

# An Application of Deep Belief Networks in Early Warning for Cerebrovascular Disease Risk

Qiuli Qin, Beijing Jiaotong University, China

Xing Yang, China Unicom Research Institute, China

Runtong Zhang, Beijing Jiaotong University, China

Manlu Liu, Rochester Institute of Technology, USA

Yuhan Ma, Beijing Jiaotong University, China

## ABSTRACT

To reduce the incidence of cerebrovascular disease and mortality, identifying the risks of cerebrovascular disease in advance and taking certain preventive measures are significant. This article was aimed to investigate the risk factors of cerebrovascular disease (CVD) in the primary prevention and to build an early warning model based on the existing technology. The authors use the information entropy algorithm of rough set theory to establish the index system suitable for the early warning model. Then, using the limited Boltzmann machine and direction propagation algorithm, the depth trust network is established by building and stacking RBM, and the back propagation is used to fine-tune the parameters of the network at the top layer. Compared with the LM-BP early-warning model, the deep confidence network model is more effective than traditional artificial neural network, which can help to identify the risk of cerebrovascular disease in advance and promote the primary prevention.

## KEYWORDS

Cerebrovascular Diseases (CVD), Deep Beliefs Network, Deep Learning, Early Warning, Index System, Neural Network, Prevention, Rough Set Theory

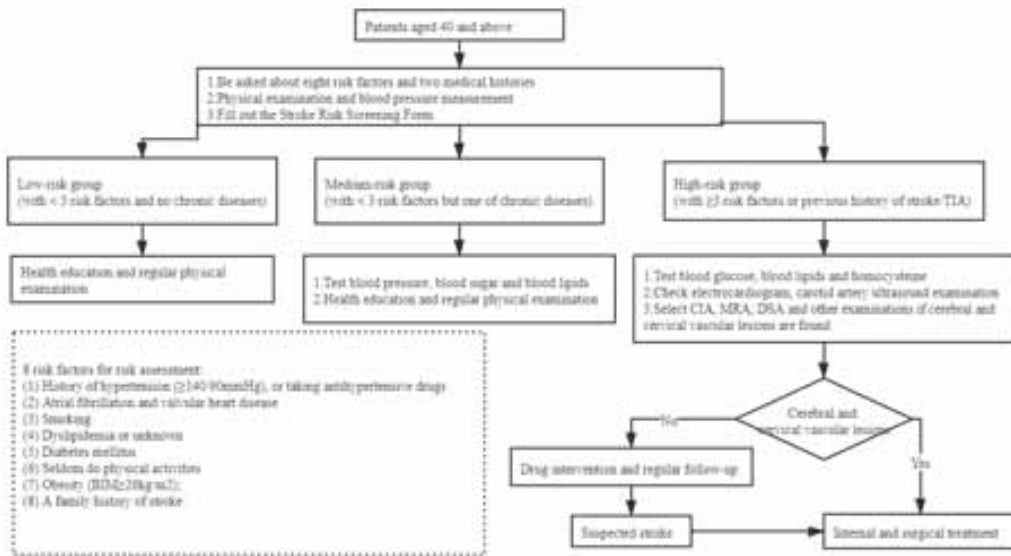
## INTRODUCTION

Cerebrovascular disease (CVD) has the characteristics of high mortality and high morbidity. In advance to identify the risk of CVD, taking certain preventive and control measures can reduce the incidence and mortality of CVD. Diagnosing CVD is a complex task, but the process is commonly based on the experience and assumptions of doctors. Prakash and Karthikeyan (2021) mentioned that it is possible to make misjudgments because of the doctors' subjective diagnosis and machine learning technology could become the universal language in emergencies to effectively predict the risk. The diagnosis of CVD has the following problems: (1) The discovery of the disease is too late to do the early detection and treatment; (2) The traditional screening process of CVD risk is not meticulous enough with eight risk factors and the boundary of classification is not obvious, as shown in Figure 1. To optimize the traditional screening process, the risk early warning model is necessary which could select relatively complete indicators and detect the possibility of CVD timely.

DOI: 10.4018/JOEUC.287574

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Figure 1. The Traditional Screening Process of a high-risk group of CVD



In this article, the authors use the information entropy algorithm of rough set theory to reduce the initial index system, extract high generalization and high sensitivity indexes, and finally establish the index system suitable for an early warning model. Then, using the limited Boltzmann machine and direction propagation algorithm, the depth trust network is established by stacking Restricted Boltzmann Machine (RBM), and the Back Propagation (BP) is used to fine-tune the parameters of the network at the top layer.

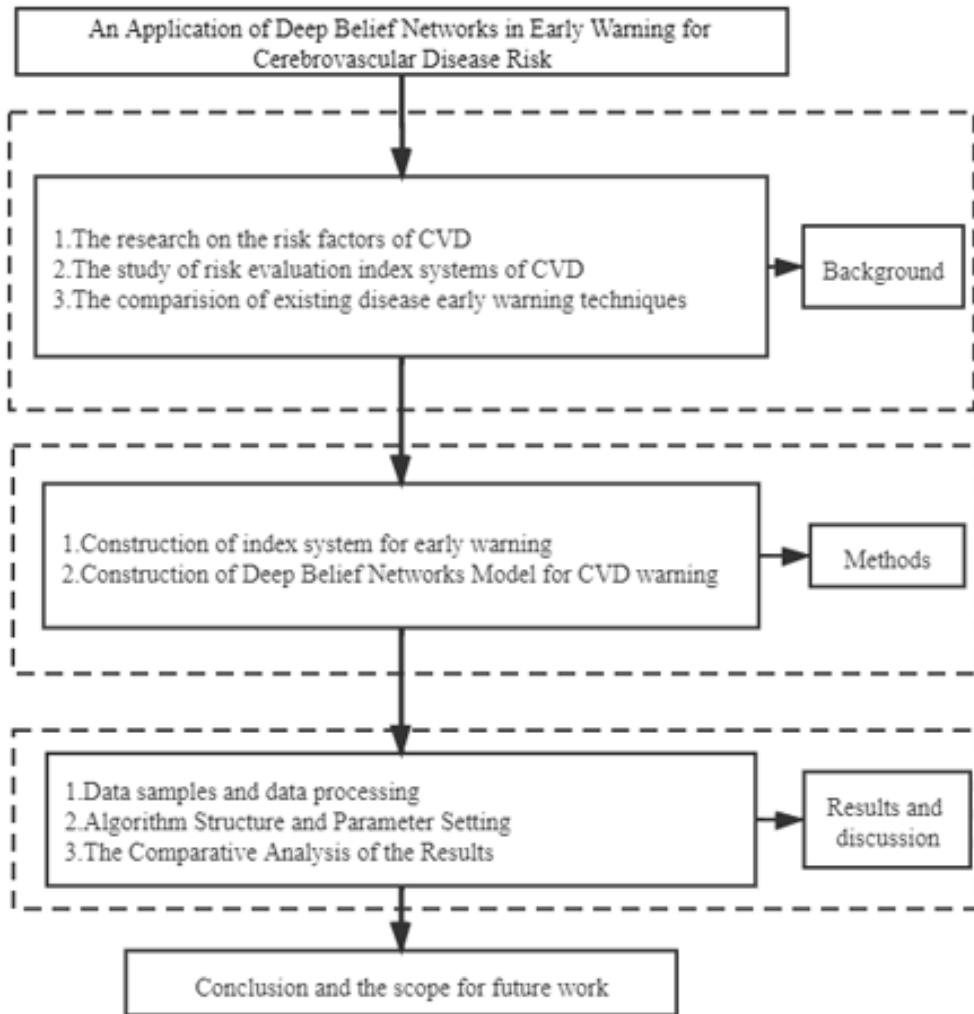
To analyze its applicability, the deep confidence network model was evaluated by the simulation data, and the results were compared with the results of the LM-BP early-warning model. It turned out that the early warning model of CVD risk based on a deep confidence network is more effective than a traditional artificial neural network, which can help identify the risk of CVD in advance.

The aim of this study is to investigate the risk factors of CVD in the primary prevention of CVD, and build an early warning model based on the existing technology. According to the traditional risk screening process, it aims to reduce the misdiagnosis rate caused by patients' self-report and doctors' subjective diagnosis. In the background, the authors summarize relevant researches about early warning for CVD risk. In the methods, they complete the design of the risk index system using rough set theory and establish an early warning model for CVD risk based on Deep Belief Networks (DBNs). In the results and discussion, medical data is used to test the model from the dimension of accuracy, training time, etc. The final section concludes this work. The overall frame diagram of the article is shown in figure 2.

## Background

CVD is characterized by hemorrhagic or ischemic injury of the brain, with diverse and complex etiological factors, which can still lead to partial loss of body function or death even after perfect treatment (Stone et al., 2014). Although some developed countries have taken effective measures to CVD causes and risk factors, which reduced morbidity and mortality to a certain extent. However, the result is not very gratifying. So the risk recognition and prevention measures of CVD are still the world's major medical challenges. It is an urgent and important task to establish a practical

Figure 2. The Overall Frame Diagram of the Article



and scientific early warning system to capture indicators of CVD crisis and formulate reasonable intervention measures at the present.

Early warning for CVD risk aims at forecasting the possibility or risk level in a period of time by identifying the related risk factors and calculating degree by scientific methods. The results of early warning can guide the individual to take certain precautions. The academic research of early warning for CVD risk is one of the hotspots in the field of CVD disease prevention and control at home and abroad, which has transferred from the risk factors of disease to auxiliary diagnosis, to the risk early warning.

A lot of useful researches on the risk factors and early warning model for CVD have been done. AHAASA (Stone et al., 2014) published the guidelines of primary and secondary prevention for stroke, classified the risk factors and hierarchy of stroke separately for those who have no history of CVD and those with a history of stroke. Liu (2011) has found that the risk of CVD increases linearly with

the rise of blood pressure levels. The correlation proportion is logarithmically related in the process of study. Stone et al. (2014) have discovered that the low-density lipoprotein cholesterol (LDL-C) and CVD risk is positively correlated according to the Meta-analysis of the random experiment. In addition, some studies (Fang et al., 2015) have confirmed that the risk of CVD for patients with diabetes is as much 1.5 to 3 times as the general population. Gu et al. (2020) have discovered that the physical inactivity and smoking have strong correlations with CVD employing the weighted regression models. Climent et al. (2020) have proved that lipid-lowering therapy is a kind of effective secondary prevention of ischemic CVD.

During the study of CVD risk and early warning, scholars establish a variety of risk evaluation models and index systems. The score table based on Framingham study of CVD (Stone et al., 2014) is a most widely used representative tool. In 2013, China established the applicable “Recurrence Risk Assessment at High Risk of Stroke and Stroke Patients” (Wu, 2003), choosing the main risk factors which include age, blood pressure, smoking, diabetes, Serum total cholesterol (TC), and Body Mass Index (BMI) as the evaluation basis. Shi and Sun (2014) divided the risk factors of CVD into intervention and non-intervention factors.

Early warning models contain statistical models and artificial intelligence models. Machine learning and neural networks are widely used in a variety of fields, including financial systems and disease prediction. Srivastava et al. (2021) adopted a support vector machine(SVM), random forest (RF), gradient boosting, and deep neural networks to forecast the index price trend of the National Stock Exchange of India (NSE) on the following day. The deep neural networks approach performed better than the machine learning classifiers. Most of the disease early warning technologies are based on classification and regression algorithms. Faizal et al. (2021) pointed out that the commonly traditional risk prediction models of CVD include Framingham risk score, thrombolysis in myocardial infarction (TIMI) and GRACE, while the risk prediction methods involve linear regression, logistic regression, Cox proportional risk and so on. Due to the complex relationship of nonlinear interaction between risk factors of CVD, the old approaches are restricted in representing complex functions. Framingham heart study (Kernan et al., 2014) was the most famous in statistical models. Besides, the artificial neural network model is the most commonly used in machine learning models for disease warnings. Bigi et al. (2005) forecasted the possibility of using an artificial neural network and Bytes classifier for 500 cases of patients with acute myocardial infarction, and finally proved the efficacy of the neural network model and Bytes theory in disease prediction. Wu et al. (2003) established CVD assessment methods and tools suitable for physical characteristics and disease features by using the COX proportional hazards model in the tenth five-year project preliminary. Song (2008) established the BP neural network model for CVD using Levenberg-Marquardt (LM) algorithm. He had improved the training speed of the network based on retrofitting the defect of slow convergence speed on the gradient descent method and proved that the artificial network method was feasible and effective in disease diagnosis. Guo (2013) constructed three kinds of the statistical prediction model, respectively logistic, BP and Elman, proving that the neural network model is accurate and effective in forecasting of myocardial infarction based on large amounts of data and variables. Jamthikar et al. (2020) used three kinds of MCML algorithms and CCVRC to compare the performance of CVD risk assessment, which showed that RF had better performance than SVM and XGBoost. Kim et al. (2021) proposed a model to predict the possibility of CVD by applying ML algorithm to NHIS-healS data in Korea, which turned out that XGBoost, RF and gradient enhancement algorithm had higher accuracy. The large quantity and complexity of medical data lead to the application of data mining techniques, which are used to discover hidden patterns or build classification models. Sung et al.(2020) used Association Rule Mining (ARM) to discover Antiepileptic drugs(AED) or combinations of them that are associated with newly diagnosed diabetes, obtaining related prescription rules to predict the consequences of taking AED and reduce the risk of suffering from diabetes.

To sum up, although the research of early warning for CVD risk has obtained specific achievements, the mainstream early warning models are still based on traditional quantitative evaluation

and artificial neural network that have problems on feasibility, accuracy, and practicability. DBNs can conduct layered representation of the original data by unsupervised learning, and finally realize the ability to learn essential characteristics from a mass of sample data through multiple complex function simulation. Le et al. (2011) established a total of 1 billion parameters of the Deep Neural Networks (DNNs) on the ImageNet data set and made classification accuracy of the model increased by 70% through the training. Dan et al. (2011) used the Convolution Neural Network (CNNs) method to identify traffic signs. The experimental results show that unsupervised feature learning was better than human identification on accuracy and efficiency. Lu et al. (2018) improved the DBN by the reconstruction error and the combination of unsupervised training with supervised optimization, suggesting that accuracy of the improved model could be around 90 percent. Smolander et al. (2019) used the combined classifier in which DBN is a learner and SVM is the input. DBN model has solved the defects of the existing mainstream model. But the algorithm is usually used in speech recognition, image recognition and natural language processing, less used for disease early warning.

According to the literature review, this article proposed a solution by using rough set and DBNs in an early warning for CVD, aiming to achieve timely warnings and provide effective services in the primary prevention of CVD. Through the establishment of index systems and DBNs, the authors collected sample data from a medical cloud platform to study the basic characteristics and build the early warning model. Different from the existing researches: (1) In order to make up for the non-quantitative and redundant problems of the existing index system, this article proposed to extract the key risk factors by using rough set and established a comprehensive index system. Based on the existing warning indicators of CVD, the initial risk factors were reduced by Attribute Reduction with reference to the stroke risk quantification table. (2) It set up an early warning model of CVD based on DBNs and the experiments proved that the DBNs model had a better predictive effect than the LM-BP neural network model for CVD warning.

## Methods

### Construction of Index System for Early Warning

The existing index systems have some deficiencies: some indicators are difficult to assess for lack of quantitative, others are redundant. Too many risk factors for CVD lead to the overlap of the indexes collected from different sources. To deal with that problem, this article proposes the theory of extracting the key risk factors using a rough set to establish a general index system, which is operational and sensitive. It sorted out the early warning indicators of CVD comprehensively in reference to stroke risk quantization tables and the existing research results, and then it reduced the initial risk factors using the method of attribute reduction based on the theory.

The basic idea of rough set theory is extracting key information and reducing knowledge of the sample data set without the change of classification, and then setting the decision rules (Guo, 2013). Rough set is usually combined with genetic algorithm and neural network to foster strengths and circumvent weaknesses of different models. Its validity has been confirmed by extensive research. Based on the rough set theory and information entropy algorithm, the process of attribute reduction is as follows:

- A: constructing a decision table. The decision table contains the object set and the attribution set. The object set is a sample participating in the study, while the attribute set represents the description of the object, which can be divided into two categories: condition attributes and decision attributes;
- B: calculating  $P(A)$ . The A represents an attribute,  $P(A) / |U| = |X_i|, (i = 1, 2, \dots, M)$ ; where  $|X_i|$  represents the base of the set  $|U|$ ,  $X_i$  represents the base of the collection;  
computing information entropy  $H(A)$  and  $H(A - \{a\})$  of an attribute set A

$$A = \{a_1, a_2, \dots, h, a_n\}$$

$$H(A) = -\sum_{i=1}^m P(A_i) \times \ln P(A_i), H(A - \{a\}) = H(A) + P(a_i) \times \ln P(a_i)$$

- D. calculating  $S_A(a_i), S_A(a) = |H(A) - H(A - \{a\})|$ , by deleting properties where  $S_A(a_i) = 0$ , and saving properties where  $S_A(a_i) > 0$ , these existing properties are recorded as  $C = \{c_1, c_2, \dots, c_s\}$ , in which s is less than n;
- E. calculating the importance degree  $S_C(c_j)$  of the index set. The related degree between the various attributes in C is the selection standard, when two attributes related to a value are greater than a specific value, the less important one will be deleted.

In the initial construction of the early warning index system, the main risk factors of CVD are divided into six first-order indicators, including demographic characteristics, lifestyle, physical index, disease history, family history, and disease index (Stone et al., 2014). On this basis, factors are divided into 51 second-order indicators. The demographic characteristic (includes age, gender, and race); way of life (includes smoking, drinking, exercise, diet); physical index (such as height, weight, and waist circumference); history of diseases (such as hypertension disease history, diabetes history, and high blood fat disease history); history (such as family history of CVD, family history of hypertension, and family history of diabetes); disease index (such as heart murmur, left ventricular hypertrophy, and transient ischemic attack). After using the information entropy algorithm based on rough set theory, the second indicators are reduced from 51 to 19. The reduced index early system is as shown in the table 1:

### Construction of Deep Belief Networks Model for CVD Warning

DBNs are a deep learning algorithm composed of multilayer RBM and a layer of BP neural network. The underlying layer of DBNs is composed of a plurality of RBMs, which stack together to realize an abstract representation of the data. The top layer is the BP network which can adjust the whole network by learning labeled data. DBNs are based on a generative probabilistic model as a whole, in which the calculation model is based on the joint distribution of data and features.

The undirected connected network at the bottom of DBNs constitute a restricted Boltzmann machine. In the calculation, RBM is the basic unit to constitute a DBN. The whole network finish unsupervised feature learning by calculating the restricted Boltzmann machine and training the parameters weight of each layer in DBNs. An RBM structure contains a visual layer and a hidden layer. Each visual node is only related to n hidden nodes and independent of other visual nodes. For each hidden node, it is influenced only by m visual node. For state (V, H), the function of RBM is shown as:

$$E(v, h) = -\sum_{i=1}^n \sum_{j=1}^m W_{ij} h_i v_j - \sum_{i=1}^m a_i v_i - \sum_{j=1}^n b_j h_j \dots \quad (1)$$

This formula (1) indicates the weights between the i node in visual layer and the j node in hidden layer; expresses the bias size of node i in visual layer and the bias size of node i in hidden layer. We maximized the following likelihood function:  $H(A - \{a\})$  Approximate gradient is obtained by approximate calculation of the gradient with an algorithm called CD-k (Bigi et al., 2005). Gradient descent method is used to update RBM parameters:

Table 1. Reduced Index System for Early Warning

Index System	The First Order Indicators	The Second Order Indicators
	A demographic characteristics	A1 age
		A2 gender
	B way of life	B1 smoking
		B2 drinking
		B3 exercise
		B4 diet
	C physical index	C1 BMI
		C2 resting diastolic blood pressure
		C3 resting systolic blood pressure
		C4 fasting blood glucose
		C5 triglyceride TG
		C6 total cholesterol TCHO
	D history of diseases	D1 history of hypertension
D2 history of diabetes		
D3 history of hyperlipidemia		
D4 history of heart disease		
D5 history of CVD		
E family history	E1 family history of CVD	
F disease index	F1 transient ischemic attack(TIA)	

$$\frac{\partial \ln P(v, \theta)}{\partial W} = E_{P_{data}} [vh^T] - E_{P_{model}} [vh^T] \dots \quad (2)$$

$$\frac{\partial \ln P(v, \theta)}{\partial a} = E_{P_{data}} [v] - E_{P_{model}} [v] \dots \quad (3)$$

$$\frac{\partial \ln P(v, \theta)}{\partial b} = E_{P_{data}} [h] - E_{P_{model}} [h] \dots \quad (4)$$

The formulas above indicate the expectation based on the distribution of input variables and the expectation based on the determined density distribution of the model.

The gradient descent method is used to update the parameters of RBM, where the learning rate of parameters updating and i is the number of iterations.

$$\theta^{(i+1)} = \theta^{(i)} + \eta \frac{\partial \ln P(v; \theta)}{\partial \theta} \dots \quad (5)$$

Stacks of Multiple RBMs constitute the underlying structure of DBNs. RBM can be trained by an unsupervised greedy algorithm, in which the weight learned from above is passed to the DBNs for memory. The state of the hidden layer in RBM calculated from the training data is passed as input data to an RBM to study further the relationship between RBM units in the hidden layer. After repeating this learning process several times, RBM will stack to learn the structure information of complex data. After the completion of the pre-training of the RBM layer, RBM stack to form a deep networks. After adding a classification layer on the top of it, a bottom-up feed-forward neural network is finished. In this article, BP algorithm is selected to distinguish the performance adjustment through labeled data.

For a given sample (A1, A2, B1, B2, B3, B4...F1, y), in which y represents the grade of CVD risk, all of the activation values in all layers' network are calculated according to the forward conduction formula, as well as the output value. For each node in the output layer, the gap between activation values and actual values in the network can be directly calculated. The gap is defined as the residuals, in which the NL layer represents output layer (Benavente et al., 2013). It can be calculated using the following formula:

$$\delta_i^{(l)} = \left( \sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}) \dots \quad (6)$$

For  $l = (n_1 - 1, n_2 - 1, n_3 - 1, \dots, n)$ , the calculation for the residual error of node i in layer l is as follows:

$$\delta_i^{(l)} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \|h_{w,b}(x) - y\|^2 = -(y_i - a_i^{(n_l)}) \times f'(z_i^{(n_l)}) \dots \quad (7)$$

The calculation of partial derivatives is as follows:

$$\frac{\partial}{\partial W_{ij}^{(l)}} J(Q, b; x, y) = a_j^{(l)} \delta_i^{(l+1)} \dots \quad (8)$$

$$\frac{\partial}{\partial b_i^{(l)}} J(Q, b; x, y) = \delta_i^{(l+1)} \dots \quad (9)$$

In summary, algorithms of DBNs can be summarized as follows:

- (1) Pre training stage: through the RBM self -training algorithm, all the RBM structures can be trained out unsupervised from the low level to the high level;
- (2) The fine-tuning stage: the output information of the last layer in RBM is transferred to input layer. BP neural network is used as supervised classifier. The spread error of network seems like the revised standard, which is transferred from the output end to the input end to modify the parameters of DBNs.

## Results and Discussion

In this article, the neural network toolbox in MATLAB is used to compile the experiment.



## Data Samples and Data Processing

The data comes from the cloud platform of CVD in the medical association model. The research is based on the project of “The Construction and Demonstration Applications of Medical Association Model of Critical Illness Collaborative Control in Health Care Reform of Beijing.” The project has established collaborative prevention composed of a first-grade hospital, emergency center and community hospital- based for CVD. The subject concentrates on information data from different hospitals on cloud platforms, collecting a total of more than 40000 people information who have taken part in the model of medical association for CVD from 2013 to 2015. The age of these people distributes between 30 to 74 years old. The platform has recorded age, gender, height, weight and other 72 indicators.

According to the index system for early warning established above, training data requires the following indicators: age, gender, family history, ethnicity, BMI, smoking, drinking, eating habits, exercise habits, blood pressure, blood glucose, blood lipids, hypertension, diabetes, cardiovascular disease, CVD, history of hyperlipidemia and transient ischemic attack. Because both information of normal population and abnormal population are on the data cloud platform, the expected output of the model is divided into five categories, namely no risk, low risk, middle-high risk, high risk and diagnosis. The determination of the risk level is based on the cerebral vascular hemodynamic study. The score of the evaluation function is the basis of dividing: above 75 represents no risk of CVD, between 70 and 74 represents a low risk; between 50 and 70 represents a middle-high risk; less than 50 represents a high risk, as shown in table 2.

Table 2. The Expected Output of the Model

The score of evaluation function	The category of output
>75	No risk
70~74	Low risk
50~70	Middle-high risk
<50	High risk

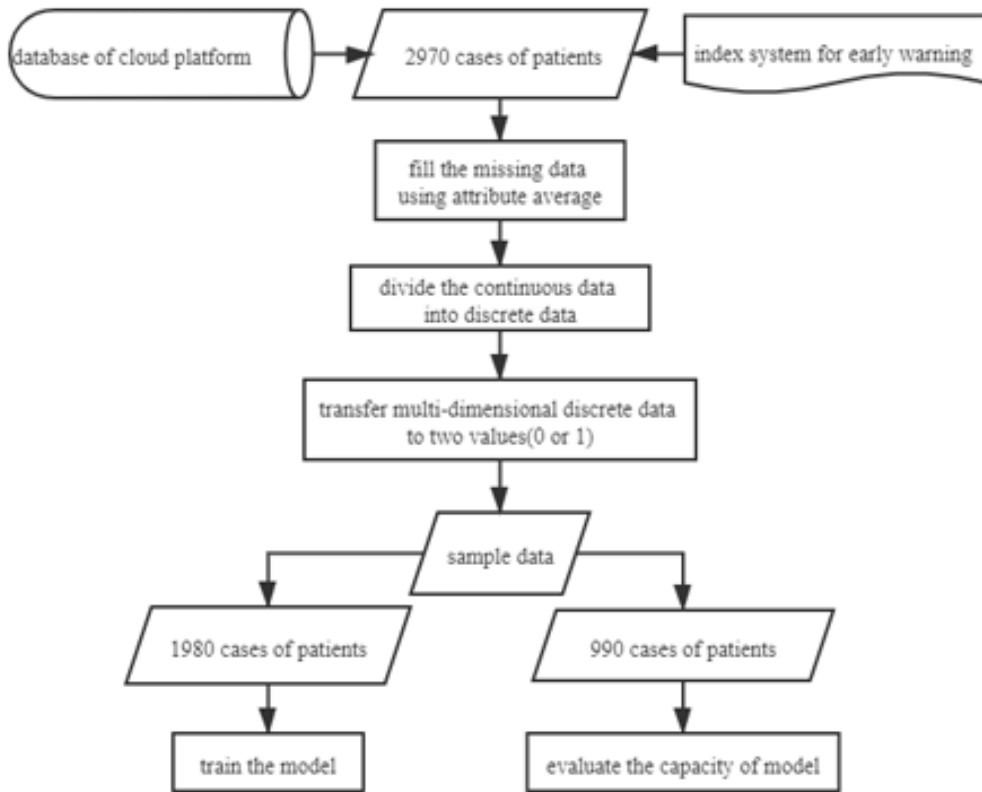
2970 cases were extracted from the database of the cloud platform according to the risk index system of CVD constructed above. 19 indexes of patients with or without cerebral vascular disease were used for early warning model training. Data preprocessing included three steps: Firstly, the missing data was filled using attribute average; secondly, the continuous data was divided into discrete data; thirdly, the multi-dimensional discrete data was transfer to two values, which are 0 and 1.

In this study, two-thirds of sample data are selected for model training, and the remaining data is for evaluating the capacity of model. The dimension of training data is [1980 x 31] and [1980x5], including 31-dimensional input features and 5-dimensional target features of 1980 patients. The dimension of test data is [990x31] and [990x5], including 31-dimensional input features and 5-dimensional target features of 990 patients. The acquisition and processing of the data sample is shown in figure 3.

## Algorithm Structure and Parameter Setting

According to the features on section data, the input nodes of the neural network are 31. Since the training target class has five types, the value of output nodes is set to 5. The increase of the node number in hidden layer can improve the discrimination performance of the network. To avoid the blindness selection and increase the effectiveness, the number is selected through the experimental method according to the reconstruction error after comprehensive comparison.

Figure 3. The Acquisition and Processing of Data Sample



Accordingly, reconstruction error is calculated when the number of hidden nodes is 13,14,15,16, and 17. When the node number in hidden layer 1 is 14, the reconstruction error is the smallest. Similarly, calculate the experimental reconstruction error when the node number in hidden layer 2 is separately 8,9,10. When the node number in hidden layer 2 is 8, the reconstruction error is the smallest. So the structure of the DBN is: 31 input nodes, 14 nodes in hidden layer 1, 8 nodes in hidden layer 2 and 5 output nodes, as shown in table 3.

Table 3. The Network Structure Parameters of DBNs

The network structure parameters	The value
the number of nodes in the input layer	32
the number of nodes in the first hidden layer	14
the number of nodes in the second hidden layer	8
The number of nodes in the output layer	5

To sum up, when building a DBN for early warning, the structure is set as shown above so that the network can get the best accuracy, namely 82.1%.

### The Comparative Analysis of the Results

BP network is a kind of supervised learning feed-forward neural network. 80% of the neural network model is based on the basic idea of BP. The learning-ability and adaptability of BP make it can be applied in the prediction and disease fields. Considering that BP neural network often appears slow convergence caused by the redundancy structure, and easily falls into local minimum, which could lead to the error of warning results, the LM algorithm is chosen in the process, which is better than the gradient convergence method for fast convergence and robustness. The speed of BP neural network is improved when using the LM algorithm. LM-BP is established as an early warning model through training and learning. An accuracy comparison of DBNs model with LM-BP network model is shown in table 4.

Table 4. Accuracy Comparison of DBNs and LM-BP

	DBNs	LM-BP
Accuracy of model	82.1%	74.17%

From the results, the early warning model based on DBNs has good prediction effect when based on the mentioned indicators to predict the risk of CVD. The accuracy of DBNs model can reach 82.1%, while LM-BP can reach 74.17%. It is obvious that the practicability of DBNs model is stronger.

Table 5. Error Rate Comparison of DBNs and LM-BP on Different Data

Error rate	100	300	500	700	900
LM-BP	0.382	0.317	0.225	0.336	0.229
DBNs	0.146	0.144	0.143	0.141	0.147

As shown in Table 5 above, when the quantity of forecast data is separately set up 100, 300, 500, 700 and 900, the DBNs and BP network models respectively show different accuracy values. From the overall trend, the error rate of DBNs model is not improved when magnitude increases, but LM-BP network model shows instability. From the result of the comparison, when the number of forecast data is less than 1000, the error rate of DBNs model is lower than LM-BP neural network model.

In addition to the model accuracy, the DBNs train the weights of RBM through an unsupervised learning feature. Through the pre-training of the hidden layer, it can obtain better parameter values compared to random assignment of BP neural network. It can also improve the learning speed. This learning method overcomes the defects of BP, which can fall into local optimum in the training process.

Table 6. Training time comparison of DBNs and LM-BP

Training time	100	300	500	700	900
LM-BP	6.5s	6.51s	6.58s	6.59s	6.63s
DBNs	0.36s	0.38s	0.43s	0.45s	0.47s

As is shown in Table 6 above, when processing a large amount of data, the DBNs reflect better than LM-BP network in the training speed. When the quantity of forecast data is more than 100, the training time of LM-BP network was over 6s, and the DBNs model of the time was less than 0.5s. So it is proved that the DBNs model is able to train large data faster than LM-BP networks within a certain range.

In summary, DBNs model shows a good effect in the empirical process of risk early warning for CVD. Through the comparison with LM-BP network, the early warning model based on DBNs shows higher accuracy. Considering the characteristics of DBNs, the model will reflect a more obvious advantage when the amount of data increases and the data dimension expands.

## **Conclusion**

CVD seriously endangers the health and brings a heavy burden to the economy. It can reduce morbidity and mortality of CVD by identifying the risk for individuals before the onset and taking certain control measures. This article focuses on the early warning in the process of primary prevention to assess the risk of CVD accurately and timely by constructing the risk early warning model.

(1) The initial index system for early-warning is established comprehensively through literature analysis and summary, which includes six first-level indexes and 51 second-level indexes. Then attributes are reduced by using information entropy algorithm based on rough set theory, containing 19-second level indexes at last. (2) The risk early warning model for CVD is constructed based on DBNs by building and completing the stack of restricted Boltzmann, and adjusting parameters in the whole network using back propagation algorithm. (3) The comparison results of the experimental with the prediction model of LM-BP neural network prove the superiority of the early warning model for CVD based on DBNs. The index system of early warning for CVD constructed in this article is suitable for primary prevention.

However, the early warning model established in the article has its limitation: Considering the difference of secondary prevention due to the pathogenesis of patients, the effectiveness of the model in the secondary prevention of CVD needs to be verified. More comparisons with other existed early warning models need further experiments. Also, whether the proposed method is applicable to other diseases besides CVD is a future research direction, such as COVID-19. When discussing the role of information technology in emergency response and pandemic management, Shiau et al. (2021) put up with that deep learning has been applied to the modeling and identification of disease outbreaks, such as the identification of the disease according to its symptoms, the predication of COVID-19 prognosis and so on. Thus, deep learning has been widely used in the prognosis and prediction of diseases. But different from chronic diseases, the ever-changing and evolving nature of epidemics puts greater demands on real-time processing data and making predictions quickly.

## **ACKNOWLEDGMENT**

This research was partially supported by a Key Project of National Natural Science Foundation of China (NSFC) [grant number 71532002].

This research was supported by the Beijing Tian Ton Hospital, Capital Medical University for provision of related experimental materials and collaborative experimental work. And it is also supported by Goodwill Company for technical assistance.

RZ has been involved in revising it for important intellectual content and acquisition of funding;

QQ made substantial contributions to conception and design, and has revised it critically;

XY made substantial contributions to acquisition of data, analysis and interpretation of data, and also drafted the manuscript;

YG has revised the manuscript and helped perform the final version.

YM has revised the manuscript and took responsibility of the submission process.

ML has been involved in revising it critically for important intellectual content.

## REFERENCES

- Benavente, O. R., McClure, L. A., Coffey, C. S., Conwit, R., Pergola, P. E., & Hart, R. G. (2013). The secondary prevention of small subcortical strokes (SPS3) Trial: Results of the blood pressure intervention. *Lancet*, 382(2), 507–515. PMID:23726159
- Bigi, R., Gregori, D., Cortigiani, L., Desideri, A., Chiarotto, F. A., & Toffolo, G. M. (2005). Artificial neural networks and robust Bayesian classifiers for risk stratification following uncomplicated myocardial infarction. *International Journal of Cardiology*, 101(3), 481–487. doi:10.1016/j.ijcard.2004.07.008 PMID:15907418
- Climent, E., Benaiges, D., & Pedro-Botet, J. (2020). Lipid-lowering treatment in secondary prevention of ischaemic cerebrovascular disease. *Clínica e Investigación en Arteriosclerosis*, 32(4), 175–182. doi:10.1016/j.arteri.2019.12.002 PMID:32035666
- Dan, C., Meier, U., Masci, J., & Schmidhuber, J. (2011). A Committee of Neural Networks for Traffic Sign Classification. *IJCNN*, 42(4), 1918–1921.
- Faizal, A. S. M., Thevarajah, T. M., Khor, S. M., & Chang, S. W. (2021). A review of risk prediction models in cardiovascular disease: Conventional approach vs. artificial intelligent approach. *Computer Methods and Programs in Biomedicine*, 207, 106190. doi:10.1016/j.cmpb.2021.106190 PMID:34077865
- Fang, H., Shuang, L., Cao, K., & Hospital, D. (2015). Interpretation of Guidelines for The Prevention of Stroke in Patients with Stroke and Transient Ischemic attack (2014), *Chinese Journal of Integrative Medicine On Cardio-Cerebrovascular Diseases (Basel, Switzerland)*, 11(13), 28–32.
- Gu, J., Charles, L. E., Fekedulegn, D., Allison, P., Ma, C. C., Violanti, J. M., & Andrew, M. E. (2020). Temporal trends in prevalence of cardiovascular disease (CVD) and CVD risk factors among U.S. older workers: NHIS 2004-2018. *Annals of Epidemiology*, 55(30), 78–82. doi:10.1016/j.annepidem.2020.10.002 PMID:33049395
- Guo, J. (2013). *Study on the application of neural network model in prediction of acute myocardial infarction and comparison of the predictive ability of the model* [Ph.D. Dissertation]. Peking Union Medical College.
- Jamthikar, A. D., Gupta, D., Mantella, L. E., Saba, L., Laird, J. R., Johri, A. M., & Suri, J. S. (2020). Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: A 500 participants study. *The International Journal of Cardiovascular Imaging*. Advance online publication. doi:10.1007/s10554-020-02099-7 PMID:33184741
- Kernan, W. N., Ovbiagele, B., Black, H. R., Bravata, D. M., Chimowitz, M. I., Ezekowitz, M. D., Fang, M. C., Fisher, M., Furie, K. L., Heck, D. V., Johnston, S. C. C., Kasner, S. E., Kittner, S. J., Mitchell, P. H., Rich, M. W., Richardson, D. J., Schwamm, L. H., & Wilson, J. A. (2014). Guidelines for the prevention of stroke in patients with stroke and transient ischemic attack: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 45(7), 2160–2236. doi:10.1161/STR.0000000000000024 PMID:24788967
- Kim, J. O., Jeong, Y. S., Kim, J. H., Lee, J. W., Park, D., & Kim, H. S. (2021). Machine learning-based cardiovascular disease prediction model: A cohort study on the Korean national health insurance service health screening database. *Diagnostics (Basel)*, 11(6), 943. Advance online publication. doi:10.3390/diagnostics11060943 PMID:34070504
- Le, Q. V., Monga, R., Devin, M., Corrado, G., Chen, K., Ranzato, M. A., . . . Ng, A. Y. (2011). *Building high-level features using large scale unsupervised learning*. arXiv preprint arXiv:1112.6209.
- Liu, L.S., & Writing Group of 2010 Chinese Guidelines for the Management of Hypertension. (2010). Chinese guidelines for the management of hypertension. *Zhonghua Xin Xue Guan Bing Za Zhi*, 39(7), 579–615. PMID:22088239
- Lu, P., Guo, S. D., Zhang, H. P., Li, Q. H., Wang, Y. C., Wang, Y. Y., & Qi, L. X. (2018). Research on improved depth belief network-based prediction of cardiovascular diseases. *Journal of Healthcare Engineering*, 2018, 1–9. Advance online publication. doi:10.1155/2018/8954878 PMID:29854369
- Prakash, V. J., & Karthikeyan, N.K. (2021). Enhanced evolutionary feature selection and ensemble method for cardiovascular disease prediction. *Interdisciplinary Sciences: Computational Life Sciences*, 1-24.10.1007/s12539-021-00430-x

Shiau, W. L., Shiau, K., Yu, Y., & Guo, J. (2021). Research commentary on IS/IT role in emergency and pandemic management: Current and future research. *Journal of Database Management*, 32(2), 67–75. doi:10.4018/JDM.2021040105

Smolander, J., Dehmer, M., & Emmert-Streib, F. (2019). Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio*, 9(7), 1232–1248. doi:10.1002/2211-5463.12652 PMID:31074948

Song, Z. G. (2008). *Research on the cerebrovascular diseases diagnosis expert system based on the back propagation neural network* [Ph.D. Dissertation]. Harbin University of Science and Technology.

Srivastava, P. R., Zhang, Z. J., & Eachempati, P. (2021). Deep Neural Network and Time Series Approach for finance systems: Predicting the movement of the Indian stock market. *Journal of Organizational and End User Computing*, 33(5), 204–226. doi:10.4018/JOEUC.20210901.0a10

Stone, N. J., Robinson, J. G., Lichtenstein, A. H., Merz, C. B., Blum, C. B., & Eckel, R. H. et al.. (2014). 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: A report of the American college of cardiology/American heart association task force on practice guidelines. *Circulation*, 129(25), S1–S45. doi:10.1161/01.cir.0000437738.63853.7a PMID:24222016

Sung, S., Lee, P., Hsieh, C., & Zheng, W. (2020). Medication use and the risk of newly diagnosed diabetes in patients with epilepsy: A data mining application on a healthcare database. *Journal of Organizational and End User Computing*, 32(2), 93–108. doi:10.4018/JOEUC.2020040105

Wu, Y. F. (2003). A Study on evaluation of the risk of ischemic cardiovascular diseases in Chinese and the development of simplified tools for the evaluation. *China J Cardiol*, 31(12), 893–901.

*Qiuli Qin is an associate professor in School of Economics and Management, Beijing Jiaotong University, she is also a director of China Information Economy Association. She has been to University of Arizona AI LAB as a visiting scholar. Her research interests include theory and practice of Enterprise Informatization, data mining and data analysis, IT Project Management and so on. She was awarded as “Advanced Individual of Three Educations of Beijing Jiaotong University”. Her works has been appeared in several SCI journals including BioMed Research International, Complexity and others. Also, she has published several papers in academic/professional Meeting proceedings.*

*Xing Yang was born in 1991. She received a master’s degree in information management from Beijing Jiaotong University. She is currently doing research work in China Unicom Research Institute. Her main research interests include machine learning, customer operations and so on. As a core member, she once participated in the project named “Construction and Demonstration Application of Medical Consortium Model for Collaborative Prevention and Treatment of Major Diseases Based on Cloud Computing.*

*Runtong Zhang is a professor of Beijing Jiaotong University, Beijing, China. His research interests include e-commerce, e-health, online health and big data. His works appeared in several international journals including Journal of Medical Internet Research, Decision Support Systems, IEEE Intelligent Transportation Systems Magazine, and others.*

*Manlu Liu is the Full Professor of Management Information Systems and Accounting and the Director of Master in Business Analytics Program at the Saunders College of Business at Rochester Institute of Technology (RIT). Her research interests include community-based open source (community source), accounting and financial analytics, blockchain and auditing, and open innovation in health IT. She started to examine community source phenomenon in 2005. She is one of the researchers who introduced this phenomenon to MIS academic field and developed research models for community source evolution and sustainability. Dr. Liu is passionate about how business analytics applies to accounting and finance. She co-initiated the Advanced Certificate of Accounting and Financial Analytics.*

*Yuhan Ma is a postgraduate student in School of Economics and Management, Beijing Jiaotong University. Her major is information management and information system. Her mentor is Qiuli Qin, and her research direction is medical data information management.*