

Retraction

Retracted: Research on the Use of Neural Network for the Prediction of College Students' Mental Health

Scientific Programming

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

Copyright © 2023 Scientific Programming. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] H. Liu and J. Xu, "Research on the Use of Neural Network for the Prediction of College Students' Mental Health," *Scientific Programming*, vol. 2022, Article ID 5759239, 8 pages, 2022.

Research Article

Research on the Use of Neural Network for the Prediction of College Students' Mental Health

Haoyue Liu ¹ and Jilin Xu²

¹Jiangsu College of Nursing, Huaian, Jiangsu 223005, China

²Huaiyin Institute of Technology, Huaian, Jiangsu 223005, China

Correspondence should be addressed to Haoyue Liu; 2013004@jscn.edu.cn

Received 22 September 2021; Revised 10 December 2021; Accepted 18 December 2021; Published 10 January 2022

Academic Editor: Rahman Ali

Copyright © 2022 Haoyue Liu and Jilin Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A healthy mental status of students plays an important role in getting quality of education. Hence, research on the prediction of college students' mental health status is of great importance and considered as a hot area of research. In this specific research study, back propagation (BP) algorithm is adopted to learn verities of characteristics of different students from the historical data of the students including: psychological characteristics, basic personal characteristics, and socio-economic characteristics. In the initial stage of the modeling, data preprocessing steps are used to prepare the data to be used by the BP algorithm for building model. The rationales behind the use of BP algorithm are its capability of handling heterogeneity of data and exploring correlations among different characteristics. The proposed model enhances the capability of BP algorithm for risk prediction of psychological problem of the students and achieves higher precision of psychological problem prediction. The results obtained show that the error between the predicted and measured values is 0.88%.

1. Introduction

Education departments at all levels in China attach great importance to the psychological early warning work of college students; however, there is a general lack of awareness of the application of information technology in the actual work, which increases the passivity of the work. In addition, the mental health level of college students is influenced by many internal and external factors, and there are complex correlations among the factors, which is a nonlinear problem [1].

In recent years, with the arrival of big data era and cloud computing, deep learning has made a great breakthrough in practical application, which not only has a deeper network structure and is a kind of intelligent machine learning, but also enables computers to learn laws from massive historical data through a series of algorithms, establish analytical models, realize the whole process from feature extraction to feature classification end-to-end, and perform intelligent recognition of new samples or make predictions for the

future [2, 3]. To perform intelligent recognition or make predictions for the future, many scholars propose that we use the machine learning and other algorithms to analyze the influencing factors that affect students' psychological problems, and through machine algorithms or building analytical models, the original data start with each layer of representation (or features) through continuous training and gradually converts them into higher-level representations to obtain the maximum data value for classification and prediction, so as to achieve the purpose of psychological problem prediction [4, 5].

As a special group carrying high expectations from society and parents, college students face significantly higher psychological pressure than other groups of their age, so it is important to study the psychological status of contemporary college students. Various aspects of psychological health are specifically expressed, including physical, intellectual, and emotional harmony; adapting to the environment and being humble to each other in interpersonal relationships; having a sense of well-being; and being able to give full play to their

abilities and lead an efficient life in work and occupation [6, 7]. It has a positive impact to explore the inner laws of various factors affecting mental health in order to cope with future mental health stress and provide new ideas for mental health work.

We seek to find out what can be explored in the irregularity of factors affecting mental health. If we use methods such as time series analysis and nonlinear systems analysis to predict, there are great difficulties and uncertainties in theoretical research and practical application [8, 9]. The BP network is a network containing an implicit layer, and its algorithm consists of forward and backward propagation. If the output layer does not get the desired output result, it then enters the error back propagation stage, and the network modifies the weights of each layer according to the back-propagation error signal to minimize the error signal, which is also the most widely used in practical applications.

Therefore, the statistical values of the factors affecting mental health in the six years from 2003 to 2008 were used as the sample to predict the weight values of the factors affecting mental health in 2010. The factors affecting college students' mental health were divided into nine categories, namely, high pressure on personal future, unsatisfactory results in college entrance exams, personal emotional frustration, conflict with classmates, high parental expectations, family financial difficulties, poor parental relationship, low professional interest, and other factors, among which high pressure on personal future and personal emotional frustration had the largest weights [10].

In summary, shortcomings of the existing studies, in respect of prediction of students mental health, include: issues in identifications of most influential characteristics of the students, and lack of mechanism for finding relationship between the characteristics and efficient modeling.

The key contributions of the study include the following:

- (i) Identification of different types of influential characteristics of different types of students from their historical data, such as psychological characteristics, basic personal characteristics, and socio-economic characteristics.
- (ii) Efficient data preprocessing for effective BP modeling.
- (iii) Effective handling of heterogeneity of data and exploring correlations among different characteristics.
- (iv) Design and development of efficient back propagation prediction model for predicting risk of psychological problem in college students.

2. Methodology

In this study, the main factors affecting the psychological problems of college students are taken as sample input, and the data are mainly reflected in the SCL-90 scale by adding factors such as basic personal information and social support, using BP neural network to establish a prediction

model of psychological problems, and learning training for the input test sample through the mapping relationship between each influencing factor and psychological problems, as long as the data of influencing factors are continuously input to complete the training analysis. The prediction model is made closer to the actual characteristics of the survey sample to achieve the purpose of psychological problem prediction [11].

Roadmap of the proposed research includes the key steps of data collection from various data sources, data preprocessing to remove inconsistencies, identifications of key characteristics for building the prediction model, training the proposed back propagation algorithm and learning hidden patterns from the historical data of the targeted students, testing the learned model with by using test dataset and some real-world examples to know the models predictive accuracy, and finally evaluating the model by comparing its results with the results of the state-of-the-art models, which is shown in Figure 1.

3. Data Acquisition and PreProcessing

The data used in this study were derived from the psychological census data of new students, basic student profile forms, and social support questionnaire statistics for the past 5 years from three representative university colleges and universities, Hunan Institute of Technology, Hunan Institute of Science and Technology, and Hunan Engineering Vocational Technology, with a total of 1036 students' historical data. In the data collection process, the psychological census data mainly refer to the 90-item symptom inventory (SCL-90 scale), which uses a five-level scale to reflect the mental health status (F1–F10) from 10-factor scores with a specific reference range of normal values; the basic student profile form is based on the fact that students in university colleges and universities have the same basic attribute characteristics: such as age, education level, education experience, etc., but also have different attribute characteristics, such as gender, major, place of birth, family status, etc. In this study, five categories of data [12], namely gender, age, major, place of birth, and family economic income, are mainly referred to; in the external influencing factors, two categories of data, namely satisfaction with material support and satisfaction with mental support, are mainly referred to.

3.1. Data Feature Extraction. In this study, based on various attribute characteristic factors related to psychological problems affecting students in colleges and universities, the adjusted results were adjusted by weighting these factors to more accurately reflect the mental health of students in university colleges and universities [13, 14]. It is divided into: firstly, collecting a large amount of small data of three categories of internal and external factors affecting psychological problems, such as basic personal situation, social support situation, etc.; secondly, classifying them according to performance characteristics, grading them according to the size of their influence, and developing the range values of weighted influence scores of each characteristic score, as well

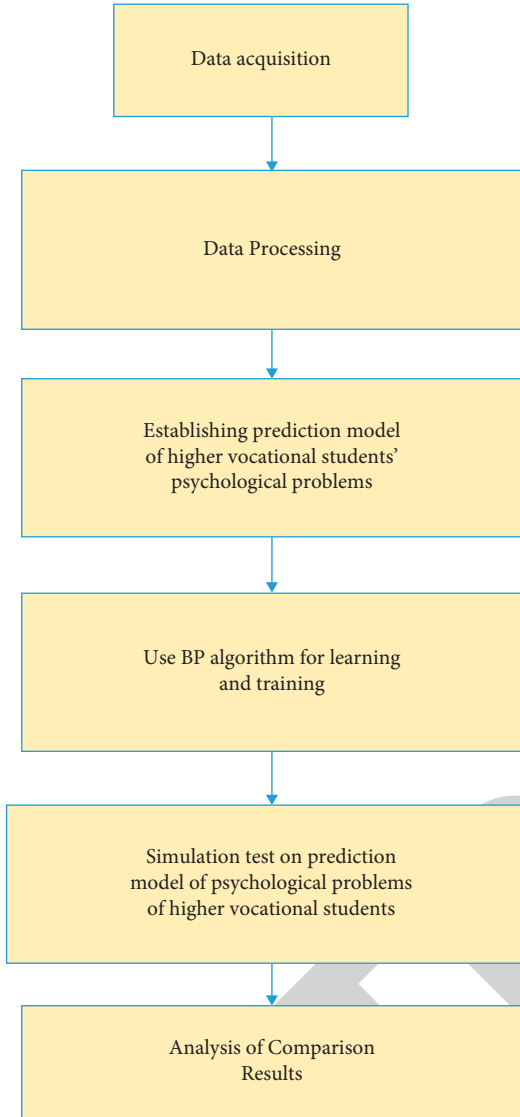


FIGURE 1: Research roadmap.

as progressive and decreasing rules. For example, there are 10 factors in the Symptoms Checklist-90 (SCL-90) scale [15], each factor contains different items, and the data are processed for factor averaging, and the results are shown in Table 1. SCL-90 is a method to evaluate psychological problems and identify symptoms, which is used in psychiatry, mental health, and educational fields for monitoring the patient's progress or treatment outcomes.

3.2. Data Normalization Process. When the sample data are trained for network, the sample data present fragmentation, diversity, and single dimension, etc. The large amount of small data that exists must be normalized to conform to the large data characteristics of traceability, quantifiability, comparability, analysis, and high dimensionality, and mapped to the $[-1, 1]$ interval for easier processing. The network input and output matrices are set as shown in Table 2.

TABLE 1: Partial results of SCL-90 data processing.

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	Other
1.45	2.11	1.91	1.87	2.09	1.23	2.23	1.09	2.09	2.15
1.79	1.23	2.34	1.12	1.86	1.83	1.96	2.10	2.11	1.34
2.01	1.89	1.56	1.67	1.34	2.13	1.65	1.67	1.23	1.66
1.32	2.56	1.23	2.09	1.03	1.44	1.23	1.23	1.78	1.76
1.68	1.09	2.34	1.02	2.78	1.61	1.67	1.56	1.22	2.31
2.21	2.45	1.57	1.89	1.56	2.01	2.22	1.20	1.41	1.21
1.23	1.56	1.23	2.62	1.45	1.89	1.23	1.84	2.01	1.77

Then the network input and output data are restricted to the interval of $[0, 1]$ or $[-1, 1]$ with the following formula:

$$\bar{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

X_i indicates the input or output data, X_{\min} indicates the minimum value in the data, and X_{\max} indicates the maximum value. The input and output data are set at $[-1, 1]$, when the original data of the middle value are converted to 0, the maximum value and the minimum value are 1 and -1 , and the formula is as follows:

$$X_{\text{mid}} = \frac{X_{\max} + X_{\min}}{2}, \quad (2)$$

$$\bar{X}_i = \frac{X_i - X_{\text{mid}}}{1/2(X_{\max} - X_{\min})}$$

4. Building a BP Network Prediction Model

Taking the construction of the BP neural network model as a tool, we optimize the BP algorithm according to the internal and external factors affecting college students' psychological problems and simulate and train the model to achieve the purpose of optimization [16]. BP learning network model refers to the design of the neural network structure in the prediction model, this study will affect the students' mental health problems structured data (training sample data) as input nodes, from the perspective of psychological prediction set filtering conditions, i.e., the number of layers (hidden layer) and number of nodes model structure is set in the middle, and the output node is the preliminary expected value of psychological problem prediction.

4.1. Determination of the Number of Neurons in the Input Layer. In the selection of the number of neurons in the input layer of the BP network, the variables that express the overall characteristics of the sample must meet the selected input variables that do not have a strong linear relationship with each other and that have a strong impact on the output and can be detected.

Combining the foundation of the previous study, the characteristic layers of the predictors of psychological problems of college students were analyzed from the internal and external perspectives, including three, namely, psychological self-assessment, basic personal situation, and

TABLE 2: Network input and output matrix.

Function	Run settings	Notes
Mapminmax	sr, ps = mapminmax (r rmin rmax)	r is the normalized input matrix
	sc, ps = mapminmax (c rmin cmax)	c is the normalized output matrix

social support. The number of 17 neurons was determined synthetically and is shown in Table 3.

For each of the 1 characteristic layer, we select the indicators with strong influence and association, and assign values to the indicators of the basic layer, with values from 0 to 10, and determine the coefficients of the basic layer, with values from 0 to 1.

$$A_i = \frac{\sum_{i=1}^3 a_y \partial_y}{\sum_{i=1}^3 y}. \quad (3)$$

4.2. Determination of the Number of Neurons in the Output Layer. The number of neurons in the output layer is determined by how the standard output is set. The output nodes represent the functional goals to be achieved by the system. In this study, the desired output of the research subject is the prediction of the risk of psychological problems of college students. Therefore, the number of neurons in the output layer was chosen to be 3 (zhengchang: equal to 1, indicating in the normal range; qingdu: equal to 1, indicating psychological confusion but within the normal range; and yanzhong equal to 1, indicating psychological abnormality.)

4.3. Determination of the Number of Hidden Layer Neurons. In the actual structural implementation of the number of hidden layer nodes, choosing a reasonable number of hidden layer nodes is a complex problem that relates to the accuracy of the whole network. In this study, the golden partition method is used to determine the number of hidden layer nodes, that is, the number of input layer nodes in this BP network is set to A , the number of output layer nodes to B , and the number of hidden layer nodes to L . The value range of L is $[A, B]$, and the specific formula is as follows:

$$a = \frac{A+B}{2} \leq L \leq (A+B) + 10 = b. \quad (4)$$

First, we calculate $2 \leq L \leq (A+B) + 10 = b$ and then calculate the formula $I_1 = 0.712 \times (b - a) + a$, I_1 as the number of hidden layer nodes counted in the training model, the error after training $E = (I_1)$; the same set I_2 as the number of hidden layer nodes to calculate the model, the error after training $E(I_2)$ If $E(I_1) < E(I_2)$, then delete (a, I_2) and repeat the above steps to determine the optimal number of hidden layer nodes L in a shorter time.

5. Deep Network Training

BP (back propagation) algorithm is a class of guided learning algorithm, mainly used in pattern recognition, classification, function approximation, etc. The basic idea of BP algorithm

is to establish a BP network to continuously adapt the network weights to correct the error function along the gradient direction down, divided into two parts: forward propagation and error back propagation. Specifically, firstly, the input sample data are passed through the input layer, and the actual output value of each unit is calculated through the implicit layer; secondly, when the actual output of the output layer is in error with the desired output value, the weights need to be continuously adjusted to carry out the learning and training process of the network, and when the error reaches the desired error, the network learning process is finished.

5.1. Training Parameter Setting. The input vector X in the BP network is given as

$$X = (x_1, x_2, x_3, \dots, x_n)^T. \quad (5)$$

Similarly, the hidden layer output vector Y is given as

$$Y = (y_1, y_2, y_3, \dots, y_n)^T. \quad (6)$$

The output layer output vector is

$$o = (o_1, o_2, \dots, o_k)^T, \quad (7)$$

and the desired output vector d is as follows:

$$d = (d_1, d_2, d_3, \dots, d_n)^T. \quad (8)$$

The weight matrix between the input node and the hidden layer is denoted by v and given as follows:

$$v = (v_1, v_2, v_3, \dots, v_n)^T. \quad (9)$$

The weight matrix between the hidden layer and the output layer is denoted by W , and W_k is the weight vector corresponding to the k th neuron of the hidden layer. The learning process is specifically divided into assigning random numbers to the weights of each node in the hidden layer and the output layer—provide training set samples—set the input layer node X_1 , calculate the output value $Y = f(\text{net}_1)$ —set the output of the hidden layer node y , the weights of the hidden layer and the output layer W_k , calculate the output value $Y = f(\text{net}_1)$ —calculate the network output error—adjust the weights—take the next sample in turn, verify that the total error of the network is less than the expected error—record the weights and end the training, which is shown in Figure 2.

5.2. Specific Training Process. The training sample data are read in, and after preprocessing such as deletion and addition to the sample data, the initial values are assigned to all the management data. The load function is selected to read the sample data as shown in Table 4.

TABLE 3: The list of the synthetically determined neurons.

Neuron	Description	Scale
A1	Gender (xb)	
A2	Age (nl)	
A3	Major (zy)	
A4	Place of birth (syd)	
A5	Family economic income (jtsr)	
B1	Degree of material support (wzzc)	
B2	Degree of spiritual support (jszc)	
C1	Somatization factor (qth)	1, 4, 12, 27, 40, 42, 48, 49, 52, 53, 56, 58
C2	Obsessive-compulsive factor (qp)	3, 9, 10, 28, 38, 45, 46
C3	Interpersonal sensitivity	6, 21, 34, 36, 37, 41, 61, 69, 73
C4	Depression factor (yy)	5, 14, 15, 20, 22, 26, 29, 30, 31, 32, 54, 71 79
C5	Anxiety factor (jl)	2, 17, 23, 33, 39, 57, 72, 78, 80, 86
C6	Hostility factor (dd)	11, 24, 63, 67, 74, 81
C7	Terror factor (kb)	13, 25, 47, 50, 70, 75, 82
C8	Paranoia factor (pz)	8, 18, 43, 68, 76, 83
C9	Psychoticism factor (jsbx)	7, 16, 35, 62, 77, 84, 85, 87, 88, 90
C10	Other factors (qt)	19, 44, 59, 60, 64, 66, 89

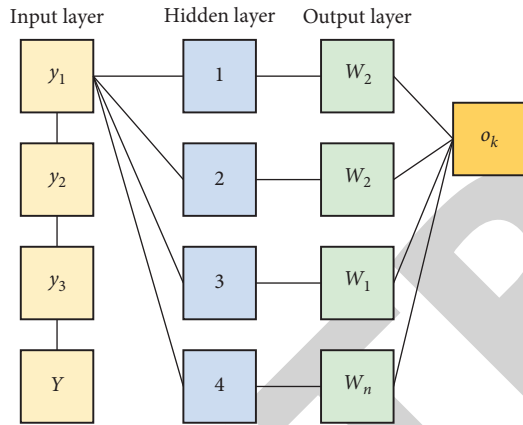


FIGURE 2: BP algorithm operation process diagram.

TABLE 4: Selection of load function to read sample data.

Function	Run settings
Load	$Y = \text{lad}(\text{input}')$ $O = \text{load}(\text{output}')$

About 100 sets of data are set as training samples and 20 sets as prediction samples, all data were trained according to their progressive rules, and the results were compared with the actual values to obtain the best matching model under the above training data, which can become more accurate and realistic with the increase of the number of training specimens and the increase of the genus data in the specimen data, as shown in Table 5.

We use the network training algorithm to select the training algorithm and set the learning training network as 17 neurons, and the neural network model structure is shown in Table 6.

The network is trained using the BP algorithm and the parameters are continuously adjusted.

```

Net.trainparam.epochs = 980;
Net.trainparam.ir = 0.12;
Set a training goal of 0.000000001.
Thennet.trainparam.goal = 0.000000001.
    
```

In this study, the neural network is established in the continuous learning training, the error is constantly reduced in the expected 900 steps, the error rate control at 3% can be completed in a relatively short period of time to fit the data, and basically the prediction of college students' psychological problems research is achieved.

6. Simulation Test

6.1. Example Analysis of Prediction of Factors Affecting Mental Health. The statistical values of the factors affecting mental health for the six years starting from 2002 to 2007 were used as a sample to predict the values of the factors in 2008. We used statistical values of the factors affecting mental health from 2003 to 2008 as a sample to predict the value of 2009 for each factor, and calculated the error so that the square root error between the predicted value and the real value was limited to 1%. At this point, the coefficients of the BP neural network were trained accurately, and finally the statistical values of the mental health factors from 2004 to 2009 were used as samples to predict the values of 2010 for each factor. It should be noted that since the "other" factor is the sum of other factors and 1 by subtraction, it has no effect on the overall error of the network, so it can be excluded when making the error, that is, the square root error of the predicted value and the true value does not include the "other" factor. " factor.

As an example, the values of the factors affecting mental health in 2009 were predicted. The structure and model of the neural network are as follows: (1) input layer and output layer: the statistical values of the factors from 2003 to 2008 were used as the input vector, with 6 nodes, and the statistical values of the factors in 2009 were used as the training sample, with 1 output node; (2) the hidden layer: the nodes

TABLE 5: Data classification.

Function	Run settings	Notes
Divdevc	Ixl, cs Fdivdevc (r.c. 20%)	XL is the training data, CS is the prediction data, and 20% is the percentage of prediction data

TABLE 6: BP network establishment settings.

Function	Run settings	Notes
Newfftrainlm	net = newff (sr, sc, 17) net = trainlm (net, sr, sc)	The hidden layer consists of 17 neurons

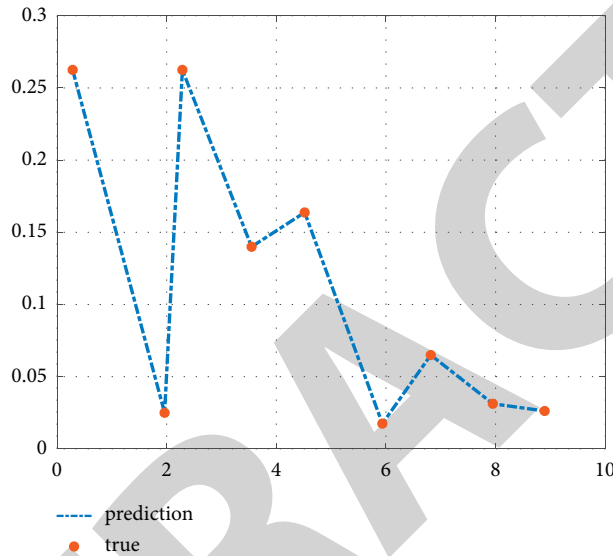


FIGURE 3: Predicted versus predicted values of mental health factors.

TABLE 7: Predicted values of factors affecting mental health in 2009.

Shadow response factor	Great pressure on personal future	Gao cheng is not ideal for performance appraisal	Personal feelings are frustrating	Tongguan spearology department shield	Father's period is too long and mother's expectation is high	Family positive difficulties	Other
Weight	0.268	0.023	0.266	0.142	0.154	0.033	0.026

of the hidden layer are not unique, and the prediction results of 2009 are compared with the error size, and the model is trained to contain 4 hidden layer nodes; and (3) parameter determination: initial weights $w = -0.2 \sim 0.2$, learning rate LP. $r = 0.1$, inertia coefficient = 0.6, correction period $T = 10$, and number of training steps = 500.

The command to build the BP network is `net = newff (a, b, {'tansig', 'purelin'}, 'traingd')`, and the command to train the network is `[net, tr] = train (net, p, t)`. The training samples are substituted back into the network for simulation, and the command `b = sim (net, p1)` is implemented, where b is the prediction result and $p1$ is the training sample. The training error curve of the network is shown in Figure 1, and it can be seen that the network has good training results. The comparison curve between the predicted and true values for 2009 is shown in Figure 4, and the overall square root

error is within 1%, thus indicating that the model can be used to predict the factors affecting mental health in the next year. The predicted results of the factors affecting mental health in 2009 are shown in Table 7.

The training error curves of the network all show that the network is well trained, and the square root error of the sum of the predicted and true values of the factors affecting mental health is within the acceptable range of the results, leading to the prediction results for 2010. The final prediction results are shown in Figures 3 and 4 and Table 8.

6.2. Real Case. The test data were entered into the constructed neural network model for simulation testing, and it can be concluded through several groups of data simulation tests that the error of each group of prediction results is

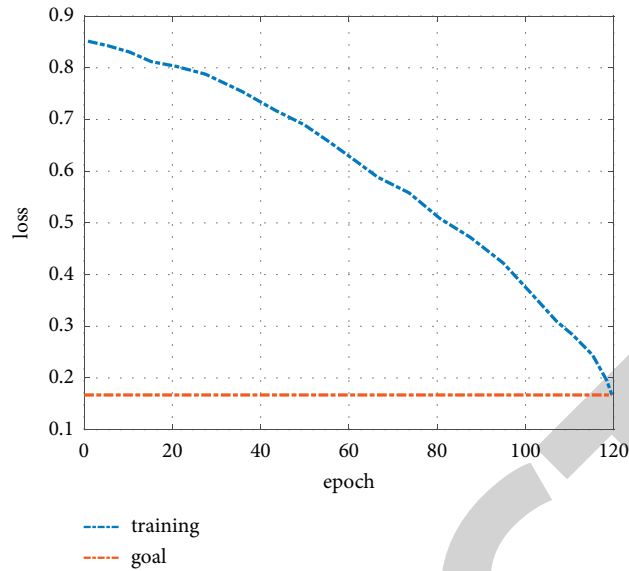


FIGURE 4: Training error curve.

TABLE 8: Predicted values of factors affecting mental health in 2010.

Factor	Great pressure on personal future	The score of college entrance examination is not ideal	Personal emotional frustration	Contradiction of classmate relationship	Parents expect too much	Other
<i>b</i>	0.274	0.011	0.274	0.142	0.161	0.031

TABLE 9: Simulation test.

Function	Run settings
Sim	An = sim (net, sr test)
mapminmax	Bpsc-mapminmax (reverse, an, s)

TABLE 10: Analysis of results for some sample data.

S. no.	Prediction output	Standard value	Absolute error	Relative error (%)
1	1.78	1	-0.132	-8.81
2	2.21	2	-0.089	-5.34
3	1.97	1	0.075	2.09
4	1.76	1	0.121	3.67
5	2.08	2	-0.097	-2.45
6	2.12	2	0.924	4.56
7	1.89	1	-0.216	2.09

relatively small and basically meets the requirements, proving that the model can predict the problem for the studied sample condition, as shown in Table 9.

In this study, the factors affecting college students' psychological problems are mapped to the corresponding mental health conditions through BP neural network learning training, and after feature selection, continuous learning training, and self-training, such that the initial weights and thresholds are adjusted to achieve problem prediction and complete result comparison, as shown in Table 10.

7. Conclusion

Combined with the mental health characteristics of college students at the present stage in China, the data of internal and external influencing factors are reprocessed to establish a neural network prediction model, and the BP algorithm is used for continuous learning and training. The multilayer forward BP network applied in the model itself still has some defects, for example, the learning speed is very slow, it is easy to fall into local minimal values, and the contradiction between the prediction ability of the network and the

training ability. However, compared with the traditional methods for predicting the mental health status of college students, the prediction method of BP neural network avoids the tedious conventional modeling process. This study establishes a mathematical prediction model based on the principle of BP neural network, which can change the parameters of the network, making the calculation of the prediction system simple and flexible, and greatly improving the model prediction efficiency and prediction accuracy. To a certain extent, it provides a reference for the study of college students' mental health.

In future, we plan to include and consider more characteristics, such as family history, social circle, drug history, educational background, and many more. Apart from these characteristics, different correlational models will be tested and for the prediction, deep learning-based models will be tested.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would gratefully like to acknowledge the support provided by the Jiangsu College of Nursing teaching reform project (no. CJB2019Y28).

References

- [1] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of "big data" on cloud computing: review and open research issues," *Information Systems*, vol. 47, pp. 98–115, 2015.
- [2] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97–107, 2013.
- [3] C. A. Bail, "The cultural environment: measuring culture with big data," *Theory and Society*, vol. 43, no. 3-4, pp. 465–482, 2014.
- [4] Z. Zhang, C. Zhang, M. Li, and T. Xie, "Target positioning based on particle centroid drift in large-scale WSNs," *IEEE Access*, vol. 8, pp. 127709–127719, 2020.
- [5] K. Sherren, J. R. Parkins, M. Smit, M. Holmlund, and Y. Chen, "Digital archives, big data and image-based culturomics for social impact assessment: opportunities and challenges," *Environmental Impact Assessment Review*, vol. 67, pp. 23–30, 2017.
- [6] Y. Sun, "Construction and research of digital archives cloud platform based on big data management," *Journal of Physics: Conference Series*, vol. 1881, no. 4, Article ID 042094, 2021.
- [7] W. Yong, L. Liming, and Q. Yongsheng, "Improvement of big data retrieval algorithm in the intelligent archives management," vol. 1, pp. 487–491, in *Proceedings of the 2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, vol. 1, IEEE, Qingdao, China, July 2015.
- [8] S. Ma, "The function of the establishment of digital archives in colleges and universities to the optimization of resources," *IOP Conference Series: Materials Science and Engineering*, vol. 750, no. 1, Article ID 012057, 2020.
- [9] T. Xie and C. Zhang, "Utilizing active sensor nodes in smart environments for optimal communication coverage," *IEEE Access*, vol. 7, no. 99, pp. 11338–11348, 2018.
- [10] C. Zhang, T. Xie, K. Yang et al., "Positioning optimisation based on particle quality prediction in wireless sensor networks," *IET Networks*, vol. 8, no. 2, pp. 107–113, 2019.
- [11] T. Owens and T. Padilla, "Digital sources and digital archives: historical evidence in the digital age," *International Journal of Dental Hygiene*, vol. 1, no. 3, pp. 325–341, 2021.
- [12] G. Taylor, "Finding 'anonymous' in the digital archives: the problem of Arden of Faversham," *Digital Scholarship in the Humanities*, vol. 34, no. 4, pp. 855–873, 2019.
- [13] T. Birkner, E. Koenen, and C. Schwarzenegger, "A century of journalism history as challenge," *Digital Journalism*, vol. 6, no. 9, pp. 1121–1135, 2018.
- [14] C. Gao and X. Huang, "Research on the security evaluation system of digital archives in the cloud computing environment," *Archives Science Study*, vol. 1, pp. 77–83, 2018.
- [15] L. R. Derogatis, R. S. Lipman, and L. Covi, *SCL-90. Administration, Scoring and Procedures Manual-I for the R (Revised) Version and Other Instruments of the Psychopathology Rating Scales Series*, Johns Hopkins University School of Medicine, Chicago, IL, USA, 1977.
- [16] H. Stančić and V. Bralić, "Digital archives relying on blockchain: overcoming the limitations of data immutability," *Computers*, vol. 10, no. 8, p. 91, 2021.