

Received February 7, 2020, accepted February 15, 2020, date of publication February 18, 2020, date of current version February 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974887

Research on Risk Prediction of Dyslipidemia in Steel Workers Based on Recurrent Neural Network and LSTM Neural Network

SHIYUE CUI¹, CHAO LI¹, ZHE CHEN¹, JIAOJIAO WANG¹, AND JUXIANG YUAN^{1,2}

¹School of Public Health, North China University of Science and Technology, Tangshan 063210, China

²Hebei Province Key Laboratory of Occupational Health and Safety for Coal Industry, North China University of Science and Technology, Tangshan 063210, China

Corresponding author: Juxiang Yuan (gwxyjxb@ncst.edu.cn)

This work was supported by the National Key Research ad Development Program of China under Grant 2016YFC0900605.

ABSTRACT With the development of medical digitization technology, artificial intelligence and big data technology, the medical model is gradually changing from treatment-oriented to prevention-oriented. In recent years, with the rise of artificial neural networks, especially deep learning, great achievements have been made in realizing image classification, natural language processing, text processing and other fields. Combining artificial intelligence and big data technology for disease risk prediction is a research focus in the field of intelligent medicine. Blood lipids are the main risk factors of cardiovascular and cerebrovascular diseases. If early prediction of abnormal blood lipids in iron and steel workers can be carried out, early intervention can be carried out, which is beneficial to protect the health of iron and steel workers. This paper around the steel workers dyslipidemia prediction problem for further study, firstly analyzes the influence factors of the steel workers dyslipidemia, discusses the commonly used method for prediction of disease, and then studied deep learning related theory, this paper introduces the two deep learning algorithms of RNN (Recurrent Neural Network) and LSTM (Long Short-Term Memory). Use the basic principle of Python language and the TensorFlow deep learning framework, establishes a prediction model based on two deep learning networks, and makes an example analysis. Experimental results show the LSTM prediction effect is superior to traditional RNN network, It provides scientific basis for the prevention of iron and steel dyslipidemia.

INDEX TERMS Dyslipidemia, risk prediction, deep learning, RNN, LSTM.

I. INTRODUCTION

As the pillar industry of the secondary industry, iron and steel industry has made an indelible contribution in the period of China from agricultural economy to industrial economy. The contribution of front-line steel workers in the processing and production of steel is even more obvious. Therefore, the physical health of front-line steel workers is directly related to the economic benefits of each steel production unit, the financial revenue of the entire city and the comprehensive strength of the country [1]. Along with the advancement of society and technology, the working environment of steel workers has also been greatly improved, which has gradually been transformed from mechanical work to manual operations [2]. However, there are still some jobs that require workers to

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei¹.

be under high temperature conditions and pay attention for a long time to ensure the successful completion of production work, such as the temperature control of molten iron in front of the furnace, the casting machine, etc., and also require workers to concentrate on standing or sitting for a long time in high temperature and noise. Hence, in addition to occupational diseases, there are also a series of chronic non-infectious diseases in the course of work [3], [4]. Dyslipidemia is one of the major risk factors for a variety of chronic non-infectious diseases, and a major cause of stroke and heart disease [5]. A series of physiological reactions will occur in the human body during high-temperature operation, mainly including changes in body temperature regulation, water and salt metabolism, circulatory system, neuroendocrine system, and urinary system. The mechanism of noise on blood lipids and glucose is not very clear, but there are reports suggesting that noise stimulation can not only damage hearing, but also

be introduced into the cerebral cortex and autonomic nervous center through hearing, triggering a series of reactions in the central nervous system. Changes in neuroregulatory functions have led to disturbances in central regulation of vascular movement, resulting in disorders of lipid metabolism [6]. In the noise environment for a long time, the auditory nerve would activate the upward activation system of the brain stem network structure, excite the cerebral cortex, increase the activity of the sympathetic nervous system, and increase the secretion of catecholamines. Epinephrine and norepinephrine can increase the synthesis of light methyl glutaramide coenzyme A reductase in liver, thus promoting the synthesis of cholesterol. High temperature workers tend to be irritable and nervous, which can also increase TC (Total Cholesterol) [7].

The health of steel workers is related to the healthy development of the industry. The prediction of blood lipid and early intervention can effectively reduce the prevalence of cardiovascular disease in steel workers and ensure the health of steel workers and the healthy development of steel enterprises. Therefore, this article focuses on the prediction of blood lipid risk for steel workers. At present, disease prediction mainly uses the traditional machine learning method, which requires the establishment of prediction model for data. However, with the increase of data types and complexity, the establishment of model becomes more difficult, so the deep learning method emerges [8]–[10]. Nowadays, deep learning has been widely accepted and successfully applied to many fields in daily life, such as natural language processing, face recognition, target tracking, video parsing, etc., [11]. In the medical field, there are also successful applications of deep learning, such as image segmentation in medical imaging, based on a large amount of medical image data, to help doctors quickly find disease pathology and perform accurate classification [12], and the use of deep learning for disease risk prediction, based on the medical data, as a diagnostic tool to assist doctors diagnose [13]. Therefore, the application of deep learning in medical big data can mine useful information in medical big data to assist doctors in diagnosis, improve the accuracy of pathological diagnosis and the quality of medical service.

Neural network experts Jordan, Pineda, Williams, Elman are equivalent to a RNN (Recurrent Neural Network) proposed in the late 1980s [14]. The essential characteristic of this kind of network is that there are both internal feedback connections and feedforward connections between processing units. It is a feedback dynamic system, which reflects the dynamic characteristics of the process in the calculation process and has stronger dynamic behavior and computing capacity than the feedforward neural network. Thus, it has become one of the important research objects of neural network experts in the world. In 1997 Sepp Hochreiter and Jurgen Schmidhuber proposed the LSTM network (Long Short-Term Memory) [15], which is a special RNN network. This article introduces the basic principles of two deep learning algorithms: RNN and LSTM. By using Python language and TensorFlow deep learning framework, a prediction model based on two kinds of deep learning networks is built, and an

example is analyzed and studied to compare the prediction effects of the two models.

II. ANALYSIS OF INFLUENCING FACTORS

OF DYSLIPIDEMIA

Cholesterol (TC) and triglyceride (TG) are the main lipid components in plasma, accounting for about one third of the total plasma fat. Most cholesterol is synthesized by the body, and a fraction of it comes from food. Triglycerides, which account for about a quarter of the total plasma fat, are obtained mostly from food and a small part from the body. Phospholipids (PL) are mainly lecithin, cephalin, serine phospholipids, neurolipids, etc., accounting for about one third of the total plasma fat. Free fatty acid (FFA), also known as non-esterified fatty acid, accounts for about 5% to 10% of total plasma lipids. It is the main source of energy for the body [16].

Blood lipid is an important indicator for monitoring health status. Dyslipidemia, also known as hyperlipidemia, is caused by an increase in total cholesterol and triglyceride levels in the plasma and a decrease in low-density lipoprotein. It may cause complications such as hypertension, fatty liver, and atherosclerosis. If the blood lipid level of steel workers can be accurately predicted and intervened in advance, the occurrence of these conditions will be greatly reduced. At present, the detection of dyslipidemia has been realized medically. The data of TC, TG, high-density protein (hdl-c), low-density protein (ldl-c) and other data can be collected to check the level of dyslipidemia in workers and to intervene in workers with dyslipidemia [17]. However, if workers do not undergo frequent physical examinations, it cannot make early predictions, leading to blood lipid accumulation in the body, which will undoubtedly have a great impact on the health of steel workers. Recent studies have shown that behavior-based dyslipidemia prediction and early intervention are cost-effective interventions to reduce cardiovascular morbidity [18].

Dyslipidemia, a few are systemic diseases, and most are the result of interactions between genetic defects and environmental factors. Studies have shown that genetic and environmental factors play an important role in the etiology of dyslipidemia. It can be a component of the metabolic syndrome and is closely related to a variety of diseases such as obesity, type 2 diabetes, hypertension, coronary heart disease, and stroke. Long-term dyslipidemia can lead to atherosclerosis and increase the morbidity and mortality of cardiovascular vessels. With the improvement of living standards and lifestyle changes, the number of dyslipidemia patients has increased significantly [19]. According to reports, the prevalence of dyslipidemia in Chinese adults is about 18.6%, and it is estimated that the number of patients can be as high as 160 million. Therefore, preventing and treating dyslipidemia is of great significance for prolonging life and improving quality of life [20], [21]. Prediction of blood lipids and early intervention can effectively reduce the prevalence of cardiovascular and cerebrovascular diseases in steel workers and ensure the

health of steel workers and the healthy development of steel companies.

III. SAMPLE SET CONSTRUCTION AND DATA ANALYSIS

A. THE RESEARCH OBJECT

In order to better study the influence of working environment in steel mills on the lipid status of workers, this paper selected 4000 male workers who worked in the stainless steel, hot rolling department, power department, cold rolling department and maintenance department of an iron and steel group for more than one year as the research object. The age, height, body quality, length of service, marital status, education level, economic level, alcohol consumption, smoking, workshop, type of work, shift type and exposure to high temperature and noise in the occupational environment were investigated by questionnaire.

B. SAMPLE COLLECTION METHOD

The steel workers under test had fasted overnight or for at least 12 hours. Then in the morning, their 2 to 3 ml of forearm venous blood would be collected and centrifuged for 10 minutes (1500 r / min). The serum was measured on the same day. Total cholesterol and triglyceride were digested by enzyme method. Apolipoprotein A, apolipoprotein B, high-density lipoprotein and low-density lipoprotein were determined by direct method. The instrument test parameters were set in strict accordance with the specification of each indicator. Before testing the serum samples, the laboratory quality control should be done first. And the samples can be tested after the quality control was qualified. The instrument is a Hitachi-7020 automatic biochemical analyzer, with $TC > 6.2 \text{ mmol}$ or $TG > 2.3 \text{ mmol}$ as anomalies [22]. In this study, smoking was defined as smoking 1 > a day for 6 months or more. Drinking referred to those who drink more than once a week for more than 6 consecutive months. Tea drinking referred to those who drink tea more than once a week for more than 6 consecutive months. Body mass index (BMI) = body mass (kg) / height (m) 2. BMI < 24.0 is normal, $24.0 < \text{BMI} < 28.0$ is overweight, and $\text{BMI} > 28.0$ is obesity [23]. The statistical results are shown in the table1:

From the table 1, it can be seen that the rate of dyslipidemia in workers exposed to high temperature, high noise and long-term shift work is significantly higher than that of other workers, and the age, height, weight, length of service, marital status, education level, economic level, alcohol consumption, smoking and other factors of workers also have a certain impact on the blood lipid. Therefore, 10 factors, including worker age, BMI, marital status, education level, family income, drinking, smoking, exposure to high temperature, noise, and shift work, were selected as independent variables and stored in the database, and brought into RNN and LSTM two deep prediction networks. The values of TC and TG in the blood were taken as the output to construct the data sample set. Training is performed in the model to achieve early prediction of blood lipid abnormalities in steel workers.

TABLE 1. Analysis of influencing factors of blood lipid abnormality in steel workers.

| group | number | t ratio (%) | constituen | | TC | | TG | |
|---|--------|-------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|
| | | | Abnormal number | abnormal rate (%) | Abnormal number | abnormal rate (%) | Abnormal number | abnormal rate (%) |
| age | | | | | | | | |
| <30 | 350 | 8.8 | 50 | 14.3 | 101 | 28.9 | | |
| 30-40 | 1280 | 32.0 | 260 | 20.3 | 542 | 42.3 | | |
| 40-50 | 1621 | 40.5 | 390 | 24.1 | 691 | 42.6 | | |
| More than 50 | 749 | 18.7 | 215 | 28.7 | 352 | 47.0 | | |
| standard of culture | | | | | | | | |
| Junior high school and below | 769 | 19.2 | 238 | 30.9 | 385 | 50.1 | | |
| High school or junior college | 1726 | 43.2 | 351 | 20.3 | 574 | 33.3 | | |
| College degree or above | 1505 | 37.6 | 278 | 18.5 | 359 | 23.9 | | |
| Marital status | | | | | | | | |
| unmarried | 208 | 5.2 | 25 | 12.0 | 59 | 28.4 | | |
| | 3792 | 94.8 | 843 | 22.2 | 1599 | 42.2 | | |
| Per capita household income (yuan) | | | | | | | | |
| | 794 | 19.9 | 181 | 22.8 | 318 | 40.1 | | |
| | 1126 | 28.2 | 299 | 26.6 | 486 | 43.2 | | |
| | 2080 | 52.0 | 427 | 20.5 | 842 | 40.5 | | |
| BMI | | | | | | | | |
| normal | 1433 | 35.8 | 296 | 20.7 | 397 | 27.7 | | |
| overweight | 1789 | 44.7 | 419 | 23.4 | 768 | 42.9 | | |
| obesity | 778 | 19.5 | 182 | 23.4 | 445 | 57.2 | | |
| smoking | | | | | | | | |
| no | 1628 | 40.7 | 323 | 19.8 | 582 | 35.7 | | |
| yes | 2372 | 59.3 | 560 | 23.6 | 1203 | 50.7 | | |
| drinking | | | | | | | | |
| no | 2516 | 62.9 | 462 | 18.4 | 967 | 38.4 | | |
| yes | 1484 | 37.1 | 408 | 27.5 | 668 | 45.0 | | |
| High temperature contact | | | | | | | | |
| no | 2110 | 52.8 | 310 | 14.7 | 360 | 17.1 | | |
| yes | 1890 | 47.3 | 560 | 29.6 | 795 | 42.1 | | |
| Noise exposure | | | | | | | | |
| no | 1945 | 48.6 | 440 | 22.6 | 322 | 16.6 | | |
| yes | 2065 | 51.6 | 760 | 36.8 | 836 | 40.5 | | |
| Shift work | | | | | | | | |
| no | 817 | 20.4 | 161 | 19.7 | 203 | 24.8 | | |
| yes | 3183 | 79.6 | 890 | 28.0 | 1324 | 41.6 | | |

IV. LSTM NEURAL NETWORK

Deep learning is a new research direction in the field of artificial intelligence. It is developed on the basis of shallow neural networks with the improvement of computer hardware

levels and the explosive growth of the current data volume. Deep learning and shallow neural network structure are both layered. Each layer will process the data input to the model and combine low-level features into potential high-level features by learning data rules. Compared with shallow models, deep learning can express complex high dimensionality such as high-variable functions and find the true relationships within the original data better. In the 1980s, artificial neural network back propagation algorithm was born. This method can automatically learn data rules from a large amount of training data without manual intervention. At present, deep learning is the most concerned research direction in the field of artificial intelligence, which completely subverts the shallow model in traditional machine, proposes a deep learning network model, and elevates it to a new height from theory to application [24]–[27]. CNN (convolutional neural network) and RNN are two types of classical deep learning network structures now.

A. RNN MODEL STRUCTURE

Deep neural networks is a network with automatic adjustment of network parameters, which can iteratively calculate data according to the set coordinates and models. The training process of deep learning model is actually a process of constantly tuning the ownership values of nodes which are all used as tools to describe data features. The key to whether the model can describe the features of things lies on the final training results of each weight. The deep neural network takes the neural network as the carrier and focuses on the depth. It can be said to be a general term, including the recurrent neural network with multiple hidden layers, the all-connected network and the convolutional neural network. Recurrent neural networks are mainly used for sequence data processing, and have a certain memory effect. The long-term and short-term memory networks derived from them are better at processing long-term dependencies. Convolutional neural networks focus on spatial mapping. And image data is particularly suitable for feature extraction of various networks. When the input data is dependent and sequential, the results of CNN are generally not good. There is no correlation between the previous input of CNN and the next input. The RNN network appeared in the 1980s. It is designed a different number of hidden layers. Each hidden layer stores information and selectively forgets some information. In this way, the data characteristics of the sequence changes of the data can be extracted. RNN has not only achieved many results in the fields of text processing and speech processing, but also been widely used in the fields of speech recognition, machine translation, text generation, sentiment analysis, and video behavior recognition. Therefore, this paper will use RNN modeling to predict the risk of abnormal blood lipids in steel workers.

RNN is good at processing time series data, and can describe the context of the data on the time axis. The RNN structure is shown in the Fig.1.

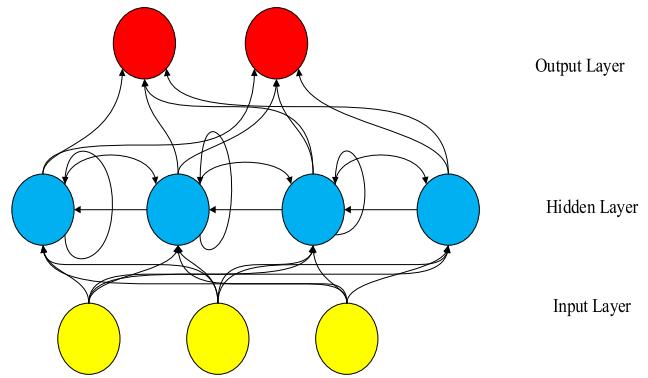


FIGURE 1. Mine phase diagram recognition classification model flow chart.

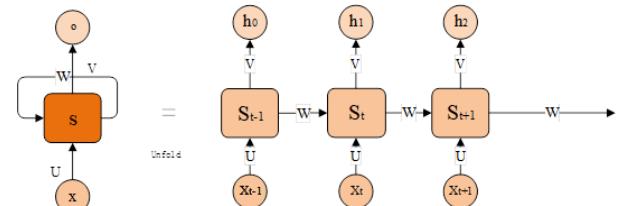


FIGURE 2. Hierarchical expansion diagram.

As can be seen from the figure above, the RNN structure is relatively simple. It mainly consists of an *Input Layer*, a *Hidden Layer*, an *Output Layer*, and an arrow in the *Hidden Layer* represents the cyclic update of data, which is the method to realize the time memory function. The input levels of this paper were: age, BMI, marital status, education level, family income, alcohol consumption, smoking, exposure to high temperature, noise, shift work. The 10-dimensional data were normalized and input into the RNN model. After the extraction of hidden layer depth features, the output layer output the sequence of lipid health status, in which 1 represented normal lipid status and 0 represented abnormal lipid status.

1) FORWARD PROPAGATION OF RNN

Fig. 2 shows the hierarchical expansion of the *Hidden Layer*. $T - 1, t, t + 1$ represent the time series. X represents the input sample. S_t represents the memory of the sample at time t . W represents the weight of the input. U indicates the weight of the input sample at this moment, and V indicates the weight of the output sample.

$$s_t = f(W * s_{t-1} + U * x_t) \quad (1)$$

At time $t = 1$, generally initialize the input, randomly initialize W , U , V , and perform the following formula calculation:

$$h_1 = Ux_1 + Ws_0 \quad (2)$$

$$s_1 = f(h_1) \quad (3)$$

$$o_1 = g(Vs_1) \quad (4)$$

Among them, f and g are activation functions, where f can be activation functions such as tanh, relu, sigmoid.

G is usually a softmax function. Time advancing, the state s_1 as the memory state of time l will participate in the prediction activity of the next time, that is:

$$h_2 = Ux_2 + Ws_1 \quad (5)$$

$$s_2 = f(h_2) \quad (6)$$

$$o_2 = g(Vs_2) \quad (7)$$

And so on, the final output value can be obtained as:

$$h_t = Ux_t + Ws_{t-1} \quad (8)$$

$$s_t = f(h_t) \quad (9)$$

$$o_t = g(Vs_t) \quad (10)$$

Here, W , U , and V are equal at every moment which is weight sharing.

2) BACK PROPAGATION OF RNN

The back propagation process of RNN is the updating process of weight parameters W , U , and V . Each output value o_t will produce an error value E_t , then the total error value can be expressed as: Since the output of each step depends on not only the network of the current step but also the state of the previous steps, the Backpropagation Through Time (BPTT) algorithm is used to pass the error value at the output back, and the gradient descent method is used to perform the weight update.

$$\nabla U = \frac{\partial E}{\partial U} = \sum_t \frac{\partial e_t}{\partial U} \quad (11)$$

$$\nabla V = \frac{\partial E}{\partial V} = \sum_t \frac{\partial e_t}{\partial V} \quad (12)$$

$$\nabla W = \frac{\partial E}{\partial W} = \sum_t \frac{\partial e_t}{\partial W} \quad (13)$$

The model training process of RNN is shown in figure 3.

B. LSTM MODEL STRUCTURE

Because there are connections between neurons in the RNN layer, the network can learn the change law of sequence data before and after, and the internal sequence rules of data is easy to be mined. Thus RNN is widely used in the field of sequence data processing such as speech recognition and machine translation. However, this structure also has some problems. When data is transmitted backward, the problem of gradient disappearance or gradient explosion is unavoidable, which limits its processing of long-term dependencies. The LSTM network changes the way of gradient transmission during backpropagation by adding multiple special computing nodes in the hidden layer of RNN, which effectively slows the problem of gradient disappearance or gradient explosion. Its model structure is shown in figure 4.

Where h_{t-1} represents the output of the previous cell, and x_t represents the input of the current cell. σ represents the sigmoid function. The difference between LSTM and RNN is that it adds a “processor” to the algorithm to determine the

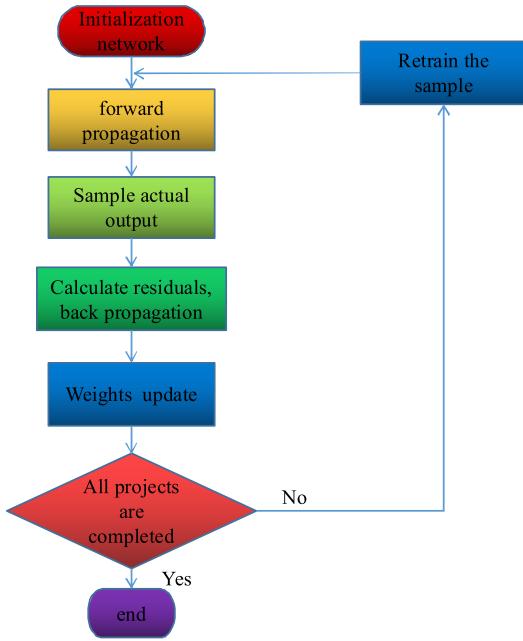


FIGURE 3. RNN training process.

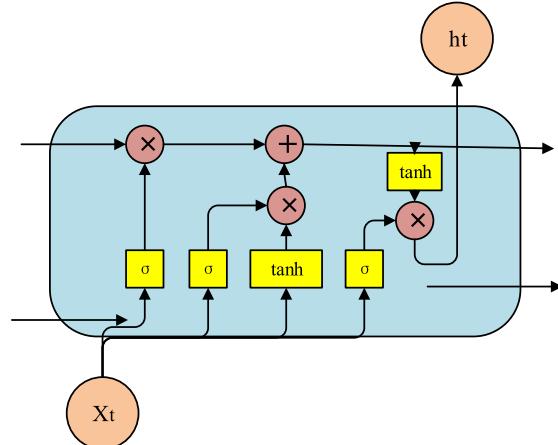


FIGURE 4. LSTM model structure.

usefulness of the information. The structure of this processor is called a *cell*. Three gates are placed in a *cell*, which are called *Input gate*, *Forget gate*, and *Output gate*. A piece of information enters the LSTM network, and it can be judged whether it is useful according to the rules. Only the information that meets the algorithm authentication will be left, and the non-conforming information will be forgotten through the *Forget gate* [28].

1) FORGET GATE

The first step for data entering the LSTM is to decide what information should be lost and what retained. This decision is made by the *Forget gate*, which reads h and x and outputs a value between 0 and 1, where 1 means “complete reserved”, 0 means “completely discarded”. *Forget gate* is calculated

as:

$$f_t = \sigma (W_f * [h_{t-1}, x_t] + b_f) \quad (14)$$

In the formula, f_t is the calculation result of the *Forget gate* which is mainly used to control the retention of the information transmitted from the unit state at the previous moment to the unit state at the current moment. $[]$ indicates that the two vectors are spliced, h_{t-1} is the output of the unit at the previous moment, and are the weight and bias of *Forget gate*, W_f and b_f are Sigmoid activation functions.

2) INPUT GATE

Input gate determines the addition of new information, and its operation process includes sigmoid layer and tanh layer. The sigmoid layer determines the information that needs to be updated. The calculation formula is:

$$i_t = \sigma (W_i * [h_{t-1}, x_t] + b_i) \quad (15)$$

In the formula, i_t is the calculation result of the input gate, and the input gate also has independent weight and bias.

The role of the tanh layer is to generate a vector of candidate update information. Its calculation formula is:

$$\tilde{C}_t = \tanh (W_c * [h_{t-1}, x_t] + b_c) \quad (16)$$

\tilde{C}_t is the unit state of the current input, the unit state of the current moment is C_t , and its calculation formula is:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (17)$$

3) OUTPUT GATE

Output gate is roughly the same as the *Input gate*, and its operation flow includes sigmoid layer and tanh layer. The sigmoid layer determines the output part of the information, and the calculation formula is:

$$o_t = \sigma (W_o * [h_{t-1}, x_t] + b_o) \quad (18)$$

Finally get the output of the current moment h_t :

$$h_t = o_t * \tanh (c_t) \quad (19)$$

The forward propagation of LSTM calculates the cell state C_t and h_t the output of the current moment, and completes the forward propagation calculation of the network. The back-propagation of LSTM is similar to the back-propagation principle of RNN. Finally, the weights and biases of all parts of the network are updated to complete the model training.

V. CASE ANALYSIS

The essence of dyslipidemia risk prediction based on physical indicators and working environment factor parameters is to build a model that reflects a mapping relationship, that is, the mapping relationship between the probabilities of dyslipidemia of workers in the future and workers' physical indicators, as well as working-environment factor parameters. This mapping relationship changes over time. The RNN model is characterized in that it can use its internal memory to process input sequences of arbitrary timing, with internal

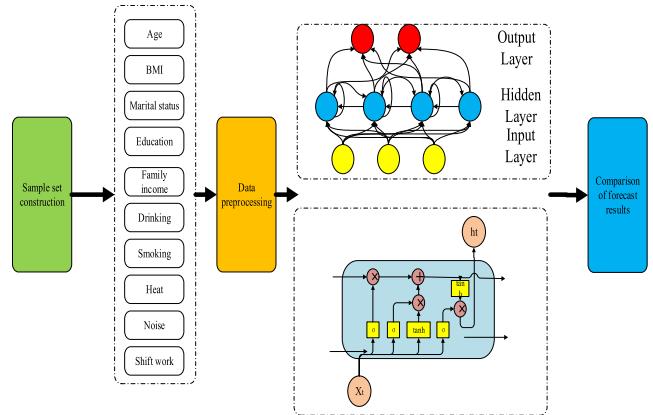


FIGURE 5. Risk prediction model for dyslipidemia based on RNN and LSTM.

feedback connections and feedforward connections between its processing units. Compared with the traditional prediction method, the input data of RNN adds the change factor of time, thus achieving better prediction effect. Dyslipidemia is a chronic disease caused by long-term poor diet, living habits and the environment, which is a problem in this category. In summary, this paper proposes a risk prediction model for dyslipidemia based on RNN and LSTM as shown in Fig.5.

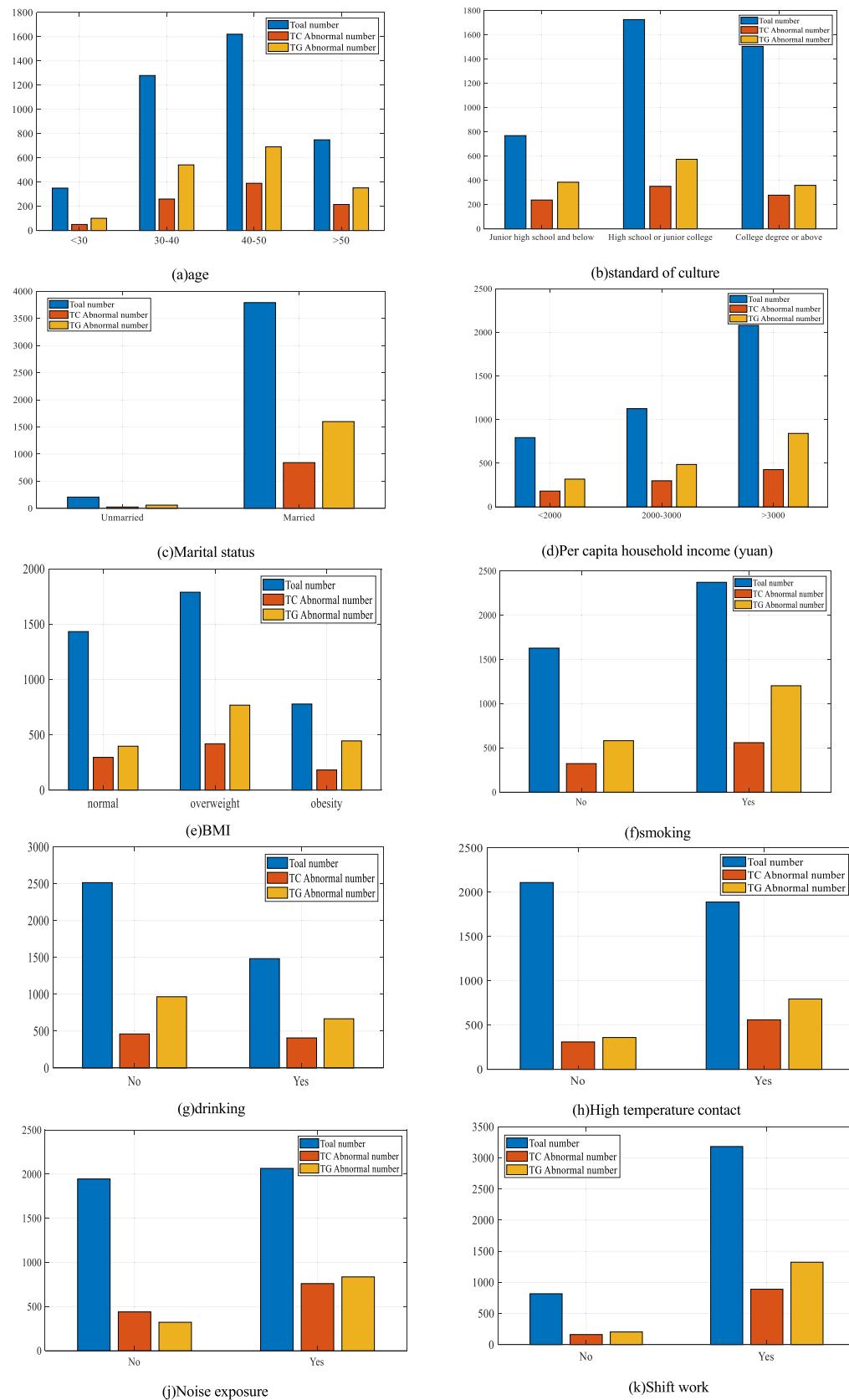
This article uses 4,000 selected steel workers as experimental samples, of which 3,000 are training sets and 1,000 are test sets. The 10-dimensional data of selected workers' age, BMI, marital status, education level, family income, alcohol consumption, smoking, exposure to high temperature, noise, and shift work are stored in the database and normalized. Fig. 6 shows the sample distribution of each indicator.

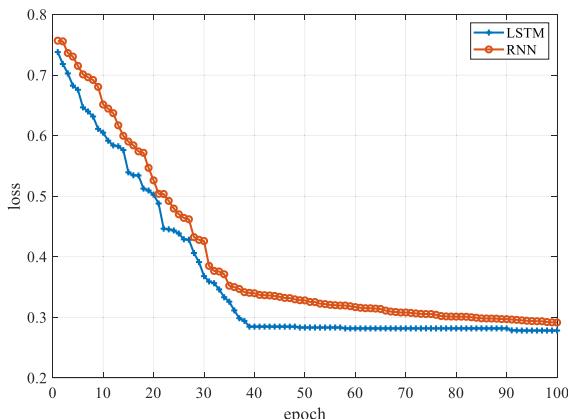
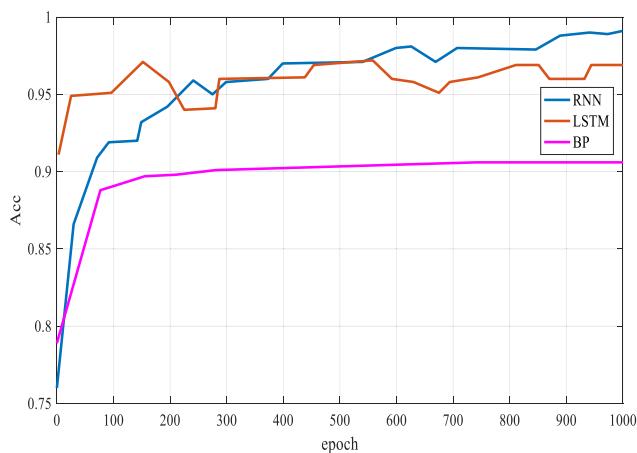
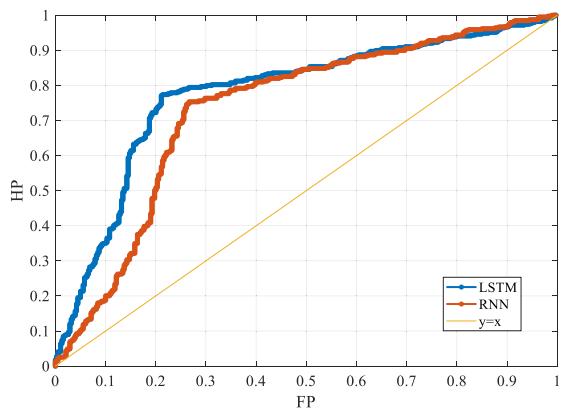
Enter the RNN model and LSTM model for training and prediction, respectively. The change curve of the RNN and LTSM loss function is shown in Fig.7. LSTM model converges faster than RNN model in the training data set, with smaller loss function values and higher precision.

As being compared the prediction accuracy of the two models is, the results are shown in the figure 8. We can see from the Fig.8 that the accuracy of the LTSM model is basically higher than that of the RNN model, and the accuracy can be maintained above 95%. Therefore, better prediction results can be obtained by applying RNN model to the prediction of dyslipidemia risk in steelworkers.

In order to test the accuracy of the model better, BP algorithm is used to compare with RNN and LSTM models in this paper, and the comparison results are shown in figure 8.

ROC (Receiver Operating Characteristic Curve) is a component method that reflects the sensitivity and specificity of continuous variables and reveals the relationship between them. It calculates a series of sensitivity and specificity by setting some different thresholds on continuous variables. With sensitivity on the ordinate and specificity on the abscissa, the larger the curve is at the bottom of the area, the higher the diagnostic accuracy is. On the ROC curve, the point closest to the upper left of the coordinate graph is the critical value with

**FIGURE 6. Histogram of sample index distribution.**

**FIGURE 7.** Loss function change of RNN and LSTM.**FIGURE 8.** Comparison chart of model accuracy.**FIGURE 9.** ROC curve of LSTM and RNN models.

high sensitivity and specificity. The ROC curves of RNN and LSTM are shown in figure 9.

In Fig.9, the area of the ROC curve of LSTM is larger than that of RNN model. The area of the ROC model of LSTM is 0.895, and the ROC area of RNN model is 0.853, which is slightly lower than that of LSTM model. From the perspective of ROC curve, the prediction performance of LSTM model is better than that of RNN model.

VI. CONCLUSION

In this paper, a risk prediction model for dyslipidemia in steel workers based on RNN and LSTM networks was established. A survey was conducted on the long-term living habits and working environment of 4000 workers in a steel enterprise with ten key factors influencing dyslipidemia being extracted. Two prediction networks, RNN and LSTM, were established to predict the risk of dyslipidemia in steel workers. Experimental results showed that the prediction effect of LSTM is significantly better than that of traditional RNN network, with an accuracy of more than 95%. Disease risk prediction refers to the discovery of potential risks and trends of diseases, which plays an important role in the prevention, intervention and management of diseases. In medicine, the ideal goal of disease risk prediction is to find the potential risks and trends of diseases before doctors diagnose diseases, and take effective measures to prevent and intervene diseases. The work in this paper can better predict the risk of dyslipidemia in steelworkers, provide a scientific basis for protecting the health of steelworkers, and expand the application scope of deep learning theory in the field of medicine. However, due to time constraints, we fail to obtain more sample data, which may reduce the robustness of the model. This content will be gradually improved in the future work.

REFERENCES

- [1] X. Xiao, S. Le Berre, D. G. Fobar, M. Burger, P. J. Skrodzki, K. C. Hartig, A. T. Motta, and I. Jovanovic, "Measurement of chlorine concentration on steel surfaces via fiber-optic laser-induced breakdown spectroscopy in double-pulse configuration," *Spectrochimica Acta B, Atomic Spectrosc.*, vol. 141, pp. 44–52, Mar. 2018.
- [2] Z. Huang, L. Li, G. Ma, and L. C. Xu, "Local information, and commanding heights: Decentralizing state-owned enterprises in China," *Amer. Econ. Rev.*, vol. 107, no. 8, pp. 2455–2488, 2017.
- [3] Z. An, F. Shan, L. Yang, R. Shen, X. Gu, F. Zheng, and Y. Zhang, "Unusual effect of temperature on direct fluorination of high temperature vulcanized silicone rubber and properties of the fluorinated surface layers," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 25, no. 1, pp. 190–198, Feb. 2018.
- [4] S. M. Kopeikin, "Millisecond and binary pulsars as nature's frequency standards - II. The effects of low-frequency timing noise on residuals and measured parameters," *Monthly Notices Roy. Astronomical Soc.*, vol. 305, no. 3, pp. 563–590, May 1999.
- [5] H. Tada, M.-A. Kawashiri, and M. Yamagishi, "Comprehensive genotyping in dyslipidemia: Mendelian dyslipidemias caused by rare variants and mendelian randomization studies using common variants," *J. Hum. Genet.*, vol. 62, no. 4, pp. 453–458, Jan. 2017.
- [6] K. N. Burns, S. K. Sayler, and R. L. Neitzel, "Stress, health, noise exposures, and injuries among electronic waste recycling workers in ghana," *J. Occupational Med. Toxicol.*, vol. 14, no. 1, Jan. 2019, pp. 77–80.
- [7] S. Hur, B. Min, K. Nam, E. Lee, and D. Ahn, "Effect of dietary cholesterol and cholesterol oxides on blood cholesterol, lipids, and the development of atherosclerosis in rabbits," *Int. J. Mol. Sci.*, vol. 14, no. 6, pp. 12593–12606, Jun. 2013.
- [8] P. Schulam and S. Saria, "A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1145–1170.
- [9] J.-H. Wu, W. Wei, L. Zhang, J. Wang, R. Damasevicius, J. Li, H.-D. Wang, G.-L. Wang, X. Zhang, J.-X. Yuan, and M. Wozniak, "Risk assessment of hypertension in steel workers based on LVQ and Fisher-SVM deep excavation," *IEEE Access*, vol. 7, pp. 23109–23119, 2019.
- [10] N. An, H. Ding, and J. Yang, "Deep ensemble learning for Alzheimers disease classification," *Mach. Learn.*, vol. 5, pp. 583–590, May 2019.
- [11] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.

- [12] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: A deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018.
- [13] A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, Z. Zhou, P. Gong, and C. Zhang, "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC Bioinf.*, vol. 18, no. S14, pp. 523–525, Dec. 2017.
- [14] K. K. Sarma and A. Mitra, "A class of recurrent neural network (RNN) architectures with SOM for estimating MIMO channels," in *Advances in Computing and Communications (CCIS)*, vol. 192. Berlin, Germany: Springer, 2011.
- [15] G. Jain, M. Sharma, and B. Agarwal, "Optimizing semantic LSTM for spam detection," *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 239–250, Apr. 2018.
- [16] O. A. Ismaiel, R. G. Jenkins, and H. Thomas Karnes, "Investigation of endogenous blood lipids components that contribute to matrix effects in dried blood spot samples by liquid chromatography-tandem mass spectrometry," *Drug Test. Anal.*, vol. 5, no. 8, pp. 710–715, Oct. 2012.
- [17] A. Mohammadbeigi, E. Moshiri, N. Mohammadsalehi, H. Ansari, and A. Ahmadi, "Dyslipidemia prevalence in iranian adult men: The impact of population-based screening on the detection of undiagnosed patients," *World J. Men's Health*, vol. 33, no. 3, p. 167, 2015.
- [18] J. S. Khan, J. M. Hah, and S. C. Mackey, "Effects of smoking on patients with chronic pain: A propensity-weighted analysis on the collaborative health outcomes information registry," *Pain*, vol. 160, no. 10, pp. 2374–2379, Oct. 2019.
- [19] P. Bramlage, S. T. Azar, O. Okkeh, P. Brudi, B. M. Ambegaonkar, H. A. Hantash, S. Jambari, M. El Zaheri, R. Rachoin, A. Chafoun, and L. Lahoud, "Factors influencing dyslipidemia in statin-treated patients in Lebanon and Jordan: Results of the dyslipidemia international study," *Vascular Health Risk Manage.*, vol. 10, p. 225, Apr. 2014.
- [20] Y. Jiao, G. Li, Y. Xing, D. Nie, and X. Liu, "Influencing factors of hemorrhagic transformation in non-thrombolysis patients with cerebral infarction," *Clin. Neurol. Neurosurg.*, vol. 181, pp. 68–72, Jun. 2019.
- [21] N. R. Balian, C. B. Alonso, and M. C. Zurrú, "Clinical predictors of hemorrhagic transformation in non lacunar ischemic stroke," *Medicina*, vol. 77, no. 2, pp. 100–104, 2017.
- [22] Y. L. Guo and L. I. Jian-Jun, "Highlights of the guidelines for prevention and treatment of dyslipidemia in Chinese adults (revised edition 2016)," *Chin. J. Frontiers Med. Sci.*, vol. 31, pp. 937–953, Jul. 2017.
- [23] N. Demirel, S. Özbay, and F. Kaya, "The effects of aerobic and anaerobic training programs applied to elite wrestlers on body mass index (BMI) and blood lipids," *J. Edu. Training Stud.*, vol. 6, no. 4, p. 58, Mar. 2018.
- [24] X. Yu, H. Lu, and D. Wu, "Development of deep learning method for predicting firmness and soluble solid content of postharvest korla fragrant pear using Vis/NIR hyperspectral reflectance imaging," *Postharvest Biol. Technol.*, vol. 141, pp. 39–49, Jul. 2018.
- [25] K. Cui and X. Qin, "Virtual reality research of the dynamic characteristics of soft soil under metro vibration loads based on BP neural networks," *Neural Comput. Appl.*, vol. 29, no. 5, pp. 1233–1242, Jan. 2017.
- [26] Y. Sun, J. Xu, H. Qiang, and G. Lin, "Adaptive neural-fuzzy robust position control scheme for maglev train systems with experimental verification," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8589–8599, Nov. 2019.
- [27] S. Hussain, J. Keung, A. A. Khan, A. Ahmad, S. Cuomo, F. Piccioli, G. Jeon, and A. Akhunzada, "Implications of deep learning for the automation of design patterns organization," *J. Parallel Distrib. Comput.*, vol. 117, pp. 256–266, Jul. 2018.
- [28] D. Wang, J. Fan, H. Fu, and B. Zhang, "Research on optimization of big data construction engineering quality management based on RNN-LSTM," *Complexity*, vol. 2018, pp. 1–16, Jul. 2018.



SHIYUE CUI was born in Baoding, Hebei, China, in 1990. She graduated from the North China University of Science and Technology, where she is currently pursuing the Ph.D. degree in public health and preventive medicine with the School of Public Health. Her main research is on occupational epidemiology.



CHAO LI was born in Lvliang, Shanxi, China, in 1996. She graduated from Hainan Medical University. She is currently a Graduate Student in public health and preventive medicine with the School of Public Health, North China University of Science and Technology. Her main research is on occupational epidemiology.



ZHE CHEN was born in Jining, Shandong, China, in 1998. She graduated from Jining Medical University. She is currently a Graduate Student in public health and preventive medicine with the School of Public Health, North China University of Science and Technology. Her main research is on occupational epidemiology.



JIAOJIAO WANG was born in Zhangjiakou, Hebei, China, in 1994. She graduated from Hebei North University. She is currently a Graduate Student in public health and preventive medicine with the School of Public Health, North China University of Science and Technology. Her main research is on occupational epidemiology.



JUXIANG YUAN graduated from the Tangshan Coal Mine Medical College, in 1978. He received the Diploma degree in occupational health and safety from the McMaster University Medical Center, Canada, in 1985, and the Ph.D. degree from the Peking Union Medical College, in 2000. He is currently a Professor with the North China University of Science and Technology. His main research interest is on occupational epidemiology.