

Retraction

Retracted: News Text Classification Method and Simulation Based on the Hybrid Deep Learning Model

Complexity

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Complexity. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

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

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

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

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

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

References

- [1] N. Sun and C. Du, "News Text Classification Method and Simulation Based on the Hybrid Deep Learning Model," *Complexity*, vol. 2021, Article ID 8064579, 11 pages, 2021.

Research Article

News Text Classification Method and Simulation Based on the Hybrid Deep Learning Model

Ningfeng Sun¹ and Chengye Du² 

¹School of Humanities, Southwestern University of Finance and Economics, Chengdu, Sichuan 610036, China

²School of Film and Television, Yunnan Arts University, Kunming, Yunnan 650500, China

Correspondence should be addressed to Chengye Du; 290016@ynart.edu.cn

Received 28 April 2021; Revised 6 June 2021; Accepted 11 June 2021; Published 18 June 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Ningfeng Sun and Chengye Du. 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 paper uses the database as the data source, using bibliometrics and visual analysis methods, to statistically analyze the relevant documents published in the field of text classification in the past ten years, to clarify the development context and research status of the text classification field, and to predict the research in the field of text classification priorities and research frontiers. Based on the in-depth study of the background, research status, related theories, and developments of online news text classification, this article analyzes the annual publication trend, subject distribution, journal distribution, institution distribution, author distribution, highly cited literature analysis, and research hotspots. Forefront and other aspects clarify the development context and research status of the text classification field and provide a theoretical reference for the further development of the text classification field. Then, on the basis of systematic research on text classification, deep learning, and news text classification theories, a deep learning-based network news text classification model is constructed, and the function of each module is introduced in detail, which will help the future news text classification of application and improvement provide theoretical basis. On the basis of the predecessors, this article separately studied and improved the neural network model based on the convolutional neural network, cyclic neural network, and attention mechanism and merged the three models into one model, which can obtain local associated features and contextual features and highlight the role of keywords. Finally, experiments are used to verify the effectiveness of the model proposed in this paper and compared with traditional text classification to prove the superiority of the network news text classification based on deep learning proposed in this paper. This article aims to study the internal connection between news comments and the number of votes received by news comments, and through the proposed model, the number of votes for news comments can be predicted.

1. Introduction

With the rapid development of internet technology and mobile communication technology, online news has become one of the important information sources for people's daily life, study, and work [1]. Compared with other traditional media, online news is rich in content and diverse in forms and can be updated in real time. Users can obtain and read online news without being restricted by time and space, and it can be obtained at any time. Users can obtain them from the internet according to their needs [2]. Online news can

bring users a better three-dimensional and all-round information receiving experience [3]. As one of the core technologies of information resource organization and management, text classification can solve the problem of information clutter to a large extent, help users narrow the scope of information retrieval, and make users more convenient and efficient to filter through the massive information resources to meet their own needs. The information is a powerful means to deal with massive information resources [4]. Especially in the context of the big data era, the importance of text classification technology for the effective

organization and utilization of massive text information is more prominent, with very important research significance and broad application prospects [5].

However, in the face of massive and chaotic network news information, the problem that users face is no longer how to find news resources, but how to efficiently and accurately obtain information that can meet their own needs from the massive news resources [6]. Therefore, in order to meet the diverse and personalized needs of online news users in the context of the big data era, effective information organization and management of online news have become an urgent problem to be solved [7]. News text classification based on traditional machine learning mainly includes the following problems: on the one hand, the use of traditional news text representation methods to represent news texts will ignore the word order and semantic information contained in the news text; on the other hand, it will ignore the word order and semantic information contained in the news text. The process of feature extraction of news text requires manual participation, which is subjective and consumes more time and energy; at the same time, using vector space model to represent news text will cause high latitude and data sparsity problems, although the feature selection method can be used for dimensionality reduction, but this will further aggravate the problem of news text feature loss and make the entire news text classification process more complicated. These problems will directly affect the effect of news text classification. Therefore, it is necessary to seek more ingenious text feature representation and text feature extraction models.

This paper mainly conducts research from two aspects: text feature representation and text feature selection in network news text classification. Traditional text classification methods have many difficult problems to solve. Aiming at the high-dimensional sparseness of traditional text representation methods in natural language, the embedding layer converts text data into low-dimensional dense vectors, avoiding the dimensional disaster caused by high-dimensional input. At the same time, by using word vectors, the impact of word segmentation errors on the accuracy of text classification tasks is avoided, and the performance of the classifier is improved. The process of feature extraction requires manual participation, which will affect the accuracy of the final extracted text features; using vector space model for text representation will ignore the word order and semantics in the text. Information affects the performance of text classification; in the face of high latitude and data sparsity problems, although feature selection methods can be used to reduce dimensionality, this will further aggravate the problem of text feature loss and make the entire text classification process more difficult. In order to solve the above problems, this article reexamines the traditional process of online news text classification based on the research of online news text classification and uses deep learning methods as the theoretical basis to reconstruct the process of online news text classification using deep learning related theories and models to achieve the purpose of solving the problems of traditional text classification and improving the effect of text classification.

2. Related Work

Some related research on text classification can be traced back to the 20th century. Abdi et al. [8] proposed the concept of word frequency and applied the concept of word frequency statistics to text classification for the first time, laying a foundation for text classification and creating a precedent for text classification research. The method of text classification using keyword classification technology proposed by Wu et al. [9] was of milestone significance, which has greatly pioneered and promoted the research work of text classification. Subsequently, more and more researchers have carried out a series of studies and achieved many results, and text classification technology has received more research and applications. Many well-known intelligence scientists, such as Shon et al. [10], have carried out research in the field of classification and have achieved remarkable research results. At that time, text classification was mainly based on word matching, which determined the category of the document according to the frequency of common words in the text and category name. This classification method is simple and intuitive, the classification rules are very mechanical, and the classification effect is not good. In the future, text classification has gradually transitioned from the idea of word frequency statistics to a method based on knowledge engineering. This method needs to rely on experts to compile classification rules. Compared with the word matching method, it has a higher understanding of logical rules and the classification effect is also improved. Because documents may be different in length (that is, the number of words is different), in order to be able to feed to a fixed-dimensional neural network, we need to set a maximum number of words. For documents with the number of words less than this threshold, we need to use "unknown words" to go filling. For example, the word with index 0 in the vocabulary can be set as "unknown word," and 0 is used to fill in the part less than the threshold. However, the formulation of such classification rules needs to rely on domain experts to manually compile for specific fields, which will consume a lot of time and energy, and cannot process text information with a large amount of data. And these rules are usually oriented to specific fields and between fields. The versatility is poor, and the range that can be covered is very limited. Therefore, methods based on knowledge engineering have not been widely used.

With the rapid growth of text information resources on the Internet, text classification methods have received unprecedented attention. However, the text classification technology based on knowledge engineering has been completely unable to meet the demand, has been gradually replaced by the emerging statistics-based machine learning text classification method, and has quickly become the new mainstream method in the field of text classification, and it is still the focus of many scholars' research. The text classification method based on statistical machine learning learns the sample data of known categories. Girgis et al. [11] use the learned category features to construct a classifier, then use the classifier to classify the text information to be classified, and finally obtain the text information. Compared with the method based on

knowledge engineering, the construction of the classifier in this method does not require human involvement, which greatly reduces manpower and material resources. Pasupa and Ayutthaya [12] fused deep learning algorithms in different ways and compared it with several other fused combinations. In terms of text classification efficiency and accuracy, Buabin [13] proposes that they have a very significant improvement. It is precisely because of the more reliable theoretical basis and better classification results that the text classification method based on machine learning has received extensive attention from scholars [14], and it is still the focus and mainstream of researchers' application and research, with a wide range of applications, such as text mining, pattern recognition, information retrieval, data mining, learning systems, and other fields [15–18]. Among the machine learning methods, the more commonly used are class center vector method, K -nearest neighbor method, and support vector machine method [19]. The research on text classification is relatively late. The earliest scholars summarized and introduced classification technology and research status, which drove scholars in the field of library and information research on Chinese text classification [20, 21]. As one of the core technologies of information resource organization and management, text classification has received great attention from many researchers [22–24]. Since then, researchers have combined the specific knowledge of Chinese texts and absorbed the results of English text classification. Reconstruction and optimization make it suitable for Chinese text classification. After continuous exploration and development by researchers, a Chinese text classification research system has been formed [25].

With the development of text classification technology, not only has it attracted the attention of many researchers, universities, research institutes, and enterprises at all levels in the country have also paid great attention to the research of text classification technology, not only related researchers and scholars who study text classification technology. With the strong support of many funds, there are more and more papers related to text classification, which promotes the rapid development of text classification research. The Chinese Academy of Sciences, Peking University, Harbin Institute of Technology, Tsinghua University, Shanghai Jiaotong University, and other universities and research institutes have conducted a lot of research in the field of text classification. After continuous exploration and research by researchers, my country has obtained fruitful research results in the field of Chinese text classification. Typical representative systems include the Zhiduoxing Chinese text classifier of the Institute of Computing Technology, the text classification of Fudan University, and the text classification of Peking University. Some of them have been successfully promoted and applied [26–28].

3. Construction of the News Text Classification Model Based on Hybrid Deep Learning

3.1. News Text Feature Selection. Feature selection is to select some of the most representative features of the text content

from the original feature space for text classification without affecting the nature of the original feature space. The basic principle of feature selection is to rank the original text word sequence with the aid of the evaluation function. By selecting some relatively high score features as the final text feature, the dimensionality of the text feature space is reduced, thereby improving the Chinese language in the field of online news.

$$\begin{aligned} u &= s \cdot \frac{ax + b}{a + b}, \\ v &= s \cdot \frac{s \cdot x + t}{s \cdot x - t}. \end{aligned} \quad (1)$$

Aiming at the problem of poor network news text classification caused by defects such as high vector latitude, sparse data, and lack of semantics in traditional machine learning news text classification methods, the word vector method is used to represent news text word sequences, which can be effective to solve these problems. In the optimization process, the introduction of momentum can speed up convergence and reduce meaningless oscillations. When the gradient points to the actual moving direction, the momentum term increases; when the gradient is opposite to the actual moving direction, it decreases. This ensures that the direction of optimization is always towards the minimum point, reduces unnecessary updates, and improves the optimization effect while reducing computational overhead. The main idea of the word vector is to map each word to the low-latitude space. In the new low-latitude feature space after the mapping, the positional relationship between the word vectors corresponding to different feature words represents the semantics between them. The association on the level can solve the problems of vector sparseness and semantic lack. Assuming that any piece of online news text in the experimental data set has k words,

$$s(x) = \frac{e^x - 1}{e^x + 1}. \quad (2)$$

The convolutional neural network used in this paper draws on the model design of Kim's paper, and the specific structure is shown in Figure 1. For each input element s , use $s(x)$ to represent the value obtained after x is processed by the function. The leftmost input layer of the convolutional neural network structure is a $k \times n$ two-dimensional word vector matrix, where k represents the length of a news text composed of words $k_1, k_2, k_3, \dots, k-i$, and n represents the length of each word vector dimension. The feature extraction part of the convolutional neural network structure mainly includes operations such as convolution and pooling. Finally, a corresponding model is constructed on the task of news text classification.

$$y(x) = \sum u(x) \cdot \lg \frac{u(x)}{v(x)}. \quad (3)$$

In the convolutional layer, the news text is input into the convolutional neural network in the form of a two-

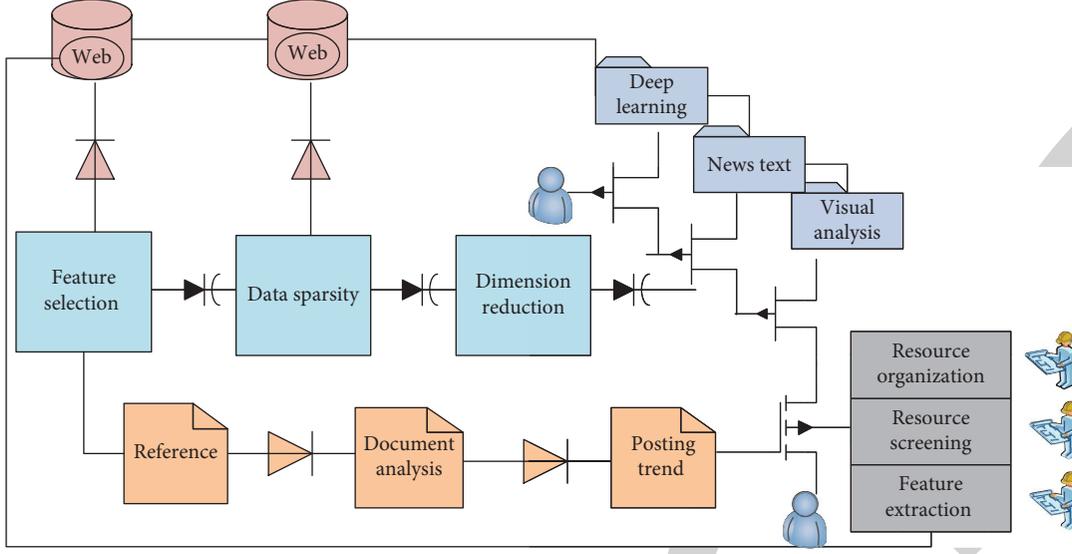


FIGURE 1: Hierarchical distribution of news text.

dimensional matrix, where u represents the result of encoding, and then u is decoded in the decoding module, and the convolution operation is used to extract the features of the news text. The convolution operation of the convolutional neural network is as follows:

$$w(x) = \sum \sum \max\{0, 1 - u(x) + v(x \cdot s(x))\}. \quad (4)$$

Among them, $c - i$ represents the result of the convolution operation from the i to $i + m - 1$ word vectors; f selects the ReLu function as the activation function, which can speed up the training convergence speed of the convolutional neural network; it represents the convolution operation, that is, feature extraction; $X - i: i + m - 1$ represents the word vector input to the i to $i + m - 1$ in the convolutional neural network; and W and B represent the weight matrix and bias of the convolutional neural network, respectively.

After the news text features are obtained through the convolution operation, if these features are directly used for classification, since there are still a lot of feature data, the amount of calculation will be very large and the training process will be slow. On the premise of preserving useful features, simplifying the number of parameters of convolutional neural networks and reducing computational complexity, it is necessary to use pooling operations to compress and merge semantically similar news text features, which can play a role in secondary extraction of features. Adaptive Moment Estimation can automatically select the appropriate learning rate. This algorithm stores the exponential decay mean of the gradient used in the optimization during the optimization process, which is similar to the momentum method. The pooling operation of convolutional neural network is shown in formula (4). In the fully connected layer, the results of the pooling operation are summarized, as shown in the following equation:

$$w(x) = \sum \sum \max\{0, 1 - u(x) + v(x \cdot s(x))\}, \quad (5)$$

$$z(x) = \frac{e^{w(x) \cdot a + b}}{\sum e^{ax + b}}. \quad (6)$$

In order to prevent the convolutional neural network from overfitting during the training process, the dropout strategy is adopted; that is, some hidden nodes are discarded with a certain probability during model training, so that it is impossible to train complex ones through the joint action of all hidden nodes. The output layer will generate CLASS_NUM probability values, which respectively represent the probability that this document belongs to the current class. The network model of each training is different, which improves the generalization ability of the network model, thus effectively preventing the phenomenon of overfitting. In the final output layer, all news text features are mapped into category information according to formula (6).

3.2. Hybrid Deep Learning Algorithm. First, a large number of special characters in the news text are filtered through regular expressions, leaving the information format of the plain text; then, as the most basic semantic unit, words need to be used to represent the characteristics of the text in the news text classification.

$$g(x) = \begin{cases} \frac{f(x) - x}{f(x)^a}, & x < \max(x), \\ 1, & x \geq \max(x). \end{cases} \quad (7)$$

For the vocabulary in each news text, use Word2Vec, a word vector open-source tool based on deep learning launched by Google to train and generate news text word vectors and convert news text features into dense word vectors of specific dimensions suitable for convolutional

neural network processing. It indicates that the training corpus comes from a 1.02 GB network-wide news text data package (SogouCA) released by Sogou Lab. The text is of variable length. The method of expressing text as numbers or vectors that a computer can operate is generally called the Word Embedding method. Word embedding transforms text of variable length into a space of fixed length, which is the first step in text classification. In this paper, the word vector dimensions of news text are set to 50, 100, 150, 200, 250, and 300, respectively, to select the best experimental parameters. News text word vectors of each dimension are trained and generated by Word2Vec.

Currently, commonly used feature selection methods include document frequency, chi-square statistics (χ statistics), and information gain. Compared with feature selection from Figure 2, feature extraction uses a more advanced method, transforming the original feature space of the text through mapping to generate a new feature space with lower dimensions and more independent between each dimension, thereby realizing the optimization of text features and dimensionality reduction. The Gradient Descent algorithm is a commonly used optimization algorithm, and the purpose of the optimization algorithm is to use an improved training method to maximize (or minimize) the loss function $J(X)$. There are some parameters inside the model; these parameters are used to calculate the true value and predicted value of the target value Y in the test set the degree of difference. However, because feature extraction is closely related to the analysis of text semantics, and the related technologies in this area are not yet mature, the dimensionality reduction effect is not ideal. Compared with feature extraction methods, the text feature set selected by feature selection is a subset of the original feature space. It has the advantages of semantic background, easy understanding, simple model, and easy implementation. Therefore, it has attracted the attention of many scholars and has become a feature reduction method in main use method of dimension.

3.3. Structure Optimization of the Text Classification Model. Before conducting the comparative experiment, we first train and adjust the relevant parameters of the convolutional neural network in the experiment on the training set and the validation set and make a final test on the effect of news text classification on the test set. The parameters mainly include the word vector dimension of the news text input into the convolutional neural network, the size of the convolution kernel of the convolutional neural network, the dropout parameter, and the total iteration rounds for training the convolutional neural network structure. DESEM-AE gave the worst results. Because of the feature extraction of stacked denoising autoencoders, DESEM-AE has achieved a significant improvement in accuracy and recall. The accuracy of the TextCNN model in the 10th category is 100%, but the Attention model is only 87%. The Attention model has a recall rate of 91% in the second category and only 77% in the TextRCNN. This shows the use of autoencoders from the original. The weighted mixed vector of the key elements of

the event in the news article is used to extract the importance of the hidden event feature. The text segmentation is a relatively basic and important part. Due to the flexibility and open source of the Chinese word segmentation database, this article uses it to segment news text and convert the original news text into the word sequence of the news text, so that the word2Vec training and word vectors can be used to represent the news text. Finally, according to the stop word list, the words and non-Chinese characters that have no real meaning but appear frequently in the news text are filtered to obtain the news text word sequence after the stop words are removed, so that the remaining words can be better to express the characteristics of the news text.

This shows that although the word embedding vector can capture the syntactic and semantic rules to some extent, a higher-level feature extraction method is needed if a better event line extraction performance is to be achieved. Due to frequent iterative calculations and parameter updates, stochastic gradient descent will eventually fall into a certain local minimum and fluctuate around this minimum. At the same time, experiments have proved that when we slowly reduce the learning rate Q during the training process of stochastic gradient descent, the convergence process of the parameters is the same as the standard gradient descent. Compared with AE + k -means, we can see that DESEM has a significant improvement in the recall rate, which shows that compared to the k -means method, the neural network-based method for event line extraction can be more effective of extracting the real event line. The influence of the size of the convolution kernel in the convolutional neural network on the accuracy of news text classification is shown in Figure 3. It can be seen from the figure that when the size of the convolution kernel is 3, the accuracy of news text classification reaches the maximum. The smaller the size of the convolution kernel is, the fewer parameters and calculations required to train the convolutional neural network are. Therefore, the choice of the size of the convolution kernel is the most appropriate under the premise of ensuring high accuracy. Compare the classification effect with the traditional news text classification model that uses vector space model for news text representation, TF-IDF for text feature selection, and SVM classifier for classification to verify the news text classification based on deep learning proposed in this article for the validity of the model.

4. Application and Analysis of the News Text Classification Model Based on Hybrid Deep Learning

4.1. News Text Preprocessing. In this paper, the classification effect of the network news text classification method based on deep learning is experimentally verified. Taking Sina News as an example, the experiment is c on Google's open-source deep learning system TensorFlow. The experimental data set used in this article comes from a subset of the public data set THUCNews Sina News of the Natural Language Processing Laboratory of Tsinghua University. The experimental data set contains 10 categories: sports, finance, real

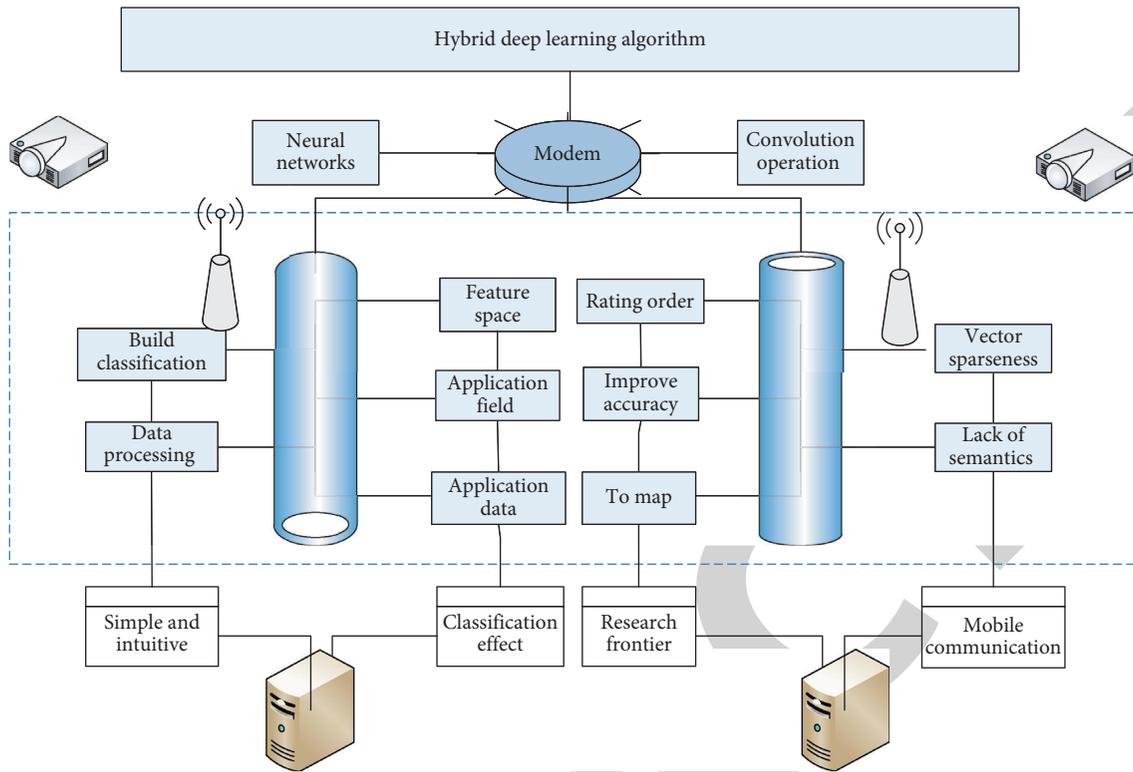


FIGURE 2: Hybrid deep learning algorithm process.

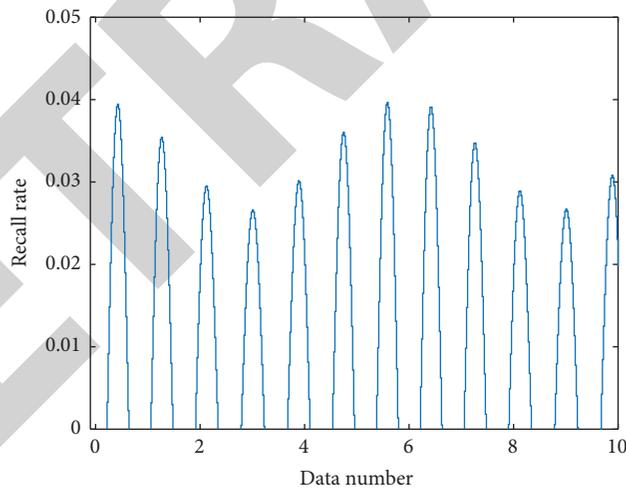


FIGURE 3: Ladder diagram of recall rate changing with data.

estate, home furnishing, education, technology, fashion, current affairs, games, and entertainment, 6500 data in each category; the experimental data sets of the labeled categories are divided as follows: 5000*10 training set, 500*10 validation set, 1000*10 test set, and one of these 3 data sets. By analyzing the annual trends in the text classification field, we can clarify the development and research situation of the text classification field from the concept of time, grasp the research status of the text classification field, and predict the development trend of the text classification field to a certain extent. In this paper, the number of relevant documents in

the text classification field changes over time in lines of statistics, as shown in Figure 4.

A total of 40 salient words are obtained by calculation. It can be clearly seen that the top three keywords with salient value are convolutional neural network, deep learning, and word vector. The salient value is document frequency of attribute reduction, feature dimensionality reduction, rough set, genetic algorithm, topic model, natural language processing, sentiment analysis, word2vec, etc., and word vectors are keywords with high prominence values in recent three years. Therefore, the current research fronts in the field of

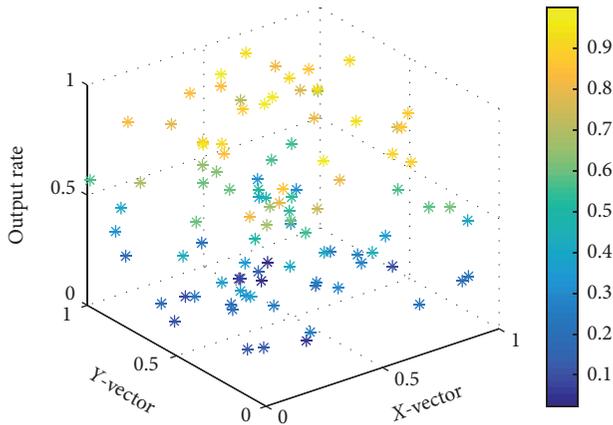


FIGURE 4: Three-dimensional scatter distribution of news text output rate.

text classification mainly include text classification models, text classification applications, word vector text representation methods, and the extraction of text features based on deep learning.

4.2. Deep Learning Module Simulation. From Figure 5, we can intuitively see that text classification belongs to a multidisciplinary field. Among all the statistical documents, the number of related documents in the field of text classification is the largest in the field of computer science. A total of 1,827 documents have been published, accounting for the total number of documents which is 87.88%, indicating that the development of text classification is inseparable from the support of computer technology. We choose convolution kernels of different scales for the title and text. The length of the title text is shorter, so the convolution kernel scale used is 2 to 5, and the length of the body text is longer and mostly sentence information, so choose 3 to 9. The information extracted from the title and text is combined using a linear layer with activation function and sent to the final output layer. Computer technology is an important pillar of text classification research. Many scholars in the computer field have made important contributions to the development of text classification. The next in order is control engineering that has published 244 articles in the field of text classification, accounting for 11.74% of the total number of documents, 148 articles in library and information archives accounted for 7.12%, 23 articles in education accounted for 1.11%, and so on. Among them, the related papers in the field of text classification in the discipline of library, information, and archives accounted for 7.12% of the total literature, second only to the discipline of computer and control engineering, indicating that library, information, and archive science is an important research field in the field of text classification. The subject's research on the technical links involved in text classification is relatively weak. It mainly focuses on related research in text classification theory, such as the development of the text classification field, the basic theories and methods of text classification, and the application in digital libraries. If it can

be deeply integrated with other disciplines and developed comprehensively, there will be greater room for improvement and development prospects.

The influence of dropout parameter changes on the accuracy of news text classification is shown in Figure 6. Use the dropout strategy to avoid overfitting. At the same time, for the internal structure of the model, we have introduced batch normalization, which speeds up the convergence of model parameters and enhances the stability of the model. Set the dropout parameter from 0.1 to 0.9. It can be seen from Figure 6 that as the value of the dropout parameter increases, the accuracy of news text classification is gradually improving, indicating that in deep learning. The introduction of the dropout strategy in the process is very effective. When the value of the dropout parameter is 0.4–0.6, the accuracy is 0.88, 0.895, 0.91, 0.925, 0.94, 0.955, 0.97, 50, 100, 150, 200, 250, and 300. The dimensionality of the news text word vector is 0.89, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 1, 2, 3, 4, 5, 6, and 7. The core size rate is relatively high, and the accuracy of news text classification reaches the highest when the dropout parameter value is 0.5. At this time, the effect of dropout to prevent overfitting is the best.

This also shows that the application value of text classification technology in various fields has not been reflected. The current research in the field of text classification is mainly based on theoretical research. Therefore, more research institutes and enterprises and institutions need to invest more in the field of text classification. The research energy promotes the conversion rate and application rate of research results in the field of text classification.

4.3. Example Results and Analysis. The impact of the word vector dimension of news text on the accuracy of news text classification is shown in Figure 7. This article sets the word vector dimension of news text to 50, 100, 150, 200, 250, and 300, respectively. It can be seen from Figure 7 that with the increase in the dimension of the word vector of the news text at the beginning, the accuracy of the classification is also improving. When the dimension of the word vector reaches 150, the accuracy of the news text classification reaches the local maximum, indicating that the semantic distribution of each word in the current news text in the low-latitude space can be more accurately represented. As the dimensionality of the word vector continues to increase, the accuracy of news text classification has decreased. This may be because the increase in the dimensionality of the word vector increases the parameters of the network structure, increases the complexity of the network structure, and increases the amount of calculation. In TextCNN, a 300-dimensional embedding dimension is used, each convolutional layer uses 256 convolution kernels, the number of hidden nodes in the linear layer is 256, the learning rate is 0.0005, and the training is of 6 rounds. The 300-dimensional embedding dimension is used in TextRCNN, the number of built-in hidden nodes in LSTM is 256, the number of hidden nodes in the linear layer is 256, the learning rate is 0.0007, and the training is 5 rounds. To a large extent, the training time of the network is prolonged, resulting in a decrease in accuracy.

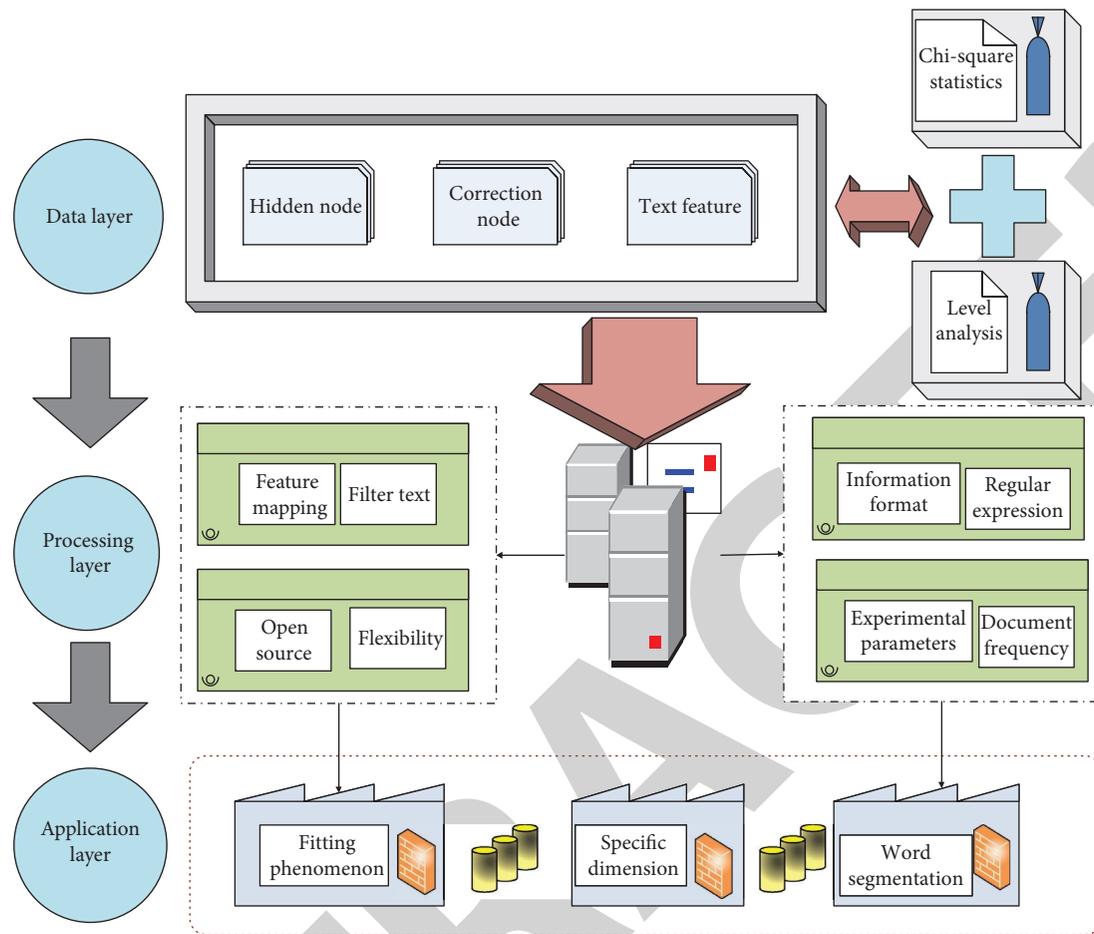


FIGURE 5: News text classification model framework based on hybrid deep learning.

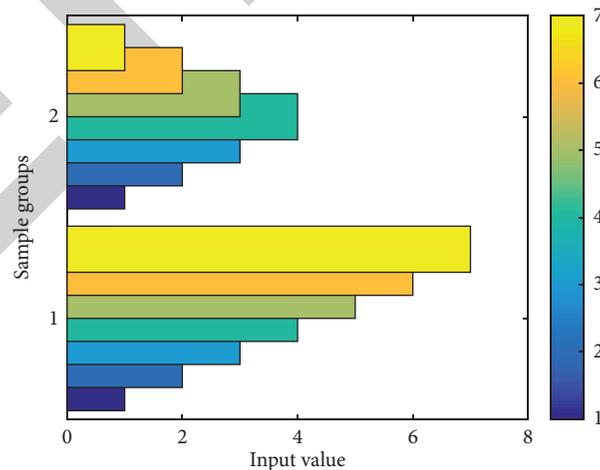


FIGURE 6: Histogram of input values for different sample groups.

Even when the word vector dimension reaches 300, the accuracy rate has been greatly improved, the training time at this time is nearly three times longer than that of the word vector of 150, so 150 is selected as the optimal word vector dimension of the news text. After the model is trained, the accuracy of the algorithm is improved.

It can be seen from Figure 8 that the number of related documents published in the field of text classification in the past ten years has shown a law of first decline and then rise. It is mainly divided into three stages: the first stage is a period of rapid growth. In this stage, the number of articles published in the field of text classification has increased year by

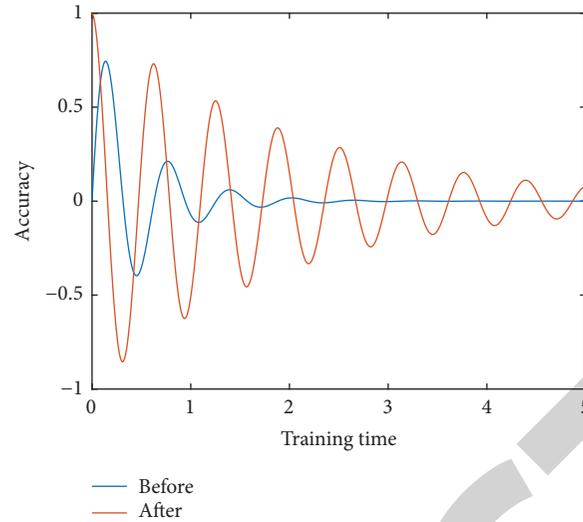


FIGURE 7: Comparison of training accuracy curves before and after model simulation.

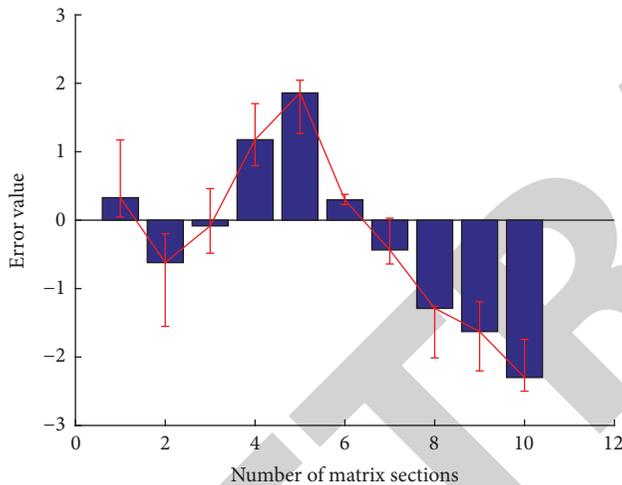


FIGURE 8: The change trend of the algorithm matrix error value.

year, reaching the highest value of 268 in 2019. After adjusting the model parameters through the training set and validation set of experimental data to obtain the optimal model, use the test set that has never participated in the training to test the classification performance of the optimal model.

The average precision and recall rate of news text classification methods based on deep learning is about 97%, and the average precision and average recall rates of traditional news text classification methods are about 84%. Compared with the traditional news text classification method, the classification method has a large improvement in precision and recall. The F1 values of sports, finance, real estate, education, technology, games, and entertainment have increased greatly, while the F1 values of furniture, fashion, and current affairs have increased relatively little. The news text classification method based on deep learning is compared with traditional news text. The classification method improves the average F1 value by about 13%.

It can be seen from Figure 9 that among the top 16 core journals in the field of text classification, 11 journals belong to the computer field and occupy a dominant position. After adjusting the model parameters through the training set and the validation set of the experimental data, the accuracy of the training set and the validation set is obtained. The 300-dimensional embedding dimension is used in the Attention model. The number of built-in hidden nodes of LSTM is 256, the dimension of attention vector is 256, the number of hidden nodes in the linear layer is 256, the learning rate is 0.0003, and the training is 8 rounds. After the accuracy of the training set is stable, it is basically above 96%, and the accuracy of the validation set is stable at about 95%, as shown in Figure 10. The total sample set contains 5477 news texts of emergencies and 2815 general news texts. 6699 news texts are randomly selected as the training set and 1593 as the test set, and the training set is input into the CNN model. The main parameters of the CNN model include text length, word vector dimension, number of convolution kernels, and number of neurons in the fully connected layer.

The effect of text classification is very good on the training set and validation set. It can be seen from the data on the amount of papers in core journals that computer science journals are the mainstream journals published in the field of text classification. It is worth noting that 5 of the top 16 journals with publication volume belong to the field of library and information, such as data analysis and knowledge discovery, information science, library and information work, information magazines, and information theory and practice. The number of publications is 63. Chapters 38, 37, 31, and 27 show that the field of library and information is the second source of text classification research, and it plays a prominent role in promoting the theoretical development and practical application of the text classification field.

Therefore, if relevant scholars need to conduct research on text classification, they can focus on selecting the relevant documents of text classification published in these high-

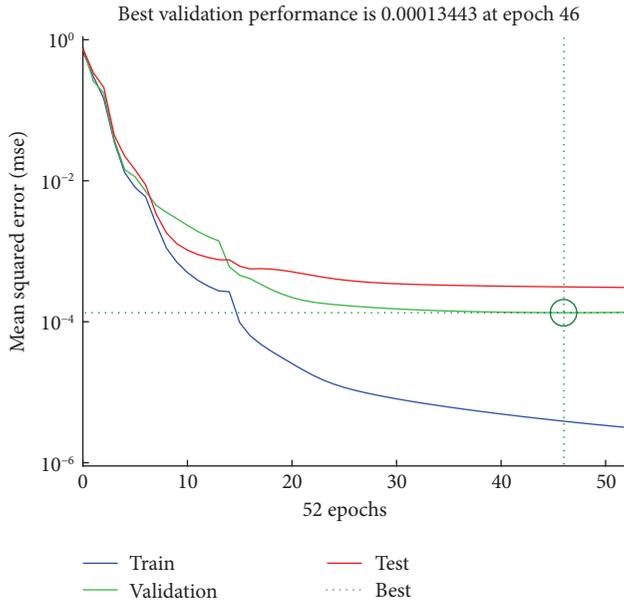


FIGURE 9: Mean square error curve of the hybrid deep learning training test.

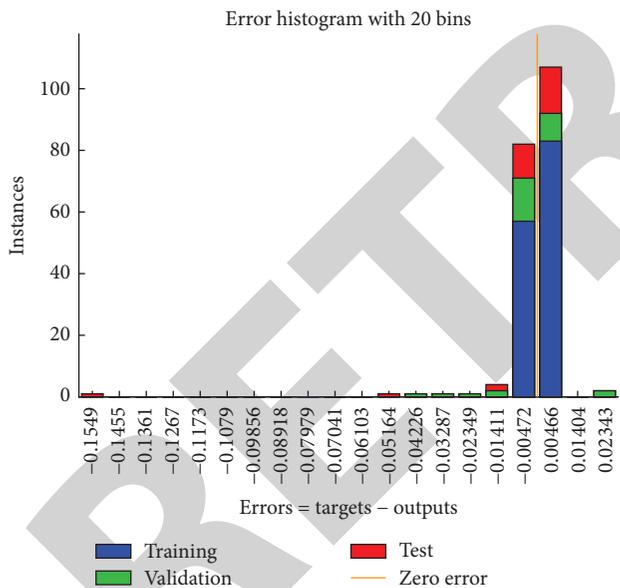


FIGURE 10: Error histogram of the news text simulation example.

volume journals, which will help researchers quickly and accurately obtain the current research status and the latest in this field research dynamics. Then change the single model parameter and repeat the simulation 3 times to get the average of the accuracy, recall, and comprehensive value of the model. The parameter with the largest average value is the best parameter. It can be seen that the accuracy, recall, and comprehensive value of the CNN model reached 99.55%, which can effectively identify common event news and emergency news and lay the foundation for the next step of text classification of emergency news.

5. Conclusion

The model is mainly composed of four parts, namely, news text preprocessing, word vector-based news text representation, news text feature extraction and classification, and text classification result evaluation. In the experiment process, the dense Word2Vec word vector representation of the news text was first obtained through learning, so as to solve the problems of high latitude, sparse data, and lack of semantics of traditional text representation. Then, the news text word vector was used as input, and the convolutional neural network news text features are automatically learned and extracted, so as to avoid the time-consuming, laborious, and error accumulation shortcomings of traditional news text classification methods for manually participating in feature extraction. Experimental results show that this method can effectively improve the efficiency of network news text classification, using word vectors. To solve the problems of high latitude, data sparseness, and lack of semantics in traditional news text representation, the convolutional neural network of the deep learning model automatically extracts news text features and classifies them, which can avoid the problems of manual feature extraction in traditional methods. It analyzes with the traditional news text classification method. Select recall rate, precision rate, and F1 value as evaluation indicators. Finally, we propose a hybrid model, which puts the feature construction modules of the above three models into the same model and uses a shared embedding layer and linear layer, which integrates the advantages of the three models while reducing computational overhead. The experimental results show that the network news text classification method based on deep learning has better comprehensive performance than the traditional news text classification method and promotes the realization of more effective information organization and management in the news field. The network news text classification model based on deep learning proposed in this paper is feasible, can provide users with better news information services, and provides a certain reference value for the development of network news text classification technology.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Ruchansky, S. Seo, and Y. Liu, "CSI: a hybrid deep model for fake news detection," *Information and Knowledge Management*, vol. 31, pp. 797–806, 2019.
- [2] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: a hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021.

- [3] Y. Li, X. Wang, and P. Xu, "Chinese text classification model based on deep learning," *Future Internet*, vol. 10, no. 11, p. 113, 2018.
- [4] X. She and D. Zhang, "Text classification based on hybrid CNN-LSTM hybrid model," *Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, pp. 185–189, 2018.
- [5] J. Zheng and L. Zheng, "A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification," *IEEE Access*, vol. 7, pp. 106673–106685, 2019.
- [6] M. N. Asim, M. U. G. Khan, M. I. Malik et al., "A robust hybrid approach for textual document classification," in *Proceedings of the Document Analysis and Recognition (ICDAR)*, pp. 1390–1396, Sydney, NSW, Australia, September 2019.
- [7] N. Bansal, A. Sharma, and R. K. Singh, "An evolving hybrid deep learning framework for legal document classification," *Ingénierie des Systèmes d'Information*, vol. 24, no. 4, pp. 221–231, 2019.
- [8] A. Abdi, S. Hasan, S. M. Shamsuddin, N. Idris, and J. Piran, "A hybrid deep learning architecture for opinion-oriented multi-document summarization based on multi-feature fusion," *Knowledge-Based Systems*, vol. 213, p. 106658, 2021.
- [9] L. Wu, Y. Rao, H. Yu, Y. Wang, and A. Nazir, "False information detection on social media via a hybrid deep model," in *Lecture Notes in Computer Science. Social Informatics*, vol. 4, pp. 323–333, Springer, Berlin, Germany, 2018.
- [10] H. S. Shon, E. Batbaatar, K. O. Kim, E. J. Cha, and K.-A. Kim, "Classification of kidney cancer data using cost-sensitive hybrid deep learning approach," *Symmetry*, vol. 12, no. 1, p. 154, 2020.
- [11] S. Girgis, E. Amer, and M. Gadallah, "Deep learning algorithms for detecting fake news in online text," *Computer Engineering and Systems (ICCES)*, vol. 3, pp. 93–97, 2018.
- [12] K. Pasupa and T. S. N. Ayutthaya, "Hybrid deep learning models for Thai sentiment analysis," *Cognitive Computation*, vol. 1, pp. 1–27, 2021.
- [13] E. Buabin, "Boosted hybrid recurrent neural classifier for text document classification on the Reuters news text corpus," *International Journal of Machine Learning and Computing*, vol. 2, no. 5, p. 588, 2019.
- [14] A. Elnagar, R. Al-Debsi, and O. Einea, "Arabic text classification using deep learning models," *Information Processing & Management*, vol. 57, no. 1, p. 102121, 2020.
- [15] Y. Liu, J. Ma, Y. Tao et al., "Text classification based on hybrid neural network," *Computational Science and Engineering (CSE)*, vol. 4, pp. 24–29, 2020.
- [16] L. Guo, D. Zhang, L. Wang, H. Wang, and B. Cui, "CRAN: a hybrid CNN-RNN attention-based model for text classification," *Conceptual Modeling*, vol. 5, pp. 571–585, 2018.
- [17] L. J. Vaickus, A. A. Suriawinata, J. W. Wei, and X. Liu, "Automating the Paris System for urine cytopathology-A hybrid deep-learning and morphometric approach," *Cancer Cytopathology*, vol. 127, no. 2, pp. 98–115, 2019.
- [18] Y. Hu, J. Ni, and L. Wen, "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction," *Physica A: Statistical Mechanics and Its Applications*, vol. 557, p. 124907, 2020.
- [19] Z. H. Kilimci and S. Akyokus, "Deep learning-and word embedding-based heterogeneous classifier ensembles for text classification," *Complexity*, vol. 2, pp. 14–18, 2018.
- [20] F. Yin, X. Xue, C. Zhang et al., "Multifidelity genetic transfer: an efficient framework for production optimization," *SPE Journal*, pp. 1–22, 2021.
- [21] J. Zhang and G. Qu, "Physical unclonable function-based key sharing via machine learning for IoT security," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 7025–7033, 2019.
- [22] J. Yang, J. Wen, B. Jiang, and H. Wang, "Blockchain-based sharing and tamper-proof framework of big data networking," *IEEE Network*, vol. 34, no. 4, pp. 62–67, 2020.
- [23] S. Yang, J. Wang, N. Zhang, B. Deng, Y. Pang, and M. R. Azghadi, "CerebellumMorphic: large-scale neuromorphic model and architecture for supervised motor learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2021.
- [24] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, and M. Woźniak, "Accurate and fast URL phishing detector: a convolutional neural network approach," *Computer Networks*, vol. 178, p. 107275, 2020.
- [25] W. Wang, X. Zhao, Z. Gong, Z. Chen, N. Zhang, and W. Wei, "An attention-based deep learning framework for trip destination prediction of sharing bike," in *Proceedings of the IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, Macau, China, September 2020.
- [26] M. Chen, S. Lu, and Q. Liu, "Uniqueness of weak solutions to a Keller-Segel-Navier-Stokes system," *Applied Mathematics Letters*, vol. 121, p. 107417, 2021.
- [27] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.
- [28] Y. Narayan and S. Kumar, "Pattern recognition of sEMG signals using DWT based feature and SVM Classifier," *International Journal of Advanced Science and Technology*, vol. 29, no. 10s, pp. 2257–2266, 2020.