

## ORIGINAL ARTICLE

# Application of back propagation neural network model optimized by particle swarm algorithm in predicting the risk of hypertension

Yan Yan MM<sup>1</sup> | Rong Chen Postgraduate<sup>1</sup> | Zihua Yang Postgraduate<sup>1</sup> |  
Yong Ma Postgraduate<sup>1</sup> | Jialu Huang Postgraduate<sup>1</sup> | Ling Luo Postgraduate<sup>1</sup> |  
Hao Liu Postgraduate<sup>1</sup> | Jian Xu Postgraduate<sup>1</sup> | Weiyang Chen Postgraduate<sup>1</sup> |  
Yuanlin Ding PhD<sup>1</sup> | Danli Kong PhD<sup>1</sup> | Qiaoli Zhang MM<sup>2</sup> | Haibing Yu PhD<sup>3,4</sup>

<sup>1</sup>Department of Epidemiology and Medical Statistics, School of Public Health, Guangdong Medical University, Dongguan, Guangdong, China

<sup>2</sup>Preventive Medicine and Hygienics, Dongguan Center for Disease Control and Prevention, Dongguan, Guangdong, China

<sup>3</sup>The First Dongguan Affiliated Hospital, Guangdong Medical University, Dongguan, Guangdong, China

<sup>4</sup>Key Laboratory of Chronic Disease Prevention and Control and Health Statistics, School of Public Health, Guangdong Medical University, Dongguan, Guangdong, China

## Correspondence

Haibing Yu, The First Dongguan Affiliated Hospital, Guangdong Medical University, Dongguan, Guangdong, China; Key Laboratory of Chronic Disease Prevention and Control and Health Statistics, School of Public Health, Guangdong Medical University, Email: [hby616688@gdmu.edu.cn](mailto:hby616688@gdmu.edu.cn)

Qiaoli Zhang, Preventive Medicine and Hygienics, Dongguan Center for Disease Control and Prevention, Dongguan 523808 Guangdong, China. Email: [zql@dg.gov.cn](mailto:zql@dg.gov.cn)

Danli Kong, Department of Epidemiology and Medical Statistics, Guangdong Medical University, Dongguan, Guangdong 523808, China. Email: [kdlgdmu@gdmu.edu.cn](mailto:kdlgdmu@gdmu.edu.cn)

## Funding information

Discipline Construction Project of Guangdong Medical University, Grant/Award Numbers: 4SG21276P, 1003K20220004; Dongguan City Science and Technology Correspondent Project, Grant/Award Number:

## Abstract

The structure of a back propagation neural network was optimized by a particle swarm optimization (PSO) algorithm, and a back propagation neural network model based on a PSO algorithm was constructed. By comparison with a general back propagation neural network and logistic regression, the fitting performance and prediction performance of the PSO algorithm is discussed. Furthermore, based on the back propagation neural network optimized by the PSO algorithm, the risk factors related to hypertension were further explored through the mean influence value algorithm to construct a risk prediction model. In the evaluation of the fitting effect, the root mean square error and coefficient of determination of the back propagation neural network based on the PSO algorithm were 0.09 and 0.29, respectively. In the comparison of prediction performance, the accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve of the back propagation neural network based on PSO algorithm were 85.38%, 43.90%, 96.66%, and 0.86, respectively. The results showed that the backpropagation neural network optimized by PSO had the best fitting effect and prediction performance. Meanwhile, the mean impact value algorithm could screen out the risk factors related to hypertension and build a disease prediction model,

Yan Yan and Rong Chen contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *The Journal of Clinical Hypertension* published by Wiley Periodicals LLC.

20221800500342; Dongguan Social Development Science and Technology Project, Grant/Award Number: 20221800905642; Natural Science Foundation of Basic and Applied Basic Research Foundation of Guangdong Province, Grant/Award Numbers: 2022A1515012407, 2021A1515010061; Guangdong science and technology research project of traditional Chinese Medicine, Grant/Award Numbers: 20221209, 20211215; Medical Scientific Research Foundation of Guangdong Province, Grant/Award Number: A2021395; Innovation and entrepreneurship training program for college students of Guangdong Medical University, Grant/Award Numbers: GDMU2021112, GDMU2021138; Basic and Applied Basic Research Foundation of Guangdong Province Regional Joint Fund Project, Grant/Award Number: 2020B1515120021; Natural Science Key Cultivation Project of Scientific Research Fund of Guangdong Medical University, Grant/Award Number: GDMUZ2020008; Zhanjiang City science and technology development special fund competitive allocation project, Grant/Award Numbers: 2020A01031, 2018A01029; Characteristic Innovation Project of Guangdong Province General University, Grant/Award Numbers: 2020KTSCX042, 2019KTSCX046, 2019KTSCX047

which can provide clues for exploring the pathogenesis of hypertension and preventing hypertension.

#### KEYWORDS

back propagation neural network, hypertension, logistic regression, mean impact value, particle swarm algorithm optimization

## 1 | INTRODUCTION

As one of the most common diseases in the world, hypertension and its complications are the leading 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.<sup>1</sup> With the improvement of people’s living standards and changes in lifestyle, the proportion of the global disease burden of hypertension continues to rise. It is estimated that 1.6 billion adults in the world will suffer from hypertension in 2025.<sup>2</sup> The high prevalence rate consumes many financial and material resources, which challenges the management of the health system and causes a large economic burden.<sup>3</sup> Before the occurrence of hypertension, lifestyle and diet interventions to avoid or delay the occurrence of the disease are the most effective and economical way to control the disease,<sup>4</sup> but the factors affecting hypertension have not been completely elucidated, and the measures to prevent and control hypertension have yet to be optimized. The disease prediction model can help to identify the high-risk population of the disease and screen out the main potential risk factors affecting the occurrence and development of the disease. Therefore, it is particularly important to construct prediction models that can accurately identify the potential risk factors, protective factors, and risk of hyperten-

sion, which will play a beneficial role in the prevention and control of hypertension.

To date, most disease risk models are based on several risk factors to predict the disease or through logistic regression or Cox proportional hazards regression to further determine the risk factor category with respect to the prescreened risk factors and establish a multivariate statistical model to better predict the risk of disease. However, it should be noted that the collinearity between the influencing factors increases the difficulty of model fitting and leads to some errors in the application of classical models, thus exposing certain limitations of the traditional method. In the era of big data, artificial intelligence (AI) methods such as machine learning algorithms that parse complex, multidimensional, and multiscale data have been increasingly applied in disease prediction and engineering. Machine learning algorithms can imitate the human thinking process for knowledge learning and storage,<sup>5</sup> and have better accuracy and predictive ability than traditional regression models.<sup>6</sup> A backpropagation (BP) neural network is a multilayer feedforward neural network trained according to the error backpropagation algorithm, which is the most widely used artificial neural network (ANN) at present. It has advantages unmatched by traditional statistical methods in complex model fitting and distribution approximation and has great potential compared with other machine learning test algorithms.<sup>7</sup> The algorithm is currently widely used in clinical, physicochemical, and engineering fields and has a certain promotion ability and generalization ability. Weng coworkers used four machine learning algorithms to predict the risk of cardiovascular events, and the

results showed that the BP neural network has the highest sensitivity rate and the best prediction performance.<sup>8</sup> Ahmadi coworkers used a variety of machine learning algorithms, including ANNs, to predict the porosity and permeability of oil reservoirs. The results showed that the difference between the model estimates and the actual data is very small, indicating that the application of machine learning can make the prediction results more reliable.<sup>9</sup> Moosavi coworkers demonstrated the excellent prediction performance of ANNs in CO<sub>2</sub>-foam flooding.<sup>10</sup> Moreover, ANNs are also widely used in the prediction of the thermal conductivity ratio of nanofluids and urban planning prediction.<sup>11,12</sup> When the sample size is small, the BP neural network can always show good performance, but when the sample set size increases, the training time of the algorithm will become lengthy and the efficiency will drop sharply.<sup>13</sup> In addition, in the training of neural networks, there may be defects such as local minima and slow convergence rates caused by unsatisfactory network generalization, especially in the process of fine-tuning the control parameter set weight and deviation.<sup>14</sup>

In recent years, many scholars have adopted hybrid algorithms to improve the parameters of BP neural networks, among which the particle swarm optimization (PSO) algorithm has been widely used. The PSO algorithm is a method for optimizing continuous nonlinear functions, by modifying adjustable parameters to make the error between the predicted output and the expected output as small as possible. Its advantage is that it can significantly improve the performance of recognition tasks and optimize the structure of deep neural networks without adjusting redundant parameters.<sup>15</sup> In addition, the calculation process of the BP neural network contains a certain degree of randomness, which has poor restrictions on the objective function and has the characteristics of strong local search ability and poor global search ability. The PSO algorithm can find the optimal solution closest to the global through cooperation and information sharing among individuals in the group, with strong global search ability but poor local search ability.<sup>16</sup> Therefore, the construction of the combined model is expected to overcome the limitations of each component.<sup>17,18</sup> Ramezanizadeh coworkers predicted the dynamic viscosity of nanofluids based on the least squares support vector machine model optimized by the PSO algorithm, and found that the model improved the accuracy of predicting the dynamic viscosity of nanofluids.<sup>19</sup> Ahmadi coworkers used the PSO algorithm to optimize an ANN and compared it with a traditional ANN. The results showed that the PSO algorithm is effective in predicting reservoir permeability.<sup>20</sup> Liu coworkers also constructed a BP neural network model based on the PSO algorithm for intelligent emergency risk avoidance of sudden financial disasters and compared it with other algorithms. The results showed that the hybrid PSO-BP algorithm is superior to other algorithms in simulation and prediction.<sup>21</sup> The above research showed the good application prospects of the PSO algorithm, which provides the basis and ideas for applying the PSO algorithm to the evaluation and prediction of diseases.

Based on the above, this study attempted to complete the construction of the hypertension risk model by constructing a BP neural network optimized by PSO. Furthermore, the mean impact value (MIV)

algorithm was used to screen the risk factors for hypertension in the PSO-BP neural network model, and the risk factor analysis of hypertension in the Guangdong region was completed, thereby providing new insights into the prevention and control of the occurrence and development of hypertension.

## 2 | METHODS

### 2.1 | Research object

This study was based on the monitoring data of chronic disease risk factors in the Guangdong region from 2016 to 2018. The research subjects were randomly sampled by cluster. After data cleaning, 3012 people (cross-sectional research subjects, including the normal population and hypertensive patients) were ultimately included. Patients with hypertension were diagnosed using the 1999 World Health Organization (WHO) - International Society of Hypertension (ISH) criteria: systolic blood pressure  $\geq 140$  mmHg and/or diastolic blood pressure  $\geq 90$  mmHg.<sup>22</sup>

### 2.2 | Data collection and pretreatment

By consulting the literature and combining the monitoring data of the subjects (including the on-site questionnaire survey, physical examination, etc.), 17 independent variables (X1-X17) were preliminarily screened out, including gender, age, educational status, occupation, smoking, drinking, sleep duration, body mass index (BMI), heart rate, blood glucose, hemoglobin, cholesterol, triglycerides, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, daily salt intake, and daily oil intake, and while high blood pressure status was the dependent variable (Y). The data were sorted, abnormal values were eliminated, and missing values were filled with mode (categorical variables) or mean (continuous variables). The data were normalized by Min-Max normalization. In the process of model construction, 80% of the samples were randomly selected as the training set to train the classification model, and 20% were selected as the test set for evaluating the quality of the model. The number of hypertension and nonhypertension cases in the training set were 316 and 2094, respectively, while in the test set, these values were 56 and 546, respectively. In the process of neural network training, the training set was further divided into three subsets, namely, the training set, validation set, and test set. The training set was used to determine the model parameters, the validation set was used for the final optimization of the model, and the test set was used to test the generalization ability of the trained model.

### 2.3 | Establishment of the hypertension prediction model

After preprocessing the data, a BP neural network was constructed by MATLAB. The optimal weights and thresholds were assigned to the BP

neural network by the PSO algorithm, and the BP neural network prediction model optimized by PSO was constructed. The input vector of the BP neural network was the 17 variables (X1–X7) initially screened, and the output vector (Y) was whether the patient had high blood pressure. The PSO-BP neural network was consistent with the BP neural network. The parameters of the BP neural network and PSO-BP neural network are shown in Table S1. The construction of the logistic model was completed through SPSS version 21.0.

## 2.4 | Evaluation of prediction models

The root mean square error and coefficient of determination were used to evaluate the fitting ability of the model. The accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC) were used to evaluate the prediction effect of the model.

## 2.5 | Screening of risk factors for hypertension

As one of the best indicators to evaluate the correlation of variables in the neural network algorithm, the MIV algorithm does not change the total information extracted from the network in the process of variable screening, and can exclude the influence of variables with interference in the model. Moreover, the MIV algorithm can reflect the matrix change of the weight of each variable in the neural network and quantitatively evaluate the importance of the influence of independent variables on the dependent variable to improve the analysis effect of the model.<sup>23</sup> Therefore, based on the PSO-BP neural network, this study explored the risk factors related to hypertension in the Guangdong region by the MIV algorithm and ultimately constructed a hypertension risk prediction model.

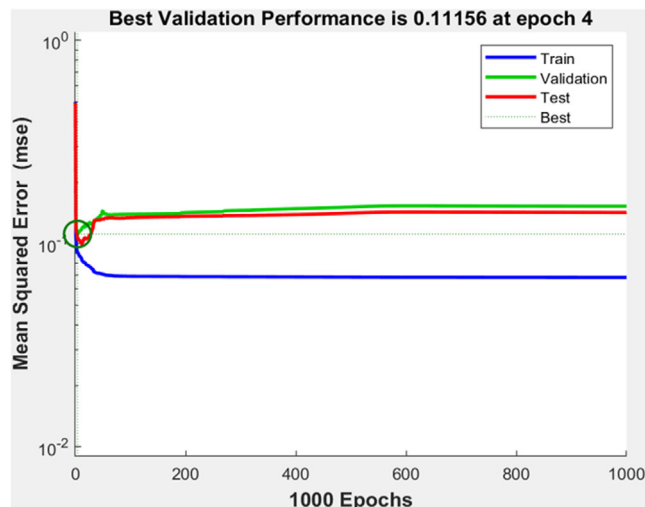
## 2.6 | Statistical analysis

The MATLAB version R2019b (mathematical software produced by MathWorks in the United States) was used to complete the construction of the BP neural network model, the optimization of the BP neural network model by particle swarms, and the analysis of data and screening of risk factors. Through SPSS version 21.0, binary multivariate logistic regression analysis was used to complete the construction of the logistic regression model and the comparison of related models.

# 3 | RESULTS

## 3.1 | Baseline characteristics

In this study, 3012 participants were included at baseline. The total prevalence of hypertension each year from 2016 to 2018 was 18%, 15.8%, and 18.6%, respectively (Table S2).



**FIGURE 1** Mean square error curve of the BP neural network. BP, backpropagation

## 3.2 | BP neural network

### 3.2.1 | Model training effect

#### (1) Effect of the BP neural network

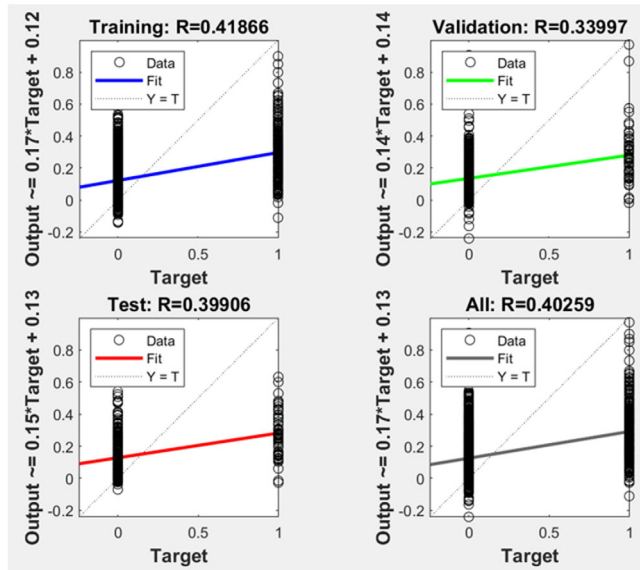
In the model training process of the BP neural network, with the gradual increase in the number of iterations, consisting of approximately 50 steps, the model prediction errors of the validation set and the test set gradually tended towards stability, and the model prediction error of the training set gradually decreased. In the green circle in the figure (approximately four steps), although the model error of the training set is relatively large, the model predictions for the validation set and test set are close to expectations, and the minimum mean square error is 0.11156 (Figure 1).

#### (2) Effect of the BP neural network-training state

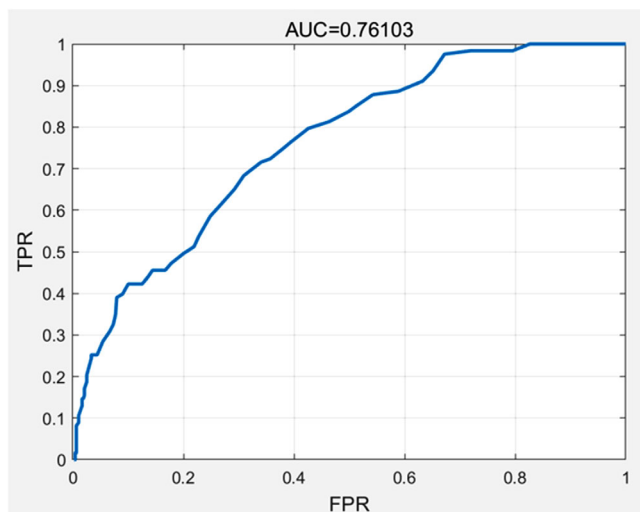
The BP neural network showed gradient descent in the training, the learning rate was 0.0001, and the effective test times were 996 times stopped, indicating that the error curve of the confirmed sample did not decline for 996 consecutive iterations during the training process of the network using the training samples (Figure S1).

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

The correlation coefficient of the training set of the BP neural network was 0.41866, the correlation coefficient of the validation set was 0.33997, the correlation coefficient of the test set was 0.39906, and the total correlation coefficient was 0.40259, indicating low degrees of positive correlation. Moreover, the association line did not fall on the diagonal, which comprehensively indicates that the fitting ability of the BP neural network was relatively general, and there was no overfitting phenomenon (Figure 2).



**FIGURE 2** Training regression of the BP neural network. BP, backpropagation



**FIGURE 3** Area under ROC curve of the BP neural network. BP, backpropagation; ROC, receiver operating characteristic

### 3.2.2 | Prediction performance of the BP neural network

The AUC value of the BP neural network was 0.76103, which is in the range of 0.7 to 0.85, indicating that the effect of the BP neural network model was general (Figure 3). The comparison diagram of the prediction effect of the BP neural network test set shows the prediction category of the real category and the BP neural network for the test set. The greater the overlap, the better the model effect was (Figure S2). For the convenience of analysis, the model confusion matrices of the test set was counted (Table S3). In light of the fact that there was no overfitting phenomenon in the BP neural network in this study, the prediction evaluation was completed through the test set. In the

**TABLE 1** Performance indicators of the three models

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

Abbreviations: AUC, area under the curve; BP, backpropagation.

test set, 17 out of 28 positive cases (hypertension) were correctly predicted and 11 were misjudged, for a correct rate of 60.71%. Among the 574 negative cases (nonhypertension), 468 were correctly predicted, and 106 were misjudged, for a correct rate of 81.53%. Therefore, the overall accuracy of the training set prediction was 80.56% (Table 1).

### 3.2.3 | Risk factors for MIV assessment in the BP neural network

The absolute value of MIV was greater than 0.002 as the screening limit, and the average value of multiple runs was taken. After feature selection, the risk factors were ranked from heavy to light according to the MIV weight. If the MIV value were positive, these factors would be positively correlated with the dependent variable, and the prevalence of hypertension would increase with an increase in these influencing factors. If the MIV value were negative, these factors would be negatively correlated with the dependent variable, and the prevalence of hypertension would increase with a decrease in these influencing factors. The results showed that factors positively associated with hypertension included low-density lipoprotein cholesterol, cholesterol, triglycerides, daily oil intake, daily salt intake, smoking, age, BMI, and drinking, and negatively associated factors included sleep duration, heart rate, high-density lipoprotein cholesterol, and hemoglobin (Table 2).

## 3.3 | PSO-BP neural network

### 3.3.1 | Model training effect

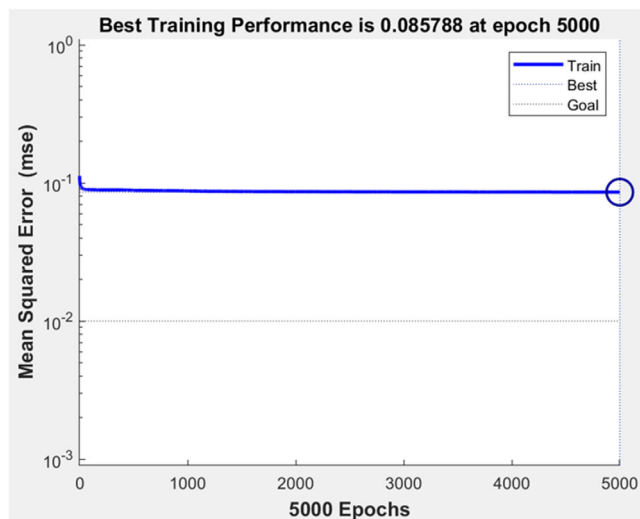
#### (1) Effect of the PSO-BP neural network

In the model-training process of the PSO-BP neural network, as the number of iterations increased to the initial maximum iteration number of 5000 steps, the model prediction errors of the validation set and the test set gradually tended toward stability, and the model prediction error of the training set gradually decreased. At that time, the minimum mean square error was 0.085778, indicating that the model no longer showed evidence of overfitting or underfitting, and the model reached the optimal level (Figure 4), but there are more iterations

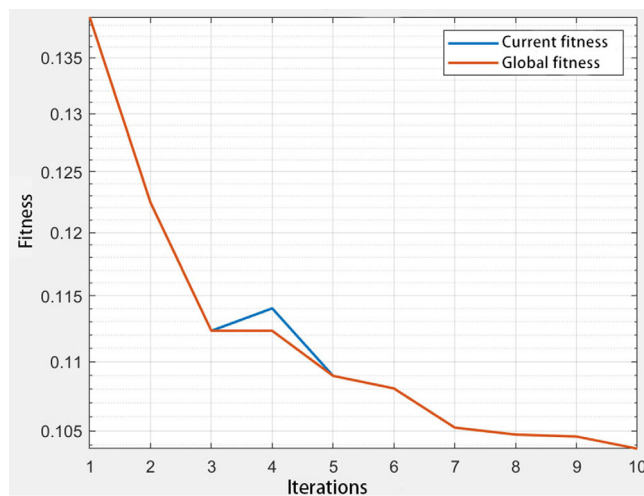
**TABLE 2** MIV screening results in the BP neural network and PSO-BP neural network

Numbers	Variables	BP neural network		PSO-BP neural network	
		MIV value	Weight	MIV value	Weight
1	LDL-C	0.022848	0.16202	0.027307	0.069858
2	Cholesterol	0.0092056	0.10992	0.63461	0.61935
3	Sleep duration	-0.010039	0.10027	-0.021362	0.072334
4	Triglycerides	0.0017972	0.079886	0.011243	0.030626
5	Daily oil intake	0.007317	0.078942	0.0036201	0.032021
6	Daily salt intake	0.0089765	0.074292	0.011187	0.029155
7	Smoking	0.0062768	0.060133	0.0058964	0.026019
8	Age	0.0072	0.053676	0.0062726	0.015869
9	BMI	0.0069582	0.052572	0.0036371	0.012329
10	Blood glucose	-0.0012076	0.043836	0.00047164	0.01562
11	Heart rate	-0.0053262	0.040962	-0.0015596	0.0052576
12	Gender	0.0011152	0.036111	-0.0054725	0.026225
13	HDL-C	-0.0045044	0.033973	-0.0034072	0.014852
14	Drinking	0.0034115	0.028692	0.0026665	0.0085437
15	Hemoglobin	-0.0028802	0.021162	-0.0052897	0.013865
16	Occupation	-0.0010221	0.018237	-0.0015046	0.0053823
17	Educational status	-0.000166	0.0053088	-0.00060963	0.0026992

Abbreviations: BMI, body mass index; BP, backpropagation; HDL-C, high-density lipoprotein cholesterol, LDL-C, low-density lipoprotein cholesterol, MIV, mean impact value; PSO, particle swarm optimization.

**FIGURE 4** Mean square error curve of the PSO-BP neural network. BP, backpropagation; PSO, particle swarm optimization

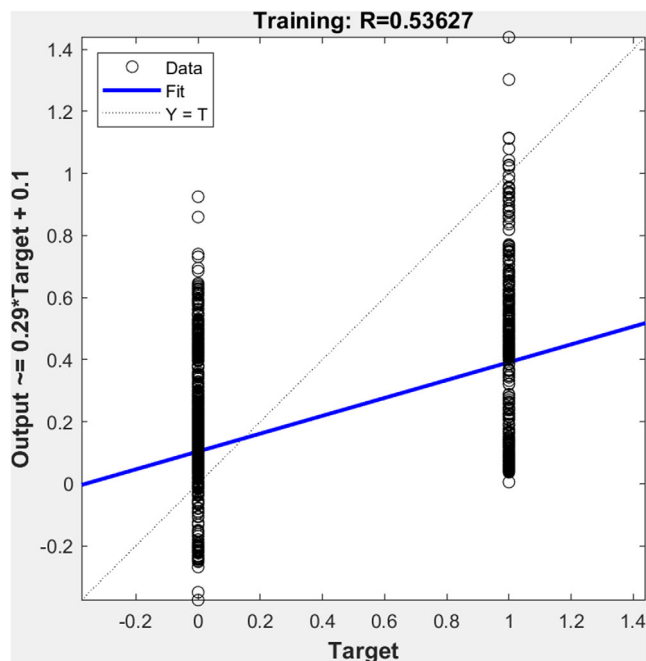
and long training times. In addition, with the gradual increase of the number of iterations, the overall trend of fitness gradually decreased. Fitness is used to describe the model error; the smaller the fitness is, the smaller the model error. After the fifth iteration, the current fitness and global fitness of the model begin to unify, and the error converged. After the number of iterations reached 10, the minimum fitness was

**FIGURE 5** Fitness curve of the PSO-BP neural network. BP, backpropagation; PSO, particle swarm optimization

approximately 0.104, and the optimal solution under this parameter was obtained (Figure 5).

## (2) Effect of PSO-BP neural network-training status

The PSO-BP neural network showed gradient descent during training. Compared with the gradient descent curve of the BP neural



**FIGURE 6** Training regression of the PSO-BP neural network. BP, backpropagation; PSO, particle swarm optimization

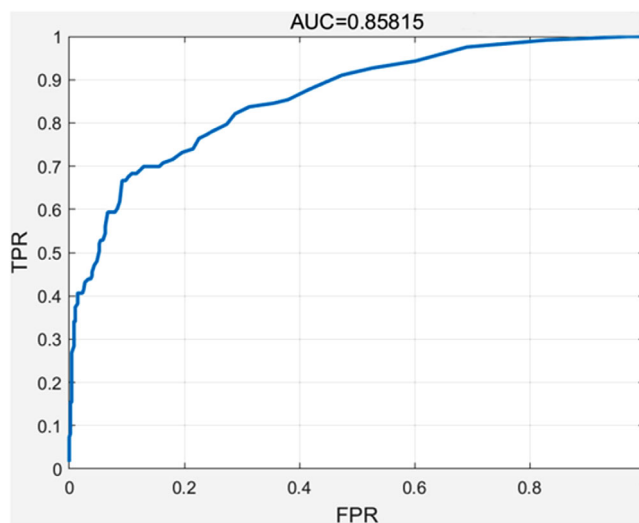
network, the descent curve of the PSO-BP neural network had less fluctuation. The learning rate was 0.0001, and the effective test times were 0 times stopped, indicating that in the process of training the network using the training samples, the error curve of the confirmation sample did not decrease at the beginning and stayed relatively stable (Figure S3).

### (3) Effect of PSO-BP neural network-fitting ability

The correlation coefficient of the PSO-BP neural network was 0.53627, indicating that the degree of correlation was moderately positive (Figure 6).

### 3.3.2 | Prediction performance of the PSO-BP neural network

The AUC value of the PSO-BP neural network was 0.85815, which is in the range of 0.85 to 0.95, indicating that the BP neural network model had a good effect (Figure 7). The comparison diagram of the prediction effect of the PSO-BP neural network test set showed the prediction category of the real category and the PSO-BP neural network for the test set (Figure S4). At the same time, the model confusion matrices of the training set and the test set were also counted (Table S3). In the test set, 49 of the 63 positive cases (hypertension) were correctly predicted and 14 were misjudged, for a correct rate of 76.56%. Among the 539 negative cases (nonhypertension), 465 were correctly predicted, and 74 were misjudged, for a correct rate of 86.27%. Therefore, the overall accuracy of predictions on the test set was 85.38%. According to the AUC value, accuracy, sensitivity, and specificity of the PSO-BP neu-



**FIGURE 7** Area under ROC curve of the PSO-BP neural network. BP, backpropagation; PSO, particle swarm optimization; ROC, receiver operating characteristic

ral network model, it could be shown that the prediction effect of this model was better (Table 1).

### 3.3.3 | Risk factors for MIV assessment in the PSO-BP neural network

The factors positively related to hypertension included cholesterol, low-density lipoprotein cholesterol, daily oil intake, triglycerides, daily salt intake, smoking, age, BMI and drinking. The factors negatively related to hypertension included sleep duration, high-density lipoprotein cholesterol and hemoglobin (Table 2).

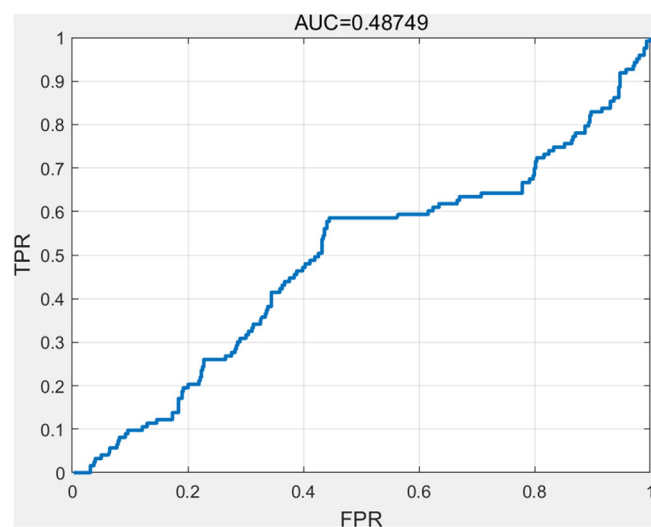
### 3.4 | Logistic regression model

#### (1) Establishment of the risk prediction model

Consistent with the neural network, this study took 17 variables as independent variables and high blood pressure status as the dependent variable to establish a logistic regression model through SPSS software, and 17 variables were entered into the MATLAB training set to evaluate the prediction performance of the logistic regression (Table S4).

#### (2) Logistic regression prediction and evaluation

The AUC value of the logistic regression model was 0.48749, which was less than 0.5, so it was inferred that the model had worse predictive performance than random guessing and thus had no predictive value (Figure 8). At the same time, the comparison chart of the prediction effect of the logistic regression model for the test set was obtained (Figure S5), and the model confusion matrices of the



**FIGURE 8** Area under ROC curve of the logistic regression. ROC, receiver operating characteristic

training set and the test set were counted (Table S3). In the test set, 72 of 287 positive cases (hypertension) were correctly predicted, and five were misjudged, for a correct rate of 25.08%. Among 315 negative cases (nonhypertension), 51 were correctly predicted, and 264 cases were misjudged, for a correct rate of 16.19%. Therefore, the overall accuracy of the training set prediction was 55.81%. According to the AUC value, accuracy, sensitivity and specificity of the logistic regression model, it could be shown that the prediction effect of this model was not satisfactory (Table 1).

## 4 | DISCUSSION

As a common disease in the population, hypertension is an important risk factor for cardiovascular and cerebrovascular diseases such as heart failure, coronary heart disease, aortic dissection, and stroke, which have a great impact on the health of the population. With the increasing proportion of patients diagnosed with this condition, the disease burden caused by hypertension has also been increasing year by year. Due to the interaction and multicollinearity between the risk factors affecting the incidence of hypertension, there will be errors in the fitting of traditional models, while emerging machine algorithm research can provide more possibilities for the prevention and treatment of the disease. The BP neural network has great potential in research in the medical field, and its prediction and risk assessment performance can be further improved by optimizing the neural network, which has certain research significance.

### 4.1 | Construction and comparison of neural network models

In this study, the neural network prediction model of hypertension was constructed based on the monitoring data of chronic disease risk

factors, and the PSO algorithm was further used to complete the optimization of the neural network. Considering that the excessive number of neurons in the input layer will have higher requirements with regard to sample size, this study preliminarily screened out the independent variables by referring to the literature and combining them with the existing content of monitoring data and constructed a neural network model with hypertension as the dependent variable. To prevent model overfitting and ensure the accuracy of testing, this study divided the dataset into a training set and testing set at a ratio of 4:1. The establishment of a neural network is flexible, and there is no unified value of functions and parameters in the establishment process. By comparison, the results showed that the performances of the BP neural network and the BP neural network optimized by PSO are different in the process of modeling than they are after modeling. Therefore, after many raining iterations, the BP neural network and the PSO-BP neural network were selected as the best prediction indicators for horizontal comparison.

In the construction model, the iteration times and running time of the BP neural network and the PSO-BP neural network were quite different. The BP neural network iterated 4 times and ran for approximately 10 s to achieve the best performance, while the PSO-BP neural network needed to iterate the maximum number of times initially set to achieve the best performance, and its running time is approximately 500 s. Cao coworkers also believed that although the PSO-BP neural network algorithm improves the performance of the traditional BP neural network algorithm, as the data size increases, the training time of the BP neural network increases as well, eventually raising efficiency issues.<sup>13</sup> However, Zhang Yijun's research showed that the optimization algorithm will lead to a shorter modeling time,<sup>24</sup> which contrasts with the findings from this study. This may be because when the two algorithms are combined to achieve local and global optimization, more iterations and running time are required to make the model reach optimal performance. In addition, the error of the PSO-BP neural network converged at the 10th iteration to obtain the optimal solution of this parameter, but it needed 5000 iterations to obtain the lowest root mean square error after combination, which comprehensively verifies that the PSO-BP neural network may have poor local search ability and need more iterations.

In the performance comparison of the training set of the model, the root mean square error of the BP neural network was 0.34, and the root mean square error of the PSO-BP neural network was smaller at 0.09. This result is consistent with the research results of most scholars. Li coworkers used the logarithmic distance loss model, the BP neural network, and the PSO-BP neural network to train the Bluetooth received signal strength indication distance model, and their results showed that the PSO-BP neural network had the lowest error rate, which was only 0.25, and had good robustness.<sup>25</sup> The research results of Liu coworkers also showed that after optimizing the BP neural network through PSO, the root mean square error of the model was reduced from 0.629 to 0.210, which significantly improved the prediction performance of the model.<sup>26</sup> In evaluating the performance ability of network fitting, the total coefficients of determination of the BP neural network and the PSO-BP neural network were 0.16 and 0.29, respectively. Compared



with the BP neural network, the coefficient of determination of the PSO-BP neural network was closer to 1, indicating that the dataset of the PSO-BP model is more related to the real category. Jiang coworkers used a genetic algorithm, ant colony algorithm and PSO to optimize a BP neural network in the study of predicting fatty acid content in the flour storage process, and the results showed that the PSO-BP neural network model had the highest coefficient of determination and the best optimization effect.<sup>27</sup> Moreover, the research results of this paper were also consistent with Liu coworkers's conclusion about using the PSO-BP neural network to predict wind turbine blades, that is, the PSO-BP neural network has small error and higher fitting ability.<sup>28</sup> Through the comparison of root mean square error and the coefficient of determination, it can be concluded that the PSO-BP has better nonlinear fitting ability. In addition, the gradient descent curve of the PSO-BP neural network is more stable than that of the BP neural network, which also proves that the PSO-BP neural network has better stability.

In the comparison of prediction performance, the data for each model did not change much across multiple runs. Compared with the logistic model, the BP neural network and the PSO-BP neural network showed significant improvements in accuracy, specificity and the AUC value, which indicates that compared with the traditional model, the neural network has better prediction accuracy. In the comparison of neural network algorithms, it was found that the accuracy, sensitivity and the AUC value of the prediction model were improved after the optimization of the PSO algorithm, which comprehensively shows that the prediction performance and diagnosis performance of the PSO-BP neural network are the best. This result is consistent with the prediction performance comparison between neural networks and traditional models that are mostly discussed at present. Xin coworkers constructed a BP neural network optimized by the PSO algorithm to identify the surface texture roughness of an object. The results showed that the model has higher convergence accuracy than the BP neural network, and the accuracy of identifying samples reached 98%.<sup>29</sup> Afrakhteh coworkers applied the PSO algorithm to optimize the multilayer perceptron neural network, and the results showed that the accuracy of the optimized model increased from 89.36% to 97.66%.<sup>30</sup> Suganthi coworkers used a multilayer back propagation neural network combining ant colony optimization and PSO to classify tumors. The results showed that the accuracy of the system reached 99.5%, and the AUC value was the highest compared with other models.<sup>31</sup> However, it is undeniable that, different from other studies, the sensitivity of the neural network in this study decreased, indicating that the diagnostic ability of the neural network constructed in this study decreased when predicting the risk of hypertension. In summary, for diseases such as hypertension, which have many causative factors that may interact with each other, the neural network model has better adaptability and fitting effect. Although the modeling time of PSO-BP is prolonged, its error is smaller, the degree of correlation is higher, the nonlinear fitting ability is better and the prediction performance is better, indicating that the performance of the BP neural network is improved after optimization, and the PSO-BP neural network can be better applied to hypertension risk research.

## 4.2 | Comparison of MIV algorithm screening

The MIV algorithm was used to screen the risk factors for hypertension in the Guangdong region. The greater the absolute value of the MIV value, the greater the influence of the influencing factors on the prevalence of hypertension was, and the greater the weight of the MIV, the higher the influence ranking of this factor was. However, there is no unified standard for how much the MIV value can be considered an influencing factor. By reviewing the literature and combining professional knowledge and comparative prior knowledge, we determined that when the absolute value of the MIV in this study was greater than 0.002, it could be considered as the influencing factor of hypertension in this region, and ranked the influencing factors according to the weight of the MIV. The comparison found that in the BP neural network and the PSO-BP neural network models, the risk factors for hypertension obtained by the MIV algorithm analysis were different, and the risk factor weights also changed, indicating that algorithm optimization will produce differences in the establishment of prediction models. It is worth noting that compared with the logistic regression model, the factors screened out by the MIV algorithm in the neural network were different. The BP neural network screened more factors, such as hemoglobin, sleep duration, daily oil intake, daily salt intake and drinking than the logistic regression model, while the PSO-BP neural network screened more risk factors such as hemoglobin, sleep duration, daily oil intake, triglycerides, daily salt intake, and drinking. Furthermore, educational level was not screened out in the neural network. Some scholars have applied the BP neural network to the study of metabolic diseases and have screened risk factors with the MIV algorithm. Their results showed that the risk factors screened with the MIV algorithm were two greater in number than in the logistic regression model,<sup>32</sup> a finding similar to what was reported this study, which showed that the MIV algorithm based on the neural network was applied to the analysis of the risk factors for hypertension, and there would also be differences in the screening results. After a comprehensive comparison of fitting performance and prediction performance, the results showed that the PSO-BP neural network had the best performance. Therefore, this study ultimately selected the PSO-BP neural network as the preferred prediction model of hypertension in the Guangdong region.

## 4.3 | Risk factors and prevention and control suggestions

Studies have shown that cardiovascular-related risk factors are different between elderly and the young.<sup>33</sup> In this study, age was a positive risk factor for hypertension, and its risk increased with age. Gender was also included as a risk factor, and the results showed a statistically significant gender difference, with women having a lower incidence than men. In Tao Hong's study on the hypertensive population, the incidence of hypertension in women was also lower than that in men, which may be related to the fact that women pay better attention to hypertension and comply with hypertension-management

recommendations than men and are better able to adhere to a healthy lifestyle.<sup>34</sup> Therefore, the prevention and control work in this region can consider taking targeted measures to different groups of people to improve the effectiveness and attention-grabbing quality of publicity.

An increasing number of studies have shown that the prevalence of hypertension in the dyslipidemia population is higher than that in normal people.<sup>35</sup> The model screening results of this study showed that cholesterol was the primary positive risk factor in this region, and the risk of hypertension increased with increasing cholesterol. Furthermore, the risk of hypertension increased with the increase in low-density lipoprotein cholesterol and triglycerides, and increased with the decrease in high-density lipoprotein cholesterol and hemoglobin. The findings of this study are consistent with multiple studies showing that cholesterol, low-density lipoprotein cholesterol, and triglycerides are associated with an increased risk of hypertension, while high-density lipoprotein cholesterol has cardiovascular benefits.<sup>36,37</sup> At the same time, decreased hemoglobin usually indicates a relatively poor health status of the human body, which may lead to hypertension.<sup>38</sup> In addition, studies have shown that the incidence of dyslipidemia can be effectively reduced by controlling body weight, blood glucose, blood pressure and the consumption of meat products.<sup>39</sup> In conclusion, measures such as lifestyle change and drug intervention can be considered to reduce the occurrence of dyslipidemia to reduce the incidence of hypertension. The relationship between this risk factor and hypertension has received more attention from researchers in recent years,<sup>40</sup> and studies have suggested that adequate sleep time is conducive to the control of blood pressure,<sup>41</sup> which is further supported by this study. Moreover, daily oil intake, daily salt intake, smoking, BMI and drinking are factors positively related to hypertension, and the risk of hypertension increases with age. These risk factors also overlap with the risk factors proposed in the guidelines for the prevention and control of hypertension in China.<sup>42</sup> Therefore, from the perspective of social prevention and control, health education can be carried out in this region to guide people to choose the right lifestyle to prevent the incidence of hypertension, including promoting a healthy diet, such as one low in salt and oil,<sup>43</sup> and advocating that people reduce the proportion of high cholesterol foods in their daily diets. In addition, the industry can be pushed in the direction of minding people's health, such as opening food counters, setting upper limits on oil and salt content in finished or semifinished products, and encouraging the development of corresponding health industries, such as fitness and yoga. On the other hand, from the perspective of individual prevention, it is necessary to reduce the frequency of staying up late, develop good work and rest habits, and actively carry out healthy diet strategies to control the amount of salt and oil intake in daily life. At the same time, smoking should be controlled or the smoking environment should be avoided, drinking should be limited, and the body mass index should be kept within a healthy range. The current study has certain limitations. The data of this study are the integration of surveillance data of chronic diseases and their risk factors in the same area for three years, which may introduce

some uncertainty to our results due to inconsistent data-collection dates.

## 5 | CONCLUSION

In this study, the BP neural network, the PSO-BP neural network and the logistic regression model were constructed and compared. The results showed that the PSO-BP neural network had the best performance. Furthermore, the MIV algorithm was used to screen hypertension-related risk factors and construct a disease prediction model to provide suggestions for the prevention and control of hypertension. In addition, due to the good prediction ability of the PSO-BP neural network model, it can be extrapolated to other diseases in the future. At the same time, the field should also consider combining BP neural networks with more optimization algorithms and other algorithms to be used in the research of disease risk prediction models.

## AUTHOR CONTRIBUTIONS

Yan Yan designed the research and wrote the manuscript. Rong Chen, Zihua Yang, Yong Ma, Jialu Huang, Ling Luo, Hao Liu, Jian Xu and Weiyang Chen contributed to study design, data collection and review of the manuscript. Yuanlin Ding, Danli Kong, Qiaoli Zhang and Haibing Yu critically reviewed the manuscript and put forward modification opinions. Yan Yan and Rong Chen finished the final version. All authors approved the final version.

## ACKNOWLEDGMENTS

We appreciate all authors for their contributions and physicians and participants. The study was funded by the Discipline Construction Project of Guangdong Medical University (Grant No. 4SG21276P, Grant No. 1003K20220004), the Dongguan City Science and Technology Correspondent Project (Grant No. 20221800500342), the Dongguan Social Development Science and Technology Project (Grant No. 20221800905642), the Natural Science Foundation of Basic and Applied Basic Research Foundation of Guangdong Province (Grant No. 2022A1515012407, Grant No.2021A1515010061). The Guangdong science and technology research project of traditional Chinese Medicine (Grant No. 20221209, Grant No.20211215), the Medical Scientific Research Foundation of Guangdong Province (Grant No. A2021395), the Innovation and entrepreneurship training program for college students of Guangdong Medical University (Grant No. GDMU2021112, Grant No.GDMU2021138), the Basic and Applied Basic Research Foundation of Guangdong Province Regional Joint Fund Project (Grant No.2020B1515120021), the Natural Science Key Cultivation Project of Scientific Research Fund of Guangdong Medical University (Grant No. GDMUZ2020008), the Zhanjiang City science and technology development special fund competitive allocation project (Grant No. 2020A01031, Grant No.2018A01029), the Characteristic Innovation Project of Guangdong Province General University (Grant No.2020KTSCX042, Grant No. 2019KTSCX046, Grant No. 2019KTSCX047).

## CONFLICT OF INTEREST

The authors declare no competing interests.

## REFERENCES

- Xing L, Jing L, Tian Y, et al. Urban-Rural disparities in status of hypertension in northeast China: a population-based study, 2017–2019. *Clin Epidemiol.* 2019;11:801–820.
- 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.* 2009;9:39.
- Gheorghe A, Griffiths U, Murphy A, et al. The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: a systematic review. *BMC Public Health.* 2018;18:975.
- Kivimäki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease. *Nat Rev Cardiol.* 2018;15:215–229.
- Krittana Wong C, Zhang H, Wang Z, et al. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol.* 2017;69:2657–2664.
- Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res.* 2017;121:1092–1101.
- Lyu J, Zhang J. BP neural network prediction model for suicide attempt among Chinese rural residents. *J Affect Disord.* 2019;246:465–473.
- Weng SF, Jenna R, Joe K, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One.* 2017;12:e0174944.
- Ahmadi M A, Ahmadi M R, Hosseini S M, et al. Connectionist model predicts the porosity and permeability of petroleum reservoirs by means of petro-physical logs: application of artificial intelligence. *J Pet Sci Eng.* 2014;123:183–200.
- Moosavi S R, Wood D A, Ahmadi M A, et al. ANN-based prediction of laboratory-scale performance of CO<sub>2</sub>-foam flooding for improving oil recovery. *Nat Resour Res.* 2019;28:1619–1637.
- Ahmadi M H, Nazari M A, Ghasempour R, et al. Thermal conductivity ratio prediction of Al<sub>2</sub>O<sub>3</sub>/water nanofluid by applying connectionist methods. *Colloids Surf A: Physicochem Eng Asp.* 2018;541:154–164.
- Le L T, Nguyen H, Dou J, et al. A Comparative Study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings energy efficiency for smart city planning. *Appl. Sci.* 2019;9:2630.
- Cao J, Cui H, Shi H, et al. Big Data: a parallel particle swarm optimization-back-propagation neural network algorithm based on mapreduce. *PLoS One.* 2016;11:e0157551.
- Tarkhaneh O, Shen H. Training of feedforward neural networks for data classification using hybrid particle swarm optimization, Mantegna Lévy flight and neighborhood search. *Heliyon.* 2019;5:e01275.
- 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.* 2017;12:e188746.
- Lin S, Chen S, Wu W, et al. Parameter determination and feature selection for back-propagation network by particle swarm optimization. *Knowl Inf Syst.* 2009;21:249–266.
- Aslanargun A, Mammadov M, Yazici B, et al. Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. *J Stat Comput Sim.* 2007;77:29–53.
- Hu L, Hong G, Ma J, et al. Clearance rate and BP-ANN Model in paraquat poisoned patients treated with hemoperfusion. *BioMed Res Int.* 2015;2015:1–6.
- Ramezanizadeh M, Ahmadi M A, Ahmadi M H, et al. Rigorous smart model for predicting dynamic viscosity of Al<sub>2</sub>O<sub>3</sub>/water nanofluid. *J Therm Anal Calorim.* 2019;137(1):307–316.
- Ahmadi M A, Zendehboudi S, Lohi A, et al. Reservoir permeability prediction by neural networks combined with hybrid genetic algorithm and particle swarm optimization. *Geophys Prospect.* 2013;61(3):582–598.
- Liu L. Research on digital economy of intelligent emergency risk avoidance in sudden financial disasters based on PSO-BPNN algorithm. *Comput Intel Neurosci.* 2021;2021:7708422.
- 1999 World Health Organization-International Society of Hypertension Guidelines for the Management of Hypertension. Guidelines Subcommittee. *J Hypertens.* 1999;17(2):151–183.
- Fausett L V. *Fundamental of Neural Networks: Architectures, Algorithms, and Applications*[M]. Prentice-Hall. 1993.
- Zhang, Y J. Prediction of hepatic cirrhosis with upper gastrointestinal bleeding based on particle swarm optimization BP neural network. *Shanxi Med Univ.* (2018).
- Li G, Geng E, Ye Z, et al. Indoor positioning algorithm based on the improved RSSI distance model. *Sensors.* 2018;18(9):2820.
- Liu J, Zhang K, Ma J, et al. High-precision temperature inversion algorithm for correlative microwave radiometer. *Sensors.* 2021;21(16):5336.
- Jiang H, Liu T, He P, et al. Rapid measurement of fatty acid content during flour storage using a color-sensitive gas sensor array: comparing the effects of swarm intelligence optimization algorithms on sensor features. *Food Chem.* 2021;338:127828.
- Liu X, Liu Z, Liang Z, et al. PSO-BP neural network-based strain prediction of wind turbine blades. *Materials.* 2019;12(12):1889.
- Xin Y, Cui M, Liu C, et al. A bionic piezoelectric tactile sensor for features recognition of object surface based on machine learning. *Rev Sci Instrum.* 2021;92(9):095003.
- Afrakhteh S, Ayatollahi A, Soltani F. Classification of sleep apnea using EMD-based features and PSO-trained neural networks. *Biomed Tech (Berl).* 2021;66(5):459–472.
- Suganthi M, Madheswaran M. An improved medical decision support system to identify the breast cancer using mammogram]. *J Med Syst.* 2012;36(1):79–91.
- Qin P, Zhang L Z, Zhao X W, et al. Application of back-propagation neural network in analysis of influencing factors for metabolic syndrome. *Pract Prevent Med.* 2011;18(10):1819–1822.
- Odden M C, Tager I B, Gansevoort R T, et al. Hypertension and low HDL cholesterol were associated with reduced kidney function across the age spectrum: a collaborative study. *Ann Epidemiol.* 2013;23(3):106–111.
- Chen H R, Tao H, Lin J, et al. Effects of lifestyle interventions on prevalence, awareness of hypertension in high risk population. *J Clin Cardiol.* 2010;26(6):427–429.
- Qiu Y, Liu Y J, Zheng Z F, et al. The relation between dyslipidemia and hypertension or diabetes in men aged 40–59 years old ]. *Chin J Health Lab Technol.* 2019;29(14):1738–1741.
- Khirfan G, Tejwani V, Wang X, et al. Plasma levels of high density lipoprotein cholesterol and outcomes in chronic thromboembolic pulmonary hypertension. *PLoS One.* 2018;13(5):e197700.
- Zhang Y, Lu M, Duan D D, et al. Association between high-density lipoprotein cholesterol and renal function in elderly hypertension. *Medicine.* 2015;94(14):e651.
- Huang J Y, Ma J, Guo Y et al. Peripheral blood albumin level in significantly correlated with the severity of atherosclerosis in hypertensive patients. *J South Med Univ.* 2020;40(02):240–245.
- Han F, Yin L, Lu G T et al. Relationship between high density lipoprotein and metabolic syndrome. *Med Recapitulate.* 2020;26(02):356–360.
- Zhu L Y, Fang Y, Cao Z X, et al. Effect of sleep duration and quality on hypertension incidence in Chinese middle-aged and elderly people. *Chin J Disease Control & Prevention.* 2020;24(02):176–182.
- Han K S, Kim L, Shim I. Stress and sleep disorder]. *Exp Neurobiol.* 2012;21(4):141–150.

42. Interpretation of Chinese guidelines for the management of hypertension (2018 revised edition). *Chin J Cardio Research*. 2019;24(1):24-33.
43. Xu H, Qing T, Shen Y, et al. RNA-seq analyses the effect of high-salt diet in hypertension. *Gene*. 2018;677:245-250.

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Yan Y, Chen R, Yang Z, et al. Application of back propagation neural network model optimized by particle swarm algorithm in predicting the risk of hypertension. *J Clin Hypertens*. 2022;24:1606-1617. <https://doi.org/10.1111/jch.14597>