

Application of BP Neural Network Model Optimized by Particle Swarm Algorithm in Predicting the Risk of Hypertension

Yan Yan

Guangdong Medical University

Rong Chen

Guangdong Medical University

Jian Xu

Guangdong Medical University

Jialu Huang

Guangdong Medical University

Ling Luo

Guangdong Medical University

Hao Liu

Guangdong Medical University

Weiyng Chen

Guangdong Medical University

Chunwen Lin

Guangdong Medical University

Yuanlin Ding

Guangdong Medical University

Danli Kong

Guangdong Medical University

Qiaoli Zhang

Preventive Medicine and Hygienics, Dongguan Center for Disease Control and Prevention, Dongguan 523808, Guangdong

Haibing Yu (✉ hby616688@gdmu.edu.cn)

Guangdong Medical University

Research Article

Keywords: Hypertension, BP neural network, particle swarm algorithm optimization, mean impact value, logistic regression

Posted Date: April 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-426673/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The BP neural network was optimized by particle swarm optimization algorithm (PSO), and the PSO-BP neural network model was constructed. The prediction effect of the model was evaluated comprehensively by comparing it with BP neural network model and Logistic regression model. Based on PSO-BP model, the mean impact value algorithm (MIV) was used to screen the risk factors of hypertension, and the disease risk prediction model was established. In the evaluation of fitting effect, the root mean square error and determination coefficient of PSO-BP neural network are 0.09 and 0.29, respectively. In the prediction performance comparison, the accuracy, sensitivity, specificity and area under the ROC curve of PSO-BP neural network were 85.38%, 43.90%, 96.66% and 0.86, respectively. The results show that the BP neural network optimized by particle swarm optimization has the best fitting effect and prediction performance. The MIV algorithm can screen out the risk factors related to hypertension, and then construct the disease prediction model, which can provide a new idea for the analysis of hypertension.

1. Introduction

As one of the most common diseases in the world, hypertension and its complications occupy the first cause of death in the world, and it is called the "silent killer". Hypertension is a key factor determining the degree of atherosclerosis, which increases the risk of cardiovascular and cerebrovascular diseases, poses a great threat to the health of the population, and seriously affects the quality of life of patients with hypertension. Hypertension accounts for almost half of the morbidity and mortality of cardiovascular diseases worldwide^[1]. With the improvement of people's living standard and the change of lifestyle, the proportion of global disease burden of hypertension continues to rise. In 2025, it is estimated that 1.6 billion adults will suffer from hypertension^[2]. The high prevalence costs a lot of financial and material resources, which makes the health system face management challenges and causes a large economic burden^[3]. Before the occurrence of hypertension, it is the most effective and economical way to prevent or delay the occurrence of the disease through lifestyle and diet intervention^[4], but the factors affecting hypertension have not been fully defined, and the measures for the prevention and treatment of hypertension are still not ideal. Disease prediction models can help identify high-risk groups of diseases and screen out the main potential risk factors that affect the occurrence and development of diseases. Therefore, it is particularly important to carry out relevant studies that can accurately identify potential risk factors, protective factors and the construction of risk prediction model of hypertension, which will play a beneficial role in the prevention and control of hypertension disease.

So far, most of the disease risk models constructed by domestic and foreign scholars are based on several risk factors to predict the disease, or through logistic regression analysis or Cox proportional risk regression analysis to further determine the category of risk factors in the pre-screening risk factors, and establish multivariate statistical models to better predict the disease risk. Some of the classic models have been widely applied. For example, in cohort studies, the classic Framingham model can be used to evaluate the short-term risk of hypertension and has a good predictive ability^[5]. Genetic risk score (GRS) can be used to evaluate the relationship between genetic single nucleotide polymorphisms and hypertension^[6]. Based on the comprehensive effects of known lifestyle factors on overall mortality and attributable mortality in Chinese men, a model of the impact of risky lifestyle behavior patterns on disease was constructed^[7]. However, it should be noted that due to the existence of collinearity among the influencing factors, it will increase the difficulty of model fitting and lead to certain errors in the application of classical models, thus exposing some limitations of traditional methods.

In the era of big data, artificial intelligence (AI) methods such as machine learning algorithms to parse complex, multi-dimensional and multi-scale data have been increasingly applied in disease prediction models. Machine learning algorithm can imitate human thinking process to learn and store knowledge^[8], and has better accuracy and predictive ability compared with traditional regression model^[9]. In terms of complex model fitting and distribution approximation, BP neural network has incomparable advantages over traditional statistical methods^[10], and can perform well even when the sample size is small. However, when the sample size increases, the training time of the algorithm will become long and its efficiency will decline sharply^[11]. In addition, there may be defects such as unsatisfactory network generalization when training neural networks, resulting in local minima and slow convergence speed, especially the weights and deviations in the process of fine-tuning the control parameter set^[12]. Therefore, optimization of the algorithm can be considered to improve the parameters of BP neural network. Particle Swarm Optimization (PSO: Particle Swarm Optimization) is a method of optimizing continuous non-linear functions. Adjustable parameters are modified to make the error between the predicted output and the expected output as small as possible. Its advantage is that without adjusting redundant parameters, it can significantly improve the performance of the recognition task and realize the optimization of the structure of the deep neural network^[13]. In addition, BP neural network has a certain randomness in the calculation process, which has a poor limitation on the objective function, and has the characteristics of strong local search ability and poor global search ability. PSO can find the optimal solution closest to the global through the collaboration and information sharing between individuals in the group. The global search ability is strong but the local search ability is poor^[14]. Therefore, through the construction of a combination model, it is expected to overcome the limitations of each other^[15, 16].

Based on the above, this research will try to complete the construction of the hypertension risk model by constructing a PSO-optimized BP neural network, and further use the mean impact value (MIV) algorithm to screen out the hypertension in the PSO-BP neural network model Complete the

risk factors analysis of the risk of hypertension in Guangdong area, and provide new ideas for the prevention and control of the occurrence and development of the public health problem of hypertension.

2. Methods

2.1 Research object

Based on the monitoring data of chronic diseases and their risk factors in Guangdong Province from 2017 to 2019, this study adopted cluster random sampling, and finally included 3012 subjects (cross-sectional subjects, including normal population and hypertension patients) after data cleaning. Among them, the hypertension diagnosis was based on the hypertension diagnostic criteria published by WHO/ISH in 1999, that is, patients with hypertension whose systolic blood pressure $\geq 140\text{mmHg}$ and/or diastolic blood pressure $\geq 90\text{mmHg}$ [17]. This study was approved by the Ethics Committee, and the scheme was implemented in accordance with the relevant guidelines and regulations. All studies were conducted with the consent and informed consent of the subjects.

2.2 Data collection and pretreatment

By consulting the literature and combining the monitoring data of the research subjects (including on-site questionnaire survey and physical examination, etc.), 17 independent variables (X_1-X_{17}) were initially screened out, namely gender, age, education level, occupation, smoking status, and alcohol consumption Condition, sleep duration, BMI index, heart rate, blood sugar, hemoglobin, cholesterol, triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, daily salt intake and daily oil intake, and whether you have high Blood pressure as the dependent variable (Y).

Preprocessing the data, verifying the measurement units and measurement standards of the data, collating and cleaning the data, and excluding missing values and outliers. For missing data such as missing values, too much missing information, and unavailable data caused by temporary inaccessibility of information and human factors, the missing value will reach 70% as the limit for elimination and mode filling (categorical variables) or mean filling (Continuous variables) processing. Before model construction, Min-Max standardization (also called dispersion standardization) is used to normalize the data, so that the original data is transformed into standardized data falling in the interval [0,1]. In the process of model construction, the "CV partition" function was used and the test set proportion test size = 0.2 was set. 80% of the samples were randomly selected as the training set (2410×17 dimensions) for the training classification model, and 20% as the test set for the evaluation of the model. Among them, 316 and 2094 people in the training set were hypertensive and 2094 people in the non-hypertensive, and 56 and 546 people in the test set were respectively selected.

2.3 Establish hypertension prediction model

After data preprocessing, BP neural network was constructed through MATLAB, and particle swarm optimization was used to adjust weights and thresholds, and the optimal weights and thresholds were assigned to BP neural network to build a prediction model of BP neural network optimized by PSO. In addition, the number of nodes in the input layer of BP neural network is determined by the input data (input vector X_1-X_{17}), the number of nodes in the output layer is determined by the output data (output vector Y), and the hidden layer is determined by the trial and error method and the connection weight is determined. In this paper, the number of cells in the hidden layer is 7 layers. Input all samples into the network through batch training, calculate the network error through the expected output value and predicted output value of the network, and define whether to converge with an error tolerance of 0.01, and complete the adjustment and update of the weight and threshold through the momentum factor (0.8). And then calculate the test output, and round the result to calculate the accuracy, specificity, sensitivity, ROC and other evaluation parameters of the test set. In the construction of PSO-BP neural network, after data preprocessing, the optimal weights and thresholds are obtained by setting particle swarm parameters, individual fitness evaluation and particle swarm updating iteration, and the optimal weights and thresholds are assigned to BP neural network for training and testing. Other steps are similar to the construction of BP neural network. In this study, the BP neural network parameter configuration is shown in Table 1, and the PSO-BP neural network parameter setting is shown in Table 2. In addition, the construction of the Logistic model was completed through SPSS 21.0 version.

Table 1
Parameter description of BP neural network prediction model

The training parameters	Parameter selection
Number of neural network layers	3
Number of neurons in the input layer	17
Number of hidden layer neurons	7
Number of neurons in the output layer	1
Input layer activation function	tansig
Output layer activates the function	purelin
The training function	trainlm
Displays the maximum number of training steps	50
Maximum number of iterations	1000
Learning rate	0.01
Gradient target value	1e-7

Table 2
Parameter description of PSO-BP neural network prediction model

The training parameters	Parameter selection
Number of neural network layers	3
Number of neurons in the input layer	12
Number of hidden layer neurons	5
Number of neurons in the output layer	1
Input layer activation function	tansig
Output layer activates the function	purelin
The training function	trainlm
Velocity upper bound in particle swarm optimization	0.5
Target error in particle swarm optimization	1e-7
Maximum speed weight	0.95
Minimum velocity weight	0.25
Maximum number of iterations	10
Local influence factor	1.5
Global impact factor	1.5

2.4 Evaluation of prediction models

In this paper, the root mean square error (RMSE) and correlation coefficient (R^2) are used to evaluate the fitting ability of the model; and the accuracy, sensitivity, specificity and area under the ROC curve (AUC) between the models are used to evaluate the prediction effect of the prediction model.

2.5 Risk factor screening and construction of hypertension risk prediction model

The MIV algorithm of the BP neural network model optimized by PSO is used to screen the risk factors of hypertension in Guangdong from 2017 to 2019, thereby constructing a hypertension risk prediction model.

2.6 Statistical analysis

The mathematical software Matlab R2019b version produced by MathWorks of the United States is used to complete the construction of the BP neural network model and the optimization of the BP neural network model by the particle swarm, as well as the data analysis and the screening of risk factors. Through SPSS 21.0 version, the construction of Logistic regression model and the comparison of related models were completed

by binary multi-factor Logistic regression analysis, and the risk factors selected by stepwise regression were entered into the Matlab training set to evaluate the predictive performance of Logistic regression.

3. Results

3.1 BP neural network

3.1.1 Parameters

When it is run for 100 times to obtain the optimal performance, the corresponding parameters are shown in Fig. 1. The weights and thresholds from the input layer to the hidden layer, and the weights from the hidden layer to the output layer are shown in Table 3 and Table 4, respectively. The threshold from the hidden layer to the output layer is -0.6783.

Table 3
Weight from input layer to hidden layer of BP neural network

weight	Hidden layer						
The input layer	0.2425	-0.0225	-0.3058	0.1332	0.3879	-0.3369	-0.5324
	0.1018	0.9289	1.0490	-0.2937	0.4603	1.2348	0.0564
	0.2204	-0.3823	-0.3731	0.3296	0.4754	0.4274	0.2667
	-0.6608	0.8313	0.4074	-0.4011	0.4486	0.4611	-0.4757
	0.2033	-0.2305	0.2382	-0.7186	-0.3592	-1.1651	-0.4862
	0.3126	0.0743	-0.0046	0.5328	-0.3317	0.0247	-0.2104
	0.3772	0.1501	-0.1222	-0.1604	-0.0278	0.2180	0.05957
	0.4167	-0.1376	0.9324	0.4318	-0.2344	-0.1270	0.6619
	0.5721	-0.1387	0.5544	0.0553	0.6000	0.0779	0.1295
	0.1738	0.6132	0.4533	0.2727	-0.4571	-0.4711	-0.4056
	0.3328	-0.2157	0.1966	-0.0745	0.4474	-0.1763	-0.5612
	-0.3460	0.1834	-0.1201	-0.0874	0.3419	0.0995	-0.0551
	0.4276	-0.3685	0.3851	0.2863	-0.1547	0.4086	-0.7055
	-0.5980	-0.1587	0.2343	0.7824	-0.3389	0.0544	0.1300
	0.5251	-0.4328	0.1275	0.2206	-0.1657	0.1700	-0.1773
	0.0812	-0.0831	0.3192	0.4543	0.2891	-0.1361	-0.5478
	-0.3301	0.6049	-0.2861	0.5753	0.0800	0.5621	-0.7337

Table 4

The threshold from the input layer to the hidden layer and the weight from the hidden layer to the output layer

Input layer to hidden layer threshold	-1.5158	-1.0760	0.5800	-0.0877	0.6017	-1.0622	-1.4430
Weight from hidden layer to output layer	0.3514	-0.8881	0.7184	-0.6160	0.0342	0.8585	0.0963

3.1.2 Model training effect

(1) BP neural network effect display

In the model training process of BP neural network, with the gradual increase of the number of iterations, about 50 steps, the model prediction error of validation set and test set tends to steady state gradually, and the model prediction error of training set decreases gradually. In the green circle of the figure (about 4 steps), although the model error of the training set is relatively large, the model prediction and expectation of the verification set and the test set are relatively close, and the minimum MSE is 0.11156, indicating that the model at this time will no longer have over-fitting or under-fitting phenomenon, and the model reaches the optimal level (Fig. 2).

(2) BP neural network effect display - training state

The BP neural network showed gradient descent in training, with a learning rate of 0.0001 and an effective test number of 996 stops, indicating that the error curve of confirmed samples did not decline for 996 consecutive iterations in the training process of the network using training samples (Fig. 3).

(3) Effect display of BP neural network - network fitting ability

The correlation coefficient of the BP neural network's training set is 0.41866, the correlation coefficient of the verification set is 0.33997, the correlation coefficient of the test set is 0.39906, and the total correlation coefficient is 0.40259, indicating that the correlation degree is low positive correlation, and the lines are not on the diagonal, which comprehensively indicates that the fitting ability of the BP neural network is relatively general, and there is no over-fitting phenomenon (Fig. 4).

3.1.3 Prediction performance of BP neural network

In this study, the ROC curve area of BP neural network is $AUC = 0.76103$, within the range of [0.7, 0.85], indicating that the effect of BP neural network model is general (Fig. 5). The comparison graph of prediction effect of BP neural network on test set shows the real category and prediction category of BP neural network on test set. The more overlap, the better the model effect (Fig. 6). For the convenience of analysis, the model confusion matrix of the test set is calculated (Table 10). Since there is no over-fitting phenomenon in BP neural network in this study, the prediction evaluation was completed through the test set. In the test set, 17 of the 28 positive cases (hypertension) were correctly predicted and 11 were wrongly judged, and the correct rate reached 60.71%. Of the 574 negative cases, 468 were correctly predicted and 106 were wrongly judged, with a correct proportion of 81.53%. Therefore, the overall accuracy of training set prediction is 80.56% (Table 11).

3.1.4 Risk factors for MIV assessment in BP neural network

Risk factors were evaluated by MIV in BP neural network. The absolute value of MIV value was greater than 0.02 as the screening limit, and the average value of multiple runs was taken. After feature selection, risk factors were ranked (from heavy to light) according to MIV weight, and the risk factors affecting hypertension were obtained. If the value of MIV is positive, the independent variable is positively correlated with the dependent variable; otherwise, it is negatively correlated. Factors positively associated with hypertension included low-density lipoprotein cholesterol, cholesterol, triglyceride, daily oil intake, daily salt intake, smoking, age, BMI, and alcohol consumption, and negatively associated factors included sleep duration, heart rate, high-density lipoprotein cholesterol, cholesterol, and hemoglobin (Table 5).

Table 5
MIV screening results in BP neural network

The serial number	dependent variable	MIV value	weight
1	LDL-C	0.022848	0.16202
2	Cholesterol	0.0092056	0.10992
3	The sleep time	-0.010039	0.10027
4	triglycerides	0.0017972	0.079886
5	Daily oil intake	0.007317	0.078942
6	Daily salt intake	0.0089765	0.074292
7	Smoking	0.0062768	0.060133
8	Age	0.0072	0.053676
9	BMI index	0.0069582	0.052572
10	Blood glucose	-0.0012076	0.043836
11	Heart rate	-0.0053262	0.040962
12	Gender	0.0011152	0.036111
13	HDL-C	-0.0045044	0.033973
14	Drinking	0.0034115	0.028692
15	Hemoglobin	-0.0028802	0.021162
16	Occupation	-0.0010221	0.018237
17	Level of education	-0.000166	0.0053088

3.2 PSO-BP neural network

3.2.1 Parameters

When it is run for 100 times to obtain the optimal performance, the corresponding parameters are shown in Fig. 7. The weights and thresholds of input layer to hidden layer and the weights of hidden layer to output layer are shown in Table 6 and Table 7 respectively, and the threshold of hidden layer to output layer is 0.9853.

Table 6
The weights from input layer to hidden layer of PSO-BP neural network

weight	Hidden layer				
The input layer	93.7685	0.0164	-4.4586	17.9715	23.1298
	-4.0682	-0.5389	-146.7394	65.6651	-19.7487
	-52.1364	0.3170	2.8780	12.5227	50.8008
	-42.5850	0.2302	-18.1300	2.5535	-13.8177
	3.3955	-0.6073	2.1371	109.3244	4.9018
	-7.6784	0.0390	1.1806	-5.7198	-22.1686
	-149.1449	1.6124	-23.3165	10.0733	84.8294
	-6.1923	-0.9055	-57.5894	12.9795	-21.9905
	-93.2607	-0.7994	-9.9727	-12.3276	12.4312
	-18.4475	-0.9582	43.7620	49.2285	-26.0898
	45.7022	0.6678	-137.9323	41.0848	-40.5153
	25.2071	0.0700	41.5357	61.3890	-12.7826
	-69.8777	0.2628	10.6211	-43.1675	-28.9981
	-56.3734	0.9910	52.2766	-45.2839	11.0177
	125.2475	-2.6197	5.6143	8.7258	-25.9603
	-19.6131	0.1143	-115.8592	68.9152	-66.5466
	-30.6039	-0.0304	81.2053	-65.6769	19.6224

Table 7
PSO-BP neural network input layer to hidden layer threshold and hidden layer to output layer weight

Input layer to hidden layer threshold	13.2229	1.8184	-15.1179	-64.5822	18.1935
Weight from hidden layer to output layer	-0.2708	-0.8767	-0.1847	-0.1562	0.2266

3.2.2 Model training effect

(1) Effect display of PSO-BP neural network

During the model training process of the PSO-BP neural network, as the number of iterations increases to the initial maximum number of iterations of 5000 steps, the model prediction errors of the validation set and test set gradually become stable, and the model prediction errors of the training set gradually decrease. At this time, the minimum MSE is 0.085778, indicating that the model no longer has over-fitting or under-fitting, and the model reaches the optimal level (Fig. 8), but there are more iterations and long training time. In addition, with the gradual increase of iterations, the overall trend of fitness decreased. The fitness is used to describe the model error, and the smaller the fitness is, the smaller the model error is. After the fifth iteration, the current fitness and global fitness of the model begin to converge, and the error has converged. After the number of iterations is 10, the fitness is the minimum, about 0.104, and the optimal solution under this parameter is obtained (Fig. 9).

(2) PSO-BP neural network effect display-training status

The PSO-BP neural network presents gradient descent in training. Compared with the gradient descent curve of BP neural network, the fluctuation of the PSO-BP neural network descent curve is small. The learning rate is 0.0001 and the number of effective tests is 0 times, indicating that the error curve of confirmed samples does not decline at the beginning of the training process of the network using the training samples, which is relatively stable (Fig. 10).

(3) PSO-BP neural network effect display - network fitting ability

The correlation coefficient of PSO-BP neural network was 0.53627, indicating that the correlation degree was moderately positive(Fig. 11).

3.2.3 Prediction performance of PSO-BP neural network

The ROC curve area of PSO-BP neural network is 0.85815, within the range of [0.85, 0.95], indicating that the BP neural network model has a good effect (Fig. 12). The comparison figure of prediction effect of PSO-BP neural network test set shows the prediction category of real category and

BP neural network for test set. The more overlapping part, the better the model effect (Fig. 13). For the convenience of analysis, the model confusion matrix of training set and test set is calculated (Table 10). In the test set, 49 of the 63 positive cases (hypertension) were correctly predicted, 14 were wrongly judged, and the correct rate reached 76.56%. Of the 539 negative cases (non-hypertension), 465 were correctly predicted and 74 were wrongly judged, a correct ratio of 86.27%. Therefore, the overall prediction accuracy of the test set is 85.38%. According to the AUC, accuracy, sensitivity and specificity of the PSO-BP neural network model, the prediction effect of this model is good (Table 11).

3.2.4 Risk factors for MIV assessment in PSO-BP neural network

Risk factors were evaluated by MIV in the PSO-BP neural network. Taking the absolute value of MIV value greater than 0.02 as the screening limit, the average value of multiple runs was taken. After feature selection, risk factors were ranked (from heavy to light) according to MIV weight, and the risk factors affecting hypertension were obtained. Factors that are positively related to high blood pressure include cholesterol, low-density lipoprotein cholesterol, daily oil intake, triglycerides, daily salt intake, smoking, age, BMI index, and alcohol consumption. Factors that are negatively related include length of sleep, high-density lipoprotein cholesterol, hemoglobin (Table 8).

Table 8
MIV screening results in PSO-BP neural network

The serial number	dependent variable	MIV value	weight
1	LDL-C	0.027307	0.069858
2	Cholesterol	0.63461	0.61935
3	The sleep time	-0.021362	0.072334
4	triglycerides	0.011243	0.030626
5	Daily oil intake	0.0036201	0.032021
6	Daily salt intake	0.011187	0.029155
7	Smoking	0.0058964	0.026019
8	Age	0.0062726	0.015869
9	BMI index	0.0036371	0.012329
10	Blood glucose	0.00047164	0.01562
11	Heart rate	-0.0015596	0.0052576
12	Gender	-0.0054725	0.026225
13	HDL-C	-0.0034072	0.014852
14	Drinking	0.0026665	0.0085437
15	Hemoglobin	-0.0052897	0.013865
16	Occupation	-0.0015046	0.0053823
17	Level of education	-0.00060963	0.0026992

3.3 logistic regression model

(1) Establishment of risk prediction model

Logistic regression analysis was performed using SPSS software, and multivariate analysis of stepwise regression was used (the criteria for selection variables is $P < 0.05$; the criteria for elimination variables is $P > 0.10$), and 9 risk factors are screened out, namely gender, age, education level, smoking history, BMI index, hemoglobin, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and heart rate (Table 9).

Table 9
Results of dichotomous multivariate logistic stepwise regression analysis

Analysis of the factors	b	S.E.	Wals	Sig	Exp(b)	Exp(b) 95% C.I.	
						down	up
Gender	-0.799	0.161	24.667	0.000	0.450	0.328	0.616
Age	0.040	0.005	56.018	0.000	1.041	1.030	1.052
educational status	-0.175	0.048	13.383	0.000	0.839	0.764	0.922
Smoking	0.104	0.052	4.016	0.045	1.109	1.002	1.228
BMI	-0.153	0.017	81.518	0.000	1.165	1.127	1.205
Hemoglobin	0.007	0.004	2.986	0.084	1.007	0.999	1.014
HDL-C	-0.459	0.188	5.972	0.015	0.632	0.437	0.913
LDL-C	0.294	0.068	18.796	0.000	1.342	1.175	1.533
Heart rate	0.039	0.005	66.708	0.000	1.039	1.030	1.049

(2) Logistic regression prediction and evaluation

The early warning factors screened were taken as independent variables, and "whether patients with hypertension" was taken as dependent variable. MATLAB software was used to establish Logistic regression model. After running, it can be found that the area under the ROC curve of Logistic regression model is $AUC = 0.48749$, less than 0.5. Therefore, it can be inferred that the prediction performance of this model is worse than random guess, so it has no predictive value (Fig. 14). The comparison diagram of the prediction effect of Logistic regression model on the test set was further obtained (Fig. 15), and the model confusion matrix between the training set and the test set was statistically analyzed (Table 10). In the test set, 72 of the 287 positive cases (hypertension) were correctly predicted, 215 were wrongly judged, and the correct rate reached 25.08%. Among 315 negative cases (non-hypertension), 51 were correctly predicted and 264 were wrongly judged, with a correct rate of 16.19%. Therefore, the overall accuracy of training set prediction is 55.81%. According to the AUC, accuracy, sensitivity and specificity of Logistic regression model, the prediction effect of this model is not very good (Table 11).

Table 10. Model confusion matrix of each prediction model test set

Model		BP neural network		PSO-BP neural network		Logistic regression	
Actual value		Non-hypertensive	Hypertensive	Non-hypertensive	Hypertensive	Non-hypertensive	Hypertensive
Predictive value	Non-hypertensive	468TN	106(FN-type2)	465TN	74(FN-type2)	264TN	51(FN-type2)
	Hypertensive	11(FP-type1)	17(TP)	14(FP-type1)	49(TP)	215(FP-type1)	72(TP)

Table 11. Evaluation results of each prediction model

	BP neural network	POS-BP neural network	Logistic regression
Accuracy	80.56%	85.38%	55.81%
Sensitivity	13.82%	43.90%	58.54%
Specificity	97.70%	96.66%	55.11%
AUC	0.76103	0.85815	0.48749

4. Discussion

As a common disease in the population, hypertension is an important risk factor for heart failure, coronary heart disease, aortic dissection, stroke and other cardiovascular and cerebrovascular diseases, which has a great impact on the health of the population. As the proportion of the number of patients increased year by year, the disease burden caused by hypertension also increased year by year. The interaction and multicollinearity among the risk factors that may affect the incidence of hypertension may lead to errors in the fitting of traditional models. The emerging research of machine algorithm can provide more possibilities for the analysis of disease prevention and control. BP neural network has

great potential in the research of medical field, and the optimization of neural network can further improve its prediction and risk assessment performance, which has certain research significance.

In this study, a neural network prediction model of hypertension was established based on the monitoring data of chronic diseases and risk factors, and the POS algorithm was further used to complete the optimization of the neural network. Considering that too many neurons in the input layer will have higher requirements on the sample size, this study selected the independent variables preliminarily by referring to literatures and combining existing monitoring data, and took whether the patient had hypertension as the dependent variable to enter the neural network model. In order to prevent model overfitting and ensure test accuracy, the data set is divided into training set and test set in a ratio of 4:1. In addition, the establishment of neural network is flexible, and there is no uniform value of function and parameters in the process of establishment. Through comparison, it can be seen that the performance of BP neural network and BP neural network optimized by PSO is different during and after modeling. Therefore, after many times of training, the best prediction indexes in BP neural network and PSO-BP neural network were selected for horizontal comparison.

In the construction model, the iteration times and running time of BP neural network and PSO-BP neural network are significantly different. BP can achieve the best performance in a single iteration with four runs of about 10 seconds, while PSO-BP can achieve the best performance only when the iteration reaches the maximum number of times initially set, and the running time is about 500 seconds. However, in Zhang Yijun's study, the optimization algorithm will lead to shorter modeling time^[18], which is different from this study. This may be because when the two algorithms are combined to achieve local and global optimality, more iterations and runtimes are needed to achieve the optimal performance of the model. In addition, the error of PSO-BP neural network has converged in the 10th iteration, and the optimal solution of this parameter is obtained. However, after the combination, it needs to iterate 5000 times to get the lowest root mean square error. It is comprehensively verified that PSO-BP neural network may have poor local search ability and need more iteration times.

In the comparison of the fitting ability and the performance of the training data set constructed by the model, the root mean square error of BP and PSO-BP is 0.34 and 0.09 respectively. When evaluating the performance ability of network fitting, the total coefficient of determination of BP neural network is 0.16, and the coefficient of determination of PSO-BP is 0.29, which is closer to 1 than that of BP neural network, indicating that the data set of PSO-BP model is more correlated with the reality. This result is consistent with Liu Xin's research results in strain prediction of wind turbine blades by using PSO-BP neural network, that is, PSO-BP neural network has a small error but higher fitting ability^[19]. Through the comparison of root mean square error and determination coefficient, it can be concluded that PSO-BP nonlinear fitting ability is better. In addition, the gradient descent curve of PSO-BP neural network is more stable than that of BP neural network, which also proves that PSO-BP neural network has better stability.

In addition, in the prediction performance comparison, the data of each model does not change much in multiple runs. The optimal operating results of each model were selected for horizontal comparison, and the predictive ability of BP neural network, PSO-BP neural network and Logistic regression was observed. Combined with the current data, this study found that compared with Logistic, BP neural network and PSO-BP neural network were significantly improved in accuracy, specificity and AUC. In view of accuracy and specificity can measure the accuracy of prediction, and AUC, as a performance index to measure the advantages and disadvantages of the learner, can be used to judge the advantages and disadvantages of the prediction model. The above results show that the neural network has better prediction accuracy than the traditional model. Through the comparison of neural network algorithm, it can be found that the accuracy, sensitivity and AUC of PSO-BP neural network prediction model are improved after POS algorithm optimization. Overall, PSO-BP neural network has the best performance in prediction and diagnosis. This result is consistent with the results of prediction performance comparison between neural network and traditional model that are mostly discussed at present^[10, 18], indicating that PSO-BP neural network can also be well applied in the prediction of hypertension risk. However, it is undeniable that, unlike other studies, the sensitivity of the neural network in this study has been reduced, indicating that the diagnostic ability of the neural network constructed in this study has been reduced when predicting the risk of local hypertension. In conclusion, the neural network model has better adaptability and fitting effect for diseases such as hypertension, where there are many pathogenic factors and there may be interactions among various factors. Although the PSO-BP modeling time has been extended, the error is smaller, the correlation degree is higher, the nonlinear fitting ability is better, and the prediction performance is better. It indicates that the performance of BP neural network has been improved after optimization, and PSO-BP can be better applied to the study of hypertension risk.

In addition, since MIV algorithm is often used in engineering^[19], meteorology, circuit technology^[20]and other aspects to screen risk factors, it can complete the screening of influencing factors and has a good identification performance. In this study, MIV algorithm is further used to complete the screening of risk factors for hypertension diseases in Guangdong region in the neural network model. The greater the absolute value of MIV, the greater the influence of the influencing factors on hypertension; The greater the weight of MIV, the greater the influence of this factor on rank. However, there is no unified standard for how much the MIV value can be regarded as the influencing factor. Through literature review, combined with professional knowledge and comparative prior knowledge, the absolute value > 0.002 of the self-determined factor MIV in this study can be considered as the influencing factor of hypertension in this area, and the influencing factors are ranked according to the MIV weight. In the prediction of BP neural network, it can be obtained through the screening of MIV algorithm that the risk factors of the disease in Guangdong area

from heavy to light can be low-density lipoprotein cholesterol, cholesterol, sleep duration, daily oil intake, daily salt intake, smoking, age, BMI index, heart rate, high-density lipoprotein cholesterol, drinking, hemoglobin. The PSO-BP neural network analyzes the risk factors of the disease in the region, and the risk factors that affect high blood pressure are in order of cholesterol, sleep duration, low-density lipoprotein cholesterol, daily oil intake, triglycerides, Daily salt intake, gender (categorical variable), alcohol consumption, age, high-density lipoprotein cholesterol, hemoglobin, BMI index, alcohol consumption. Through comparison, it can be concluded that the selected risk factors obtained by MIV analysis under BP neural network and PSO-BP neural network model are different, and the weight of risk factors will also be changed, indicating that the algorithm optimization will produce differences in the establishment of prediction model.

It is worth noting that in comparison with the Logistic regression model for screening risk factors, the neural network MIV algorithm screening factors are different. The BP neural network screens cholesterol, sleep duration, daily oil intake, daily salt intake, and Drinking and other factors, while the more risk factors in the PSO-BP neural network are cholesterol, sleep duration, daily oil intake, triglycerides, daily salt intake, and alcohol consumption. However, the neural network lacks education as a risk factor, which shows that when the neural network-based MIV algorithm is used to screen risk factors in the analysis of hypertension, there will also be differences in the results of screening factors. Finally, after a comprehensive comparison between the fitting performance and the prediction performance, the PSO-BP neural network has the best performance. Based on this research, the PSO-BP neural network is finally selected as the prediction model for hypertension in Guangdong region.

Studies have shown that the risk factors for cardiovascular development are different in the elderly and the young [21]. Among the subjects in this study, age was a positively correlated risk factor, and the risk of disease increased with age. In addition, gender is also a risk factor, and the gender difference is significant. In this study, the incidence of female is lower than that of male. In Tao Hong's study of hypertension population, the incidence of hypertension in women is also lower than that in men, which may be related to the fact that women pay more attention to hypertension and have better compliance than men, and are more able to adhere to a good lifestyle than men [22]. Therefore, the prevention and control work in this region can be considered to take targeted measures for different groups of people to improve the effectiveness of publicity and the degree of concern.

Although the specific mechanism is unclear, more and more studies have shown that people with abnormal lipid indexes have a higher prevalence of hypertension than normal people [23]. The results of model screening in this study indicate that cholesterol is the primary positively correlated risk factor in this area, and the risk of hypertension will increase with the increase of cholesterol content. In addition, the risk of hypertension increased with increases in LDL cholesterol and triglycerides, and increased with decreases in HDL cholesterol and hemoglobin. This screening is consistent with multiple studies showing that cholesterol, LDL cholesterol, and triglycerides are associated with an increased risk of hypertension, while HDL cholesterol has cardiovascular benefits [24, 25]. Reduced hemoglobin usually indicates a relatively poor health status of the human body, which may lead to hypertension [26]. Studies have shown that the incidence of dyslipidemia can be effectively reduced by controlling body weight, blood sugar, blood pressure and consumption of meat products [27]. To sum up, lifestyle change and drug intervention can be considered to reduce the occurrence of dyslipidemia, so as to reduce the incidence of hypertension.

In terms of life risk factors, the sleep duration in this study model is the primary risk factor with negative correlation, indicating that the decrease of sleep duration will lead to the increase of the risk of hypertension. In recent years, more researchers have paid attention to the relationship between this risk factor and hypertension [28]. Studies believe that adequate sleep time is conducive to the control of blood pressure [29], and this study further supports this view. Based on this study, daily oil intake, daily salt intake, smoking, BMI and alcohol consumption were positively related factors, and the risk of hypertension increased with their increase. These risk factors overlap with those proposed in China's hypertension prevention and control guidelines [30].

In conclusion, based on the PSO-BP model, the risk factors of hypertension in Guangdong were analyzed from the perspective of social prevention and control. On the one hand, health education can be carried out in this region to guide people to choose the right lifestyle to prevent hypertension, including the promotion of healthy diet such as low-salt and low-oil diet [31], and the promotion of people to reduce the proportion of high-cholesterol food in their daily diet. In addition, promote the industry toward the development conducive to the health of the population, such as the establishment of food counters, set the content of oil and salt in finished or semi-finished products, etc. Encourage the development of corresponding health industry, such as fitness, yoga, etc. On the other hand, from the perspective of personal prevention, first of all, we should reduce staying up all night and develop good work and rest habits; actively develop healthy eating strategies to control salt and oil intake in daily life. At the same time, control smoking or stay away from smoking environments, limit alcohol consumption, and maintain or control body mass index within a healthy range.

5. Conclusion

In conclusion, this study constructed BP neural network, optimized BP neural network by PSO, Logistic regression model and completed the comparison, and concluded that the BP neural network optimized by PSO had the best performance. Moreover, MIV algorithm is further used to

screen the risk factors related to hypertension and construct the disease prediction model, which can provide prevention and control advice for the prevention of hypertension. In addition, considering that the PSO-BP neural network model has good predictive ability, it can be extrapolated to other diseases in the future, and it can be considered to combine BP neural network with more optimization algorithms and other algorithms to study the disease risk prediction model.

Declarations

Acknowledgements

We appreciate all authors for their contributions and physicians and participants. The study was funded by the Medical Scientific Research Foundation of Guangdong Province(Grant No. A2021395), Natural Science Foundation of Basic and Applied Basic Research Foundation of Guangdong Province (Grant No.2021A1515010061), Undergraduate Innovation Experiment Project of Guangdong Medical University (Grant No. ZZDG003), Natural Science Key Cultivation Project of Scientific Research Fund of Guangdong Medical University (Grant No. GDMUZ2020008), Zhanjiang Science and Technology Development Special Fund Competitive Allocation Project (Grant No. 2020A01031), Guangdong Provincial Colleges and Universities Characteristic Innovation Project (Grant No. 2019KTSCX047), Basic and Applied Basic Research Foundation of Guangdong Province Regional Joint Fund Project(Grant No.2020B1515120021) and Young Innovative Talents Project of Guangdong Universities(Grant No.2018KQNCX088).

Author Contributions

Y.Y. designed the research and wrote the manuscript. C.R., X.J., H.J.L, L.L., L.H., C.W.Y. and L.C.W. contributed to study design, data collection and review of the manuscript. D.Y.L., K.D.L, Z.Q.L. and Y.H.B. critically reviewed the manuscript and put forward modification opinions. Y.Y. and C.R. finished the final version. All authors approved the final version.

Competing interests

The authors declare no competing interests.

References

1. Li, Y. *et al.* Urban-Rural disparities in status of hypertension in northeast China: a population-based study, 2017–2019. *Clin Epidemiol.* **11**, 801–820 (2019).
2. Tesfaye, F., Byass, P. & Wall, S. Population based prevalence of high blood pressure among adults in Addis Ababa: uncovering a silent epidemic. *BMC Cardiovasc Disord.* **9**, 39 (2009).
3. Gheorghe, A. *et al.* The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: a systematic review. *BMC Public Health.* **18**, 975 (2018).
4. Kivimäki, M. & Steptoe, A. Effects of stress on the development and progression of cardiovascular disease. *Nat. Rev. Cardiol.* **15**, 215–229 (2018).
5. Carson, A. P. *et al.* Evaluating the Framingham hypertension risk prediction model in young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Hypertension.* **62**, 1015–1020 (2013).
6. Lim, N. K., Lee, J. Y., Lee, J. Y., Park, H. Y. & Cho, M. C. The Role of Genetic Risk Score in Predicting the Risk of Hypertension in the Korean population: Korean Genome and Epidemiology Study. *PLoS One.* **10**, e0131603 (2015).
7. Zhang, Q. *et al.* Combined Impact of Known Lifestyle Factors on Total and Cause-Specific Mortality among Chinese Men: A Prospective Cohort Study. *Sci Rep.* **7**, 5293 (2017).
8. Kittanawong, C., Zhang, H., Wang, Z., Aydar, M. & Kitai, T. Artificial Intelligence in Precision Cardiovascular Medicine. *J Am Coll Cardiol.* **69**, 2657–2664 (2017).
9. Ambale-Venkatesh, B. *et al.* Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res.* **121**, 1092–1101 (2017).
10. Lyu, J. & Zhang, J. BP neural network prediction model for suicide attempt among Chinese rural residents. *J Affect Disord.* **246**, 465–473 (2019).
11. Cao, J., Cui, H., Shi, H. & Jiao, L. Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on MapReduce. *PLoS One.* **11**, e0157551 (2016).
12. Tarkhaneh, O. & Shen, H. Training of feedforward neural networks for data classification using hybrid particle swarm optimization, Mantegna Lévy flight and neighborhood search. *Helijon.* **5**, e01275 (2019).

13. Ye, F. Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data. *PLoS One*. **12**, e188746 (2017).
14. Lin, S., Chen, S., Wu, W. & Chen, C. H. Parameter determination and feature selection for back-propagation network by particle swarm optimization. *Knowl Inf Syst*. **21**, 249–266 (2009).
15. Aslanargun, A., Mammadov, M., Yazici, B. & Yolacan, S. Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. *J Stat Comput Sim*. **77**, 29–53 (2007).
16. Hu, L. *et al.* Clearance Rate and BP-ANN Model in Paraquat Poisoned Patients Treated with Hemoperfusion. *BioMed Res Int*. **2015**, 1–6 (2015).
17. 1999 World Health Organization-. International Society of Hypertension Guidelines for the Management of Hypertension. Guidelines Subcommittee. *J Hypertens*. **17**, 151–183 (1999).
18. Zhang, Y. J. Prediction of hepatic cirrhosis with upper gastrointestinal bleeding based on particle swarm optimization BP neural network. *Shanxi Med Univ*. (2018).
19. Zhang, X. *et al.* Determination of zinc oxide content of mineral medicine calamine using near-infrared spectroscopy based on MIV and BP-ANN algorithm. *Spectrochim Acta A Mol Biomol Spectrosc*. **193**, 133–140 (2018).
20. Xie, H., Lv, X. X. & Zhang, P. Contribution Analysis of Influential Factors for Wind Power Curtailment Caused by Lack of Load-Following Capability Based on BP-MIV. *Distributed Energy*. **4**, 9–14 (2019).
21. Odden, M. C. *et al.* Hypertension and low HDL cholesterol were associated with reduced kidney function across the age spectrum: a collaborative study. *Ann Epidemiol*. **23**, 106–111 (2013).
22. Chen, H. R., Tao, H., Lin, J., Chen, L. X. & Cen, M. Effects of lifestyle interventions on prevalence, awareness of hypertension in high risk population. *J Clin Cardiol*. **26**, 427–429 (2010).
23. Qiu, Y. *et al.* The relation between dyslipidemia and hypertension or diabetes in men aged 40–59 years old. *Chin J Health Lab Technol*. **29**, 1738–1741 (2019).
24. Khirfan, G. *et al.* Plasma levels of high density lipoprotein cholesterol and outcomes in chronic thromboembolic pulmonary hypertension. *PLoS One*. **13**, e197700 (2018).
25. Zhang, Y. *et al.* Association Between High-density Lipoprotein Cholesterol and Renal Function in Elderly Hypertension. *Medicine*. **94**, e651 (2015).
26. Huang, J. Y. *et al.* Peripheral blood albumin level is significantly correlated with the severity of atherosclerosis in hypertensive patients. *J South Med Univ*. **40**, 240–245 (2020).
27. Han, F., Yin, L., Lu, G. T., Tong, Z. H. & Ding, Y. B. Relationship between High Density Lipoprotein and Metabolic Syndrome. *Med Recapitulate*. **26**, 356–360 (2020).
28. Zhu, L. Y. *et al.* Effect of sleep duration and quality on hypertension incidence in Chinese middle-aged and elderly people. *Chin J Disease Control & Prevention*. **24**, 176–182 (2020).
29. Guan, Y. *et al.* Association between sleep duration and hypertension of migrant workers in China: a national cross-sectional surveillance study. *BMJ Open*. **9**, e031126 (2019).
30. Interpretation of Chinese guidelines for the management of hypertension. (2018 revised edition). *Chin J Cardio Research*. **24**, 24–33 (2018).
31. Xu, H. *et al.* RNA-seq analyses the effect of high-salt diet in hypertension. *Gene*. **677**, 245–250 (2018).

Figures

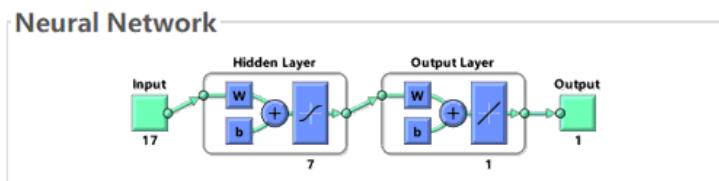


Figure 1

BP neural network parameters

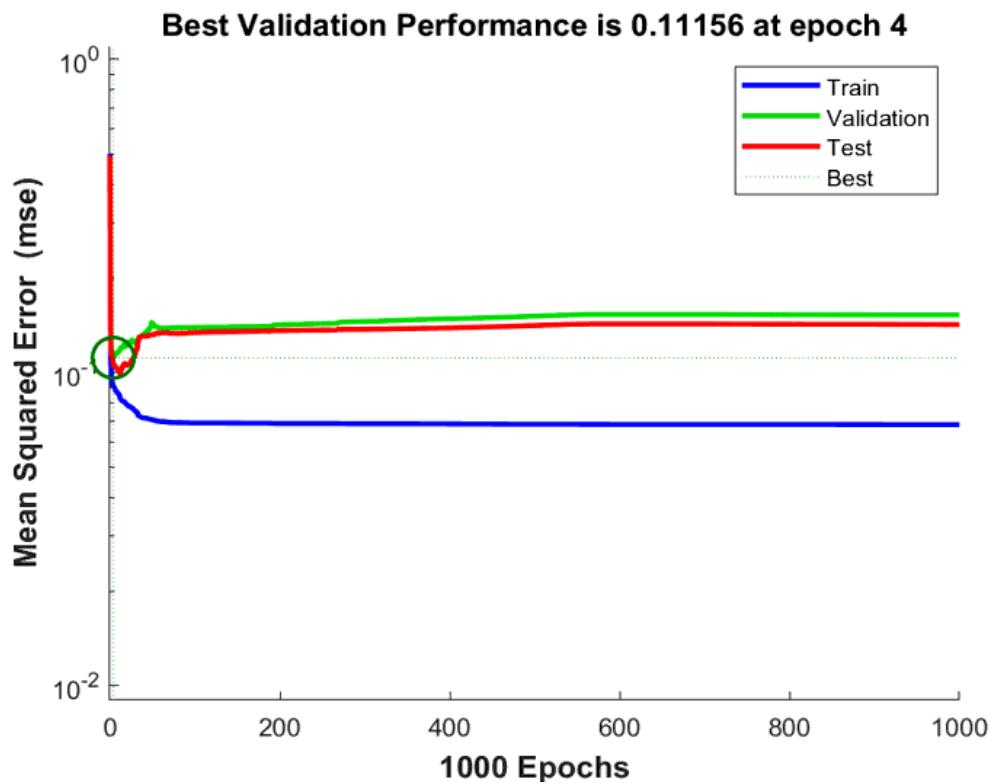


Figure 2

Mean square error curve of BP neural network.

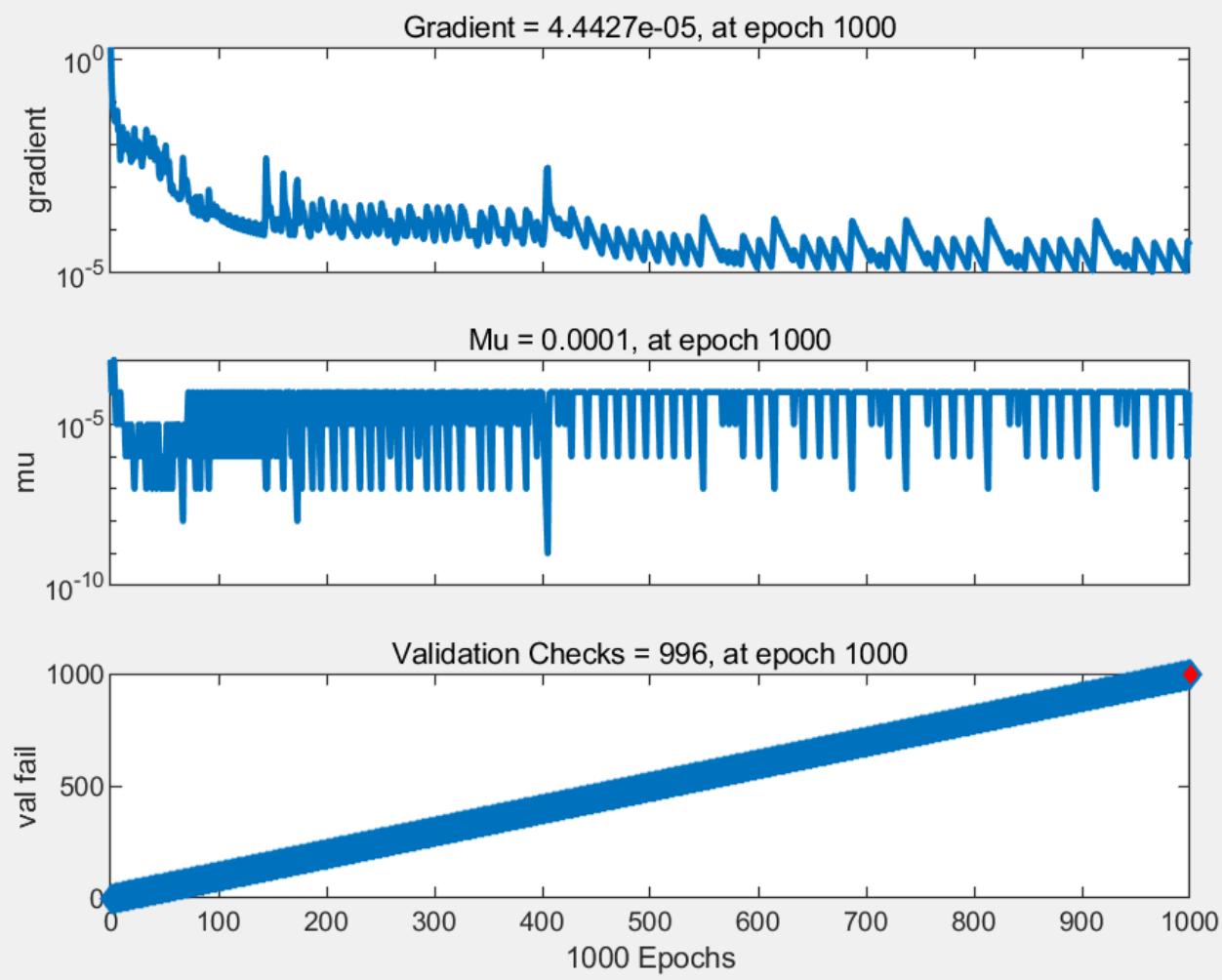


Figure 3

Training state of BP neural network

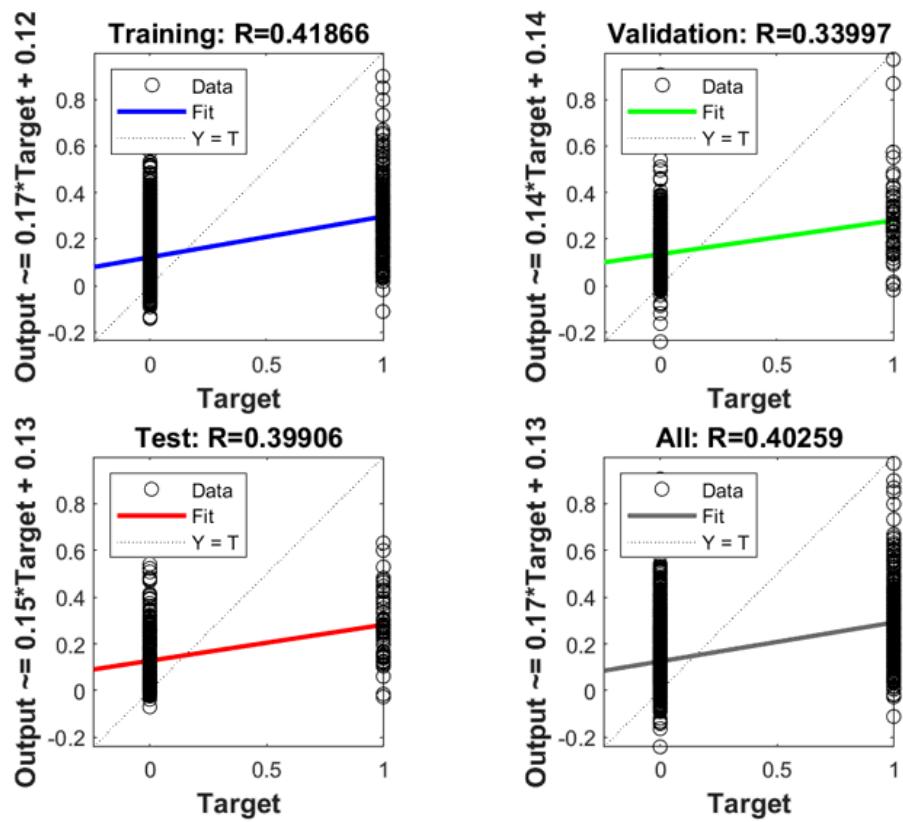


Figure 4

Training regression of BP neural network

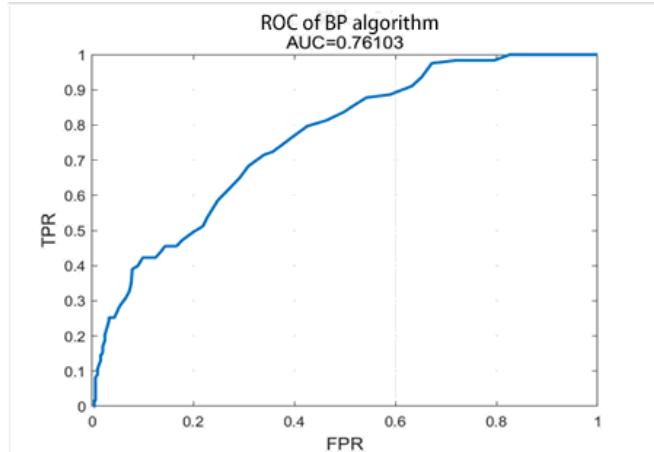


Figure 5

Area under ROC curve of BP neural network

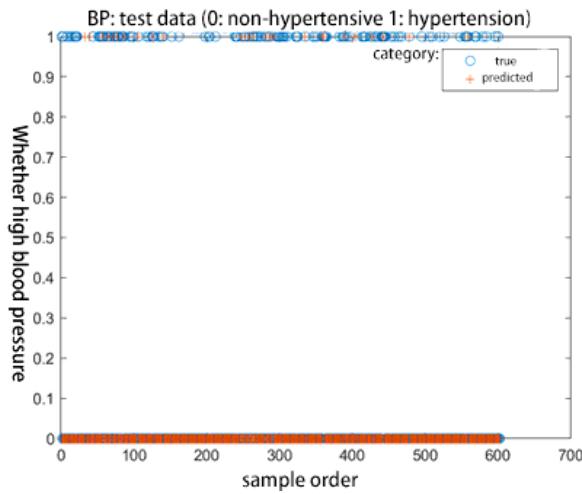


Figure 6

Comparison of prediction effect of BP neural network test set

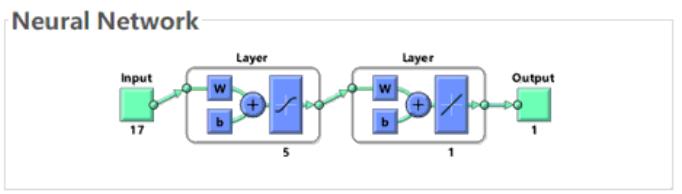


Figure 7

PSO-BP neural network parameters

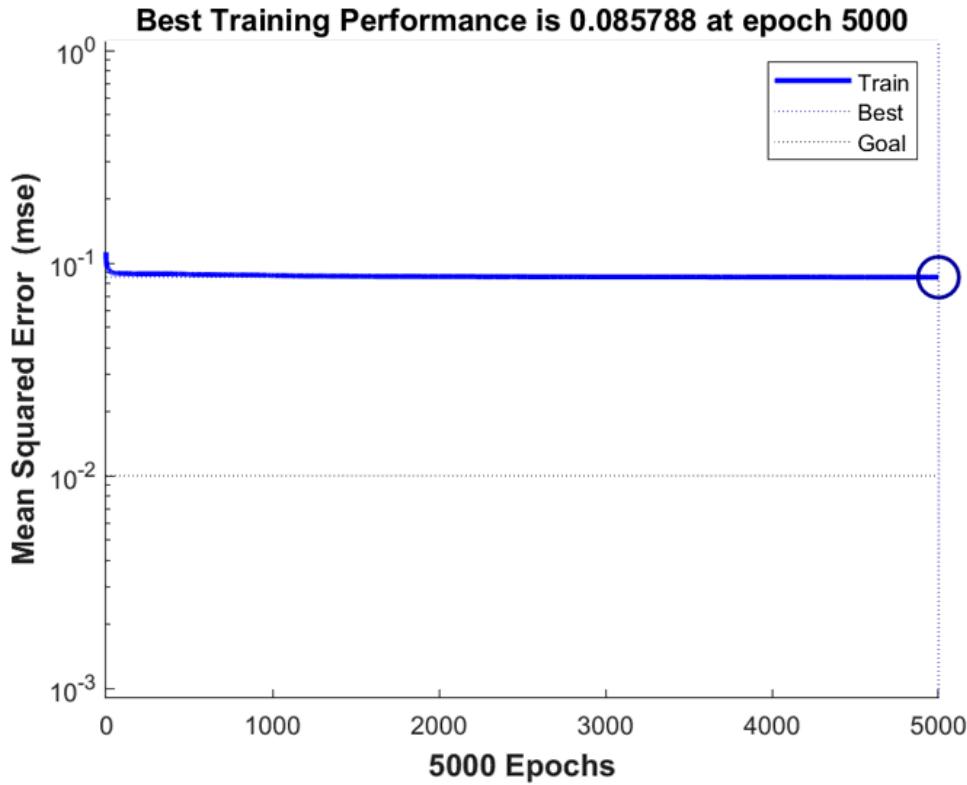


Figure 8

Mean-square error curve of PSO-BP neural network

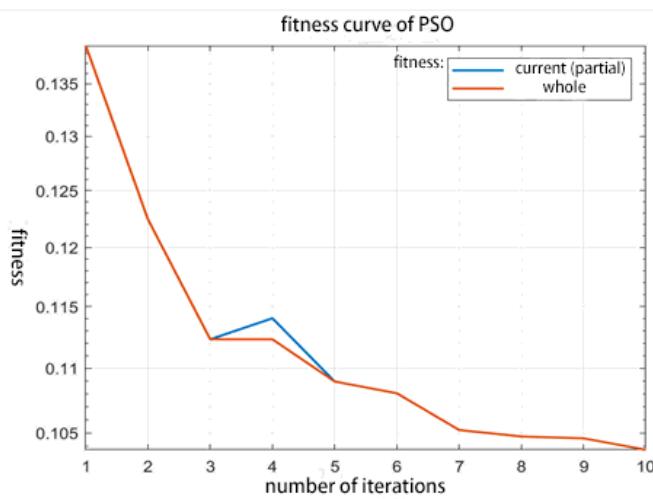


Figure 9

Fitness curve of PSO-BP neural network

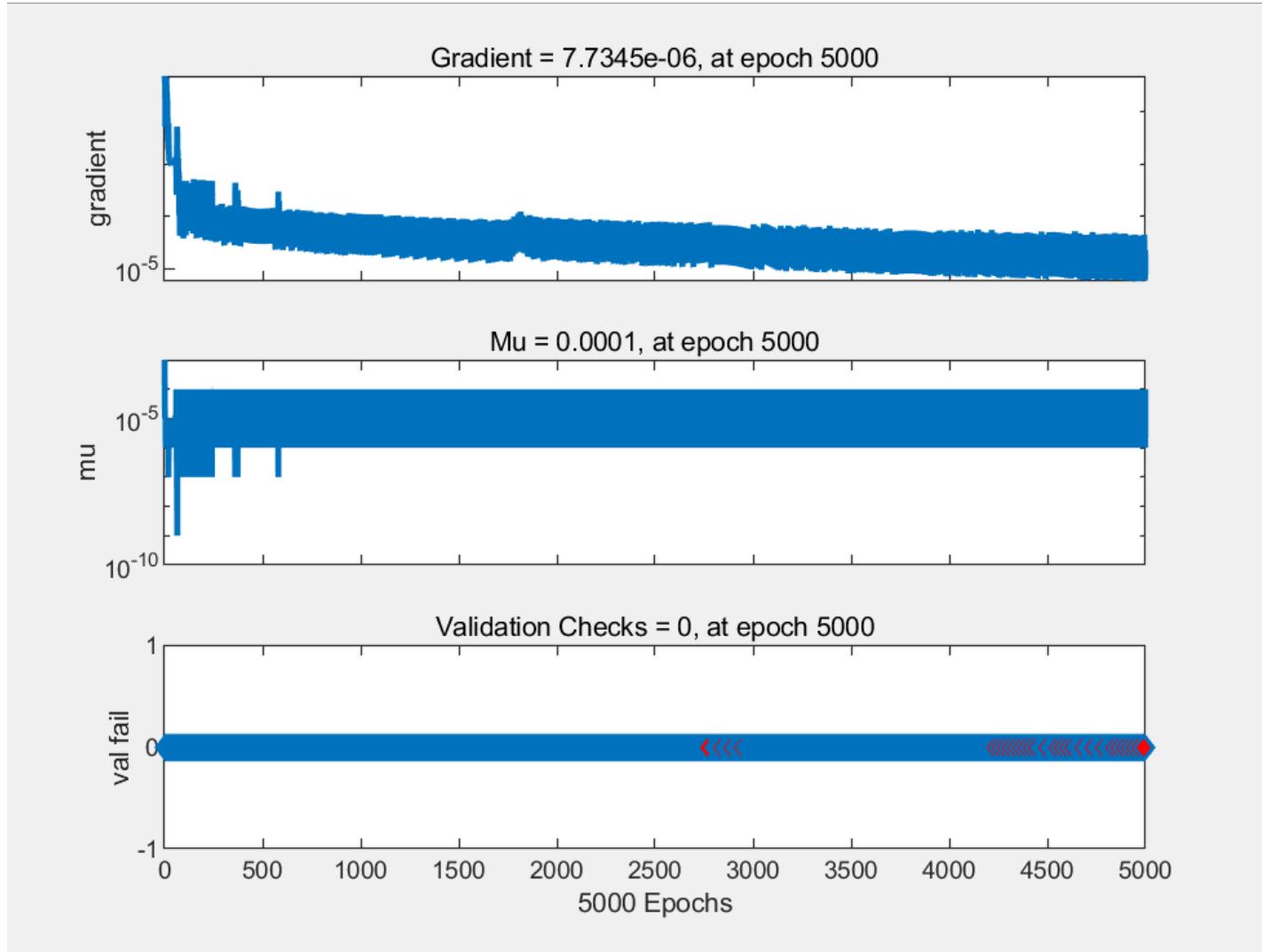


Figure 10

Training status of PSO-BP neural network

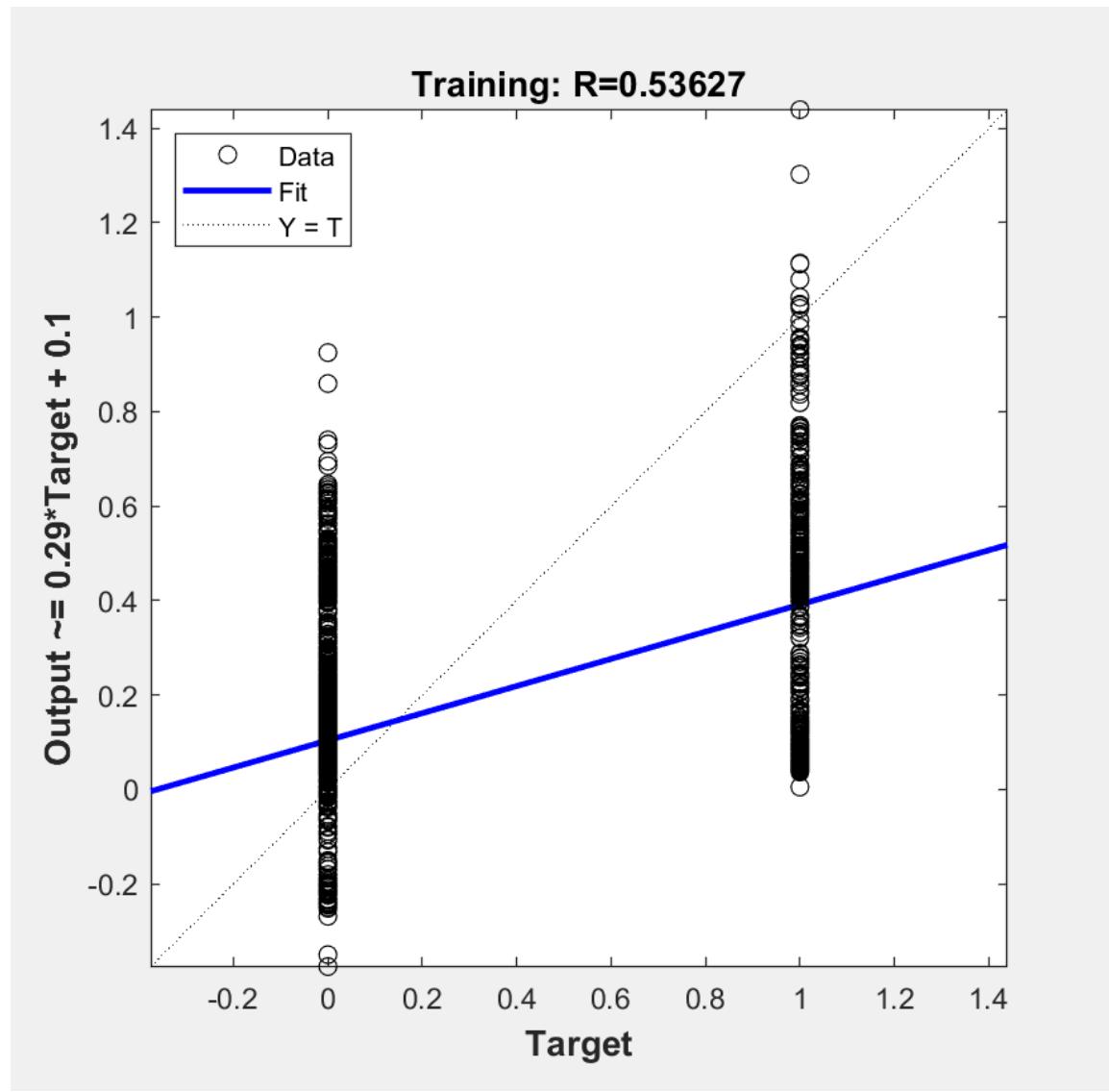


Figure 11

Regression of PSO-BP neural network

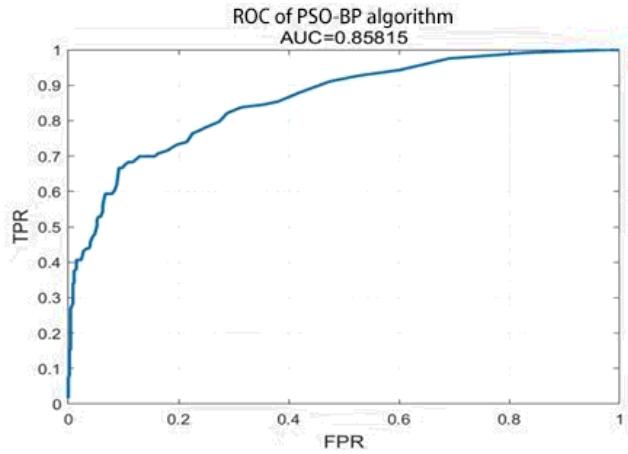


Figure 12

ROC curve and ACU under ROC curve of PSO-BP neural network

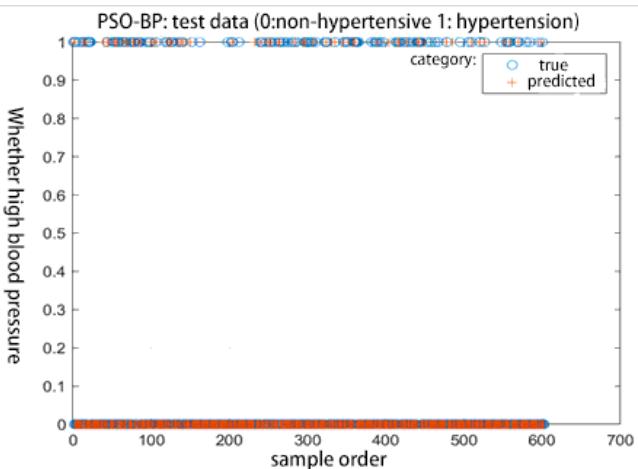


Figure 13

Comparison of prediction effect of PSO-BP neural network test set

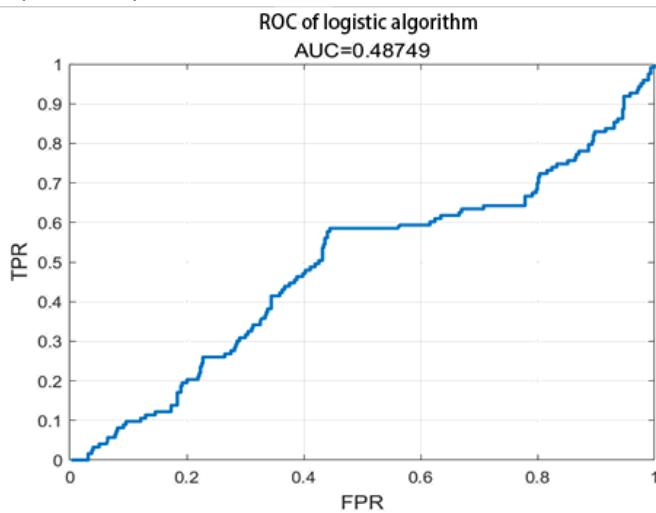


Figure 14

Area under Logistic regression ROC curve

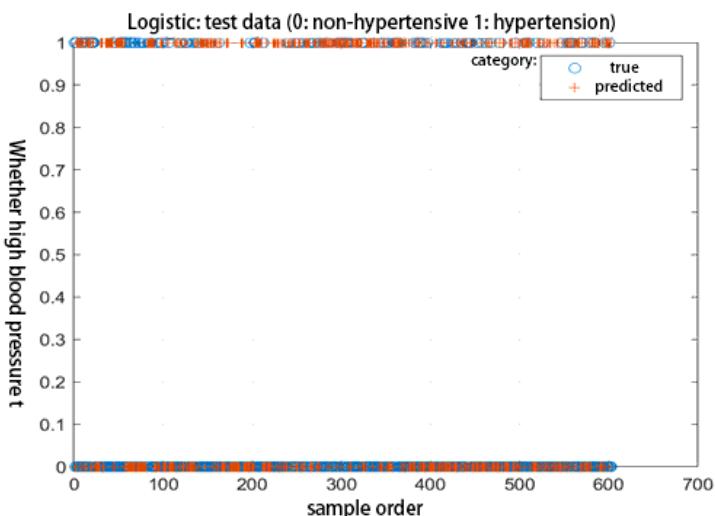


Figure 15

Comparison of Logistic Regression Model's Prediction Effect on Test Set