

Research Article

Digital Computer Modulation Signal Classification Based on Neural Network

Guisheng Wang 

Tongling University, Tongling, 244000 Anhui, China

Correspondence should be addressed to Guisheng Wang; wgs1869@tlu.edu.cn

Received 10 December 2021; Revised 25 December 2021; Accepted 10 February 2022; Published 27 February 2022

Academic Editor: Deepak Kumar Jain

Copyright © 2022 Guisheng 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.

Computer is one of the indispensable tools in the human world, and human needs for it are increasing, so the emergence and application of more advanced computers are needed. The current computers do not respond intelligently, and it is difficult to meet people's needs for information processing in the era of big data. In order to solve these problems, this paper proposes the application of a neural network-based data classification algorithm in computers, aiming to study the practical application of the algorithm in computers. The research method of this paper is to introduce the BP neural network, select the appropriate method of classification features, and then study the data classification algorithm. The function of the research method is to compare the classification error and convergence speed of the BP network composed of different hidden layer nodes, to study whether a certain feature item of the data exists and the difference in the amount of information classification of the entire document, and to select high efficiency, accuracy, and scalability algorithm. This paper compares the forward reasoning time of the model before and after cutting through experiments based on neural network model design, algorithm design, and man-machine dialogue model design. The results show that, in terms of computing speed, the adaptive model compression method based on the accuracy and redundancy ratio compresses the model after the forward reasoning time is greatly reduced, and the reasoning time becomes 35% of the original, and in terms of calculation accuracy, the absolute error after using the SOM method in this article has not reached 0.5.

1. Introduction

1.1. Background. In the long-term natural selection and biological evolution process, the human brain has formed a powerful thinking ability and excellent intelligent perception. Computer technology has been developed for decades, and from the early integration of transistors to the era of rapid development of Internet technology, computer technology includes more and more definitions, including chip-level technology, system-level technology, and software technology. In recent years, jargon such as neural networks and classification algorithms has been flooded with communication methods and entire personal lives. The basic principle of neural network is to imitate the human brain neuron network, trying to have the same intelligent learning ability and logical analysis ability as human; however, the human brain has billions of neurons, which cannot be simulated by current computers. However, this article can start from

the structure of the human brain and learning principles and try to simulate a network that can think like a brain as much as possible. And with the improvement of computer performance, many problems that are currently regarded as bottlenecks may be solved. Combining theoretical and experimental result analysis, this paper finds that the traditional neural network algorithm is expected to break through its computing bottleneck in the future Internet of Things system and usher in rapid development.

1.2. Significance. In order to adapt to a more complex computer application environment, people need to build models with strong learning capabilities. Due to its unique two-dimensional data processing method, convolutional neural networks are particularly suitable for data classification and processing. Therefore, the research theme of this study not only has important academic significance but also has the importance of practical leadership in the engineering field.

The automatic recognition and classification of data is of great significance for both military and civilian scenarios, and the modulation recognition task is simply a method of determining the modulation pattern for some characteristics of the data signal after receiving the data signal. The modulation recognition has an irreplaceable position in both military and civilian fields, when the modulation style of the data signal is correctly determined, it can help managers better manage the increasingly scarce spectrum resources, better identify illegal intrusion signals, and strengthen the utilization and management of spectrum, so the task of data recognition and classification has strong practical significance.

1.3. Related Work. Neural networks and data classification algorithms have always been hot topics in the computer field, and comments around them are endless. Dong and Wang studied the robust index problem based on output feedback control [1]. Secor et al. believe that artificial neural networks are used to solve the Schrödinger equations of the single-well and double-well potentials of hydrogen-bonded molecular systems for proton transfer. The artificial neural network mapping is trained to predict the lowest five proton vibration energy, wave function, and density according to the proton trend and predict the excited state proton vibration energy and density according to the proton ground state density [2]. Chan and Elsheikh proposed that neural network technology is increasingly being considered in the field of geology, characterized by high-dimensional spatial data dominated by multipoint statistics. Because the current method does not provide parameterization of the condition generation process, he proposed a method to obtain the parameterization of the direct generation condition realization [3]. Takano et al. explored the use of classification algorithms to test the modification of specific sentences by developing a computer classifier that distinguishes specific and nonspecific memories. He extracted a large number of grammatical features, such as past tense, negation, adverbs, and phrases related to time and place [4]. Lemke et al. believe that the analysis currently used in the health system requires an access classification algorithm based on management data. His goal is to show expanded and revised versions of existing algorithms and use this tool to characterize ED usage patterns in a sample of American hospitals and a large number of health plan registrants [5]. Aminu and Ahmad proposed the least square discriminant analysis method, a well-known chemometric feature extraction and information classification technology. The accuracy of feature learning and extraction depends on the degree of capture of the discriminative subspace, and the comparison performance is measured based on the distinction and classification of these data sets [6]. Kundu and Ari proposed an interface system for character recognition based on the P300 signal, and the autoencoder proposed an ensemble of weighted artificial neural networks for P300 detection to minimize the difference between different classifiers. In order to provide more data support for better classifiers, the algorithm assigns higher weights to better-performing classifiers [7]. Aiming at the problem of real-time arrhythmia classification, Cai et al. propose a low-latency, high-practical, and high-reliability deep learning

real-time arrhythmia classification algorithm, which can be easily applied to the real-time arrhythmia classification system. His research results show that the algorithm is very accurate in predicting patient location [8].

1.4. Innovation. This experiment studies and summarizes the latest technologies and theories in the field of brain-inspired computing, effectively absorbs previous research work, and compares and analyzes the latest brain-inspired computing architecture and brain-inspired computing chips. This process needs to solve the key technical points required to operate computer research like a brain, such as data packet multicast transmission, efficient interconnection between boards, and neuron simulation. This paper studies the computer network fault diagnosis, using SOM method and LM method to simulate the computer network fault. The SOM neural network and the BP neural network are effectively combined by adding weights, and the parallel algorithm is used to make certain improvements to the LM algorithm, and these examples show that this research has a certain significance through case diagnosis.

2. Neural Network and Classification Algorithm

2.1. BP Neural Network. BP refers to the most widely used training algorithm in various feedforward neural networks is the error back propagation algorithm [9]. In this paper, the feedforward neural network trained using the BP algorithm is usually called the BP neural network. Therefore, the BP neural networks mentioned in the following chapters of this article are all expressed as feedforward neural networks using the BP algorithm. In many learning problems of multilayer feedforward neural networks that have been plagued by scholars, the BP algorithm has a good performance [10]. The learning process of the BP algorithm can generally be divided into two stages: the forward propagation stage of the training input data and the back propagation stage of the error signal, as shown in Figure 1.

After understanding the basic principles of the BP algorithm, this section specifically introduces the derivation of the algorithm in the three-layer feedforward neural network model [11]. The energy function used for derivation is the formula for calculating the mean square error, as shown in Formula (1).

$$W(\bar{x}) = 1/2 \sum_{k=1}^k [y_k(\bar{x}) - t_k(\bar{x})]^2. \quad (1)$$

In the standard BP learning algorithm, during the forward propagation stage of training data, the input information of each layer of neuron in the network is directly taken from the previous neuron, and the calculated output is directly sent to the latter neuron. If the network output error does not meet the standard, enter the back propagation stage of the error signal, and use Equations (2) and (3) to adjust the threshold and weight of each neuron node in the neural network [12]. Among them, the partial derivative of each connection weight of the neural network and the

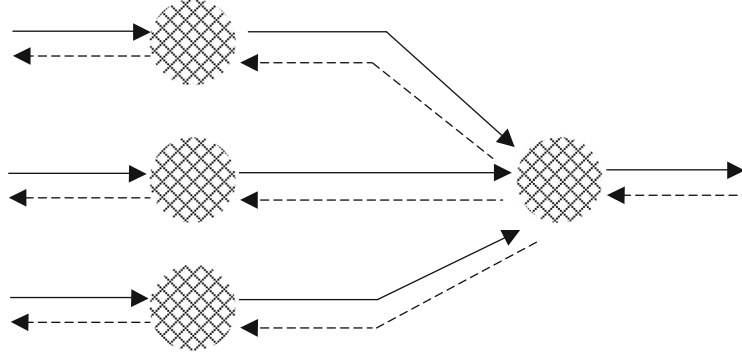


FIGURE 1: Signaling of the algorithm.

energy function of the neuron threshold are calculated by Equations (4) and (5).

$$Q_{jk} = Q_{jk} + n \frac{\partial W(\bar{x})}{\partial Q_{jk}}, \quad (2)$$

$$\theta_j = \theta_j + n \frac{\partial W(\bar{x})}{\partial \theta_j}, \quad (3)$$

$$\frac{\partial W(\bar{x})}{\partial Q_{jk}} = \frac{\partial W(\bar{x})}{\partial y_k(\bar{x})} \frac{\partial y_k(\bar{x})}{\partial net_k(\bar{x})} \frac{\partial net_k(\bar{x})}{\partial Q_{jk}}, \quad (4)$$

$$\frac{\partial W(\bar{x})}{\partial Q_{jk}} = \sum_{k=1}^k \frac{\partial W(\bar{x})}{\partial y_k(\bar{x})} \frac{\partial y_k(\bar{x})}{\partial net_k(\bar{x})} \dots \frac{\partial y_j(\bar{x})}{\partial net_j(\bar{x})} \frac{\partial net_j(\bar{x})}{\partial Q_{jk}}. \quad (5)$$

Under normal circumstances, the conventional BP algorithm always approaches the minimum point from one direction and has no ability to jump out, so the error value does not change much as the training continues, as shown in Figure 2:

According to the comparison of the classification error and convergence speed of the BP network composed of different hidden layer nodes, the BP network topology used in the classification experiment of the data set in the computer network is finally determined: 12 input layer nodes, 36 hidden layer nodes, and 12 output layer nodes [13]. Carrying on the final classification experiment with the selected network structure, the result is shown as in Figure 3.

Figure 3 shows the error of the classification result, which is a direct error, that is, the error obtained by directly subtracting the actual result from the root of the classification result [14]. It can be seen that under the best BP neural network determined by the experiment, the results of this article are still very ideal, and the mean square error MSE is about 0.015.

Convolution operations have an important position in neural networks, such as continuous convolution operations and discrete convolution operations [15]. Next, in this article, this article will introduce the formula of convolution operation in neural network.

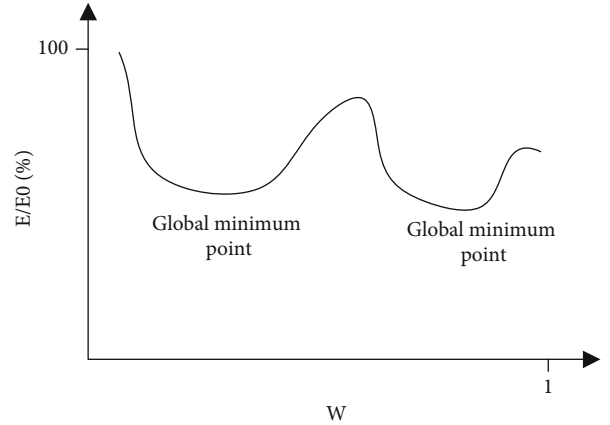


FIGURE 2: Local minimum and global minimum of error.

Continuous convolution operation formula:

$$W(n) = \int_{-\infty}^{\infty} x(p)h(n-p)dp = x(n) * h(n). \quad (6)$$

Discrete convolution operation formula:

$$W(n) = \sum_{-\infty}^{\infty} x(i)h(n-i)dp = x(n) * h(n). \quad (7)$$

The convolution operation of convolutional neural network belongs to discrete convolution, but there are some differences in the definition of discrete convolution in analytic mathematics. Convolution operation is actually a linear operation, not a real convolution operation. The corresponding convolution kernel can also be called a filter [16], the size of the convolution kernel determines the size of the subregions involved in the operation in the image, and the size of the convolution kernel parameters determines the increase in the voting power of the pixels corresponding to the image area covered by the final convolution kernel, and the higher the weight, the greater the result of the convolution.

2.2. Selection Method of Classification Features. The information acquisition method (IG for short) selects features based on the amount of information provided by the feature

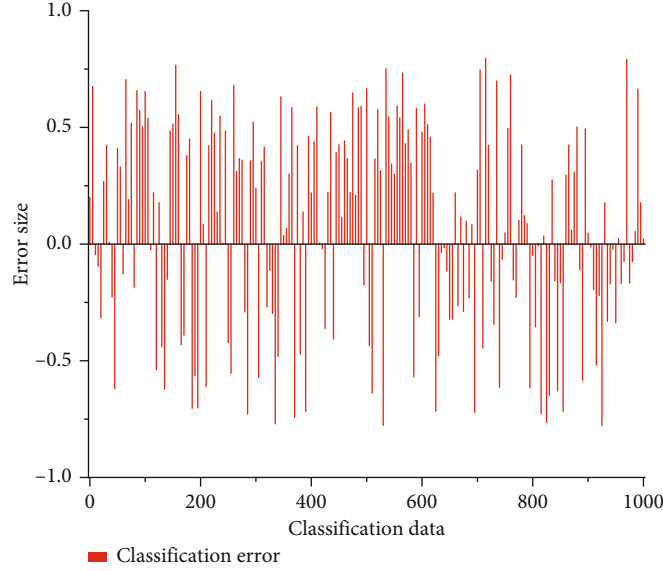


FIGURE 3: Error statistics of classification results.

items for the entire document classification and the amount of information in the entire document classification with or without feature items. In this process, the difference in classification information is called the gain of feature items, and entropy determines the amount of information [17]. The information gain is given in Equation (8), which is the difference between the entropy of the document when there are no features and the entropy of the document when there are features.

$$IG(n) = - \sum_{i=1}^m Y(Ci) \log(Ci) + Y(n) \sum_{i=1}^m Y(Ci | n) \log(Ci | n) + Y(\bar{n}) \sum_{i=1}^m Y(Ci | \bar{n}) \log(Ci | \bar{n}). \quad (8)$$

Among them, $Y(Ci)$ represents the probability of appearing in document, Ci , $Y(n)$ represents the probability of the document including the feature item t , $Y(Ci | n)$ represents the conditional probability of the document including the feature item t , $Y(\bar{n})$ represents the probability of a document that does not include the feature item t , $Y(Ci | \bar{n})$ represents the conditional probability of belonging to Ci when the document does not include the feature item t , and m represents the number of categories. From the above definition and information gain equation, it can be seen that the information gain of a feature is the best feature selection method in theory. But in actual engineering, most of the features with high information gain are usually very low values. This is because it shows that there are problems with frequency and data sparseness, leading to poor classification results [18].

The TC statistic (X2) detects the degree of correlation between the feature item t and the category c , assuming that t obeys the X2 distribution and the first degree of freedom

[19]. In a specific category, the higher the TC statistics of a feature, the better, and the greater the correlation between the feature and the category, the more useful information the feature has for classification. The definition of TC statistics is shown in Formula (9):

$$TC(n, Ci) = \frac{N(AD - BC)^2}{(A + C)(B + C)(A + B)(D + C)}, \quad (9)$$

where N represents the total number of documents, A represents the number of documents that belong to class Ci and contains t , B represents the number of documents that does not belong to class Ci and contains t , and C represents the number of documents that belongs to class Ci and does not contain t , D represents the number of documents that does not belong to class Ci and do not contain t . For the problem of data classification, the function of TC statistics provides two solutions [20].

One is to calculate the TC value of t for each category separately and calculate Equation (10) in the training set.

$$CHI_{\max}(n) = \max_{1 \leq i \leq m} \{CHI(n, Ci)\}. \quad (10)$$

If m represents the total number of categories, it can be known that the method gives a threshold. Then, delete the functions below the threshold and select the functions above the threshold [21].

Second, using Formula (11) to calculate the average value of each category and function.

$$CHI_{\text{avg}}(n) = \sum_{i=1}^m Y(Ci) CHI(n, Ci). \quad (11)$$

This method takes the average value as the TC value of each category.

The third point is to find the amount of mutual information (MI for short), and the basic idea is that the higher the frequency of the function t and the category c , the greater the amount of mutual information. Because A, B, C, and N are the same as the definitions in the previous section, the performance of mutual information is shown in Equation (12).

$$MI(n, Ci) = \log \frac{Y(n, Ci)}{Y(n)Y(Ci)} \approx \log \frac{AN}{(A+C)(A+B)}. \quad (12)$$

If feature t and category Ci are independent of each other, then $Y(n, Ci) = Y(n)Y(Ci)$, $MI(n, c) = 0$. Like the X2 statistics, there are many ways to find the maximum and average values of multiple classification problems in the text [22].

2.3. Data Classification Algorithm. In the entire data classification process, the key part is to select the correct data classification algorithm [23]. In general, the choice of data classification algorithm in this paper should consider the robustness, efficiency, accuracy, scalability, and comprehensibility of the algorithm for analysis. A large number of literature studies have shown that the following classification algorithms are more effective.

Bayesian classification is a data classification method based on Bayes' theorem [24]. This article can use Bayesian classification method to estimate the probability of a specific sample type. In other words, this article can predict the probability that a sample belongs to a specific class member.

The basic idea of Bayesian classification is as follows. For a sample group X with m possible classifications (marked as $C1, Ci-1, Ci, Ci+1$), any sample individual vector X has n attributes $X = \{X1, X2, \dots, Xn\}$ and belongs to the marked class only when inequality (13) is established.

$$Y(Ci | \bar{x}) > Y(Cj | \bar{x}), 1 \leq j \leq m, j \neq i. \quad (13)$$

Among them, $Y(Ci | \bar{x})$ and $Y(Cj | \bar{x})$ are calculated by Formula (14).

$$Y(Ci | \bar{x}) = \frac{Y(\bar{x} | Ci) \times Y(Ci)}{Y(\bar{x})}. \quad (14)$$

The key to solving Equation (2) is that the calculation of $Y(\bar{x} | Ci)$ can adopt the naive assumption that the class labeling conditions are independent of each other and use Equation (15) for calculation.

$$Y(\bar{x} | Ci) = \prod_{k=1}^n Y(x_k | Ci). \quad (15)$$

The BPNN algorithm is currently the most widely used neural network and has successfully solved practical problems in many fields [25]. The sample set can be defined as Formula (16).

$$D = \{(Xi, \omega_j) | Xi \in X, \omega_j \in \Omega, 1 \leq i \leq n, 1 \leq j \leq c\}. \quad (16)$$

Among them $Xi = [x1, x2, \dots, xd]^T$, d is the number of attributes of the sample set. The set of categories the sample belongs to is set to $\Omega = \{\omega1, \omega2, \dots, \omega_j, \dots, \omega_c\}$, then Ω is a set of label vectors with c categories, of which $\omega_j = [y1, y2, \dots, yk]^T$.

In Bayesian decision-making based on the minimum error rate, it is assumed that each attribute in the naive Bayes algorithm is independent of each other and the class conditional probability of continuous attributes obeys Gaussian distribution. When the prior probability and the class conditional probability density are known, the Bayes formula is used to compare the posterior probabilities of the samples belonging to each category [26]. Under the condition that the total error rate is minimized, the category to which any unknown pattern in the test sample belongs is determined as the category with the largest posterior probability. The Bayesian decision algorithm based on the minimum error rate used in this paper can be simply described as the definition of sample set D that is shown in Formula (16). Supposing any unknown pattern in the test set is x , N is the total number of training samples, and Ni is the number of samples belonging to ω_i categories in the training samples, then there is a prior probability of $Y(\omega_i) \approx Ni/N$, ($i = 1, 2, \dots, c$) for each category. Assuming that the class conditional probability $Y(x/\omega_i)$ obeys the Gaussian distribution, in the d -dimensional feature space, the multivariate Gaussian probability density function is defined as shown in Equation (17):

$$Y\left(\frac{x}{\omega_i}\right) = \frac{1}{(2\pi)^{d/2} |\sum_i x|^{1/2}} \exp\left(-1/2(x - u_i)^T \sum_i^{-1} (x - u_i)\right), \quad (17)$$

where $u_i = Q(x)$ is the mean value of the class, ω_i is the covariance matrix of x times x , \sum_i is the value of the determinant of $|\sum_i|$, and \sum_i^{-1} is the inverse matrix of \sum_i^{-1} . Using the Bayesian formula to obtain the posterior probability $Y(x/\omega_i)$ of the unknown mode x , the calculation formula is shown in Formula (18):

$$Y(x/\omega_i) = Y(\omega_i/x) Y(\omega_i) / \sum_{j=1}^c Y\left(\frac{x}{\omega_j}\right) Y(\omega_j). \quad (18)$$

Suppose the calculation formula of discriminant function $g_i(x)$ corresponding to category ω_i is shown in Formula (19):

$$g_i(x) = Y\left(\frac{\omega_i}{x}\right), i = 1, 2 \dots c. \quad (19)$$

Secondly, the category corresponding to the discriminant function with the highest verification probability is the category to which the unknown mode x belongs [27]. Because the denominator $\sum_{j=1}^c Y(x/\omega_j) Y(\omega_j)$ in Equation (18) is the same for all classes of $Y(x/\omega_i)$, it can be

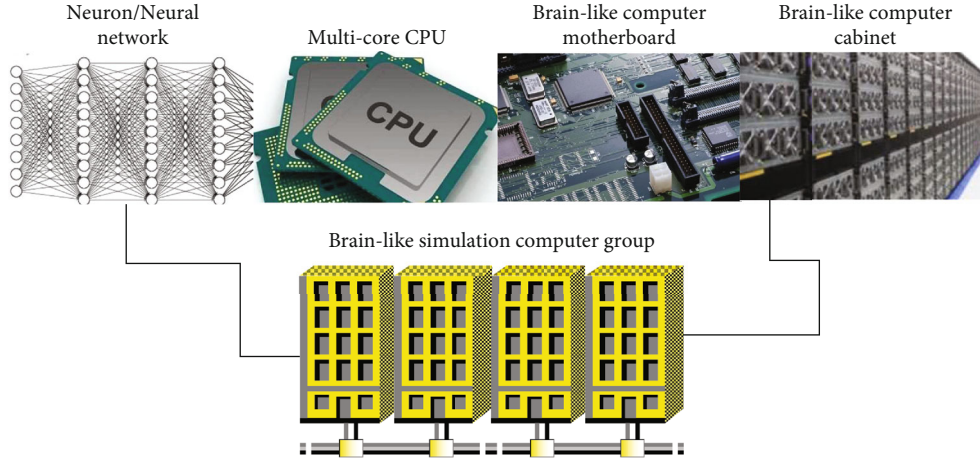


FIGURE 4: Schematic diagram of the overall system architecture.

omitted, and then the discriminant function equation (19) can be rewritten as Equation (20):

$$g_i(x) = Y\left(\frac{x}{\omega_i}\right)Y(\omega_i), i = 1, 2 \dots c. \quad (20)$$

Then, the policy rule for the category of unknown pattern x in the test sample set is: if $g_k(x) = \max(g_i(x)), i = 1, 2, \dots, c$, then unknown pattern $x \in \omega_k$.

3. Computer Application Experiment of Data Classification Algorithm

3.1. Model Design Based on Neural Network. At present, research on large computers such as the brain is in full swing, but it is difficult to achieve. In this paper, a neural network simulation program is proposed, and a CAM-based multicast transmission mechanism of large-pulse neuron data packets is designed, which successfully solves the board-level interconnection problem of large-scale brain-like computers [28].

The large-scale brain-like simulation computer platform designed in this experiment connects a large number of general-purpose processors through an efficient custom network, providing a wide range of communication possibilities to manipulate the neural network model used [29] Each entity has a small and independent state, such as a neural network, and the overall architecture of the system is shown in Figure 4.

The large-scale brain-like simulation computer platform developed in this paper has high-efficiency large-scale parallel processing capabilities, low system power consumption, and good system stability [30]. In the experimental environment, a computer motherboard is like a brain, which can simulate more than one million neurons, which provides a feasible solution for the development of brain-like computers.

In addition, neural networks can also be used to detect computer failures and analyze computer network interface failures. The reasons for the failure of the interface are generally divided into the following four types: B1 controls the security of the interface, B2 controls the stability of the net-

TABLE 1: Input sample.

| B1 | B2 | B3 | B4 | B5 |
|-----|--------|--------|--------|--------|
| 0.7 | 0.1160 | 0.3480 | 0.1160 | 0.0387 |
| 0.4 | 0.1940 | 0.5820 | 0.1164 | 0.0146 |
| 0.5 | 0.8000 | 0.8000 | 0.1333 | 0.0190 |
| 0.2 | 0.6250 | 4.3750 | 0.8750 | 0.8750 |
| 0.1 | 0.6230 | 4.3610 | 1.0903 | 0.2181 |
| 0.8 | 0.7250 | 2.9000 | 0.4833 | 0.0604 |
| 0.2 | 0.2290 | 1.3740 | 0.4580 | 0.1145 |
| 0.8 | 0.1200 | 0.8400 | 0.8400 | 0.4200 |
| 0.2 | 0.3830 | 0.3830 | 0.0958 | 0.0479 |
| 0.2 | 0.4400 | 0.8800 | 0.4400 | 0.0733 |

work, B3 equipment is congested, and the B4 communication protocol is incompatible, and they are used as the output nodes of the SOM neural network. MIB-2 is subordinate to the 5 status signs of the 2 interface, The sample input for this experiment is shown in Table 1.

Using the training function in the MATLAB neural network toolbox to train the SOM neural network of this article, it is found that as the number of training steps increases, the distribution of neurons gradually becomes more reasonable. After the network training is completed, the weight will be corrected, and when the experiment inputs each value, the network will automatically cluster the value. Using the combination of the above SOM and LM methods to retrain the sample data and train it, and the training process is shown in Figure 5.

It can be seen from the figure that the combination of the SOM method and the LM method has a faster convergence rate than the general BP neural network.

After using the sample data to construct and train the neural network, the absolute error of the simulation results is shown in Table 2.

Table 2 corresponds to the absolute error of the sample simulation results using the combination of the SOM method and the LM method. It can be seen that the absolute error after using the SOM method does not reach 0.5, but

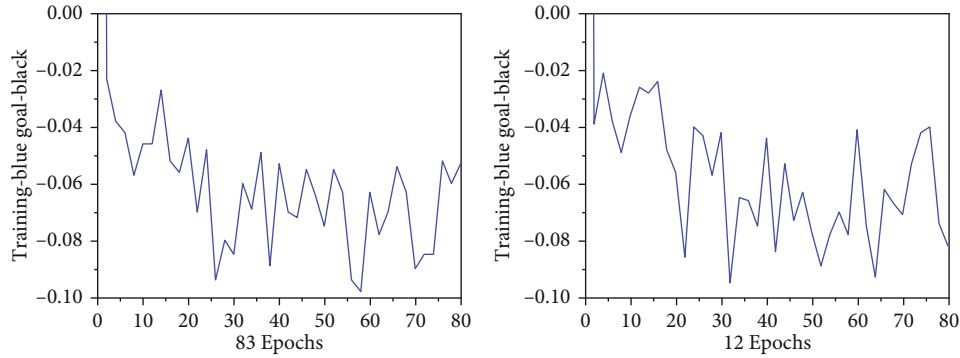


FIGURE 5: Comparison of the training process of the combined algorithm and the original algorithm.

TABLE 2: Absolute error.

| B1 error | B2 error | B3 error | B4 error | B5 error |
|----------|----------|----------|----------|----------|
| -0.0112 | 0.2130 | 0.4129 | 0.0609 | -0.0101 |
| 0.2587 | 0.2433 | 0.3105 | 0.3450 | 0.2573 |
| 0.1349 | 0.0152 | 0.5488 | 0.1421 | 0.1337 |
| 0.1103 | 0.1012 | 0.4885 | 0.1195 | 0.1093 |
| 0.0649 | 0.0941 | 0.4682 | 0.1014 | 0.0648 |

the absolute error of the unused method reaches -0.0112 the smallest. It can be seen that the combination of SOM method and LM method has a better effect on improving the accuracy of sample simulation. Using the combination of SOM method and LM method not only improves the accuracy but also greatly reduces the number of training.

3.2. Algorithm Design Based on Neural Network. Through the introduction of the previous article, this article has a deep understanding of the logical composition, software architecture, and technical route of the system, and this section will use the data classification algorithm to collect and store data.

GoogLeNet has 22 network layers, its model has been iteratively updated in four versions so far, and now it has reached the v4 version. Although the number of network layers is almost 3 times that of AlexNet, the amount of parameters is only 1/12 of AlexNet. In addition to the application of the very innovative inception network structure, GoogLeNet basically cancels the use of the fully connected layer and is more compact and excellent in the design of the model. GoogLeNet shows a very compact and ingenious network structure, which can fully extract and merge features at different scales, and this network model avoids the use of a fully connected layer, which reduces the number of model parameters to a certain extent. The specific network structure of GoogLeNet and the number of parameters are shown in Table 3:

GoogLeNet was originally designed to solve the classification problem of image data sets, so this experiment uses it to classify the modulation styles of the Fourier time-frequency graph data sets. The neural network structure used in this experiment is basically similar to the GoogLeNet

structure, but the auxiliary classifier is removed, and the network structure of the input and output layers is fine-tuned to make it more suitable for the purpose of signal modulation style classification. Figure 6 shows the network structure loss curve of the training set and the recognition accuracy curve of the verification set.

The residual network model is another innovation in the field of neural networks in recent years. Usually, as the network model continues to deepen, the neural network will become more difficult to train, and the loss of the training set does not decrease but rises, resulting in failure to train, which is the degradation of neural networks. Until the emergence of ResNet, the training of large-scale neural network models became possible.

Comparing the experimental results, it can be found that the recognition and classification effects of ResNet 3.0 and ResNet 5.0 on the short-time-frequency map data set are basically the same. From the analysis of the experimental results, this article believes that the 40-layer model basically has the ability to fully fit the function of distinguishing signal types. Compared with ResNet 3.0, the 50-layer ResNet 5.0 network is basically identical to the redundant part of ResNet 3.0, so there is no overfitting, but the result of the classification ability of the two is almost the same. From the experimental results, it can be further verified that the residual network has the adaptive ability. Figure 7 is the variation curve of the classification accuracy of the validation set of ResNet 3.0 and ResNet 5.0:

In this section of the experiment, the AlexNet network model of the classified IQ modulation signal trained in this experiment is tailored, and random sampling is also carried out from 800 samples of 15 modulation modes of the IQ original signal data set. Each modulation method selects 60 samples as the training data set used in the compression process, and divides them according to the ratio of 6:2:2, and uses the improved adaptive model cropping method based on the accuracy and redundancy ratio to crop the model. In the fixed-parameter cropping experiment, the initial cropping channel number is 32, and the performance recovery Epoch is 15 after cropping, and a total of 60 Epochs are cropped. Using cross-entropy loss function, SGD is used as the optimizer for training, learning rate $lr=0.001$, and momentum = 1. The main parameters of the adaptive channel model compression method based on the accuracy rate

TABLE 3: Network structure and the number of parameters.

| Type | Patch size | Output size | Depth | 5x5red uce | Pool proj | Params | Ops |
|----------|-----------------|----------------------------|-------|------------|-----------|--------|------|
| Incep-4a | $7 \times 5/2$ | $112 \times 56 \times 256$ | 0 | | | | |
| Incep-4b | $14 \times 7/6$ | $56 \times 112 \times 56$ | 0 | | | | |
| Incep-4c | $4 \times 6/2$ | $64 \times 56 \times 28$ | 0 | | | | |
| Incep-4d | | $28 \times 56 \times 28$ | 1 | 153 | 85 | 175 k | 54 M |
| Incep-4c | | $112 \times 64 \times 256$ | 0 | 63 | 45 | 163 k | 50 M |
| Incep-5a | | $56 \times 28 \times 64$ | 0 | 179 | 68 | 100 k | 68 M |
| Incep-5b | $3 \times 2/2$ | $64 \times 256 \times 28$ | 1 | | | | |
| Linear | | $28 \times 64 \times 56$ | 1 | 187 | 47 | 108 k | 37 M |
| Softmax | $9 \times 7/2$ | $112 \times 64 \times 256$ | 0 | | | | |

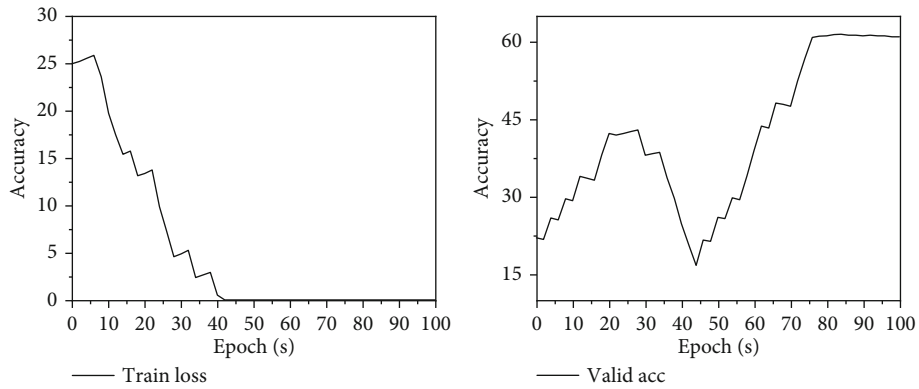


FIGURE 6: Training loss change curve and validation set accuracy change curve after fine-tuning.

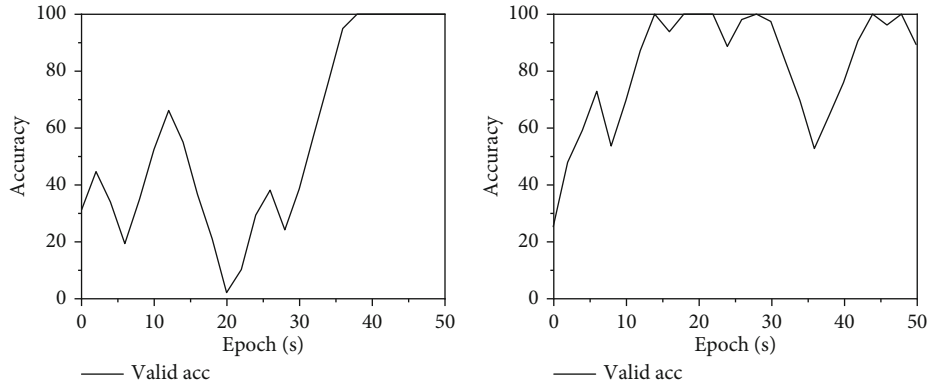


FIGURE 7: Variation curve of classification accuracy of ResNet 3.0 and ResNet 5.0 validation sets.

redundancy ratio are basically the same as those of fixed channel clipping, but the accuracy rate redundancy ratio threshold parameter is 9 and the channel drop rate is set to 1. That is, every time the accuracy rate drops by more than 4%, the number of channel cuts will drop to 86% of the original. The experimental comparison between the accuracy of the two cutting methods and the number of Epoch rounds is shown in Figure 8:

From the comparison chart of the accuracy drop curve, it can be seen that when the Epoch is greater than about 45, the performance of the fixed-parameter model clipping method begins to drop sharply, and this is because the

redundant parameters of the model are almost all cut off. At this time, if the original cutting level is continued, it will seriously affect the performance of the model. The model clipping method based on the accuracy and redundancy ratio can still maintain a high recognition accuracy in the later stage of the clipping, because the adaptive model clipping strategy reduces the clipping strength and continues to perform redundant clipping under the premise of ensuring the accuracy. Therefore, the adaptive model clipping method based on the accuracy rate redundancy ratio has better clipping performance than the fixed parameter clipping method in ensuring the accuracy index.

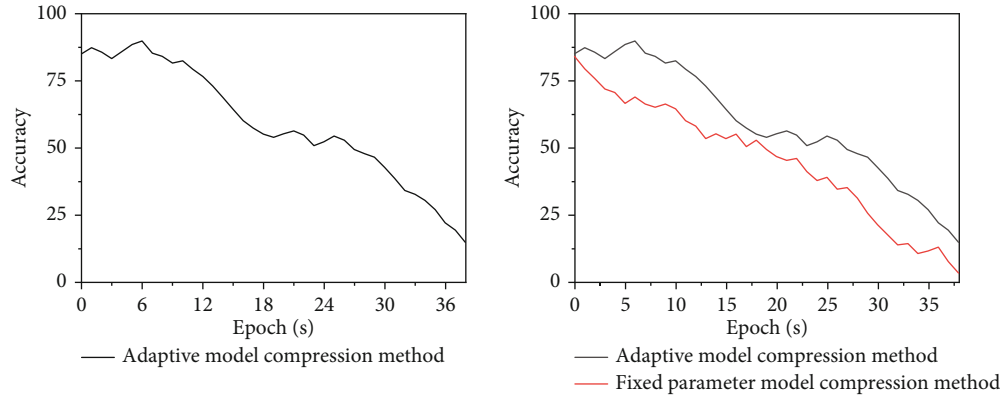


FIGURE 8: Comparison of accuracy drop curve before and after improvement.

TABLE 4: Time-consuming situation of fixed parameter model compression method.

| Platform | Original model time (s) | Compression model time (s) | Reduction ratio |
|------------------|-------------------------|----------------------------|-----------------|
| NVIDIA GTX1060 | 4.19 | 1.4665 | 0.65 |
| NVIDIA RTX2060 | 2.114 | 0.3593 | 0.83 |
| GT730M | 8.601 | 2.5803 | 0.7 |
| I3-8100 | 17.195 | 2.5792 | 0.85 |
| NVIDIA RTX A4000 | 8.676 | 1.5616 | 0.82 |

After completing the model loading with the method described above, forward inference experiments are performed on the above verification data sets, and the comparison results of the forward inference time are shown in Table 4:

According to the calculation results, it can be concluded that the classification accuracy of the uncompressed model is the highest, reaching 85%. But its model reasoning time is also the longest, 2.114 seconds on the RTX2060 platform, 8.676s on the RTX A4000 platform, and 17.195s on the I3-8100 processor platform; however, the forward reasoning time is greatly reduced after the model is compressed by the adaptive model compression method based on the accuracy and redundancy ratio. On the RTX2060 platform, it was reduced to 1.4665s, and the reasoning time became 35% of the original, and on the RTX A4000 platform, the reasoning time was 1.5616s, which was reduced to 18% of the original. It is reduced to 2.5792s on the I3-8100 processor platform, and the reasoning time is directly reduced to 15% of the original. It can be inferred from this that model compression is of great significance on edge computing devices, and the improvement effect is more obvious on devices with weaker computing power. Compressing the model can greatly reduce the inference time within the acceptable accuracy loss range, and this is of great significance for scenarios that require edge deployment and require higher real-time performance.

3.3. Human-Machine Dialogue Model Design Based on Data Classification Algorithm. Making machines understand human instructions is an important task in the computer field, as pointed out by the Turing Test proposed in 1951.

If the interaction with the computer can be considered by people as if they are communicating with a human, then it can be determined that the machine is intelligent. With the development of artificial intelligence technology, people are no longer satisfied with machines recognizing human instructions and hope that machines can recognize natural language and make a reasonable and natural response to it. Therefore, a large number of researchers have carried out research on human-machine dialogue systems. The model in this section is improved based on the Transformer structure, and the model is divided into two parts, and its structure is shown in Figure 9.

In the figure above, the encoder: the encoder is composed of n identical layers, and this experiment takes 6 modules to be superimposed. Each encoder layer contains three sublevel layers, and the first is the sequential information layer added in this section, the second is the self-attention layer, and the third is a simple, position-related fully connected network floor, and a residual connection and multi-layer normalization are added to each sublayer. Decoder: the decoder also consists of the same number of layers, but in addition to the three sublayers of each layer, the decoder also has an encoder-decoder attention sublayer, which combines the output of the encoder and decoder to perform sequence attention calculations.

Because the Transformer structure is a parallel computing network structure, there is no internal timing calculation operation but the relative position information obtained by artificially adding a mathematical calculation to the input terminal. But for rich natural language sequences, such simple injection of position information is very insufficient. Especially in dialogue tasks, the sequence of sentences is very important semantic information, if the sequence information of sentences cannot be well mined, it will have a great impact on the generation of normal sentences. The improvement of the model in this section adds an RNN network as a sequential information layer before the self-attention sublayer of each module of Transformer. It is used to capture sequence information and then transfer the output result to the Transformer structure for further learning.

The baseline model in this paper is the model HRED based on the basic hierarchical cyclic encoder decoder and the multiround dialogue model HVRED based on variational

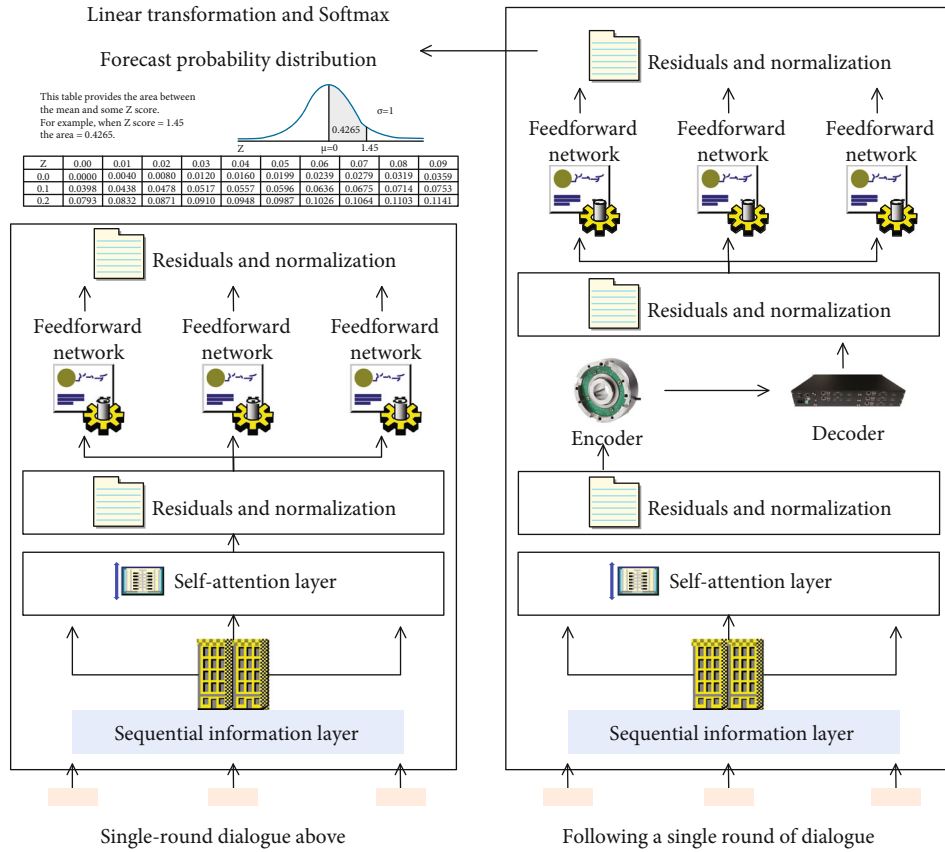


FIGURE 9: Transformer-based single-round dialogue model.

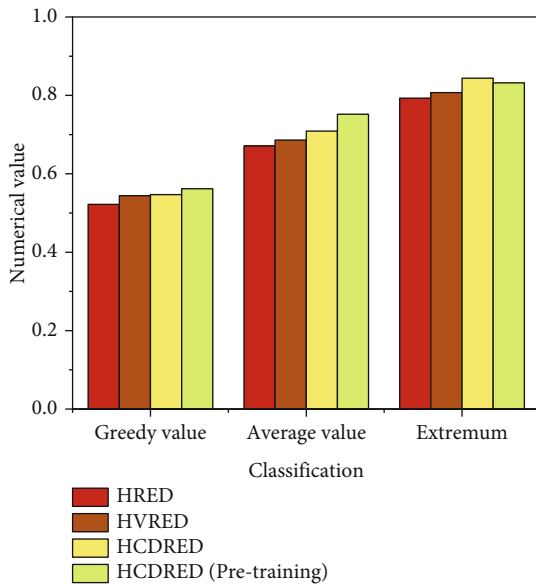


FIGURE 10: Histogram of experimental result analysis.

autoencoding. The multiround dialogue model HCVRED based on conditional variational self-encoding proposed in this paper calculates the average value, greedy value, and extreme value related to the word vector according to the generated results and label comparison of different models. In order to more intuitively understand the variation of the

evaluation indicators and better analyze the model results, the experimental results are drawn as a histogram, as shown in Figure 10.

It can be seen that in all models, HRED is the worst of all models, and HVRED has been significantly improved after introducing variational autoencoding. Through pretraining, HCVRED has further improved in terms of the degree of actual word overlap or the average value, greedy value, and extreme value calculated based on word vectors.

4. Discussion

Because computer network failure is a very complex and comprehensive problem, and there are many factors for network failure, this article does not apply the built model and system to the network management software manufactured by current manufacturers and does a good job of compatibility testing. This is also the shortcoming of the research in this paper, and it is also the direction that needs to be worked hard. In addition, the dialogue task in this article only uses text data as the training data set. With the continuous accumulation of various forms of data, especially the accumulation of pictures and video information, integrating these data into the dialogue task can further promote the intelligence of the dialogue system. Therefore, integrating a multimodal technology into dialogue tasks is also a popular technical research direction.

5. Conclusions

Although the research of neural network in the field of fault diagnosis has achieved good results, the structure and training times of neural network have a great influence on the fault diagnosis ability of neural network. A neural network with an unreasonable design may not be effective in fault diagnosis, at this stage, it is generally through multiple designs and trainings to gradually optimize the network results, which undoubtedly wastes a certain amount of resources. This paper is based on the task of automatic recognition of digital modulation signals, and on top of it, researches on the automatic recognition algorithm of digital modulation signals based on deep learning. It uses the current mainstream network model to study the signal modulation recognition task, and on this basis, some network models are improved for the signal classification task, which improves the classification accuracy and anti-interference performance. After using the SOM method, the absolute error did not reach 0.5, but the unused absolute error reached -0.0112 the smallest. It can be seen that the combination of SOM method and LM method has a better effect on improving the accuracy of sample simulation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Conflicts of Interest

The author states that this article has no conflict of interest.

References

- [1] Y. Dong and H. Wang, "Robust output feedback stabilization for uncertain discrete-time stochastic neural networks with time-varying delay," *Neural Processing Letters*, vol. 51, no. 1, pp. 83–103, 2020.
- [2] M. Secor, A. V. Soudackov, and S. Hammes-Schiffer, "Artificial neural networks as mappings between proton potentials, wave functions, densities, and energy levels," *The Journal of Physical Chemistry Letters*, vol. 12, no. 9, pp. 2206–2212, 2021.
- [3] S. Chan and A. H. Elsheikh, "Parametric generation of conditional geological realizations using generative neural networks," *Computational Geosciences*, vol. 23, no. 5, pp. 925–952, 2019.
- [4] K. Takano, M. Ueno, J. Moriya, M. Mori, Y. Nishiguchi, and F. Raes, "Unraveling the linguistic nature of specific autobiographical memories using a computerized classification algorithm," *Behavior Research Methods*, vol. 49, no. 3, pp. 1–18, 2016.
- [5] K. W. Lemke, K. Pham, D. M. Ravert, and J. P. Weiner, "A revised classification algorithm for assessing emergency department visit severity of populations," *The American Journal of Managed Care*, vol. 26, no. 3, pp. 119–125, 2020.
- [6] M. Aminu and N. A. Ahmad, "Complex chemical data classification and discrimination using locality preserving partial least squares discriminant analysis," *ACS Omega*, vol. 5, no. 41, pp. 26601–26610, 2020.
- [7] S. Kundu and S. Ari, "A deep learning architecture for P 300 detection with brain-computer interface application," *Innovation and research in biomedical en*, vol. 41, no. 1, pp. 31–38, 2020.
- [8] J. Cai, G. Zhou, M. Dong, X. Hu, G. Liu, and W. Ni, "Real-time arrhythmia classification algorithm using time-domain ECG feature based on FFNN and CNN," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6648432, 17 pages, 2021.
- [9] A. Liu, Z. Zhao, C. Zhang, and Y. Su, "Smooth filtering identification based on convolutional neural networks," *Multimedia Tools and Applications*, vol. 78, no. 19, pp. 26851–26865, 2019.
- [10] J. Q. James, Y. Hou, and V. O. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1694–1703, 2019.
- [11] A. Sebastian, A. Pannone, S. Subbulakshmi Radhakrishnan, and S. Das, "Gaussian synapses for probabilistic neural networks," *Nature Communications*, vol. 10, no. 1, pp. 1–11, 2019.
- [12] G. Rajchakit, R. Sriraman, N. Boonsatit, P. Hammachukiattikul, C. P. Lim, and P. Agarwal, "Global exponential stability of Clifford-valued neural networks with time-varying delays and impulsive effects," *Advances in Difference Equations*, vol. 2021, no. 1, 2021.
- [13] L. Li, Z. Qiao, and Z. Long, "A smoothing algorithm with constant learning rate for training two kinds of fuzzy neural networks and its convergence," *Neural Processing Letters*, vol. 51, no. 2, pp. 1093–1109, 2020.
- [14] C. Yin, L. Feng, and L. Ma, "An improved Hoeffding-ID data-stream classification algorithm," *The Journal of Supercomputing*, vol. 72, no. 7, pp. 2670–2681, 2016.
- [15] Y. Zhang, S. Wang, H. Zhao, Z. Guo, and D. Sun, "CT image classification based on convolutional neural network," *Neural Computing and Applications*, vol. 33, no. 14, pp. 8191–8200, 2021.
- [16] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.
- [17] F. Luo, W. Guo, Y. Yu, and G. Chen, "A multi-label classification algorithm based on kernel extreme learning machine," *Neurocomputing*, vol. 260, pp. 313–320, 2017.
- [18] P. Tapkan, L. Özbakır, S. Kulluk, and A. Baykasoglu, "A cost-sensitive classification algorithm: BEE-miner," *Knowledge-Based Systems*, vol. 95, pp. 99–113, 2016.
- [19] J. H. Du, "Automatic text classification algorithm based on gauss improved convolutional neural network," *Journal of Computational Science*, vol. 21, pp. 195–200, 2017.
- [20] C. Wu, L. Liu, M. Wei, B. Xi, and M. Yu, "Statistics-based optimization of the polarimetric radar hydrometeor classification algorithm and its application for a squall line in South China," *Advances in Atmospheric Sciences*, vol. 35, no. 3, pp. 296–316, 2018.
- [21] M. H. Rafiei and H. Adeli, "A new neural dynamic classification algorithm," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 28, no. 12, pp. 3074–3083, 2017.
- [22] J. I. Fernández-Michelli, M. Hurtado, J. A. Areta, and C. H. Muravchik, "Unsupervised classification algorithm based on EM method for polarimetric SAR images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 56–65, 2016.

- [23] N. Milic, B. Popovic, S. Mijalkovic, and D. Marinkovic, "The influence of data classification methods on predictive accuracy of kernel density estimation hotspot maps," *The international arab journal of information technology*, vol. 16, no. 6, pp. 1053–1062, 2019.
- [24] D. Devarriya, C. Gulati, V. Mansharamani, A. Sakalle, and A. Bhardwaj, "Unbalanced breast cancer data classification using novel fitness functions in genetic programming," *Expert Systems with Application*, vol. 140, article 112866, 2020.
- [25] A. Soula, K. Tbarki, R. Ksantini, S. B. Said, and Z. Lachiri, "A novel incremental kernel nonparametric SVM model (iKN-SVM) for data classification: an application to face detection," *Engineering Applications of Artificial Intelligence*, vol. 89, p. 103468, 2020.
- [26] Y. Li, Y. Chai, H. Yin, and B. Chen, "A novel feature learning framework for high-dimensional data classification," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 2, pp. 555–569, 2021.
- [27] C. Wang, C. Deng, Z. Yu, D. Hui, X. Gong, and R. Luo, "Adaptive ensemble of classifiers with regularization for imbalanced data classification," *Information Fusion*, vol. 69, no. 3, pp. 81–102, 2021.
- [28] A. Care, F. A. Ramponi, and M. C. Campi, "A new classification algorithm with guaranteed sensitivity and specificity for medical applications," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 393–398, 2018.
- [29] H. Rodrigues, F. J. G. Silva, L. G. Morgado, J. C. Sá, L. P. Ferreira, and R. D. S. G. Campilho, "A novel computer application for scrap reporting and data management in the manufacturing of components for the automotive industry," *Procedia Manufacturing*, vol. 51, no. 9–12, pp. 1319–1326, 2020.
- [30] A. Celesti, A. Bagula, I. De Falco, P. Brandão, and G. Sannino, "Emerging networked computer applications for telemedicine," *Journal of Network and Computer Applications*, vol. 130, pp. 104–106, 2019.