

Research Article

Research on Improved Depth Belief Network-Based Prediction of Cardiovascular Diseases

Peng Lu ^{1,2} Saidi Guo ^{2,3} Hongpo Zhang^{2,4} Qihang Li ^{1,2} Yuchen Wang ¹
Yingying Wang ^{2,3} and Lianxin Qi ^{2,3}

¹School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China

²Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou 450001, China

³Industrial Technology Research Institute, Zhengzhou University, Zhengzhou 450001, China

⁴State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou Science and Technology Institute, Zhengzhou 450003, China

Correspondence should be addressed to Peng Lu; lupeng@zzu.edu.cn

Received 15 December 2017; Revised 27 March 2018; Accepted 5 April 2018; Published 9 May 2018

Academic Editor: Vincenzo Positano

Copyright © 2018 Peng Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Quantitative analysis and prediction can help to reduce the risk of cardiovascular disease. Quantitative prediction based on traditional model has low accuracy. The variance of model prediction based on shallow neural network is larger. In this paper, cardiovascular disease prediction model based on improved deep belief network (DBN) is proposed. Using the reconstruction error, the network depth is determined independently, and unsupervised training and supervised optimization are combined. It ensures the accuracy of model prediction while guaranteeing stability. Thirty experiments were performed independently on the Statlog (Heart) and Heart Disease Database data sets in the UCI database. Experimental results showed that the mean of prediction accuracy was 91.26% and 89.78%, respectively. The variance of prediction accuracy was 5.78 and 4.46, respectively.

1. Introduction

Cardiovascular disease has become the most pathogenic disease in our country [1]. The establishment of a prediction model of cardiovascular disease and the quantitative analysis of the risk of disease can effectively reduce the incidence of the disease [2].

In the past few decades, researchers have conducted a lot of research on the computer classification of ECG, such as support vector machines (SVMs), artificial neural networks (ANNs), decision trees, Bayesian networks, support feature machines (SFMs), and regression analysis. Cardiovascular disease prediction model is divided into two categories; one is the traditional prediction model based on probability. For example, in Framingham Heart Study (FHS) [3], the model is characterized by the adoption of a mathematical formula, which has good stability, but its effect is poor and the accuracy is low in the multiclassification and nonlinear complex factors. And the other is based on shallow neural network prediction model of cardiovascular disease. In

Munster Heart Study (PROCAM) [4], two neural network models are used: a multilayer perceptron network with one hidden layer (MLP) [5] model and a probabilistic neural network (PNN) [6]. The characteristics of this kind of model are that it can effectively expand the forecasting factor, quickly process the fuzzy data and nonlinear data, and provide high rate of accuracy [7]. However, due to the randomness of initialization of shallow neural network parameters, the prediction results will be much lower than the average accuracy, and the variances of multiple prediction results will be larger.

Recently, deep learning has been widely used in different fields and has made great progress [8–11]. The deep learning model uses multiple samples to extract high-level features and learns hierarchical representations by combining low-level inputs more effectively. The learned features characterize more intrinsic features of the data, avoiding the process of artificial feature design and selection, and have the characteristics of many varieties and high accuracy [12].

This paper takes deep learning as the point of penetration and uses multilayer network architecture to abstract the characteristics of layers and establish a cardiovascular disease prediction model based on deep belief network. At the same time, the prediction model based on deep trust network is improved by using reconstruction error to achieve better prediction.

Overall, the major contributions of this work can be summarized in three aspects. First, we use the deep belief network to build predictive models of cardiovascular disease, skip the morphological feature extraction step, and classify the original ECG data directly, thus solving the problem that the cardiovascular disease prediction model is not robust due to the large difference in waveform characteristics between patients with the same disease. Second, we adopt the best network parameters trained to initialize the neural network, so as to solve the instability problem caused by stochastic initialization. Finally, we use reconstruction error to improve the prediction model which is based on the deep trust network, so that it can independently determine the network depth and achieve better predicted results.

2. Related Work

The literature related to this classification application was studied, and it can be seen that a great variety of methods were used, which reached high classification accuracies.

Algorithms for R-peak extraction tend to use wavelet transforms to compute features from the original ECG followed by a fine-tuned threshold-based classifier. Since the accurate estimation of heart rate and heart rate variability can be extracted from the R-peak feature, the specially designed algorithm is usually used for the classification of coarse-grained heart rhythm. Sundar et al. [13] proposed a prototype using data mining techniques, namely, Naïve Bayes and WAC (weighted associative classifier). The recognition rate of 84% and 78% was obtained from weighted associative classifier and Naïve Bayes. Iftikhar et al. [14] present a hybrid approach using a supervised learning model based on a well-known classifier SVM and evolutionary optimization techniques (genetic algorithm (GA) and particle swarm optimization (PSO)). The results have shown considerably improved accuracy of more than 88%.

However, because of the differences in the ECG waveforms of different people and the great differences in ECG waveform characteristics of different diseases, the feature extraction of the waveform is inaccurate. Therefore, these characteristics are not sufficient to distinguish most cardiovascular diseases.

With the rapid development of artificial intelligence, inspired by automatic speech recognition, hidden Markov model with Gauss observation probability distribution has been applied to the beat detection task [15], and the hottest artificial neural network is also used for the task of beat detection [16]. Elsayad proposed an approach which used the learning vector quantization (LVQ) neural network to establish the ECG positive anomaly model and obtained an accuracy of 74.12% [17]. Olaniyi et al. [18] designed a neural

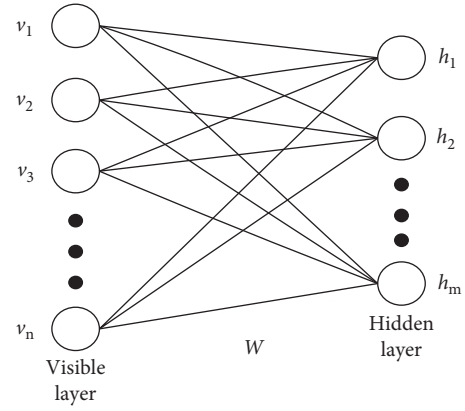


FIGURE 1: Restricted Boltzmann machine undirected configuration diagram.

network for diagnosis of heart diseases with the heart disease sample obtained from UCI machine learning repository. The system is a multilayer neural network model based on backpropagation training and is simulated on a feed-forward neural network. The recognition of 85% was obtained from testing of the network.

Although the self-learning ability of backpropagation (BP) neural network is strong, the convergence speed is slow, and the result is easily affected by the random initialization of network parameters. In particular, there has been no unified and complete theoretical guidance for the selection of BP neural network structure. Generally, it can only be selected by experience.

The DBN model not only has the self-adaptive ability of the self-adjustment of the general neural network but also avoids the defects of the BP neural network, which is easy to fall into the local minimum. DBN uses a network structure composed of multiple RBM networks, which is more effective for modeling one-dimensional data [19].

3. Description of the Proposed Approach

3.1. Deep Belief Network. Deep belief network (DBN) is one of the main tools for deep learning, which is based on the restricted Boltzmann machine (RBM) [20], to propose. The structure of RBM includes only the visible layer and the hidden layer; the neurons between two layers are fully connected, and the neurons in the same layer are not connected [21].

In Figure 1, $v(v_1, v_2, \dots, v_n)$ represents the visible layer, v_i is the visible unit; $h(h_1, h_2, \dots, h_m)$ denotes the hidden layer, h_j is the hidden unit; and W is the connection weight matrix between two layers. The data are input from the visible layer, (v_1, v_2, \dots, v_n) represents the feature set of the data, and the hidden layer data are generated by the random initialization of the weight value w and the state of each neuron. Due to the disconnection between neurons at the same level, when determining the neuron state, it has the following properties: when the visible cell state is determined, the hidden unit condition is activated independently; otherwise, if the state of the hidden cell is determined, the conditions of the visible units are activated independently.

Given a set of states (v, h) , the energy function of the RBM model can be defined by the following equation:

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

$$= -a^T v - b^T h - h^T w v,$$

where $a = (a_1, a_2, \dots, a_n)$ denotes the offset vector of the visible unit, $b = (b_1, b_2, \dots, b_m)$ denotes the bias vector of the hidden unit, and $v = (v_1, v_2, \dots, v_n)$ denotes the state vector of the visible layer, $h = (h_1, h_2, \dots, h_m)$ denotes the state vector of the hidden layer, $w = (w_{i,j})$ denotes the connection weight matrix, and $w_{i,j}$ denotes the weight of the i th visible unit and the j th hidden element.

For the state (v, h) , according to (1), the joint probability distribution can be given as follows:

$$P(v, h; \theta) = \frac{1}{Z} e^{-E(v, h)}, \quad (2)$$

$$Z = \sum_v \sum_h e^{-E(v, h)},$$

where $\theta = \{a, b, w\}$ is the RBM network parameters and Z is called the normalization factor or the partition function.

In practical applications, the probability distribution $p(v)$ of training data v is generally used, that is, the edge probability distribution of $P = (v, h, \theta)$:

$$P(v) = \frac{1}{Z} \sum_h e^{-E(v, h)}. \quad (3)$$

Similarly, the edge probability distribution $P = (h)$ of the hidden layer state can be obtained:

$$P(h) = \frac{1}{Z} \sum_v e^{-E(v, h)}. \quad (4)$$

RBM training data are obtained by solving the model optimal parameters in (3), so that the model can better fit the distribution of training data even if the sample reaches the maximum probability in the distribution. Constructing log-likelihood functions:

$$\ln P(v) = \ln \left(\sum_h e^{-E(v, h)} \right) - \ln \left(\sum_v \sum_h e^{-E(v, h)} \right). \quad (5)$$

The model parameters are respectively solved by the maximum likelihood function method:

$$\frac{\partial \ln P(v)}{\partial a} = E_{Pd}[v] - E_{Pm}[v],$$

$$\frac{\partial \ln P(v)}{\partial b} = E_{Pd}[h] - E_{Pm}[h], \quad (6)$$

$$\frac{\partial \ln P(v)}{\partial w} = E_{Pd}[vh^T] - E_{Pm}[vh^T],$$

where E_{Pd} denotes the expectation of the input conditional probability distribution of training data, and E_{Pm} denotes the expectation of the joint probability distribution of the

model. The expected computation is done by the Gibbs sampling method, while the computation cost is too large in the computation process of each iteration. Hinton proposed the contrastive divergence (CD) algorithm [21] for the approximate calculation after sampling.

According to the above formula, when the neuron state of the given layer is given, it can be inferred that the activation probability of hidden units is

$$P(h_j = 1 | v) = \sigma \left(\sum_i v_i w_{ij} + b_j \right). \quad (7)$$

After obtaining the hidden element state matrix, the reconfigurable visible element state probability can be calculated according to the CD algorithm:

$$P(v_i = 1 | h) = \sigma \left(\sum_j h_j w_{ij} + a_i \right), \quad (8)$$

where σ is a sigmoid function $\sigma(x) = 1/(1 + \exp(-x))$.

The maximum value of the likelihood function is gradually approximated by gradient ascent. The formula of the RBM parameter is updated as follows:

$$\theta^{i+1} = \theta^i + \eta \frac{\partial \ln P(v)}{\partial \theta}, \quad (9)$$

where η is the parameter learning rate for the model, and i is the current iteration. The parameters θ are iteratively updated according to the rules of (9), and the maximum value of the gradient of the likelihood function is reached quickly, which is the optimal parameter.

DBN is composed of a plurality of RBM units connected to the bottom layer of the RBM visible layer as the input layer, the underlying RBM hidden layer of the upper RBM visible layer. The tuning of global training parameters is carried out by the BP neural network.

RBM is a probabilistic neural network that determines the probability generation of DBNs, this is establishing a joint probability distribution between the feature and the labels:

$$P(v, h_1, h_2, \dots, h_l) = P(v | h_1)P(h_1 | h_2), \dots, P(h_{l-2} | h_{l-1})P(h_{l-1} | h_l), \quad (10)$$

where $P(h_k | h_{k+1})$ is the conditional probability distribution of h_k for the given h_{k+1} state; $P(h_{l-1}, h_l)$ is the joint probability distribution of h_{l-1} and h_l . $P(v, h)$ is the joint probability distribution of a single RBM. The hidden layer of low-level RBM in DBN is the visual layer of high-level RBM. So (10) is the probability distribution for the whole model.

The use of DBN to establish a deep learning-based cardiovascular disease prediction model is an important entry point to solve the problem of accuracy and stability of prediction models.

3.2. Phase 1: Forecasting Model Based on Deep Belief Network.

The use of deep trust network to establish a cardiovascular disease prediction model is divided into two stages, as shown in Figure 2, respectively, upward training and downward adjustment.

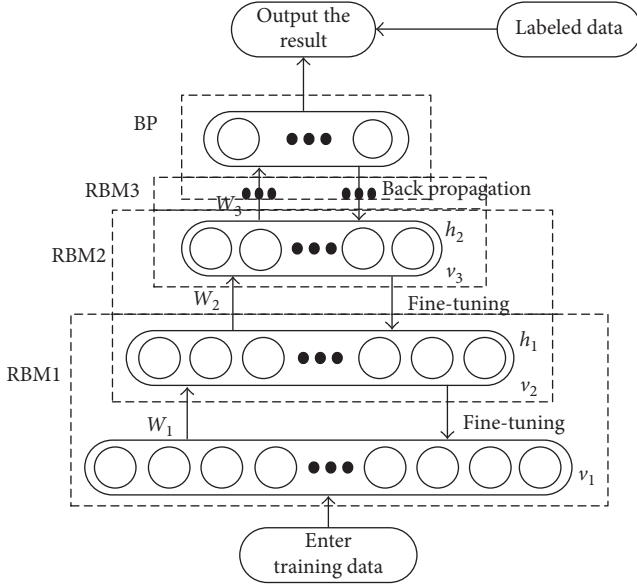


FIGURE 2: DBNs training flow chart.

- (1) Training section: use the greedy layer-by-layer training algorithm to learn the parameters of each layer of RBM $\theta = \{a, b, w\}$ in turn by unsupervised learning. First, the training data are received by the visible layer of the first layer RBM, and the state v_1 is generated. The hidden state h_1 is generated upwards by the initialized weight matrix w_1 , and the visible layer state v'_1 is reconstructed by h_1 . Generating new hidden units, the new layer is generated by w_1 remapping to the hidden unit h'_1 . The parameters are updated using the CD algorithm until the reconstruction error is least, that is, to complete the first layer RBM training. Stacked RBMs are trained layer by layer according to greedy learning rules, each layer maps different feature spaces. The topmost RBM bidirectional connections make up the associative memory layer, which can be associated with the optimal parameters of memory layers. By unsupervised learning, the DBNs gains a prior knowledge, obtains more abstract features at the top level, and better reflects the real structure information of the training data. Stacked RBM pre-training input is as follows: training data x , DBN; and output is as follows: unsupervised DBN.
- (2) Tuning section: taking the pre-trained parameters of the network as initial values, the labeled samples are used to supervise the DBN model and the top-down reverse propagation error of the network is used as the standard to further optimize the RBM parameters of various layers. The initial value of BP network is the high abstract feature set obtained by the pretraining of DBN, which solves the problem of falling into local optimum and overfitting caused by random initialization of the traditional neural

network. The parameters are finetuned based on the BP algorithm, and the input is the parameters of each layer of the DBN pretraining and the output vector of the top RBM; the output is the DBN after finetuning the parameters.

Through the above steps, a globally optimal DBN model is constructed and fully trained. To sum up the above learning phase, a complete DBN model is established, and the input is as follows: number of DBN structure layers, training samples; output is as follows: fully trained DBN.

Cardiovascular disease training samples without label values were entered into the visible layer of the bottom RBM without any characteristics of supervised learning data. The top RBM will learn the optimal characteristic parameters as the initial value of the neural network solves the defects caused by random initialization and improves the stability of the model prediction.

3.3. Phase 2: Improved Deep Belief Network Forecasting Model. The more complex the network structure of DBN, the stronger the ability to solve complex problems. Simultaneously, the higher the number of network layers, the harder the training will be, the greater the training error accumulates, and the lower the correctness of the model [22]. In application, in order to establish suitable DBN structure for specific tasks, due to lack of corresponding theoretical support and effective training mode, the depth of network and the number of hidden units need to be set by experience, which leads to the deviation in the modeling process and the high cost [23].

Aiming at the problem of determining the number of layers of DBN, based on the reconstruction error of each RBM training, this paper improves the prediction model of deep trust network and establishes a DBN which can automatically select the network depth to improve the automatic analysis ability of the cardiovascular disease prediction model. Specific methods are as follows.

In each RBM, the input data of the visible layer are reconstructed and mapped to the hidden layer again, and the reconstruction error is calculated based on the difference between the reconstructed output data and the initial training data.

$$R_{\text{error}} = \frac{\sum_{i=1}^n \sum_{j=1}^m P_{ij} - x_{ij}}{nmp_x}, \quad (11)$$

where R_{error} denotes the reconstruction error, n denotes the number of training samples, m denotes the number of features in each group of samples, P_{ij} denotes the reconstructed value of RBM training sample per layer, x_{ij} denotes the true value of the training sample, and P_x denotes the calculation of the number of values.

In order to prevent the training data from overfitting or reconstructing large deviation of the data and at the same time to balance the training cost of the network model, when the difference between the two reconstruction errors is less than the present value, the depth accumulation is stopped.

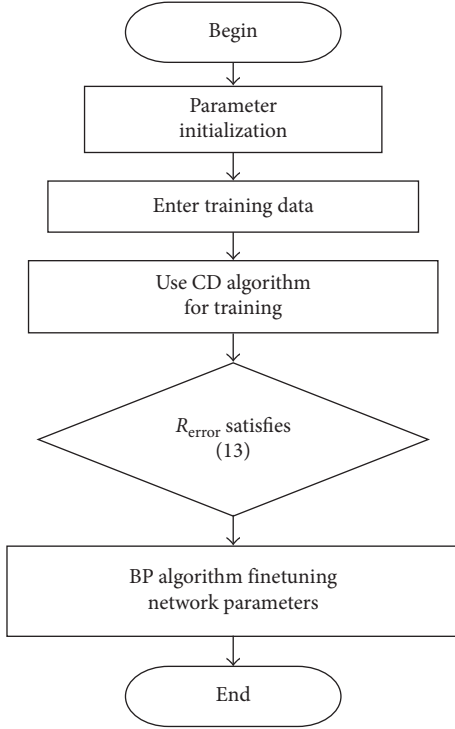


FIGURE 3: Calculation flowchart of DBNs depth.

$$\begin{aligned}
 L &= N_{\text{RBM}} + 1, \\
 |R_{\text{error}}(k-1) - R_{\text{error}}(k)| &> \varepsilon, \\
 L &= N_{\text{RBM}}, \\
 |R_{\text{error}}(k-1) - R_{\text{error}}(k)| &< \varepsilon,
 \end{aligned} \tag{12}$$

where L denotes the hidden layer number of DBN, $R_{\text{error}}(k)$ denotes R_{error} of current layer, and ε denotes the default value. The selection of the preset value is one of the keys to determining the accuracy of the model. The value of the default value ε is too large, which can cause inaccurately finding the optimal number of network layers. If the value is too small, the number of layers in the deep neural network may be too large and the calculation amount is too large. For the number of cardiovascular disease prediction model parameters and the performance of laboratory equipment, we determined that $\varepsilon \in [0.01, 0.05]$. Compared with many experimental results, when $\varepsilon = 0.03$, the prediction model can determine the network depth independently.

In the pretraining phase of the unsupervised, when it reaches the number of layers of target value, the top-level trained output is used as input of the BP algorithm and the reverse fine-tuning parameters are started. The process of building a network relies on R_{error} , as shown in Figure 3.

R_{error} is positively related to the network energy $E(v, h)$, and this coupling characteristic also proves feasibility of DBN depth with the reconfiguration error as the standard. It is proved as follows.

Let P be the calculated value, and X be the actual label value, then $P = P(v)$ and $X = P(v_1)$; according to the conditional probability formula, there is

$$P = P(v) = P(v_1)P(h | v_1)P(v | h). \tag{13}$$

According to the total probability formula, there is

$$P(v | h) = \frac{P(v, h)}{P(h)}. \tag{14}$$

According to (14), to rewrite (13), there is

$$\begin{aligned}
 P &= P(v_1) \frac{P(v_1, h)}{P(v_1)} \frac{P(v, h)}{P(h)}, \\
 &= P(v_1, h) \frac{P(v, h)}{P(h)}.
 \end{aligned} \tag{15}$$

According to (14) again, there is

$$\begin{aligned}
 P &= P(v_1 | h)P(h) \frac{P(v, h)}{P(h)}, \\
 &= P(v_1 | h)P(v, h).
 \end{aligned} \tag{16}$$

Substituting the above formula in (11) to reconstruct the error:

$$\begin{aligned}
 R_{\text{error}} &= \frac{\sum_{i=1}^n \sum_{j=1}^m (p_{ij} - x_{ij})}{nm p_x}, \\
 &= P - X, \\
 &= P(v_1 | h)P(v, h) - P(v_1), \\
 &= P(v_1)[P(v, h) - 1].
 \end{aligned} \tag{17}$$

As the energy of the neural network is proportional to the probability distribution, that is, $P(v, h) \propto E(v, h)$, there is

$$R_{\text{error}} \propto P(v, h) \propto E(v, h). \tag{18}$$

Equation (18) shows that there is a coupling relationship between R_{error} and network mechanism, and it is reasonable to rely on reconstruction error to determine the network depth of DBN autonomously. The number of neurons in each layer also has an impact on the network. At present, there is a lack of a clear theory to prove that the appropriate number of cells is set and the improvement is achieved. The DBN structure focuses on the ability to determine the depth of a network, and the number of neurons in each layer is fixed.

4. Experiment Analysis

4.1. Database Description. Experimental data select the Statlog (Heart) data set and the Heart Disease Database data set for the UCI Machine Learning Library. The Statlog (Heart) data set contains 270 sets of instances and the Heart Disease Database data set contains 820 sets of instances. The properties of both data sets contain continuous, two-category, ordered multiclass, and unordered multiclass variables. As shown in Table 1, select the same 13 attributes and 1 classified label values in two data for experiments.

The physical meaning, data unit, and order of magnitude of each attribute in the selected data set are different and need to be normalized before the experiment. Text-based data are

TABLE 1: Data set attributes.

Features	Description	Data types	Normalization	Value
Age	Age	Continuous data	Min-max scaling	16–80; 0~1
Sex	Gender	Text-based data	Direct mapping	0: female 1: male
Cp	Chest pain type	Text-based data	Direct mapping	0: typical angina 1: typical type angina 2: nonangina pain 3: asymptomatic
Trestbps	Trest blood pressure	Range data	Improved min-max scaling	MmHg on admission to the hospital
Chol	Serum cholesterol	Range data	Improved min-max scaling	(mg/dl)
Fbs	Fasting blood sugar	Hierarchical data	Hierarchical mapping	0: <120 mg/dl 1: >120 mg/dl
Restecg	Resting electrographic results	Text-based data	Direct mapping	0: normal 1: having ST-T wave abnormality 2: showing probable or definite left ventricular hypertrophy
Thalach	Maximum heart rate achieved	Range data	Improved min-max scaling	—
Exang	Exercise-induced angina	Text-based data	Direct mapping	0 = no 1 = yes
Oldpeak	ST depression induced by exercise relative to rest	Range data	Improved min-max scaling	
Slope	Slope of the peak exercise ST segment	Text-based data	Direct mapping	0: unsloping 1: flat 2: downsloping
Ca	Number of major vessels colored by fluoroscopy	Text-based data	Direct mapping	0–3
Thal	1: fixed defect 2: reversible defect	Text-based data	Direct mapping	0: normal
Num	Predicted attribute	—	—	0, 1, 2, 3, 4

directly converted to numeric data. The reference standard for medicine is the data attribute of the hierarchical classification structure. The normalized assignment is the corresponding discrete arithmetic progression or geometric progression. For the data attributes of the range type, we proposed improved min-max normalization due to the existence of data imbalances: take the average of the first k large values of the feature term as the maximum value, and take the average of the first k small values as the minimum value. The feature item is normalized to the interval (0, 1) as min-max.

In the two data sets, 70% of the instances are selected as training samples, and the remaining instances are test samples. The data set is divided into two mutually exclusive collections, and the consistency of data distribution is maintained as much as possible.

4.2. Improved DBN Model Network Depth Analysis Experiment

4.2.1. Improved DBNs Model Experiment. Using training data, improved DBN models are built and tested with test

data. Inputs are as follows: training sample risk factor data $\{x\}$, training sample tag value $\{y\}$, and testing samples $\{x', y'\}$; the output is the forecast results. The steps are as follows:

- (1) Set the initial value of the network, the learning rate is set to 1, the initial error is 0, the setting error of the reconstruction error ϵ is set to 0.03, the maximum training period of each RBM is set to 10 times. The weight (w), the visible layer bias (a), and the hidden layer offset (b) are all randomly generated values that are smaller, and the training batch is set to 100.
- (2) The training data $\{x\}$ with the label value removed is input as the first layer network and the unsupervised pretraining phase is started. The number of neurons in the input layer automatically takes the value of the sample feature dimension, that is, 13 risk factors in the data set. Perform the following steps using Gibbs sampling and CD algorithms, as shown in Table 2.

Update the parameters and calculate the error and repeat the above steps until the end conditions are met. In this case, the first layer of RBM is trained, and the principle of

TABLE 2: Algorithm implementation steps.

$h_1 = \text{sigm}(v_1^T w_1 + b)$	Use the input data to construct a hidden unit state
$v'_1 = \text{sigm}(h_1^T w_1 + a)$	Reconstruct the input using the hidden layer structure
$h'_1 = \text{sigm}(v_1'^T w_1 + b)$	Construct the hidden layer with the reconstructed input again

reconfiguring the error method to determine the depth of the network is used to calculate whether the condition is met; if it is satisfied, it stops; if it is not, h'_1 is used as the input for the next layer of training.

- (3) Use step (2) to determine the final depth of the network, and remember the optimal parameters of each layer. The trained DBN structures and the parameters are passed to the BP network to build the same depth of backpropagation network.
- (4) The top RBM output for the BP network input, while inputting the training data label value, began to monitor the tuning phase and further adjust the parameters of the DBN layers.
- (5) Put the unlabeled test data into the constructed improved DBN, and compare the value of the label value of the network to the true label value to calculate the prediction accuracy.
- (6) The algorithm ends.

4.2.2. Standard DBNs Model Experiment. In order to improve the correctness of the network depth determined by DBN autonomously, a standard DBN is established and the optimal network layer number is determined by experiment. The optimal number of cells in each layer is experimentally selected according to

$$N = \lceil \sqrt{mn} \rceil + k, \quad (19)$$

where m denotes the dimension of the input data, that is, the number of CVD risk factors; n denotes the number of output layer units and CVD predicts the probability as the output, that is, $n = 1$; N is the number of hidden units; $\lceil \cdot \rceil$ is the uplift symbol; and k is an integer between $[1, 5]$, which is used to increase the interval of units selection and avoid blind selection.

4.2.3. Experimental Results and Analysis. The improved DBN prediction model was tested in two data sets and stopped increasing when Statlog (Heart) was added to the third layer, with a depth of 4; the Heart Disease Database stopped increasing when it increased to the fourth level with a model depth of 5. The R_{error} curve in the RBM computing process of each layer is shown in Figures 4 and 5.

In order to improve the performance of DBN, a standard DBN model with the same structure was established, that is, a 4-layer neural network was established for Statlog (Heart) and a 5-layer neural network was established for the Heart Disease Database. The number of network units per layer was based on (19), and the best number of units is selected by the

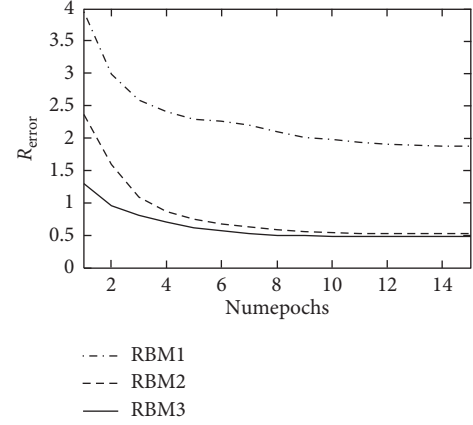
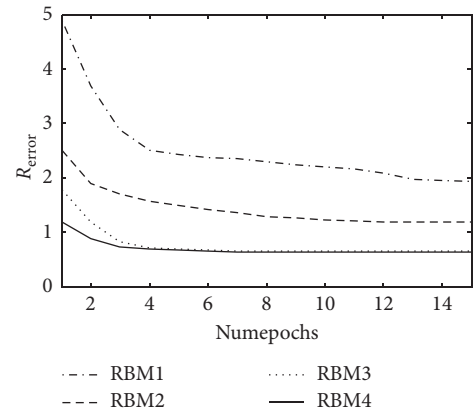
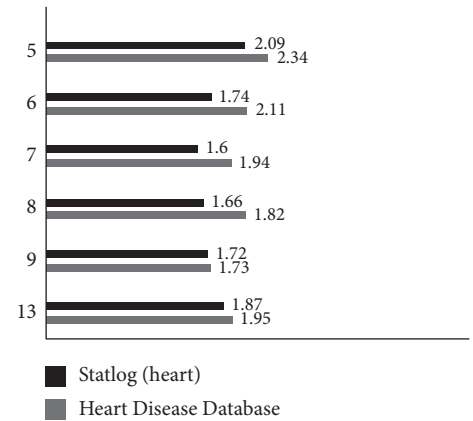
FIGURE 4: The R_{error} curve of each layer RBM on Statlog (Heart).FIGURE 5: The R_{error} curve of each layer RBM on Heart Disease Database.

FIGURE 6: Reconstruction error of RBM1 with different number of hidden units.

experimental method. The number of input layer units is equal to 13 feature latitudes of the data set, that is, $m = 13$; the network output is a label probability obtained by regression calculation, that is, $n = 1$; and the number of second layer units ranges from 5 to 9 experiments to select the smallest reconstruction error as the optimal unit number, the number of units under the reconstruction error shown in Figure 6.

TABLE 3: DBNs test results at different network depths.

Data set	Network depth	R_{error}	Accuracy (%)	Runtime (s)
Statlog (Heart)	2	1.8700	81.43	17.30
	3	0.5189	85.71	19.90
	4	0.4912	91.26	22.70
	5	0.4306	87.14	25.20
	6	0.3877	82.58	29.50
Heart Disease Database	2	1.9450	80.40	21.40
	3	1.1931	83.20	22.90
	4	0.6594	85.60	25.20
	5	0.6372	89.78	27.80
	6	0.5765	87.20	31.50

TABLE 4: Comparison of classification results of different techniques on the Statlog (Heart) data set.

Classification algorithms	Accuracy (%)
SVM linear [14]	86.62
SFM [24]	83.31
OCSFM [24]	86.73
Naïve Bayes [13]	78.93
WAC [13]	84.00
ANN [14]	86.04
PSO-SVM-SMO-RBF [14]	88.24
Naïve Bayes [25]	82.31
Decision tree [25]	84.35
LVQNN [17]	74.12
BPNN [18]	85.00
Improved DBN	91.26

As shown in Figure 6, the R_{error} of RBM1 in the Statlog (Heart) data set is the smallest at the 7th implicit unit, and the number of units is determined to be 7. The Heart Disease Database has the smallest R_{error} at the 9th implicit unit, and the number of units is determined to be 9. Similarly, the DBN structure finally determined according to the above method is Statlog (Heart): 4-layer network, the number of units of per layer is 13-7-6-4; Heart Disease Database: 5-layer network, the number of units of per layer is 13-9-8-5-4.

To further improve the correctness of the network depth determined by DBN, we increase the hidden layer number of the standard DBN model in Figure 6.

Reconstruction error of RBM1 with different numbers of hidden units turns and judges the correctness of the test data. To ensure that the number of layers is the only independent variable, the number of units in each layer is the same as that of the improved DBN model. The results are shown in Table 3.

Analysis of Table 3 shows that increasing the network hierarchy reduces R_{error} and training time will increase. The accuracy of the test data was maximized for Statlog (Heart) at depth 4, maximum for the Heart Disease Database at depth 5, and in line with the improved network depth that DBN automatically determines; it further proves that the prediction model of cardiovascular diseases based on improved DBN has better performance.

Table 4 presents the overall results of the proposed Statlog (Heart) data set evaluation using the UCI Machine

Learning Library for the proposed improved DBN prediction model and other different hybridization and non-hybrid techniques for cardiac classification and identification of relevant risk factors.

From the comparison of the tables, we can see that the traditional feature extraction algorithm is more specific to a specific data set. Based on the experimental accuracy rate, a special manually set feature combination is used. This method is to dig out the characteristics of the data set itself, not the essential characteristics of ECG data; the generalization ability of the method is weak, the portability is poor, and the accuracy is relatively poor.

The traditional classification model based on probability uses a combination of multiple feature extraction methods. However, the deep learning method can learn a kind of deep-level nonlinear network structure and can effectively obtain the deep-level essential feature representation of ECG from the sample. The effectiveness of the model based on deep learning is better than that of the traditional classification models based on probability and shallow neural networks.

This paper constructs a deep confidence network which can independently determine the network structure. The performance of the model is evaluated on two data sets, and the highest accuracy is achieved. The algorithm has strong generalization ability, and it can fully tap the deep-level characteristics of ECG and achieve an accurate and stable automatic classification of cardiovascular diseases in complex individuals and complex environments. The performance of heart disease classification is superior to other technologies.

5. Conclusion

For these issues, the probabilistic-based predictive model cannot integrate multiclass and nonlinear factors, and the stability of shallow neural network is poor. A prediction model based on deep learning is proposed and improved to enable it to independently determine the network parameters. The proposed prediction model was validated with the Statlog (Heart) data set and the Heart Disease data set, which proves that the prediction model has high accuracy and good stability.

Our further research is to apply the prediction model based on improved depth learning to actual cardiovascular disease predictions. By analyzing the prediction results in detail, we can quantify the proportion of each risk factor to the risk of cardiovascular disease and provide personalized advice to reduce the risk of cardiovascular disease.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by the National Natural Science Foundation of China, under Contract 60841004, 60971110, 61172152, and 61473265; the Program of Scientific and

Technological Research of Henan Province, China, under Contract 172102310393; the Support Program of Science and Technology Innovation of Henan Province, China, under Contract 17IRTSTHN013; the Key Support Project Fund of Henan Province, China, under Contract 18A520011; the Fund for “Integration of Cloud Computing and Big Data, Innovation of Science and Education,” China, under Contract 2017A11017; and the CERNET Innovation Project, China, under Contract NGII20161202.

References

- [1] W.-W. Chen, R.-L. Gao, and L.-S. Liu, “China cardiovascular diseases report 2015: a summary,” *Journal of Geriatric Cardiology*, vol. 1, pp. 1–10, 2017.
- [2] N. Cohnj, “Prevention of cardiovascular disease,” *Trends in Cardiovascular Medicine*, vol. 25, no. 5, pp. 436–442, 2015.
- [3] M. J. Pencina, R. B. Agostino, M. G. Larson et al., “Predicting the 30-year risk of cardiovascular disease: the Framingham heart study,” *Circulation*, vol. 119, no. 24, pp. 3078–3084, 2009.
- [4] R. Voss, P. Cullen, H. Schulte et al., “Prediction of risk of coronary events in middle-aged men in the prospective cardiovascular munster study (PROCAM) using neural networks,” *International Journal of Epidemiology*, vol. 31, no. 6, pp. 1253–1262, 2002.
- [5] A. V. Bahena, M. Mejialavalle, and J. R. Ascencio, “Toward optimal parameter selection for the multi-layer perceptron artificial neural network,” in *Proceedings of the International Conference on Mechatronics, Electronics and Automotive Engineering*, pp. 103–108, Cuernavaca, Mexico, November 2013.
- [6] M. E. Pastell and M. Kujala, “A probabilistic neural network model for lameness detection,” *Journal of Dairy Science*, vol. 90, no. 5, p. 2283, 2007.
- [7] R. Deo, “Machine learning in medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.
- [8] A. Khumoyun, Y. Cui, and L. Hanku, “Spark based distributed deep learning framework for big data applications,” in *Proceedings of the International Conference on Information Science and Communications Technologies*, pp. 1–5, Tashkent, Uzbekistan, November 2016.
- [9] M. A. Alsheikh, D. Niyato, S. Lin et al., “Mobile big data analytics using deep learning and apache spark,” *IEEE Network*, vol. 30, no. 3, pp. 22–29, 2016.
- [10] B. M. Wilamowski, B. Wu, and J. Korniak, “Big data and deep learning,” in *Proceedings of the Jubilee International Conference on Intelligent Engineering Systems*, pp. 11–16, Budapest, Hungary, June 2016.
- [11] Q. Zhang, M. Lin, L. T. Yang et al., “Energy-efficient scheduling for real-time systems based on deep Q-learning model,” *IEEE Transactions on Sustainable Computing*, vol. 99, p. 1, 2017.
- [12] Q. Zhang, L. T. Yang, Z. Chen et al., “A survey on deep learning for big data,” *Information Fusion*, vol. 42, pp. 146–157, 2018.
- [13] N. A. Sundar, P. P. Latha, and M. R. Chandra, “Performance analysis of classification data mining techniques over heart disease database,” *International Journal of Engineering Science and Advanced Technology*, vol. 2, no. 3, 2012.
- [14] S. Iftikhar, K. Fatima, A. Rehman, A. S. Almazyad, and T. Saba, “An evolution based hybrid approach for heart disease classification and associated risk factor identification,” *Biomedical Research*, vol. 28, no. 8, pp. 3451–3455, 2017.
- [15] D. A. Coast, R. M. Stern, G. G. Cano et al., “An approach to cardiac arrhythmia analysis using hidden Markov models,” *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 826–836, 1990.
- [16] S. L. Melo, L. P. Caloba, and J. Nadal, “Arrhythmia analysis using artificial neural network and decimated electrocardiographic data,” *Computers in Cardiology*, vol. 27, pp. 73–76, 2000.
- [17] A. M. Elsayad, “Classification of ECG arrhythmia using learning vector quantization neural networks,” in *Proceedings of the International Conference on Computer Engineering & Systems*, pp. 139–144, Cairo, Egypt, 2009.
- [18] E. O. Olaniyi, O. K. Oyedotun, A. Helwan et al., “Neural network diagnosis of heart disease,” in *Proceedings of the International Conference on Advances in Biomedical Engineering*, pp. 21–24, Beirut, Lebanon, September 2015.
- [19] A. Mohamed, G. E. Dahl, G. Hinton et al., “Acoustic modeling using deep belief networks,” *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 1, pp. 14–22, 2011.
- [20] G. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [21] F. Asja and I. Christian, “Training restricted Boltzmann machines: an introduction,” *Pattern Recognition*, vol. 47, no. 1, pp. 25–39, 2014.
- [22] J. Schmidhuber, “Deep learning in neural networks: an overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [23] P. Guangyuan, C. Wei, and Q. Junfei, “Depth determination method of DBN network,” *Journal of Control and Decision*, vol. 30, no. 2, pp. 256–260, 2015.
- [24] A. Adeli and M. Neshat, “A fuzzy expert system for heart disease diagnosis,” in *Lecture Notes in Engineering and Computer Science*, Springer, Berlin, Germany, 2010.
- [25] V. Chaurasia and S. Pal, “Data mining approach to detect heart disease,” *International Journal of Advanced Computer Science and Information*, vol. 2, no. 4, pp. 56–66, 2013.

