

Received April 12, 2020, accepted April 24, 2020, date of publication April 29, 2020, date of current version May 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991202

# Stream of Unbalanced Medical Big Data Using Convolutional Neural Network

WEIWEI GAO<sup>1</sup>, LI CHEN<sup>2</sup>, AND TAO SHANG<sup>3</sup>

<sup>1</sup>College of Information and Technology, Wenzhou Business College, Wenzhou 325035, China

<sup>2</sup>College of Mathematics Physics and Information Engineering, Jiaxing University, Jiaxing 314001, China

<sup>3</sup>College of Mechanical and Electrical Engineering, Jiaxing University, Jiaxing 314001, China

Corresponding author: Li Chen (cl\_zjx@163.com)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY15F020040, and in part by the Humanity and Social Science Youth foundation of Ministry of Education of China under Grant 17YJCZH033.

**ABSTRACT** In order to address the problem that the traditional algorithm can not predict the network link load effectively, which leads to high packet loss and energy loss, long turnaround time, slow stream rate and poor anti-attack ability, the paper proposes the stream algorithm of unbalanced medical big data based on convolutional neural network (CNN). The proposed algorithm included two stages: In the first stage, the decomposition-prediction model was constructed, the combined wavelet analysis and neural network analysis were used to complete the network link load prediction; In the second stage, based on the network link load situation, we analyzed the structure of each layer of convolution neural network, constructed the medical big data stream optimization model, introduced the ReLu function to calculate the convolution neural network, solved the optimization model, and completed the stream processing of unbalanced medical big data. The experimental results show that the network link load prediction accuracy of the proposed stream algorithm is as high as 93%, the lowest packet loss rate is only 2.0%, the energy loss of the stream process is low, the rate is fast, and the anti-attack efficiency is high, which is more conducive to the realization of data stream.

**INDEX TERMS** Convolutional neural network, unbalanced medical big data, network load forecasting, ReLu function, data stream.

## I. INTRODUCTION

Unbalanced data often exists in the medical field, that is, the types of data are not evenly distributed [1]–[3]. The amount of data in one category is abnormally larger than that in other categories [4], [5], occupying a dominant position. This kind of data is called unbalanced data [6].

Intelligent medical treatment is a hot trend in medical development at present. It mainly uses technology to integrate medical data to form a resource sharing platform, promotes medical resource sharing, and solves the problems of waste of medical resources and inefficient medical progress due to news blocking

During the development of intelligent medical treatment [7]–[9], a large amount of medical data will be generated, with different data sources and data uses [10], and there are also huge differences in computer communication interfaces and related communication protocols. The resulting big data

The associate editor coordinating the review of this manuscript and approving it for publication was Honghao Gao<sup>1</sup>.

has various attributes [11], [12]. Due to different link lengths and uneven load capacity, it is easy to cause data loss and excessive average turnaround time during data sharing [13]. In order to better store and share medical data resources, suppress data packet loss, and ensure data integrity, medical big data needs to be streamed. Unbalanced medical data processing as a special data learning method has become a key research issue in the medical field [14].

Convolutional neural network is a commonly used artificial neural network model. It is widely used in data processing, artificial intelligence and pattern recognition [15]. It is currently used in medical fields with significant effects. Compared with other data stream methods, the convolutional neural network has a special structure and partial connection between layers. It can independently select data features, fully and flexibly process medical data, and obtain better data research results.

Yumeng *et al.* [16] used convolutional neural networks to conduct user interest recognition research, and used convolutional neural networks to fuse diverse user information. Based

on this, they used maximum likelihood estimation to implement interest recognition. Wanjun *et al.* [17] proposed an adaptive enhanced convolutional neural network recognition method, constructed an adaptive enhanced model, combined with the convolutional neural network operation process analysis, optimized the performance of the neural network, and obtained good recognition results. Yongping *et al.* [18] used a convolutional neural network to adaptively set the window size, extract the semantic features of short text, and predict the emotional tendency, which significantly improved the short text classification effect.

Based on the above research, this paper proposes a stream algorithm of unbalanced medical big data based on convolutional neural network. The algorithm first performs network link load prediction analysis, and on this basis, uses convolutional neural networks to stream medical data traffic.

Finally, to test the performance of the algorithm stream, we carried out test and the results showed that compared with the traditional stream algorithm of unbalanced medical data. The link load prediction accuracy under the stream algorithm is high, the data packet loss rate is reduced, the average data turnaround time is shortened, the stream efficiency is high, the energy consumption is low, and the anti-attack ability is strong. Thus it showed that the algorithm has better stream effects and is conducive to medical data balance and sharing.

The main contributions of this paper are as follows:

(1) A decomposition-prediction model was constructed, which combines wavelet analysis and neural network analysis to complete the network link load prediction together, and provides the basis for data stream.

(2) Constructed an optimization model for medical big data stream;

(3) ReLu function was introduced in the process of stream treatment of unbalanced medical big data, which increased the data stream rate in a powerful way.

(4) The diversified index was used to verify the proposed algorithm, which improves the reliability of the research results.

## II. RELATED WORK

For the time being, the medical data stream technology is centered on intelligent algorithm, such as the clustering algorithm, ant colony algorithm and tabu search algorithm.

Literature [19] proposed a new attribute weighting method. This method treats each attribute as a single classifier and uses the area under the ROC curve (AUC) to measure its discriminative ability. Then, each AUC value is used for weighting. Finally, this new attribute weighting method is incorporated into Naive Bayes for unbalanced data classification, but the method has weaker attack resistance. Literature [20] focused on analyzing an unbalanced data classification method, and experimentally verified the performance of the method, and obtained good results, but the method has a higher data packet loss rate; Literature [21] discussed the use of GPUs to speed up gravity calculations, and used data classification methods in the process. Classify the data of the gravity calculation

process and assign them to the GPU threads in parallel to calculate the gravity at the same time. The calculation results are better, but the energy loss during the calculation process is higher;

Literature [22] designed a large-scale unbalanced data classification algorithm with the differential twin neural network, which enhanced the depth structure expressing ability of convolutional neural network through the differential convolution in order to increase the discriminative ability of convolutional neural network. On this basis, we optimized the class feature maps of each set of unbalanced data, linked each set of data with multiple hyper planes and judged the class labels of data in accordance with distance between data and hyper plane so as to realize data stream. However, this method has the problem that it takes a long time to split the data. Reference [23] designed a new distributing prediction algorithm of unbalanced data based on threshold optimization algorithm. This algorithm has added the information classification process based on the traditional KSVM classifier and effectively classified the information near the hyper plane. Meantime, as for the fixed defects of time threshold of unbalanced data, we carried out dynamic adjustment to realize data stream with dynamic threshold optimization algorithm. However, the calculation process has a large energy loss;

Literature [24] pointed out that most traditional pattern classifiers assume that their input data performs well in similar low-level class distributions, balanced class sizes, etc. However, the actual data set shows various forms of irregularities, which are often enough to confuse the classifier, thereby reducing the ability of the classifier. Therefore, the method based on the irregular characteristics of the distribution and characteristics to analyze the data classification problem, found an effective data research method, and achieved certain results, but the calculation performance of this method still needs to be further improved.

Aiming at the problems of the above research methods, this paper proposes a stream algorithm of unbalanced medical big data based on convolutional neural network. Algorithm research has been completed from the two stages of network link complex prediction and data stream description. The experimental results show that the link load prediction accuracy of the proposed algorithm is high, the minimum packet loss rate is only 2.0%, the average stream turnaround time is only 6.82s. The energy loss during the shunting process is low, and the rate is fast. The anti-attack efficiency can reach 90%, and the overall performance is better.

At present, medical big data is gradually coming out of inter-departmental restrictions and have realized resource interconnection and sharing, thus laying an important data foundation for the development of medical undertakings [25], [26]. In the face of the huge amount of medical data that are growing at an amazing speed, however, the smooth and effective operation of data shared network is confronting huge pressure [27], [28]. They are likely to exceed the available bandwidth of network link and cause network congestion. Therefore, it is urgent to find an effective

solution. In case of such situation, the network operator has put forth two solutions. The first is to upgrade the current medical data shared network or establish a base station with a smaller coverage [29]. This solution is effective though, it will cost a high volume of time and costs and the overall efficiency is low [30]. The second is data stream. Data stream can be explained as distributing data flow of some wireless network to another or several wireless network(s) in return to share the load of a single network and reduce data transmission delay, shorten the average turnaround scheduling time, lower data package loss rate and ensure the data integrity [31]–[33].

Therefore, we designed the stream algorithm of unbalanced medical big data based on convolutional neural network and it mainly includes two parts: The first part is to establish the prediction model of network load degree to predict the network link load; the second part is to distribute the medical data flow with the stream model of convolutional neural network according to the prediction results. Next, we'd like to make a concrete analysis of the algorithm.

### III. STREAM ALGORITHM OF UNBALANCED MEDICAL BIG DATA BASED ON CNN

#### A. NETWORK LINK LOAD PREDICTION

The stream algorithm of unbalanced medical big data has two steps—link load prediction and data flow stream. When transferring the data from one node to another node and finally to the destination node and one of the nodes selects the next hop [34], [35], we need to test the cache state. Only when the remaining cache space of the next hop node is larger than the information size can it hop to the next node [36]. It will not cause congestion. This process is called network link load prediction. Network link load prediction is important for data stream [37], [38].

As different networks have different topological structures, the network loading capacities are also different [39]. Therefore, before predicting the load of medical data shared network, making clear the network topology is of great importance to the next procedure of prediction of network link load.

The commonly used network link load prediction methods include wavelet analysis and neural network analysis. The wavelet analysis in the most effective method for processing unstable time series. The wavelet can decompose the long medical data flow into multiple independent small data flows. In this way, we can carry out time series prediction for the historical data flow on the network links. Neural network analysis relies on unique structural advantages, which can input, process, and output data to complete network link load prediction. Although both methods can achieve load forecasting, the accuracy of the research results is lower by relying on a single method for prediction.

Therefore, this work combines the two methods to construct a decomposition-prediction model. The model first uses wavelet analysis to decompose the data stream, and then uses the decomposed data as the input data of the neural

network model, and runs the neural network to predict the network link load capacity. The steps are as follows:

#### 1) DECOMPOSE DATA STREAM WITH WAVELET ANALYSIS

Randomly select a wavelet basis function and use the following process to expand and translate and get a set of function families. The selected wavelet basis function  $A_{ab}(t)$  is as follows:

$$A_{ab}(t) = \frac{\left(\frac{t-b}{a}\right)}{\sqrt{a}} \quad (1)$$

where,  $a$  represents time scale;  $b$  represents time displacement;  $t$  represents time.

By expanding and translating process for formula (1), namely the discrete wavelet transform, and the transformed wavelet basis function is defined as follows:

$$A_{mn}(t) = \frac{\left(\frac{t-na_0^m b_0}{a}\right)}{\sqrt{a_0^m}} \quad (2)$$

where,  $m$  represents the specified wavelet scaling integer;  $n$  represents the specified integer of wavelet translation;  $a_0$  represents fixed scale parameters greater than 1 ( $a_0 > 1$ );  $b_0$  represents transfer-parameters.

By calculating the similarity coefficient, and the process is as follows:

$$B_{mn} = \int x(t) A_{mn}(t) dt \quad (3)$$

where, Data  $x(t)$  can be expressed as the cumulative sequence of similarity coefficient and detail coefficient:

$$x(t) = \sum_{m=1, n=1}^{\infty} B_{mn} + \sum_{m=1, n=1}^{\infty} C_{mn}(t) \quad (4)$$

where,  $C_{mn}(t)$  is the detail coefficient, which can be expressed as:

$$C_{mn}(t) = \frac{1}{\sqrt{2^m}} A_{mn}(t) \quad (5)$$

According to the transformed wavelet basis function and similarity coefficient, data stream decomposition can be completed, and the decomposed data stream can be expressed as:

$$D = O[A_{mn}(t)x(t) + C_{mn}(t)] \quad (6)$$

where,  $O$  represents the original data stream.

#### 2) NEURAL NETWORK LINK LOAD PREDICTION

Generally speaking, the neural network includes input layer, hidden layer and output layer [40], [41], and the basic thinking is as follows: The neural network includes two processes: learning and training. First, design network structure to determine the input and output layer. Here the historical flow data in the network structure are taken as the input and the hopping load as the output. And the hidden layer is more important, as the number of hidden layer neurons will directly influence

the prediction results. Therefore, the number of hidden neurons is generally calculated by the following formula:

$$s = \sqrt{q + p} + c \tag{7}$$

where,  $q$  represents the number of neurons in the input layer;  $p$  represents the number of neurons in the output layer;  $c$  represents the constants between [1, 10].

The decomposed data stream  $D$  is input to the neural network model, and the data stream is processed according to the neurons. During the process, the S-type tangent function and S-type logarithmic function are selected as the excitation function, and the formula is as follows:

S-type tangent function:

$$\tan x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

S-type logarithmic function:

$$\log x = \frac{1}{(1 + e^{-x})} \tag{9}$$

where, in formulas (8) and (9),  $e$  represents gradient, and  $x$  represents threshold.  $\theta$  is the excitation parameter.

Therefore, the neural network link load prediction value can be obtained and expressed by the following formula:

$$P_{load} = D (\log \theta + \tan \theta) \cdot s \tag{10}$$

Based on the above steps, the neural network link load prediction can be completed, the network status can be clear, and the foundation for medical big data stream can be laid. For prediction, we selected neural network toolbox in MATLAB for network training [42]–[44].

### B. USING CONVOLUTIONAL NEURAL NETWORK TO STREAM MEDICAL BIG DATA

Convolutional neural network includes input layer, pooling layer, fully connected layer and output layer. The input layer is responsible for managing the original data. The pooling layer is mainly set up to avoid dimensional disasters. Therefore, the main role of this layer is to compress large data and avoid fitting problems. To achieve the above goals in this layer, it is mainly accomplished by the maximum value pooling method, that is, the maximum flow value of the data is calculated and retained. Fully connected layer and output layer. They are located at the bottom of the entire convolutional neural network. The main role is to output the optimal medical big data stream result.

The convolutional neural network is used to reasonably stream the unbalanced medical big data, as follows:

Input: Original medical big data, and network link load prediction;

Output: Unbalanced medical big data stream results.

(1) Firstly, we calculated the stream of frequent dynamic measurement of unbalanced medical big data:

$$V = N\beta + \exp(gQ) \tag{11}$$

where,  $\beta$  represents the cross-correlation of unbalanced medical big data,  $N$  represents data volume,  $g$  represents ambiguity function of unbalanced medical big data, and  $Q$  represents data fuzzy features.

(2) As for limited unbalanced medical big data set  $X_i$ , we established vector set of three-dimensional reconstruction stream:

$$R = \begin{cases} 0 & \text{if } i < 1 \\ \text{new}i^* & \text{otherwise} \end{cases} \tag{12}$$

where,  $\text{new}i^*$  can be obtained through statistical analysis method. The process is as follows:

$$\text{new}i^* = \frac{V}{X_i} \tag{13}$$

(3) Building the automatic corrective decision function of unbalanced medical big data, such as:

$$D = R \left( \sum_{i=1}^N X_{ik} \times t \right) \tag{14}$$

where,  $t$  is time, and  $k$  is the effective amplitude of big data.

(4) On the basis of realizing automatic corrective decision of unbalanced medical big data, the optimization model of medical data flow stream is as follows:

$$\begin{cases} st \sum_{i=1}^N r_i = 1 \\ 0 \leq r_i \leq 1 \\ z\bar{r} \leq g \end{cases} \tag{15}$$

where,  $z$  represents the total transmission rate of medical data flow;  $g$  represents the max sending rate;  $r_i$  represents the ratio of the data sub-flow assigned on the  $i$  link  $N_i$  accounting for the total transmission data;  $\bar{r}$  represents the result vector of flow stream ratio.

(5) The convolutional neural network will select sigmoid function as the network activation function to solve the optimization model of the above medical data flow stream. The expression of the formula is shown in formula (9). But this will lead to the growing calculated amount and vanishing gradient. As a result, the entire neural network cannot finish the follow-up training. To this end, we will use a new motivation function, namely ReLu function. The function formula is as follows:

$$f(d) = \max(0, 1) = \begin{cases} 1, & d < 0 \\ d, & d \geq 0 \end{cases} \tag{16}$$

where,  $d$  represents the stream predictive value.

(6) After learning convolutional neural network, we need to train it, namely adjusting the values of various parameters through the error between the estimated value and actual value to increase the rationality of shared network stream of medical data and obtain unbalanced medical big data stream results.

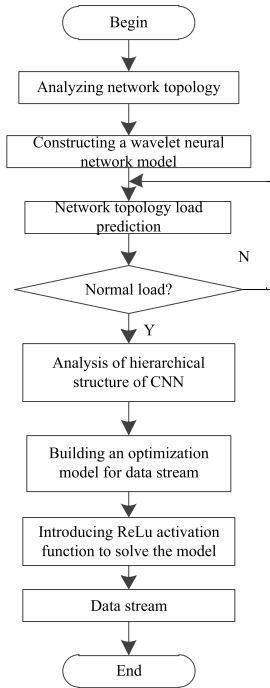


FIGURE 1. Unbalanced medical big data stream algorithm.

TABLE 1. Experimental environment.

Type	Parameter
System paramter	Intel Core i5-2400CPU、 4G memory、 8.5G hard disk
Development environment	MyEclipse2014+JDK1.7
Programming environment	Matlab R2015b
Operating system	Windows10

(7) End

The above process can complete the unbalanced medical big data stream. The overall stream process is shown in Figure1.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

##### A. EXPERIMENTAL ENVIRONMENT AND PARAMETER

In order to verify the application performance of the proposed stream algorithm of unbalanced medical big data based on convolutional neural network, we have carried out the below simulation tests. The testing environment consists of 25 virtual nodes. It is shown in Table1.

This paper completes the unbalanced medical big data stream based on the convolutional neural network. The parameters of the convolutional neural network are shown in Table2.

##### B. EXPERIMENTAL DATASET

The experiment selects the data set in the ADNI database (<http://adni.loni.usc.edu>) as training data and experimental test

TABLE 2. Convolutional neural network parameters.

Parameter	Value
Feature dimension	90
Number of convolutional layers	2
Number of iterations	500
Weight range	[-100,100]

TABLE 3. Experimental dataset.

Dataset name	Data amount / ten thousand	Attributes
Magnetic Resonance Imaging	300	3
Positron Emission Computed Tomography	300	3
Plasma from Blood	300	3
single nucleotide polymorphisms	300	3
Subject physical examination	300	3

data. The database includes a large amount of clinical data, Magnetic Resonance Imaging, Positron Emission Computed Tomography, Plasma from Blood, single nucleotide polymorphisms and Subject physical examination. The experimental data is described in Table 3.

According to Table 3, the five data sets used in this experiment are Magnetic Resonance Imaging, Positron Emission Computed Tomography, Plasma from Blood, single nucleotide polymorphisms and Subject physical examination. Each data set is 300 ten thousand, with a total of 1500 ten thousand data sets, of which 500 ten thousand data sets are used for model training and 1000 ten thousand data sets are used as test data sets.

##### C. EXPERIMENTAL INDEXES

In the experimental test, there are six test indexes, namely:

###### 1) NETWORK LOAD PREDICTION ACCURACY

In the process of completing big data stream, the network load prediction was first carried out, which laid the foundation for the offloading. In order to verify the effectiveness of the proposed algorithm, the network load prediction accuracy of the proposed algorithm and other literature algorithms need to be compared.

$$Accuracy = \frac{M}{M_{act}} \times 100\% \quad (17)$$

where,  $M$  represents the value of the algorithm network load prediction result.  $M_{act}$  represents the actual network load result value.

###### 2) DATA PACKAGE LOSS RATE

Data package loss rate refers to the ratio of the number of lost data packages in the test accounting for the number of data packages being transmitted, which is usually tested within the

throughput range.

$$v = \frac{y}{Y} \cdot 100\% \quad (18)$$

where,  $v$  represents the medical data package loss rate;  $y$  represents the number of lost data packages;  $Y$  represents the total number of medical data packages being transmitted.

### 3) AVERAGE STREAM TURNAROUND TIME OF MEDICAL DATA PACKAGES

The average stream turnaround time of medical data packages refers to the time used from data being input to completion of stream.

$$l = \frac{1}{G} \sum_{i=1}^G (H_1 - H_2) \quad (19)$$

where,  $l$  represents average turnaround time;  $G$  represents quantity;  $H_1$  represents data upload time;  $H_2$  represents the completion time of application.

### 4) ENERGY LOSS DURING STREAM

Energy loss refers to energy consumption precipitated by transmission medium during data stream. The lesser the energy loss, the higher the stream efficiency.

$$L = L_s - (L_1 + L_2) \quad (20)$$

where,  $L_s$  represents basic path loss;  $L_1$  and  $L_2$  represent the plane wave gains at the data sending and receiving terminals.

### 5) DATA STREAM RATE

The stream rate can reflect the work efficiency of the stream process.

$$\text{Stream rate} = \frac{1}{v_e \times \log 2T} \quad (21)$$

where,  $v_e$  represents symbol transmission rate;  $T$  represents the valid value state of the stream process.

### 6) ANTI-ATTACK CAPABILITY OF THE STREAM PROCESS

The anti-attack capability of the stream process can be reflected through effective anti-attack efficiency which can capture the stability of stream process. The calculation process is as follows:

$$\begin{aligned} &\text{Effective anti-attack efficiency} \\ &= \frac{\text{Number of successful defenses}}{\text{Total Attacks}} \times 100\% \quad (22) \end{aligned}$$

## D. EXPERIMENTAL RESULTS

### 1) COMPARISON OF NETWORK LOAD PREDICTION ACCURACY

Using the calculation method of formula (17), the network load prediction accuracy of the proposed algorithm and Literature [20], Literature [21], Literature [22], Literature [23], and Literature [24] are calculated, and the comparison results are shown in Figure 2.

According to the analysis in Figure 2, with the amount of data increases, the prediction accuracy curve of the proposed

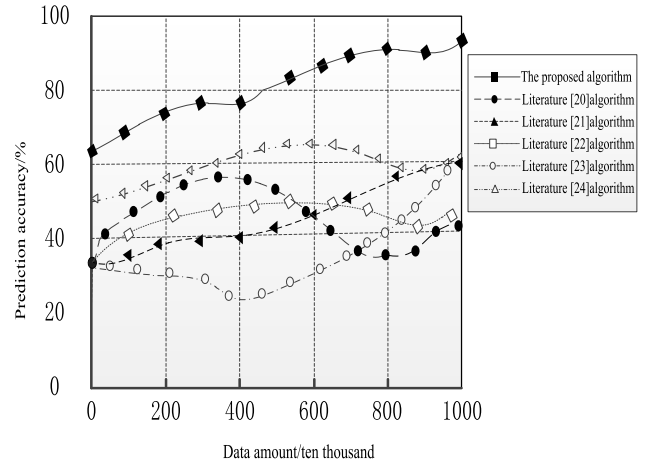


FIGURE 2. Comparison of network load prediction accuracy.

algorithm always shows an upward trend, and the accuracy curves of other literature algorithms fluctuate up and down, especially the algorithm in Literature [20] algorithm, The prediction accuracy curve of the algorithm first rises and then decreases, and the fluctuation range is large. The accuracy of Literature [24] is relatively high. The maximum prediction accuracy of Literature [21] and Literature [23] is 60%. Literature [22] is relatively stable, about 42%. It can be clearly seen from the Figure2 that the prediction accuracy of the proposed algorithm is always the highest, which can be as high as about 93%.

According to the above data analysis, it can be seen that the algorithm has significant advantages and high prediction accuracy. This is because this paper proposes a decomposition-prediction model, which combines wavelet analysis and neural network analysis to predict the network load, and obtains good results.

### 2) COMPARISON OF DATA PACKAGE LOSS RATE

Using the calculation method of formula (18), the data packet loss rate of the data classification algorithms of different literature algorithms is calculated, and the results are shown in Figure3.

In comparison to the five traditional data stream algorithms, the use of the stream algorithm of unbalanced medical big data based on convolutional neural network for stream can largely reduce the package loss rate of medical data, with the package loss rate ranging from 2.0% to 3.4%.

Among the five traditional algorithms, the data loss rates of the algorithms in Literature [21] and Literature [22] are high, both of which are above 8%. The data packet loss rate of the algorithm in Literature [20] is about 6%. With the increase of the number of iterations in the algorithm of Literature [23], the packet loss rate changes greatly, the highest is 8%.

The Literature [24] has a relatively low packet loss rate, but the lowest value is 4%, which is still higher than the highest packet loss rate of the proposed algorithm. It can be clearly seen that the proposed algorithm is significantly lower than the other five data stream algorithms. It can be seen that

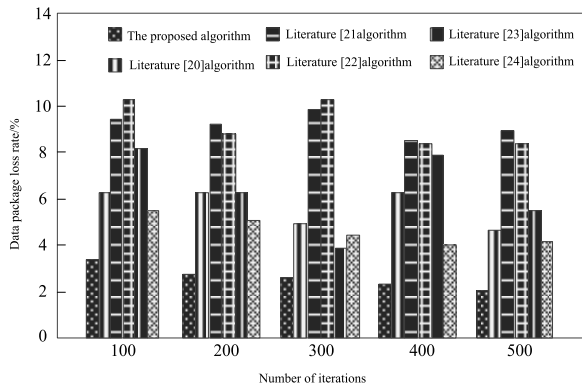


FIGURE 3. Comparison of data packet loss rates of different stream algorithms.

TABLE 4. Medical data packet average stream turnaround time test.

Iteration number	The proposed algorithm	Literature [20] algorithm	Literature [21] algorithm	Literature [22] algorithm	Literature [23] algorithm	Literature [24] algorithm
100	6.9	9.8	12.7	12.5	14.5	12.7
200	7.0	10.0	14.2	12.2	10.1	13.0
300	6.5	10.4	16.1	11.4	16.2	14.1
400	6.1	12.1	16.7	10.8	17.7	12.4
500	7.6	10.1	15.3	10.7	14.1	13.7
Average value	6.82	10.48	15.0	11.52	14.52	13.18

the unbalanced medical big data stream algorithm based on convolutional neural network can effectively guarantee the integrity of the packets with stream.

### 3) COMPARISON OF AVERAGE STREAM TURNAROUND TIME OF MEDICAL DATA PACKAGES

Using formula (19), Test and calculate the turnaround time of different algorithms during the stream of medical data packets, and calculate average value.

From Table4,in comparison to the five traditional data stream algorithms, the stream with the stream algorithm of unbalanced medical big data based on convolutional neural network can greatly reduce shorten the average stream turnaround time of medical data packages. The average stream turnaround time of medical data packages with such algorithm is 6.82s but that of the five traditional data stream algorithms is all above 10s.

Literature [21] algorithm has the highest, which is 15.0s. The average stream turnaround time of medical data packets in Literature [20] is relatively low, which is 10.48s, but it is still higher than the average value of 3.68s in proposed algorithm. Thus it can be seen that the treatment performance of stream algorithm based on convolutional neural network

is superior and has achieved the goal of this study and is more conducive to the development of medical and health undertakings.

### 4) COMPARISON OF ENERGY LOSS DURING STREAM

We tested the energy loss of stream process of different algorithms and the results are shown in Figure4.

According to Figure4 (a)-Figure3 (f), with the increase of the number of iterations, the energy loss of stream process of different algorithms is changing. The energy loss of stream process of algorithm in Literature [20] is on the rise and the algorithm in Literature [21] before the number of iterations 400, the loss polyline continued to decline, the number of iterations 400-500, the loss did not change. The algorithm in Literature [22] after 200 iterations, the energy loss in the stream process continued to decline.

The energy loss of stream process of algorithms in Literature [23] and [24] and the proposed algorithm does not have an obvious law of change. But that of the proposed algorithm is always under 20dB, which is lower than the loss rate of other five stream methods. It can be seen that the stream efficiency of stream algorithm of unbalanced medical big data based on convolutional neural network is the highest.

This is because during the operation of the convolutional neural network in this paper, the new excitation function ReLu function is used for analysis, which reduces the computational complexity and the energy consumption of algorithm operation.

### 5) COMPARISON OF DATA STREAM RATE

In order to further verify the efficiency of stream algorithm of unbalanced medical big data based on convolutional neural network, we contrasted the data stream rate of different algorithms. Use formula (21) to calculate the stream rate of different algorithms. The results are shown in Table 5.

From Table5, with the increase of the number of iterations, the data stream rate of different algorithms is changing. The data stream rate of the proposed algorithm is all above 55bts. The data diversion rate of algorithms in Literature [20], Literature [21] and Literature [23] is too low to exceed 30bts. The stream rate of the algorithms in Literature [22] and Literature [24] is relatively high, but it is still far lower than that of

TABLE 5. Comparison of medical data stream rate (bts).

Iteration number	The proposed algorithm	Literature [20] algorithm	Literature [21] algorithm	Literature [22] algorithm	Literature [23] algorithm	Literature [24] algorithm
100	55.4	21.9	24.5	42.7	22.6	35.0
200	56.1	20.7	24.2	44.4	22.7	44.7
300	55.4	20.9	26.0	50.0	26.4	30.8
400	56.7	30.0	24.0	47.5	27.7	47.0
500	56.1	29.1	29.3	40.3	24.7	41.4

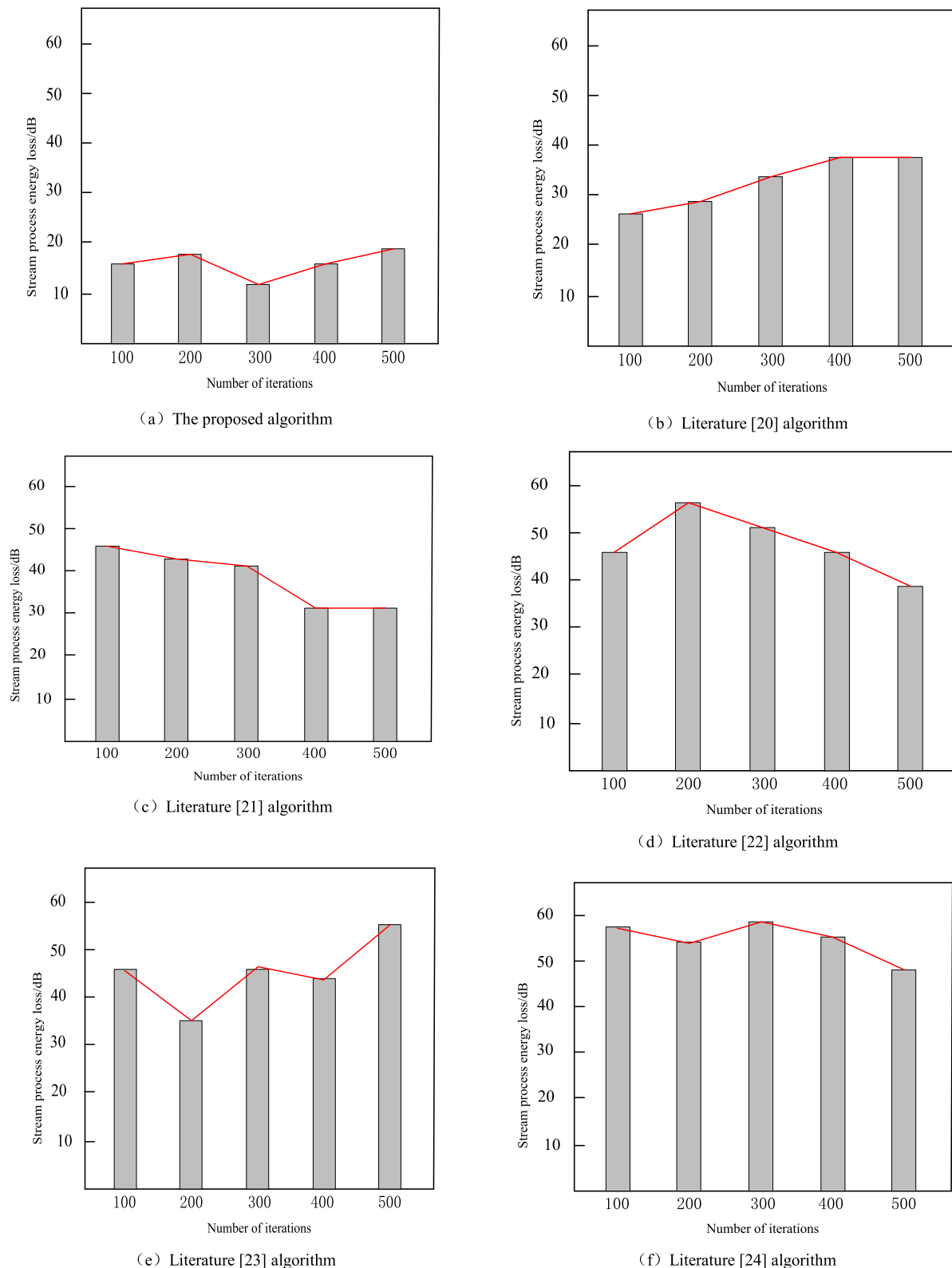


FIGURE 4. Comparison of energy loss in the stream process of different algorithms.

the proposed algorithm, and the maximum is no more than 50bps. Therefore, it can be seen that the stream algorithm of unbalanced medical big data based on convolutional neural

network has faster stream rate, fully takes advantage of the convolutional neural network, and has a high efficiency in the stream process.



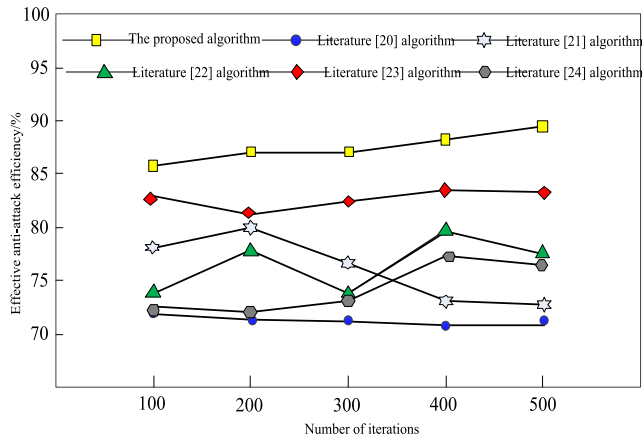


FIGURE 5. Comparison of effective anti-attack efficiency of different stream algorithms.

## 6) COMPARISON OF ANTI-ATTACK CAPABILITY OF THE STREAM PROCESS

Use formula (22) to calculate the effective anti-attack efficiency of different algorithms and verify the anti-attack ability of the algorithm stream process. The comparison results are shown in Figure 5.

From Figure 5, in multiple iterations, only the effective anti-attack efficiency of the stream algorithm of unbalanced medical big data based on convolutional neural network of the proposed algorithm is steadily on the rise, with the highest up to 90%.

The algorithm in Literature [23] has the anti attack efficiency next to that proposed paper, which is more than 80% as a whole. The algorithms in Literature [21], Literature [22], and Literature [24] have obvious fluctuations, and the overall variation varies from 72% to 80%. The algorithm in Literature [20] has the lowest effective anti-attack efficiency, with an average of about 71%. It can be seen that the stream algorithm of unbalanced medical big data based on convolutional neural network has stronger anti-attach performance and application stability.

## V. CONCLUSION

Medical undertakings have a direct bearing on the survival and development of humans. Every day, medical institutions receive numerous patients and there will be generous amounts of medical data. But the medical data are stored in the medical institutions and not shared with others. These independent data are unfavourable to the long-term development of medical undertakings. Therefore, the state has rolled out many policies to promote resource sharing. Whereas, in the process of medical data sharing, the shared network falls short of full load, which will contribute to data package loss and slow average turnaround time. To this end, this paper has proposed a stream algorithm of unbalanced medical big data based on convolutional neural network. The network link load is analyzed. Under the condition of grasping the network load, the convolutional neural network and the ReLu function are

used to perform data calculation and analysis to complete the research on balanced medical big data stream.

The experimental results show that the proposed network link load prediction accuracy of the proposed algorithm is as high as about 93%, the data packet loss rate is low, the average stream turnaround time is short, and the energy loss during the stream process is always kept below 20dB. The rate is fast and the anti-attack efficiency is high. This proves that the proposed algorithm has a strong shunting ability, is conducive to achieving data balance, and can provide effective support for the realization of medical resource sharing.

In the research, we also find that when solving the data flow stream model with convolutional neural network, if the sigmoid function is used as network activation function, it will give rise to increasing calculation amount and vanishing gradient, which not only makes it hard for the neural network to complete the subsequent training, increases energy loss of stream process and decreases stream rate. While the use of ReLu function as activation function can avoid this phenomenon.

In the future, we will consider different application environments, further collect more comprehensive unbalanced medical big data, further analyze the comprehensive characteristics of the data, find more detailed stream paths, optimize data stream steps, and obtain more effective ways to analyze medical data.

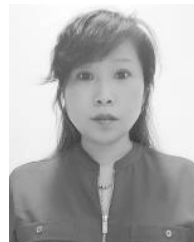
## REFERENCES

- [1] M.-A. Ziad, M. Gouskir, and B. Bouikhalene, "Classification of brain tumor from magnetic resonance imaging using convolutional neural networks," *Int. J. Adv. Sci. Technol.*, vol. 126, pp. 31–38, May 2019.
- [2] B. A. M. Ashqar and S. S. Abu-Naser, "Identifying images of invasive hydrangea using pre-trained deep convolutional neural networks," *Int. J. Control Autom.*, vol. 12, no. 4, pp. 15–28, Apr. 2019.
- [3] Y. Yin, J. Xia, Y. Li, Y. Xu, W. Xu, and L. Yu, "Group-wise itinerary planning in temporary mobile social network," *IEEE Access*, vol. 7, pp. 83682–83693, 2019.
- [4] Y. Yin, L. Chen, Y. Xu, J. Wan, H. Zhang, and Z. Mai, "QoS prediction for service recommendation with deep feature learning in edge computing environment," *Mobile Netw. Appl.*, vol. 25, no. 2, pp. 391–401, Apr. 2020, doi: 10.1007/s11036-019-01241-7.
- [5] J. Li, X. Zhang, Z. Wang, X. Chen, J. Chen, Y. Li, and A. Zhang, "Dual-band eight-antenna array design for MIMO applications in 5G mobile terminals," *IEEE Access*, vol. 7, pp. 71636–71644, 2019.
- [6] W. Yi and R. Shuxia, "Review of medical big data visualization research," *Comput. Sci. Explor.*, vol. 11, no. 5, pp. 681–699, 2017.
- [7] J. Yu, J. Li, Z. Yu, and Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2019.2947482.
- [8] J. Yu, C. Zhu, J. Zhang, Q. Huang, and D. Tao, "Spatial pyramid-enhanced NetVLAD with weighted triplet loss for place recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 661–674, Feb. 2020, doi: 10.1109/TNNLS.2019.2908982.
- [9] X. Ma, H. Gao, H. Xu, and M. Bian, "An IoT-based task scheduling optimization scheme considering the deadline and cost-aware scientific workflow for cloud computing," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–19, Dec. 2019, doi: 10.1186/s13638-019-1557-3.
- [10] D. Chen, R. Zhang, and L. Fu, "Research on medical health big data mapping and migration method under SNOMEDCT system," *J. Chin. Soc. Inf.*, vol. 37, no. 5, pp. 78–86, 2018.
- [11] K.-J. Kim, S.-H. Jung, W.-H. So, and C.-B. Sim, "A study on mushroom pest and diseases analysis system implementation based on convolutional neural networks for smart farm," *Int. J. Control Autom.*, vol. 10, no. 11, pp. 61–72, Nov. 2017.

- [12] Y. Cho and S. Woo, "Automated ROI detection in left hand X-ray images using CNN and RNN," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 7, pp. 81–92, Jul. 2018.
- [13] Z. Xiaoting, L. Yin, and T. Jinglei, "Design and simulation of big data streaming system based on optimized clustering algorithm," *Comput. Simul.*, vol. 35, no. 12, pp. 220–223, 2018.
- [14] C. Yuxiang, F. Ting, and Z. Hui, "Reasonable streaming method for mobile big data in cloud computing environment," *Sci. Technol. Eng.*, vol. 18, no. 25, pp. 194–199, 2018.
- [15] X. Yan, "Design and research of big data offload system in cloud computing platform," *Electron. Design Eng.*, vol. 27, no. 9, pp. 125–128, 2019.
- [16] D. Yumeng, Z. Weinan, and L. Ting, "User interest recognition based on topic enhanced convolutional neural network," *Comput. Res. Develop.*, vol. 55, no. 1, pp. 188–197, 2018.
- [17] L. Wanjuan, L. Xuejian, and Q. Haicheng, "Adaptive enhanced convolutional neural network image recognition," *Chin. J. Image Graph.*, vol. 22, no. 12, pp. 1723–1736, 2019.
- [18] D. Yongping, Z. Xiaozheng, and P. Bingbing, "Short text sentiment classification based on CNN-LSTM model," *J. Beijing Univ. Technol.*, vol. 7, pp. 662–670, 2019.
- [19] T. Kim, B. D. Chung, and J.-S. Lee, "Incorporating receiver operating characteristics into naive Bayes for unbalanced data classification," *Computing*, vol. 99, no. 3, pp. 1–16, 2016.
- [20] Y. Xu, Q. Wang, X. Pang, and Y. Tian, "Maximum margin of twin spheres machine with pinball loss for imbalanced data classification," *Int. J. Speech Technol.*, vol. 48, no. 1, pp. 23–34, Jan. 2018.
- [21] L. Peng, H. Zhang, H. Hassan, Y. Chen, and B. Yang, "Accelerating data gravitation-based classification using GPU," *J. Supercomput.*, vol. 75, no. 6, pp. 2930–2949, Jun. 2019.
- [22] R. Jiali and W. Wenjing, "Large-scale imbalanced data classification algorithm based on differential twin convolutional neural network," *Comput. Appl. Softw.*, vol. 36, no. 11, pp. 267–274, 2019.
- [23] J. Xin, G. Guoqing, and L. Xu, "Threshold optimization algorithm e-KSM for unbalanced data classification prediction," *Comput. Appl. Softw.*, vol. 35, no. 1, pp. 276–280, 2018.
- [24] S. Das, S. Datta, and B. B. Chaudhuri, "Handling data irregularities in classification: Foundations, trends, and future challenges," *Pattern Recognit.*, vol. 81, pp. 674–693, Sep. 2018.
- [25] S. Barreda and T. M. Nearey, "A regression approach to vowel normalization for missing and unbalanced data," *J. Acoust. Soc. Amer.*, vol. 142, no. 4, p. 2583, Oct. 2017.
- [26] L. Hailong, L. Baoan, and L. Xueqiang, "Research on image retrieval algorithm based on deep convolutional neural network," *Appl. Res. Comput.*, vol. 34, no. 12, pp. 3816–3819, 2017.
- [27] Y. Dongri, C. Yue, and L. Yi, "Research on optimizing fault tolerance strategies of storage system for medical big data," *J. China Acad. Electron. Sci.*, vol. 12, no. 5, pp. 546–550, 2017.
- [28] W. Yong, Z. Huiyi, H. Hao, Y. Miao, and K. Wenlong, "Network traffic classification method based on deep convolutional neural networks," *J. Commun.*, vol. 39, no. 1, pp. 14–23, 2018.
- [29] S. Yu, Z. Song, S. Su, W. Li, Y. Wu, and W. Zeng, "A novel dataset generating method for fine-grained vehicle classification with CNN," *Int. J. Database Theory Appl.*, vol. 9, no. 6, pp. 45–52, Jun. 2016.
- [30] Y. Jin, X. Guo, Y. Li, J. Xing, and H. Tian, "Towards stabilizing facial landmark detection and tracking via hierarchical filtering: A new method," *J. Franklin Inst.*, vol. 357, no. 5, pp. 3019–3037, 2020, doi: 10.1016/j.jfranklin.
- [31] X. Fu, J. Chen, S. Deng, J. Wang, and L. Zhang, "Layered virtual machine migration algorithm for network resource balancing in cloud computing," *Frontiers Comput. Sci.*, vol. 12, no. 1, pp. 75–85, Feb. 2018.
- [32] Y. Han, C. Liu, S. Su, M. Zhu, Z. Zhang, and S. Zhang, "A proactive service model facilitating stream data fusion and correlation," *Int. J. Web Services Res.*, vol. 14, no. 3, pp. 1–16, Jul. 2017.
- [33] J. Yu, M. Tan, H. Zhang, D. Tao, and Y. Rui, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2932058.
- [34] H. Gao, W. Huang, Y. Duan, X. Yang, and Q. Zou, "Research on cost-driven services composition in an uncertain environment," *J. Internet Technol.*, vol. 20, no. 3, pp. 755–769, 2019.
- [35] Y. Zhang and S. Liu, "A real-time distributed cluster storage optimization for massive data in Internet of multimedia things," *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 5479–5492, Mar. 2019.
- [36] T. Yang, W. Zhenyu, and Z. Fangjin, "Research on multi-service-oriented stream strategy in heterogeneous wireless networks," *Comput. Eng.*, vol. 444, no. 4, pp. 115–119, 2018.
- [37] L. Qianyu, "Big data classification and mining technology based on fuzzy convolutional neural network," *J. Chongqing Univ. Technol. (Natural Sci.)*, vol. 33, no. 10, pp. 121–126, 2019.
- [38] Y. Yuan, W. Chen, L. Yan, B. Huang, and J. Li, "A similarity-based disease diagnosis system for medical big data," *J. Med. Imag. Health Informat.*, vol. 7, no. 2, pp. 364–370, Apr. 2017.
- [39] K. Fan, W. Jiang, H. Li, and Y. Yang, "Lightweight RFID protocol for medical privacy protection in IoT," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1656–1665, Apr. 2018.
- [40] Y. Genxia, "Medical big data mining algorithm based on association rules," *Microelectron. Comput.*, vol. 36, no. 4, pp. 111–114, 2019.
- [41] Y. Jiang, Z. Huang, and D. H. K. Tsang, "Towards max-min fair resource allocation for stream big data analytics in shared clouds," *IEEE Trans. Big Data*, vol. 4, no. 1, pp. 130–137, Mar. 2018.
- [42] F. Wenjiang, J. Wenhao, and X. Yi, "Design of cellular data offload compensation mechanism based on D2D communication," *Syst. Eng. Electron.*, vol. 39, no. 9, pp. 2119–2124, 2017.
- [43] C. Yueping, Z. Wenpeng, and L. Sen, "Research on differential stream transmission control protocol for data center networks," *J. Xi'an Jiaotong Univ.*, vol. 51, no. 6, pp. 122–128, 2017.
- [44] C. Shengling, S. Siqi, and L. Dongsheng, "Integrated imbalanced learning method based on sample weight update," *Comput. Sci.*, vol. 45, no. 7, pp. 37–43, 2018.



**WEIWEI GAO** was born in 1976. She received the bachelor's degree from the Harbin Institute of Technology, China, and the master's degree from Northeastern University. She is currently working as a Professor with the Wenzhou Business College. Her research interests include artificial intelligence, information security, and big data analysis.



**LI CHEN** received the Ph.D. degree. She is currently an Associate Professor. Her main research interests include machine learning and mobile communications technology. She is a member of the China Computer Federation.



**TAO SHANG** was born in 1968. He received the Ph.D. degree. He is currently a Professor. His research interest is mainly in the area of mechatronic engineering, robotics and intelligent systems, and data mining. He has published dozens of research articles in scholarly journals in the above research areas.