivity of 92.7% and a specificity of 43.9%. Furthermore, K^{trans} derived from DCE parameters had a greater correlation with pathological differentiation grade ($r_s = 0.471$) than f derived from perfusion-related IVIM parameters (Table 3).

3.5. Interobserver Agreement. The assessment of interobserver agreement between the two radiologists by ICC revealed excellent agreements for all parameters derived from perfusion-related IVIM and DCE (Table 1).

4. Discussion

This study found that parameters f , K^{trans} , and K_{ep} significantly distinguished poorly and well-moderately differentiated rectal cancer. K^{trans} had the highest AUC value compared with other perfusion-related IVIM and DCE parameters. Furthermore, K^{trans} showed a better correlation with pathological differentiation grade than f .

Although the imaging principles of IVIM-MRI and DCE-MRI vary, the basic theories are based on capillary permeability and hemodynamics hypotheses. Therefore, the relationship between perfusion-related IVIM and quantitative DCE-MRI parameters has been a topic of interest for researchers. Previous studies [12–15] have found that the perfusion-related IVIM parameters have a moderate to strong correlation with quantitative DCE-MRI parameters,

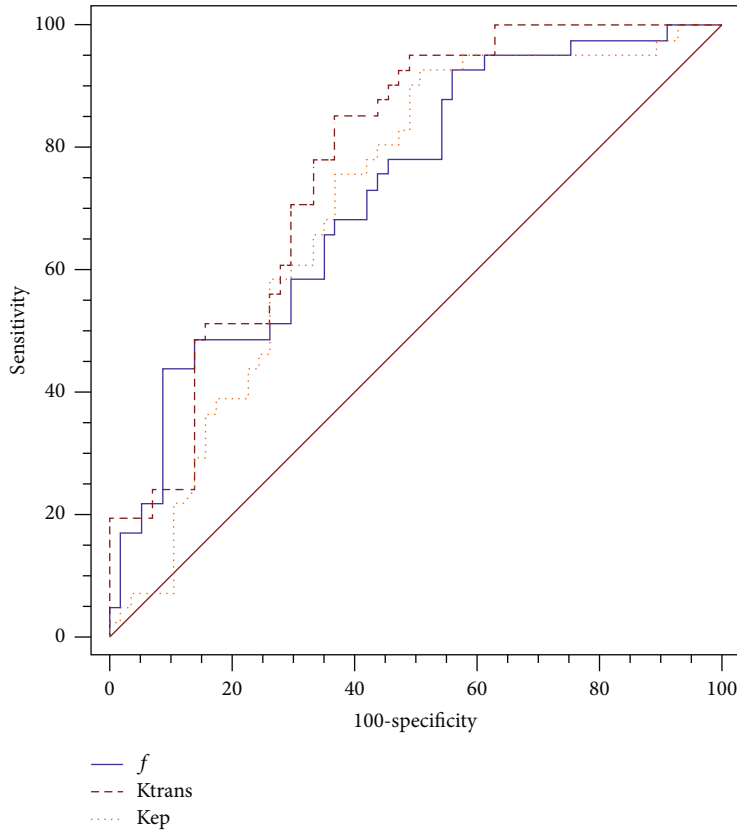


FIGURE 2: ROC curves are shown the diagnostic performance of perfusion-related IVIM and DCE parameters for discriminate poor differentiation tumor.

suggesting that they can be used to diagnose and assess various tumors. In our study, we found a weak correlation between f and K_{ep} parameters. However, there was no correlation between D^* or $f \cdot D^*$ and any quantitative DCE-MRI parameters, which may be attributed to several reasons. First, other factors such as microvessel anatomy, blood vessel permeability [16], and microvessel density [17] also affect the f value. Second, K^{trans} is affected by both blood volume and microvascular permeability. Third, the D^* value may represent more than one physiological process, such as gland secretion and fluid flow in capillaries [18].

The D^* value of perfusion-related IVIM parameters indicates a perfused capillary microcirculation in the tissue. The f value is related to the capillary blood volume and represents the rate of diffusion related to microcirculation perfusion in the respective voxel of interest compared to total diffusion. Moreover, $f \cdot D^*$ allowed the estimation of relative perfusion or blood flow in the tumor microcirculation, which was previously thought to depend on the interaction among the microvascular anatomy, vascular permeability, and blood flow dynamics [19]. Our results showed that the f value in the poorly differentiated tumors was significantly higher than the well-moderately differentiated tumors. This may be because the permeability of immature blood vessels and capillaries increased with the increased malignancy of the tumor. This is in line with the results of another study [20], which found a significant correlation between f value and blood

TABLE 3: Correlations between DCE and IVIM parameters and pathological differentiation grade.

Parameters	Differentiation grade	
	R_s	P value
f	0.289	<0.001
K^{trans}	0.471	<0.001
K_{ep}	0.358	<0.001

vessel count and microvessel density. However, Sun et al. [20] and Lu et al. [21] found that the f value gradually decreased with the increasing degree of malignancy. The differences in results between this study and others may be related to the increase in the complexity of the microvessels in the tumor as the degree of malignant increases.

The DCE-MRI quantitative parameters objectively assess the characteristics of microcirculation perfusion and microvessels in tissues. K^{trans} represents the rate of contrast media transporting through the vascular endothelium to the extravascular extracellular space (EES). K_{ep} is the opposite of K^{trans} , and K^{trans} and K_{ep} represent the amount of microvascular perfusion in the tissue as well as the penetrated area of the blood vessel. Our findings revealed that K^{trans} and K_{ep} were significantly higher in rectal cancer with poor

differentiation than that with well-moderate differentiation. This is because the structure and function of tumor neovascularization become impaired with the increase in the degree of malignancy of the tumor, resulting in higher permeability and blood flow in tumors with higher malignancy. This result is consistent with the findings of another study [2]. V_e is the size of EES, which reflects the percentage of contrast media concentration in the entire voxel in EES. The mixing of blood volume and plasma flow is represented by V_p . In our study, the V_e and V_p values showed no significant difference between poor differentiation and well-moderate differentiation. This can be attributed to scan time and the tumor-node-metastasis (TNM) stage.

The ROC curves analysis revealed that K^{trans} achieved the highest AUC value for distinguishing between tumors with poor differentiation and those with well-moderate differentiation. Furthermore, Spearman correlation analysis showed that K^{trans} had a better correlation with pathological differentiation grade than f .

Before considering the implications of the findings of this study, it is important to consider some of the limitations. First, the sample size of well-differentiation tumors is small and insufficient. Patients with moderately differentiated grades will have a different prognosis. The sample size must be large, and the group must be refined to obtain better and significant results. Second, the selection of b value affects IVIM-MRI parameters. At present, there is no standard reference for the value and number of b values. Further research is needed to optimize the b value. Third, our study did not analyze the correlation between perfusion-related IVIM and DCE-MRI parameters and immunohistochemical findings (microvessel density and vascular endothelial growth factor). We will conduct a more in-depth analysis of comparison in the follow-up study.

5. Conclusion

In conclusion, our study showed that DCE-MRI had greater diagnostic accuracy than perfusion-related IVIM parameters. Especially, the K^{trans} derived from DCE-MRI parameters was the optimal parameter for diagnostic efficiency. Moreover, the f value derived from perfusion-related IVIM offers similar diagnostic performance to DCE-MRI for patients with renal insufficiency.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest.

Authors' Contributions

Ming Li, Xiaodan Xu, and Kaijian Xia contributed equally to this work.

Acknowledgments

This work was supported in part by the Jiangsu Provincial Medical Youth Talent under Grant QNRC2016212, the Jiangsu Committee of Health under Grant H2018071, the Suzhou Clinical Special Disease Diagnosis and Treatment Program under Grant LCZX20182, the Suzhou GuSu Medical Talent Project under Grant GSWS2019077 and GSWS2020108, and the Suzhou Science and Technology Development Program under Grant SYS2020058;

References

- [1] F. Khalifa, A. Soliman, A. El-Baz et al., "Models and methods for analyzing DCE-MRI: a review," *Medical Physics*, vol. 41, no. 12, article 124301, 2014.
- [2] F. U. Shen, J. Lu, L. Chen, Z. Wang, and Y. Chen, "Diagnostic value of dynamic contrast-enhanced magnetic resonance imaging in rectal cancer and its correlation with tumor differentiation," *Molecular and Clinical Oncology*, vol. 4, no. 4, pp. 500–506, 2016.
- [3] R. A. P. Dijkhoff, R. G. H. Beets-Tan, D. M. J. Lambregts, G. L. Beets, and M. Maas, "Value of DCE-MRI for staging and response evaluation in rectal cancer: a systematic review," *European Journal of Radiology*, vol. 95, pp. 155–168, 2017.
- [4] A. Karaman, O. Araz, I. Durur-Subasi et al., "Added value of DCE-MRI in the management of cystic-cavitary lung lesions," *Respirology*, vol. 21, no. 4, pp. 739–745, 2016.
- [5] R. Hemke, C. M. Nusman, J. M. van den Berg et al., "Construct validity of pixel-by-pixel DCE-MRI: correlation with conventional MRI scores in juvenile idiopathic arthritis," *European Journal of Radiology*, vol. 94, pp. 1–5, 2017.
- [6] M. Klauss, P. Mayer, K. Maier-Hein et al., "IVIM-diffusion-MRI for the differentiation of solid benign and malign hypervascular liver lesions—evaluation with two different MR scanners," *European Journal of Radiology*, vol. 85, no. 7, pp. 1289–1294, 2016.
- [7] S. Woo, J. M. Lee, J. H. Yoon, I. Joo, J. K. Han, and B. I. Choi, "Intravoxel incoherent motion diffusion-weighted MR imaging of hepatocellular carcinoma: correlation with enhancement degree and histologic grade," *Radiology*, vol. 270, no. 3, pp. 758–767, 2014.
- [8] S. Zhu, Y. Wei, F. Gao et al., "Esophageal carcinoma: Intravoxel incoherent motion diffusion-weighted MRI parameters and histopathological correlations," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 1, pp. 253–261, 2019.
- [9] Y. Li, Z. Wang, F. Chen et al., "Intravoxel incoherent motion diffusion-weighted MRI in patients with breast cancer: correlation with tumor stroma characteristics," *European Journal of Radiology*, vol. 120, article 108686, 2019.
- [10] B. K. Sokmen, S. Sabet, A. Oz et al., "Value of intravoxel incoherent motion for hepatocellular carcinoma grading," *Transplantation Proceedings*, vol. 51, no. 6, pp. 1861–1866, 2019.
- [11] D. Le Bihan, R. Turner, and J. R. MacFall, "Effects of intravoxel incoherent motions (IVIM) in steady-state free precession (SSFP) imaging: application to molecular diffusion imaging," *Magnetic Resonance in Medicine*, vol. 10, no. 3, pp. 324–337, 1989.
- [12] L. Jiang, X. Lu, B. Hua, J. Gao, D. Zheng, and Y. Zhou, "Intravoxel incoherent motion diffusion-weighted imaging versus dynamic contrast-enhanced magnetic resonance imaging:

- comparison of the diagnostic performance of perfusion-related parameters in breast,” *Journal of Computer Assisted Tomography*, vol. 42, no. 1, pp. 6–11, 2018.
- [13] H. Wang, L. Zhu, G. Li, M. Zuo, X. Ma, and J. Wang, “Perfusion parameters of intravoxel incoherent motion based on tumor edge region of interest in cervical cancer: evaluation of differentiation and correlation with dynamic contrast-enhanced MRI,” *Acta Radiologica*, vol. 61, no. 8, pp. 1087–1095, 2020.
- [14] M. Song, Y. Yue, J. Guo et al., “Quantitative analyses of the correlation between dynamic contrast-enhanced MRI and intravoxel incoherent motion DWI in thyroid nodules,” *American Journal of Translational Research*, vol. 12, no. 7, pp. 3984–3992, 2020.
- [15] Y. Zhou, J. Liu, C. Liu et al., “Intravoxel incoherent motion diffusion weighted MRI of cervical cancer—correlated with tumor differentiation and perfusion,” *Magnetic Resonance Imaging*, vol. 34, no. 8, pp. 1050–1056, 2016.
- [16] H. Sun, Y. Xu, Q. Xu et al., “Correlation between intravoxel incoherent motion and dynamic contrast-enhanced magnetic resonance imaging parameters in rectal cancer,” *Academic Radiology*, vol. 26, no. 7, pp. e134–e140, 2019.
- [17] H. J. Lee, S. Y. Rha, Y. E. Chung et al., “Tumor perfusion-related parameter of diffusion-weighted magnetic resonance imaging: correlation with histological microvessel density,” *Magnetic Resonance in Medicine*, vol. 71, no. 4, pp. 1554–1558, 2014.
- [18] H. C. Thoeny, F. de Keyzer, R. H. Oyen, and R. R. Peeters, “Diffusion-weighted MR imaging of kidneys in healthy volunteers and patients with parenchymal diseases: initial experience,” *Radiology*, vol. 235, no. 3, pp. 911–917, 2005.
- [19] D. Le Bihan and R. Turner, “The capillary network: a link between IVIM and classical perfusion,” *Magnetic Resonance in Medicine*, vol. 27, no. 1, pp. 171–178, 1992.
- [20] H. Sun, Y. Xu, A. Song, K. Shi, and W. Wang, “Intravoxel incoherent motion MRI of rectal cancer: correlation of diffusion and perfusion characteristics with prognostic tumor markers,” *American Journal of Roentgenology*, vol. 210, no. 4, pp. W139–W147, 2018.
- [21] B. Lu, X. Yang, X. Xiao, Y. Chen, X. Yan, and S. Yu, “Intravoxel incoherent motion diffusion-weighted imaging of primary rectal carcinoma: correlation with histopathology,” *Medical Science Monitor*, vol. 24, pp. 2429–2436, 2018.

Research Article

CT Image Analysis and Clinical Diagnosis of New Coronary Pneumonia Based on Improved Convolutional Neural Network

Wu Deng ^{1,2}, Bo Yang ³, Wei Liu ³, Weiwei Song ⁴, Yuan Gao ³, and Jia Xu ^{3,5}

¹College of Electronic Information, Sichuan University, Chengdu Sichuan 610000, China

²Information Center/Engineering Research Center of Medical Information Technology, Ministry of Education, West China Hospital of Sichuan University, Chengdu Sichuan 610000, China

³Information Center/Engineering Research Center of Medical Information Technology, West China Hospital of Sichuan University, Chengdu Sichuan 610000, China

⁴Department of Radiology, West China Hospital of Sichuan University, Chengdu Sichuan 610000, China

⁵College of Physics, Sichuan University, Chengdu Sichuan 610000, China

Correspondence should be addressed to Jia Xu; scu_xujia@stu.scu.edu.cn

Received 23 April 2021; Revised 8 June 2021; Accepted 24 June 2021; Published 22 July 2021

Academic Editor: Hamidreza Mohafez

Copyright © 2021 Wu Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, based on the improved convolutional neural network, in-depth analysis of the CT image of the new coronary pneumonia, using the U-Net series of deep neural networks to semantically segment the CT image of the new coronary pneumonia, to obtain the new coronary pneumonia area as the foreground and the remaining areas as the background of the binary image, provides a basis for subsequent image diagnosis. Secondly, the target-detection framework Faster RCNN extracts features from the CT image of the new coronary pneumonia tumor, obtains a higher-level abstract representation of the data, determines the lesion location of the new coronary pneumonia tumor, and gives its bounding box in the image. By generating an adversarial network to diagnose the lesion area of the CT image of the new coronary pneumonia tumor, obtaining a complete image of the new coronary pneumonia, achieving the effect of the CT image diagnosis of the new coronary pneumonia tumor, and three-dimensionally reconstructing the complete new coronary pneumonia model, filling the current the gap in this aspect, provide a basis to produce new coronary pneumonia prosthesis and improve the accuracy of diagnosis.

1. Introduction

Novel coronavirus pneumonia (NCP) is referred to as novel corona pneumonia, which refers to pneumonia caused by 2019 novel coronavirus infection (2019 Novel Coronavirus (COVID-19)) [1]. Since December 2019, Wuhan City, Hubei Province, has successfully discovered multiple cases of NCP patients with COVID-19 infection [2]. Coronavirus is a type of RNA virus with an envelope and a linear single-stranded genome. Obvious rod-shaped particles are protruding on the shape of the corolla, so it is named coronavirus [3]. Coronavirus only infects vertebrates and is related to a variety of diseases in humans and animals. It can cause respiratory and digestive tract and nervous system diseases in humans and animals. So far, in a total of 7 coronaviruses that can infect humans, 3 coronaviruses can cause severe respiratory

diseases, namely COVID-19, SARS-CoV, and MERS-CoV, respectively. The early manifestations are mainly patchy ground-glass shadows in the double lungs [4]. As the lesion progresses, the lesions increase with consolidation, and bronchial inflation signs, “paving stone signs,” “halo signs,” and “antihalo signs” may occur. The lungs are diffusely real. During the recovery period, the lung lesions are absorbed and fibrous foci formed, and there were less hilar and mediastinal lymphadenopathy and a large amount of pleural effusion [5]. Because deep learning methods are more efficient in analyzing images, the results are more accurate.

The semantic segmentation of an image is to classify pixels according to the category to which the pixels belong, divide pixels belonging to the same category, and then realize image segmentation [6, 7]. The segmentation algorithm based on the texture primitive forest combines the context

information of the image, expresses the image as a histogram distribution of the texture image, to train the K -nearest neighbor or support vector machine classifier, and then classifies the image to achieve the purpose of image segmentation [8]. However, this algorithm is greatly affected by the changes in illumination and image rotation in the neighborhood [9]. Different neighborhood sizes have different classification performance, and different images will have different optimal neighborhood sizes [10]. The segmentation method based on random forest directly uses the low-level pixels of the image, extracts a fixed-size window from the image to obtain a feature vector according to random sampling, and then randomly selects two pixels from the window to obtain the value vector, training random forest classifier [11]. The specific method is to vote for the vector extracted from each window through the leaf nodes of the random forest, predict the pixel type according to the voting result, and complete the semantic segmentation of the image [12]. In contrast, this method can greatly improve the segmentation efficiency. In the early deep learning methods, the classification network usually has a fully connected layer whose input is a fixed-size pixel block [13]. Therefore, the semantic segmentation applied to the image classified based on image block 3, that is, the pixels around each pixel extracted from the image. The image block is composed and used to judge the category of the pixel [14]. Not only this segmentation algorithm is cumbersome to calculate but also the effect is not ideal [15]. In 2017, at the University of California, Jiang and others proposed a fully convolutional neural network (FCN) to change the fully connected layer in the previous convolutional neural network to a convolutional layer [16]. FCN does not require a fully connected layer to perform dense pixel prediction and can accept input images of any size while avoiding the problem of repeated storage and calculation convolution caused by using pixel blocks, which accelerates the processing speed and is more efficient [17]. Therefore, the neural network for semantic segmentation mostly adopted the structure of FCN. The appearance of FCN is a milestone for the application of a convolutional neural network to image semantic segmentation [18]. Based on the FCN network structure, Mazurowski et al. proposed a 3D FCN for segmentation of a new type of coronary pneumonia MRI image and further optimized the segmentation results through the deformation model algorithm to improve the accuracy of image segmentation [19]. Pan et al. first used CT images to train the 2D FCN network, input the CT images of the sagittal plane, coronal plane, and cross-section corresponding to each pixel, and then used the 3D majority-voting algorithm to perform the three-image vote on the segmentation result to obtain the final segmentation result [20]. Pons and others input the coronary MRI images in brain MRI, myocardial MRI, and cardiac CT angiography into the same FCN network, so that the final trained network can perform the segmentation tasks for the above three types of images at the same time [21].

This article focuses on the research of the new coronary pneumonia tumor image diagnosis method. Through deep learning and preoperative new coronary pneumonia tumor CT images, the new coronary pneumonia tumor image automatic segmentation, tumor detection, and diagnosis will

assist doctors in preoperative tumor identification and diagnosis, and follow-up. The main research contents of this article include segmentation of the vertebral body area and surrounding soft tissues in the CT image of the new coronary pneumonia, detection, and identification of tumor lesions of the new coronary pneumonia, and image diagnosis of the new coronary pneumonia. In this paper, firstly, the CT image of the new coronary pneumonia is segmented by the convolutional neural network, and the new coronary pneumonia area is extracted to obtain the binary image with the new coronary pneumonia area as the foreground and the remaining areas as the background. The convolutional neural network has a strong feature learning ability. Then, based on the deep convolutional neural network, feature extraction is performed on the new type of coronary pneumonia tumor image to obtain a higher-level abstract representation of the feature and then predict the lesion location and obtain its bounding box in the image. Finally, by generating an adversarial network for image diagnosis of the CT image of the new coronary pneumonia tumors, the normal new coronary pneumonia CT image output network is trained to obtain the corresponding network model, and then the tumor area in the CT image of the new coronary pneumonia tumor is added, the mask is used as the input of the model, and the output of the model is the image of the complete new coronary pneumonia diagnosed. The image after the diagnosis will be used to reconstruct the complete three-dimensional digital model of the new coronary pneumonia through a three-dimensional reconstruction algorithm, which provides a basis for the subsequent production of the new coronary pneumonia.

2. CT Image Analysis of New Coronary Pneumonia Based on Convolutional Neural Network

2.1. Improved Convolutional Neural Network Model. The traditional medical image segmentation methods used in this paper include region-based segmentation methods, such as threshold method, region growth algorithm, and level set-based segmentation methods. The threshold segmentation algorithm is the most common image segmentation method [22]. The algorithm assumes that the pixel values of the adjacent pixels in the target or background in the gray image are similar, but the pixel values of different targets are different. In the results presented on the image histogram, different targets correspond to different peaks, so the threshold at the peak and valley positions is selected to separate the different targets. In turn, the lesion site is predicted, and its bounding box in the image is obtained. Finally, the image diagnosis of the novel coronary pneumonia tumor CT images is performed by generating an adversarial network, and the normal novel coronary pneumonia CT images are output to the network for training to obtain the corresponding network model. If only one threshold is selected, the image is divided into a single target area and a background area, which is called single threshold segmentation; if multiple thresholds are selected, the image is divided into multiple target areas, which is called multithreshold segmentation. The advantage

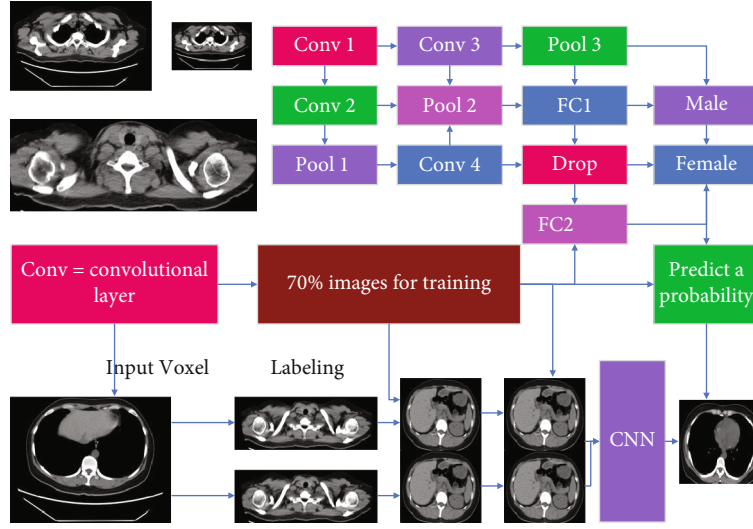


FIGURE 1: Improved convolutional neural network.

of this method is that it is relatively simple to implement. When the gray values of different categories are very different from other feature values, the image can be effectively segmented; accordingly, if the gray difference in the image is not large or the gray between regions the value ranges overlap, it is difficult for the algorithm to segment accurate results. The basic idea of the area growth algorithm is to gather pixels with similar properties: first, select a seed point and then traverse the pixels in the image, in turn, to determine the pixels about the seed [23]. If the similarity is satisfied, it is divided into seed areas at the same time, and the pixel is pushed into the stack as an alternative seed point until all pixels have a home and eventually form a different area. The algorithm is simple to calculate and particularly suitable for segmenting small structures, such as tumors and scars. The point has great impact on the segmentation results and is sensitive to noise, as shown in Figure 1.

Since the introduction of ResNet, it has proved that increasing the number of network models and the accuracy of the model can be considered. Therefore, there are endless variations of the network, and the improved model further improves performance. The 2017 CVPR best paper DenseNet also draws on the idea of ResNet and at the same time creatively proposed a brand-new structure. DenseNet is a densely connected convolutional neural network [24]. Like ResNet, the model contains multiple dense blocks. There are connections between any two layers. Specifically, each layer of input is all previous layers. The union of the output of the feature map extracted by this layer will also directly be passed to the subsequent convolutional layer. The purpose of this is to ensure that the information flow between the layers reaches the maximum.

Unlike the ResNet model, the elements between the layers added together and the connection between the different layers in DenseNet are splicing, which is the superposition of dimensions. This dense connection method will reduce the computational cost of each layer of the network and will reuse features. There is no need to relearn redundant feature maps. The operation of dimensional stitching will

add more features, compared with the model [25]. Another advantage of DenseNet is that the configuration of its structure allows each layer to directly obtain gradients from the loss function and model input, which improves the flow of information and gradients in the model, which is very beneficial for training. In ResNet, due to the residual learning module, the output of the u layer can be expressed as

$$-\operatorname{div} \left(\frac{\nabla^a u}{|\nabla^a u| + p} \right) + \lambda_e (u - u^0) = 0. \quad (1)$$

The identity map and the output of the nonlinear transformation superimposed in an additive manner, which destroys the information flow in the model to a certain extent. However, in DenseNet, the output of layer X can be expressed as

$$(a, b; X, \Phi) = |a|^{-0.5} \int_{-\infty}^{+\infty} a(\tau) g(\tau - t) e^{-jw\tau} d\tau. \quad (2)$$

According to the definition of migration learning, migration learning can be divided into three types, distribution difference migration learning, feature difference migration learning, and label difference migration learning. Distribution difference migration learning has different edge distribution or conditional probability distribution for source and target domain data, feature difference migration learning has different feature space for source and target data, and label difference migration learning refers to different data tagging space for source and target domains. Since each layer contains the output information of all previous layers, only a few feature maps are needed. It is enough and can reduce the phenomenon of gradient disappearance. When tensor stitching, the size of the feature map needs to be the same; otherwise, the dimensions cannot be superimposed. However, pooling operation is essential. To reconcile this contradiction, DenseNet adds a transition layer between dense blocks.

The loss function in the original K_{nn} is

$$k_{nn}(x) = \sum_{i=1}^N \omega_i s_i(x) = W^T S(x). \quad (3)$$

The first term does not depend on the generator and can be used as a generator loss function. However, for the discriminator, the generator is optimized by fixing the generator, and the contribution to the discriminator loss function is

$$D_\epsilon(x(t)) = [s(\|x(t) - \ell_{j1}\|)]. \quad (4)$$

The data is then randomly assigned. Since the discriminator in formula (4) maximizes the loss function on the right, it is added with a minus sign and converted to find the minimum value, so that it derives about the $Q_N(w)$ pair:

$$Q_N(w) = \frac{1}{2} \sum_{i,j} P_{ij} (1 - P_{ij})^2 + \|w^T X_b\|_2^2. \quad (5)$$

Get the optimal discriminator function:

$$\phi_{m,n} = \frac{\|q_{m,n}\|^2}{\delta^2} \exp\left(-\frac{(q_{m,n} * z)}{2\delta^2}\right) * [e^{i(q_{m,n} * z)} - e^{-\delta^2/2}]. \quad (6)$$

Intuitively, formula (4) represents the relative proportion of sample x from the true distribution and the probability of generating distribution, and it is substituted into formula (2):

$$\frac{\partial L}{\partial a} = \sum_{i=1}^n \left[y_i - \frac{\exp\left(a + \sum_{j=1}^m x_{ij} \beta_j\right)}{1 + \exp\left(a + \sum_{j=1}^m x_{ij} \beta_j\right)} \right] = 0. \quad (7)$$

It can be further transformed into

$$\begin{aligned} \frac{\partial L}{\partial \beta_j} &= \sum_{i=1}^n \left[y_i - \frac{\exp\left(a + \sum_{j=1}^m x_{ij} \beta_j\right)}{1 + \exp\left(a + \sum_{j=1}^m x_{ij} \beta_j\right)} x_{ij} \right] \\ &= 0, j = 1, 2, \dots, m, \\ g_i &= \omega_i / \sum_{j=1}^n \omega_j, \sum_{i=1}^n g_i = 1. \end{aligned} \quad (8)$$

Divergence is as follows:

$$b_{ij} = \frac{b_{ij}}{\sum_{k=1}^n b_{kj}}, (i, j = 1, 2, \dots, n) \quad (9)$$

The development of medical image segmentation technology not only affects the development of other related technologies in medical image processing, such as visualization and 3D reconstruction, but also occupies an extremely important position in the analysis of biomedical images. In recent years, medical image segmentation techniques have

made remarkable progress due to the application of some emerging disciplines in medical image processing. In this paper, we take the perspective of the specific theoretical tools applied for segmentation. As can be seen from the above derivation formula, when the training discriminator is close to the optimal, the divergence will be smaller, but when there is no overlap between the real distribution and the generated distribution or the overlapping part can be ignored, the JS divergence will be fixed and constant, which means that the gradient is zero. Therefore, when the discriminator is in the optimal situation, the generator cannot obtain the gradient information, which results in unstable training. b replaces P divergence with the Wasserstein distance to measure the distance between the real distribution and the generated distribution, which is defined as follows:

$$P(y_i) = p_i^{y_i} (1 - p_i)^{1-y_i}. \quad (10)$$

Since the lower bound value in formula (10) cannot be solved directly, it is transformed into

$$m = P(R|S) = \frac{P(S \cap M)}{P(S)} = \frac{s}{s+h}. \quad (11)$$

2.2. Design of CT Image Dataset for New Coronary Pneumonia. Before training the network, it is necessary to make a new type of coronary pneumonia CT image data set and corresponding annotation data. First, for the original image, you need to convert its format to the BMP or jpg format required by this article, remove the patient's information data from the original image, and then use the traditional segmentation method, that is, the above level set to process the original new coronary pneumonia CT. We perform segmentation to obtain labeled data and divide the data set into a training set and a test set according to a 7:3 ratio. This may lead to a high false-positive rate of 27.49% in the LIDC-IDRI dataset for physicians with benign and malignant lung nodules. Besides, the training process will be verified by a 10-fold cross-validation method. Ten-fold cross-validation is to divide all the training sets into ten parts, take one of them for testing without repeating each time, which is the so-called verification set, and use the other nine parts as training sets to train the model. Before training the network, this paper needs to enhance the data of the new coronary pneumonia CT image, which on the one hand increases the amount of data and, on the other hand, is conducive to the network's adaptability to tissue deformation. The process of data enhancement is shown in Figure 2. After using the level set method to segment the new coronary pneumonia CT image to obtain the label, the training set and the corresponding label are channel merged, then the merged image is subjected to such as rotation, translation, and image processing such as mirroring, so that the training set image and the corresponding label get the same transformation, and finally, the transformed combined image is separated into the training set and the label, thereby obtaining the expanded training set and the corresponding label.

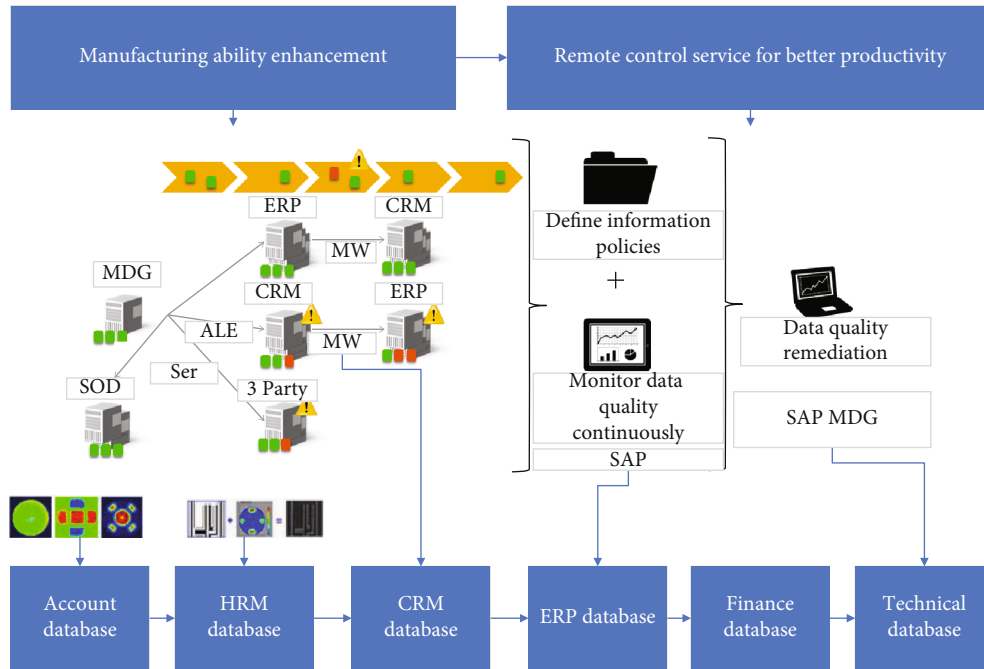


FIGURE 2: Data enhancement process.

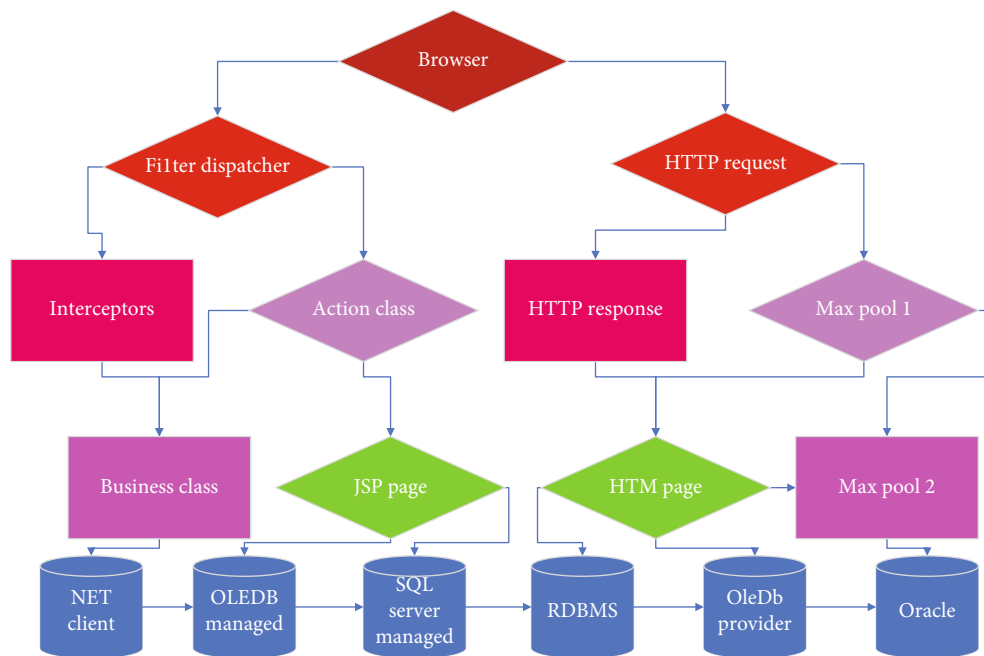


FIGURE 3: U-Net framework and specific parameter settings in this article.

Due to the outstanding performance of U-Net in medical image segmentation tasks, this paper will complete the segmentation task of new-type coronary pneumonia CT images through this type of network. U-Net is a typical end-to-end network structure. As its name indicates, its structure is U-shaped. Aiming at the task of new-type coronary pneumonia CT image segmentation in this paper, some adjustments were made to the original U-Net, as shown in Figure 3. The entire network is divided into two processes. The first is the downsampling process, which is used to capture context informa-

tion. Four sets of convolution operations are completed. After each set of convolution operations, the feature map is pooled to reduce the image to half of the original after four convolution operations, and the original image with the original size of 512×512 is reduced to 16×16 . Then, there is the upsampling process, which also uses four sets of deconvolution operations. Each upsampling expands the size of the feature map to twice the original size and stitches with the corresponding feature map during the downsampling process. The size is $256 \times 256 \times 16$ feature map. Finally, the

number of channels is reduced to 1 by a 1×1 convolution kernel. Its specific parameters are shown in Figure 3. In the downsampling process, this paper uses a convolution kernel of size 9×9 and padding to perform the convolution operation. The advantage of this is that the feature maps at each stage of upsampling have the same resolution, and upsampling and downsampling corresponding to the size of the feature map are the same, eliminating the cropping operation during splicing, and can retain all the information extracted by the convolution operation.

V-Net is a further improvement based on U-Net, which is used to segment three-dimensional data. Since the problem to be solved in this article is the segmentation of two-dimensional images, we will extract the core elements in V-Net to complete the new crown pneumonia CT image segmentation task. The encoder-decoder structure is like U-Net, except that the input and output of the convolutional layer in each stage of the encoder are added to obtain residual learning. The result of this is to be able to retain due to compress the lost information, improve the accuracy of the target boundary segmentation in the image, converge in a short time, and improve the model convergence speed. The basic idea of the region growing algorithm is to merge pixel points with similar properties together. For each region, a seed point is designated as the starting point for growth, then the pixels in the field around the seed point are compared with the seed point, and the points with similar properties are merged and continue to grow outward until no pixels that meet the conditions are included. In this way, the growth of a region is completed. And in the encoder, after each stage of the feature map processing, the resolution will be compressed by the convolution kernel of 4×4 and step size 4, thereby achieving the effect of reducing the size of the feature map to half, which is like the function of the pooling operation that not only does not lose edge information but also does not need to switch the mapping between input and output in the pooling operation, which reduces the use of memory. At the same time, in the entire network, the size of the convolution kernel in each stage of the convolution operation is 10×10 , and the PReLU nonlinear activation function is used.

2.3. Analysis and Diagnostic Evaluation Design. The main purpose of fine-grained classification is to distinguish multiple subcategories under the same basic category. Each subcategory shares very similar attribute characteristics, and the differences between subcategories are very subtle, so fine-grained classification is very challenging. In the medical background, there are few studies on the problem of fine-grained image analysis, which is mainly due to the exploration of fine-grained medical image analysis that is limited by the existing data and annotations. Data labeling requires professional knowledge, and the cost of labeling is very expensive. However, the fine-grained evaluation will bring about a reduction in the sample of each subclass of the attribute and a more subtle difference in its characteristics. At the same time, the samples of each subclass are extremely unbalanced. As shown in Table 1, the distribution is very biased. The slanted situation and the lack of coverage of minority samples will cause the algorithm to have serious problems

TABLE 1: Distribution of samples of subordinate subclasses of the nine semantic attributes of novel coronary pneumonia.

Semantic attributes	Number of subclasses	Sample distribution of each subclass
Leaf	8	5
Roundness	7	4
Glitch	10	3
Detail	10	6
Texture	1	4
Edge	9	10
Malignancy	10	1
Internal structure	2	3
Calcification	7	5

in the underrepresented minority classes. The problem of data imbalance in the fine-grained attribute subclasses of lung nodules can easily bias the learning framework. This further increases the difficulty of fine-grained medical image analysis. Traditional oversampling techniques implicitly assume that the behavior of oversampling will not deviate from the original class distribution. However, this assumption does not apply to complex high-dimensional data, because the synthetic samples of these data are inherently difficult, and simple interpolation can easily deviate from the original class distribution.

In the experiment, we implemented 5 network models based on the network structure shown in Figure 3. Basic network, where n_1, n_2 , and n_3 are 32, 64, and 128 and 64, 128, and 256, respectively, and the model, where n_1, n_2 , and n_3 are 64, 128, and 256, use bounce connection, only residual learning, and both. The initial learning rate is set to 0.001, the learning rate decays to the original 0.2 every 50 cycles, and a total of 200 cycles are trained to obtain the final model of each network. Then, we calculate the average of the SSIM and PSNR indicators of the 40 chest radiographs of the test set and the DES label. The model and corresponding evaluation results are shown in Figure 4. Both PSNR and SSIM are larger. The higher the similarity between the two images, the experimental data shows that both bounce connection and residual learning can improve the performance of the model. If used simultaneously, the model is better. Also, the network's increased number of convolution cores can improve the performance. Both PSNR and SSIM indicate that the larger the value, the higher the similarity between the two images. Experimental data shows that both bounce connection and residual learning can improve the performance of the model. If used simultaneously, the model is better. Besides, the network's increased number of convolution cores can also improve the performance.

During preprocessing, the original image is first scaled to a resolution of 1024×1024 , which is consistent with the resolution used by the model during training, and then the coordinate values of the nodule annotations are adjusted accordingly. Although the fully convolutional network does not require that the input image size must be the same, because our training data is relatively small and the source

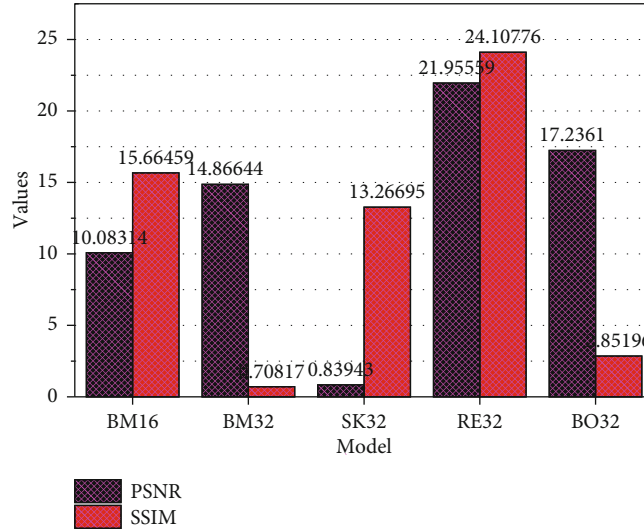


FIGURE 4: Performance comparison of lung suppression model.

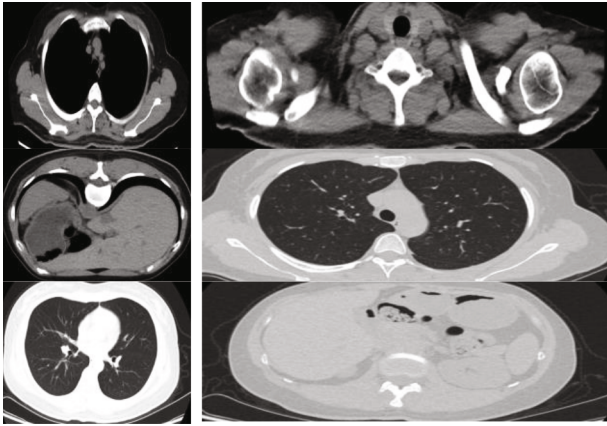


FIGURE 5: CT images of original new coronary pneumonia and 2D V-Net and improved segmentation results.

is single, to reduce the model’s “unfit” on the new data, we adjusted the new chest radiograph image that is consistent with the input image resolution when the skeleton suppression model is trained. After these images were processed by the rib suppression network, and then the size was converted to the original size, the ribless chest radiograph data was obtained. Then, we used the target-detection network RetinaNet as the detection model to verify the effect of our proposed rib suppression method on the detection of lung nodules on 5908 chest radiographs with lung nodules. We randomly selected 296 chest radiographs as the test set and the remaining 5612 samples as the training set. The two data sets were divided the same, and the image size was adjusted to 1024×1024 . The anchor aspect ratio in RetinaNet is $[0.8, 1, 1.25]$. To make the scale more compact, each aspect ratio has 5 different scales as $[20, 21/5, 22/5, 23/5, 24/5]$, with a higher scale density, as much as possible to cover all lung nodules.

The evaluation index used in training is mAP (mean average precision), which is also the most commonly used model evaluation index for target detection. Since only one

type of target object is lung nodules in this paper, it is only necessary to calculate the average accuracy of lung nodules, where AP is the area under the accuracy and recall curve (PR curve). The initial learning rate used for training is $1e - 4$. Since each Layer will have more inputs, the number of channels in each layer can be reduced, the feature utilization rate is higher, and the overall calculation will be less than ResNet’s feature for the same number of connections. When the model no longer improved for two consecutive training cycles, the learning rate is attenuated, and the attenuation factor is 0.1. After 100 training cycles, the model with the best result is selected. Two data sets were used to train the same detection network, and the data were the data after the rib suppression treatment and the original image of the chest radiograph.

3. Result Analysis

3.1. Performance Analysis. The self-attention model highlights the semantic difference of the intermediate feature map at each image scale, and then the attention units at different scales can respond to the foreground content of the image. Therefore, this paper integrates the self-attention model into the previous in the 2D V-Net network, and the schematic diagram is shown in Figure 5. In the jump connection, the feature map of the encoder first taken as XL of the self-attention mode. After using the self-attention model to improve 2D V-Net, this paper trains it on the data set, based on the same hardware configuration and software environment, and the segmentation results are shown in Figure 5, which can be seen in the improvement of 2D U-Net. Afterwards, the edge segmentation of CT image segmentation of the new coronary pneumonia is more refined than before, with loss values of 0.0021 and 0.0025 on the training and validation sets, and 0.9834 and 0.9935 of IOU, respectively. And the training time has also been reduced. The original 2D V-Net training time was 46640 seconds, and the training time after adding the self-attention model was 41744 seconds, a reduction of 5.76%.

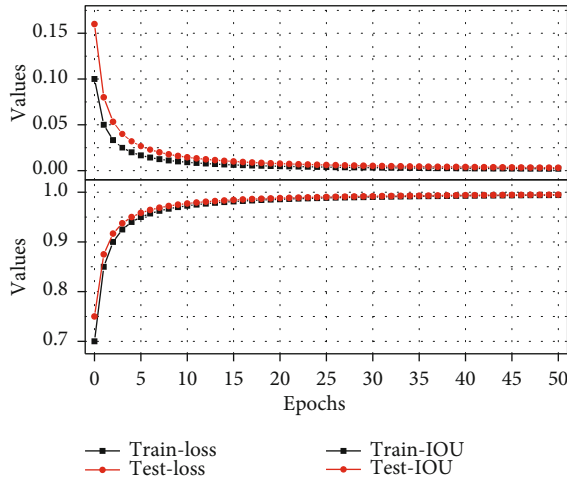


FIGURE 6: Loss and IOU curves of the training set and validation set.

By using the traditional segmentation method and deep learning-based segmentation method, the new-type coronary pneumonia CT image is subjected to binary segmentation of the new coronary pneumonia region, and the results of the level set segmentation are used as the labels for the neural network U-Net and V-Net training, and the original image is simultaneously enhanced with data to expand the data volume. From the segmentation results shown in Figure 6, V-Net performs better and more efficiently than U-Net. Also, this chapter improves 2D V-Net based on the self-attention model. The improved model can more finely segment the target edge, effectively improving the performance of the model. The segmentation results in this chapter will be used for subsequent image diagnosis and three-dimensional reconstruction. Therefore, accurate segmentation results lay the foundation for subsequent work.

In this study, multiple semantic attributes are used as the source task, and multilevel correlation transfer learning is used to assist in the diagnosis of the pathological benign and malignant of pulmonary nodules. First, we use first-order association transfer learning to see which source tasks can better support the pathological diagnosis, then through all the tenth-order transfer learning to find the source task that makes the diagnosis of benign and malignant lung nodules the best. Finally, we compared the multilevel correlation migration method with existing methods and the subjective classification performance of radiologists. We use accuracy, area-under-curve (AUC) score, and Macro-F1 score as evaluation indicators, as shown in Figure 7.

Figure 7 shows the performance of first-order transfer learning-assisted diagnosis of pathological benign and malignant pulmonary nodules on accuracy and Macro-F1 indicators. We sorted the contents in Figure 7 by Macro-F1 from small to large for easy comparison. From Figure 7, in the ten semantic source tasks, “text” is used in accuracy, AUC, and Macro-F1 in the first-order transfer learning to assist in the diagnosis of pathological benign and malignant lung nodules. The above performances are the best, indicating that the “text” feature has very important and valuable feature information for the diagnosis of pathological benign and

malignant lung nodules. In the first-order migration experiment, the semantic degree (mal) as the source task feature migrated to “pathological benign and malignant (ppm).” Since the semantic malignancy attribute is a doctor’s subjective judgment on the malignancy of lung nodules, especially when making a malignant judgment on lung nodules, general doctors tend to conservatively estimate, which may lead to a high misjudgement rate; in the LIDC-IDRI data set, the doctor’s misjudgement rate of benign and malignant lung nodules is as high as 28.56%. In our experiments, it was confirmed that using semantic benign and malignant attributes as the source task of transfer learning is not the best choice.

3.2. Forecast Results Analysis. We initialize the model with the optimal RetinaNet model trained as a parameter, and then when the data is enhanced, no other expansion methods that may change the location information are used except horizontal symmetry, because in this network, we use the location and scale information. It is not suitable for data expansion through operations such as rotation, translation, and scaling. At the same time, during training, we fixed the network parameters of ResNet in front of the network and only trained the feature pyramid and the parameter weights of the classification and regression subnetwork parts. The selection of hyperparameters for training is the same as when training RetinaNet, and the network with the highest mAP was also selected for comparison. To generate the FROC curve, we selected different score thresholds to output the test results, and then based on the test results, respectively, we calculated the sensitivity under different thresholds; the average false-positive rate and the specific statistical data are shown in Figure 8 This generates the FROC curve.

Finally, the FROC curves of the two models are shown in Figure 8. We can see from the results that the overall FROC curve of LSFNet is above the original RetinaNet model. This shows that the improvement made in this paper for the detection of lung nodules is effective. This is because the points 0.9 and 0.95 are special in nature. Then, to show the difference in effect more intuitively, we select two points with a sensitivity of 0.85 and 0.97 from the curve for specific comparison, as shown in Figure 9.

As can be seen from Figure 9, when the sensitivity is the same, the method in this paper can bring a lower false-positive rate. Next, we select one of the cases from dataset A and show the detection results of the two models separately. The sample original image and the labeling situation are shown in Figure 9. This chest radiograph has only one lung nodule, but there are many highlighted areas in the picture, which are easily mistaken as lung nodules. To ensure that the lung nodules are correctly found, it is also to more easily verify that the method proposed in this article can effectively reduce false positives. For the number of samples, we set the threshold to 0.01, so that there will be some false-positive samples in the test results.

3.3. Model Diagnosis and Result Analysis. The input of the generator is a set of random noise. The generated result is judged by the discriminator, and the discriminated result is returned to the generator. Because the new coronary artery

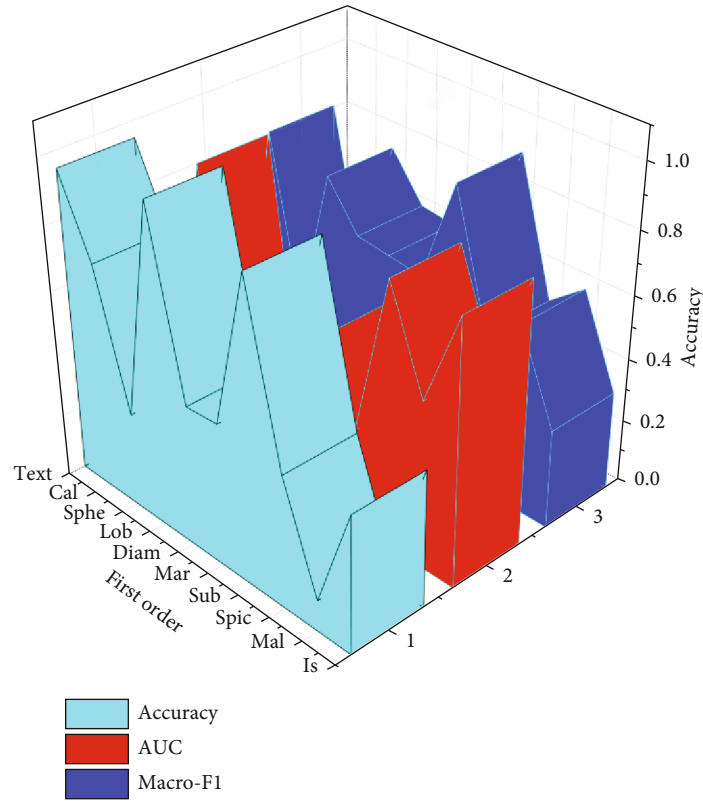


FIGURE 7: The performance of first-order transfer learning-assisted diagnosis of pathological benign and malignant pulmonary nodules on accuracy, AUC, and Macro-F1 indicators.

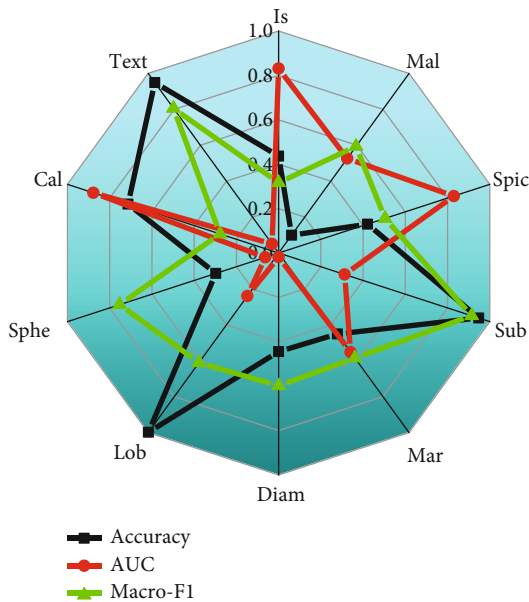


FIGURE 8: Comparison of FROC curve of data set A.

pneumonia area is smaller than the background area, to reduce the influence of the background area, the image with the initial size of 512×512 is cropped and scaled, and the size becomes 128×128 . Using the same data, DCGAN, WGAN-GP, and PGGAN were trained separately. The implementation details are shown in Figure 10. The generated new coro-

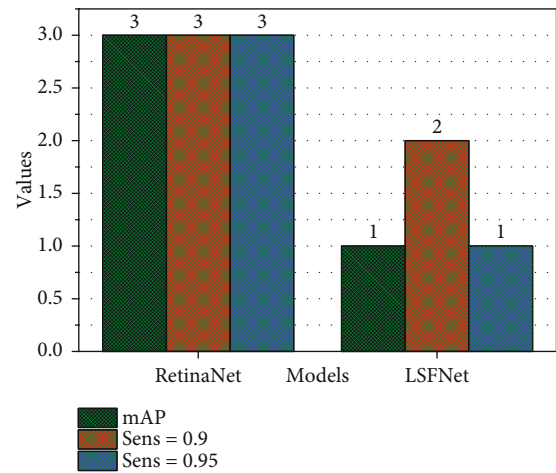


FIGURE 9: Comparison of experimental results of data set A.

nary pneumonia images are shown in Figure 10. It can be seen that the results generated by WGAN-GP and PGGAN are better than those of DCGAN. Better yet, although DCGAN has a good network framework, the gradient disappears easily during the training process, which is not easy to train. For WGAN-GP and PGGAN, the former is a lightweight network structure, the training process is stable, and there is no need to balance the training times of the generator and the discriminator. The latter is due to the progressive training method, in the same hardware and software

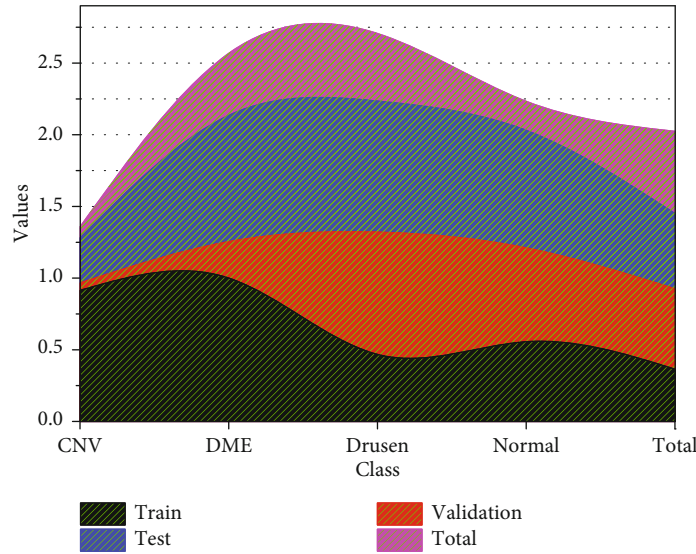


FIGURE 10: WGAN-GP model training results.

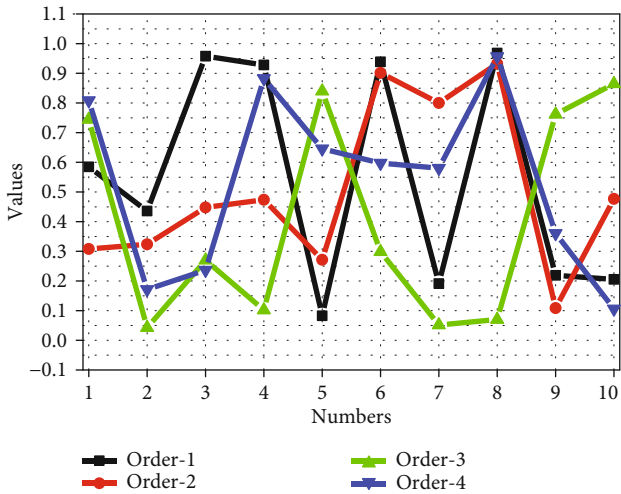


FIGURE 11: Image diagnosis results.

environment. Next, the training time is about five times that of the former.

Based on the above image diagnosis process, we first add a mask to the new coronary artery pneumonia area, as shown in Figure 11, and then use it as the input of the generation network, because the model has been trained at this time, that is, the model has saved the information of the generated image; for the image model with a mask added, the mask area will be filled with appropriate and contextual information to make it more consistent with the content of other areas, which is achieved by the loss function. Finally, when the loss function converges, the mask area is filled, that is, the final diagnosis image is obtained, as shown in Figure 11.

For the evaluation of the image quality after diagnosis, the paper chooses the largest mean difference, which was first proposed for the two-sample detection problem to determine whether the two distributions are the same. If the value is small enough, the two distributions are considered the same; otherwise, they are considered different. At the same time,

this value is also used to determine the similarity between the two distributions. This article uses it to measure the similarity between the original image and the image after diagnosis. KMMD was used for image diagnosis based on WGAN-GP. It is 0.001084, which indicates that the model is effective in diagnosing the missing image area. Volume rendering and surface rendering are two methods of three-dimensional reconstruction. Volume rendering is the direct projection of voxels onto the monitor screen, so it is also called direct rendering; surface rendering is the use of geometric figures to usually fit the target surface for triangular patches. It can also be called indirect drawing. Among them, the better algorithms are the ray casting algorithm and the moving cube algorithm. The accuracy of our model has improved by about 5% compared to the comparative model.

The basic principle of the ray casting algorithm is to emit a ray from each pixel on the screen along the line of sight. If the ray passes through the volume data, it is sampled at equal intervals along the ray direction, and the color of each sampling point is cumulatively calculated by interpolation value and opacity, when the light has passed through the voxel space or reached the set opacity, the projection ends; then, the color values and opacity of the sampling points on each ray are synthesized in order, to obtain the pixels of the projection plane. In the attribute value, this method belongs to volume rendering, which is easy to achieve a good rendering effect and is widely used. First input the original data, read a series of two-dimensional images into memory, construct volume data, and perform preprocessing operations on it. Then, select the appropriate conversion function to convert the scalar values of the volume data to color and opacity from the screen. During the process of the light passing through the entire image sequence, the image sequence is sampled to obtain color information and opacity. Finally, the image plane is rendered according to color accumulation.

The moving cube is a classic algorithm for surface rendering and the basis of all voxel-based surface rendering algorithms. The algorithm treats the surface of an object as a

closed gray-scale isosurface. If the gray-scale value of the voxel is less than the gray-scale value, then the voxel is located outside the isosurface, and accordingly, if the gray value of the voxel is greater than the gray value, then the voxel is located inside the isosurface. According to the relationship between the data values of the eight vertices of each cube unit and the given data value of the volume data, the algorithm finds the equivalent points on the twelve sides of the cube unit and then connects the equivalent points with triangles to form the isosurface. The specific approach is as follows: First, for volume data, place it in a cube that can be surrounded, then each cross-section parallel to the horizontal plane is a CT image, and then, divide the cube into small cubes, that is, voxel, used for resampling in three-dimensional space, that is, to determine whether the eight vertices of the small cube are inside the target. If a vertex is inside the target object, then the vertex is marked as 0; otherwise, it is marked as 1. The judgment method is to set a threshold, compare the pixel value of the image with it, and determine the state of the vertex. Next, determine the state of the eight vertices according to the picture. There are a total of 256 cases in the arrangement and combination. For each case, some isosurfaces can be generated in the small cube. After obtaining the isosurfaces, calculate the corresponding method.

4. Conclusion

In this paper, the U-Net series neural network and the improved 2D V-Net are used to segment the new coronary pneumonia CT images to separate the vertebral body region from other regions. It has good robustness and solves the image noise and tissue incompatibility. Evenness and other issues avoid the defects of traditional image segmentation method of poor edge segmentation. The model used for lesion detection is a network model based on the target-detection task. However, the accuracy requirements of medical images are higher, and a better model is needed; for the diagnosis of the new coronary pneumonia model, consider the spatial information of the new coronary pneumonia and the new coronary pneumonia; the diagnosis of the image can increase the feature information of the upper and lower CT slices; in this paper, based on the target-detection framework Faster RCNN, the feature extraction of the new coronary pneumonia tumor image is performed to obtain a higher-level expression of the image, predict the tumor lesion location, and determine the location of the tumor. Detection and positioning can provide support for doctors' further diagnosis.

Data Availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Consent

Informed consent was obtained from all individual participants included in the study references.

Conflicts of Interest

We declare that there is no conflict of interest.

Acknowledgments

We thank the Research on Construction and Application of Artificial Intelligence 5G Remote Diagnosis for Difficult Wound (Sichuan Department of Science and Technology, Key R&D projects, No.: 2020YFS0298).

References

- [1] T. A. Retson, A. H. Besser, S. Sall, D. Golden, and A. Hsiao, "Machine learning and deep neural networks in thoracic and cardiovascular imaging," *Journal of Thoracic Imaging*, vol. 34, no. 3, pp. 192–201, 2019.
- [2] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
- [3] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.
- [4] J. R. England and P. M. Cheng, "Artificial intelligence for medical image analysis: a guide for authors and reviewers," *American Journal of Roentgenology*, vol. 212, no. 3, pp. 513–519, 2019.
- [5] G. B. Kim, K. H. Jung, Y. Lee et al., "Comparison of shallow and deep learning methods on classifying the regional pattern of diffuse lung disease," *Journal of Digital Imaging*, vol. 31, no. 4, pp. 415–424, 2018.
- [6] M. J. Cha, M. J. Chung, J. H. Lee, and K. S. Lee, "Performance of deep learning model in detecting operable lung cancer with chest radiographs," *Journal of Thoracic Imaging*, vol. 34, no. 2, pp. 86–91, 2019.
- [7] M. Mannil, J. von Spiczak, R. Manka, and H. Alkadhi, "Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible," *Investigative Radiology*, vol. 53, no. 6, pp. 338–343, 2018.
- [8] T. Rahmat, A. Ismail, and S. Aliman, "Chest x-rays image classification in medical image analysis," *Applied Medical Informatics*, vol. 40, no. 3-4, pp. 63–73, 2018.
- [9] S. Chilamkurthy, R. Ghosh, S. Tanamala et al., "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.
- [10] X. Liu, L. Faes, A. U. Kale et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The Lancet Digital Health*, vol. 1, no. 6, pp. e271–e297, 2019.
- [11] C. P. Langlotz, B. Allen, B. J. Erickson et al., "A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop," *Radiology*, vol. 291, no. 3, pp. 781–791, 2019.
- [12] M. Dewey and U. Wilkens, "The bionic radiologist: avoiding blurry pictures and providing greater insights," *NPJ Digital Medicine*, vol. 2, no. 1, 2019.

- [13] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.
- [14] K. H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 719–731, 2018.
- [15] S. Park, S. M. Lee, N. Kim et al., "Application of deep learning-based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy," *European Radiology*, vol. 29, no. 10, pp. 5341–5348, 2019.
- [16] F. Jiang, Y. Jiang, H. Zhi et al., "Artificial intelligence in healthcare: past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017.
- [17] J. C. Seah, J. S. Tang, A. Kitchen, F. Gaillard, and A. F. Dixon, "Chest radiographs in congestive heart failure: visualizing neural network learning," *Radiology*, vol. 290, no. 2, pp. 514–522, 2019.
- [18] S. Park, S. M. Lee, K. H. Lee et al., "Deep learning-based detection system for multiclass lesions on chest radiographs: comparison with observer readings," *European Radiology*, vol. 30, no. 3, pp. 1359–1368, 2020.
- [19] M. A. Mazurowski, M. Buda, A. Saha, and M. R. Bashir, "Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 939–954, 2019.
- [20] I. Pan, S. Agarwal, and D. Merck, "Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks," *Journal of Digital Imaging*, vol. 32, no. 5, pp. 888–896, 2019.
- [21] E. Pons, L. M. Braun, M. M. Hunink, and J. A. Kors, "Natural language processing in radiology: a systematic review," *Radiology*, vol. 279, no. 2, pp. 329–343, 2016.
- [22] M. J. Willeminck, W. A. Koszek, C. Hardell et al., "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, pp. 4–15, 2020.
- [23] S. M. Humphries, K. Yagihashi, J. Huckleberry et al., "Idiopathic pulmonary fibrosis: data-driven textural analysis of extent of fibrosis at baseline and 15-month follow-up," *Radiology*, vol. 285, no. 1, pp. 270–278, 2017.
- [24] C. J. Liew, P. Krishnaswamy, L. T. Cheng, C. H. Tan, A. C. Poh, and T. C. Lim, "Artificial intelligence and radiology in Singapore: championing a new age of augmented imaging for unsurpassed patient care," *Annals of the Academy of Medicine, Singapore*, vol. 48, no. 1, pp. 16–24, 2019.
- [25] K. Y. Ngiam and W. Khor, "Big data and machine learning algorithms for health-care delivery," *The Lancet Oncology*, vol. 20, no. 5, pp. e262–e273, 2019.