Research Article

Research on Students' Mental Health Based on Data Mining Algorithms

Mengjun Luo [ORCID]

College of Preschool Education and Humanities, Dongguan Vocational and Technical College, Dongguan, Guangdong 523808, China

Correspondence should be addressed to Mengjun Luo; luomj@st.btbu.edu.cn

With the diversification and rapid development of society, people's living conditions, learning and friendship conditions, and employment conditions are facing increasing pressure, which greatly challenges people's psychological endurance. Therefore, strengthening the mental health education of students has become an urgent need of society and a hot issue of common concern. In order to solve the problems of high misjudgment rate and low work efficiency in the current mental health intelligence evaluation process, a mental health intelligence evaluation system based on a joint optimization algorithm is proposed. The joint optimization algorithm consists of an improved decision tree algorithm and an improved ANN algorithm. First, analyze the current research status of mental health intelligence evaluation, and construct the framework of mental health intelligence evaluation system; then collect mental health intelligence evaluation data based on data mining, use joint learning algorithm to analyze and classify mental health intelligence evaluation data, and obtain mental health intelligence evaluation results. Finally, through specific simulation experiments, the feasibility and superiority of the mental health intelligent evaluation system are analyzed. The results show that the system in the article overcomes the shortcomings of the existing mental health intelligence evaluation system, improves the accuracy of mental health intelligence evaluation, and improves the efficiency of mental health intelligence evaluation. It has good system stability and can meet the actual current situation, which are requirements for mental health intelligence evaluation.

1. Introduction

Due to the rapid development of modern society, people's pressure to survive has also risen. The rising pressure of social competition [1, 2] has caused more and more people to have psychological problems such as emotions and behaviors. Every year, psychological barriers [3] and even suffering appear. The number of mentally ill people began to grow substantially. Resolving people's psychological problems [4] caused by various factors has become the most important work. The directions that need to be studied are how to provide people with mental health services [5] and improve the results of mental health work [6]. In response to the above problems, it is necessary to use innovative ideas and concepts, combined with modern technology and information technology, to build a mental health intelligent evaluation system [7] to improve the scientific level of

mental work. Some scholars [8–10] have built a mental health service system for the elderly based on the internet technology, and some scholars have built an automatic evaluation system for users' mental health in online forums based on multifeature fusion. Improving users' mental health has improved the data collection and data collection of mental health. Regarding its management, its function is too one-sided, and its applicability is weak. Traditional mental health assessment methods [11–14] are mostly carried out in the form of scales or questionnaires. There will be certain limitations in terms of survey objects and the accuracy of survey results. With the rise of the Internet technology, the application of network technology to mental health assessment has become a new phenomenon. The tendency of survey respondents to self-assess in a more concealed environment can reduce their psychological pressure.

Popular internet software will record user status to varying degrees. It can be seen that the Internet evaluation system is a highly applicable way of evaluating the mental health of college students [15]. Papineni et al. [16] proposed an automatic evaluation method for online forum users' mental health, designed an evaluation framework, hierarchically analyzed the data reflecting the user's mental health, and used different methods to fuse user psychological data to obtain a more comprehensive evaluation method. The system takes a long time; Soet [17] designed a college student mental health consulting service system based on the web, carefully analyzed user needs, and divided the system into different functional categories to build the entire mental health consulting service system to help users better provide psychological services, but the recall rate is not high. David et al. [18] selected the research object for the elderly and used different theoretical models for this group to conduct mental health tests, evaluate the advantages and disadvantages of different models, and integrate the advantages of various theoretical models. A mental health assessment method [19] is proposed, but the accuracy rate needs to be improved. Yang et al. [20] research object is also the elderly, and they analyzed the system requirements, designing an overall mental health assessment framework, and analyzed them one by one, but the system is efficient in operation room for improvement. With the continuous improvement of the ability to collect, analyze, and process data in the era of big data, it is no longer necessary to use paper questionnaires to sample data. People can accurately capture various data dynamics through computers combined with big data technology. In past research, schools mainly used paper questionnaires to measure students' mental health [21]. This method consumes manpower and material resources and is costly. Using big data technology, mental health assessment can be extended to a larger scope. Students use the internet to generate massive amounts of big data and use big data technology [22] to build a mental health evaluation system, which can fully grasp the psychological dynamics of students, combined with the establishment of mental health scales, and can more accurately predict the mental health of students.

The application of machine learning [23–25] in the field of mental health has gradually become an important development trend. Machine learning is a process of automatically analyzing or predicting new and unknown data by discovering the laws of a large amount of existing data and information. Tom gave a formal definition of machine learning: assume that performance is used to evaluate the performance of a computer program on a certain task if the program obtains performance improvement on the task by using experience. The currently available data types for machine learning are (1) text data [26]. Its sources include social media, clinical assessments and records, electronic health records, and diaries. In particular, social media can provide a large amount of data reflecting individual psychological and behavioral traces, which has higher ecological validity. At present, it mainly focuses on the prediction of suicide risk, the analysis of personality traits, the prediction of mental health status, and subjective well-being. (2) Survey

data [27] is commonly used is the combination of demographic variables, psychological scales, and publicly authoritative statistical data to predict and diagnose mental illnesses such as personality types and mental disorders. (3) Brain imaging data [28]: the detection of traditional brain imaging data relies too much on expert judgment, is time-consuming and laborious, and has a high misdiagnosis or missed diagnosis rate. The accuracy of automatic classification using machine learning is better than expert judgment. Currently, it is commonly used in the prediction of cognition, memory, emotions, and addictive behaviors. (4) Behavioral and physiological data [29]: it mainly comes from eye-tracking, portable mobile devices, multichannel physiological recorders, and expression analysis systems. Machine learning algorithms are divided into three categories: (1) supervised learning [30, 31]; (2) semisupervised learning [32]; (3) unsupervised learning [33]. Supervised learning is to obtain the relationship between data with features and labels through training to judge new data. Use labeled data to train the model. The model can predict the label of new data. For example, [34] the machine learning model can automatically identify based on the brain imaging data of traumatic brain injury patients and normal individuals to distinguish individual illnesses. Unsupervised learning is to discover the relationship between unfeatured and labeled data through training to judge new data. In many cases, labeled data sets may be difficult to obtain, so the model's predictive ability can be enhanced by using unlabeled data, but it should be understood that no algorithm can solve every problem well. Hence, it is necessary to compare multiple algorithms to determine which algorithm is most suitable for a particular data set and task to be solved. Commonly used machine learning algorithms include decision trees [35], naive Bayes [36], support vector machines [37], K-nearest neighbors [38], logistic regression [39], multilayer perceptions [39], random forests [40], and K-Means [41].

To improve the effectiveness of mental health intelligence evaluation, this paper proposes a mental health intelligent evaluation system based on data mining, which can effectively promote system information management, solve problems such as the old system resources, make perfect use of internet advantages, create a new internet environment, and effectively solve the psychological problems of contemporary people.

2. Related Work

The goal of psychology [42] is to describe, explain, predict, and control behavior. At present, most psychological research focuses on describing and explaining the relationship between variables. Research that can truly achieve the prediction goal is not common, and traditional psychological research [43] is due to the small sample size. The results of some studies are contradictory and uncertain due to low data quality and lack of covariate information. And, the significance test based on p value faces a crisis of duplication, that is, the irreproducibility of research results. Strict and systematic use of machine learning cross-

validation technology [44] can provide great potential for realizing the reproducibility of psychological research. Technology-based machine learning can construct learning models from massive amounts of data, more accurately identify the underlying laws of the data, and have stronger generalization capabilities. The model can be applied to different samples or groups, thereby minimizing prediction errors to make more accurate results.

Many researchers try to use machine learning to predict complex psychological problems of individuals, such as predicting stress disorders and anxiety disorders. The research of Tian et al. [45] and Zhu et al. [46] found that only the effective features that can predict the possibility of suicide can reach thousands. They use a multilayer perceptron algorithm to build a suicide recognizer for Weibo social media to evaluate the suicide of Weibo users in real time. As regards possibility, the prediction accuracy rate can reach about 94%. Carpenter [47] used preschool psychiatric assessment data and machine learning to assess the risk of anxiety in children aged 2 to 5 years, and their prediction accuracy for generalized anxiety disorder and separation anxiety disorder was as high as 90%.

Moreover, machine learning usually does not make assumptions or establish relationships between independent variables and dependent variables. It usually makes judgments among all possible independent variables, gives prediction results, and outputs important independent variables. Walsh [48] used the random forest model to predict future suicide behaviors of suicide attempted patients and normal subjects. The independent variables used include demographic variables, medical records, and socioeconomic status, and the accuracy of the predictive indicators can reach about 89%. Moreover, the long-term prediction cycle is the most important predictor for those who have attempted suicide in the history of hospital visits; in the short-term prediction cycle, dependence on addictive drugs is an important predictor. The type indicates that the activation degree of the amygdala and the hippocampus during the cognitive reassessment task and the amygdala gray matter volume have a good predictive ability for borderline personality disorder, and its accuracy and sensitivity can reach more than 70%. Such prediction results are correct. The development and selection of personalized mental interventions are crucial. In order to enable people to have an accurate understanding of their mental health, and at the same time to promote the scientific and informatization of mental health guidance, this paper constructs a mental health intelligent evaluation system based on a decision tree algorithm, using scientific mental health evaluation. The tool provides a comprehensive and objective reflection of the user's mental health. The overall structure of the system is shown in Figure 1.

3. Methodology

3.1. Improved Decision Tree Algorithm

3.1.1. Entropy for Preprocessing. The client software of this system uses the method of psychological evaluation data mining to realize the evaluation results, processes the obtained user evaluation result data and establishes a database, analyzes the evaluation results through the decision tree algorithm, and obtains evaluation. Psychological assessment data mining includes extracting data from the database, cleaning the data, selecting the mining model, and outputting the results. The complete data mining process is shown in Figure 2. In this paper, the system uses the user's answer results to establish an initial data set. After preprocessing data integration, data extraction, data cleansing, and data conversion, a mental health assessment data set to be mined is obtained. The decision tree algorithm is applied in the mining process. The data classification is obtained through the mining results, thereby obtaining the specific classification of the user's mental health status.

The decision tree algorithm is an inductive learning algorithm, a classification rule obtained by inducing a set of chaotic examples based on examples. The decision tree mainly includes two steps when processing the classification problem. The first is to generate a decision tree classification model through the learning and training set; the second is to classify unknown types of samples through the model. In the classification process of a certain sample type, the root node is the starting point, a leaf node is an endpoint, and the sample attributes are gradually tested in the downward direction of the branch. In this paper, the C4.5 decision tree algorithm is applied, and the splitting index is the information gain rate, which solves the problem of biasing the selection of multiple attributes when using information gain to select test attributes. The definition formula of information gain rate is

$$\text{gain ratio} = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)}. \quad (1)$$

Gain(S, A) is the information gain of the attribute A and SplitInformation(S, A) is the breadth and uniformity of splitting the sample set S according to the attribute A.

The decision tree C4.5 algorithm is used to classify the mental health intelligence assessment data to provide data support for the system in this paper.

$$\text{SplitInformation}(S, A) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_2\left(\frac{|S_i|}{|S|}\right). \quad (2)$$

In the process of decision tree generation, the most important thing is to determine the split target. In order to determine the splitting index in the C4.5 algorithm, it is necessary to compare the size of the attribute information gain rate in each training sample data and select the attribute with the largest information gain rate and higher than or equal to the average value of all attributes as a branch node of the decision tree for the existence of continuous descriptive attributes; the continuity needs to be divided to obtain a discrete set of intervals. Discretization methods include the following:

(1) Discretize the continuity attribute A of the continuous interval value. The continuity attribute values are arranged in the training set S samples in
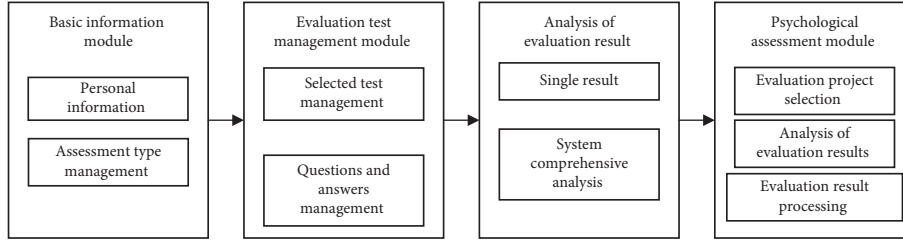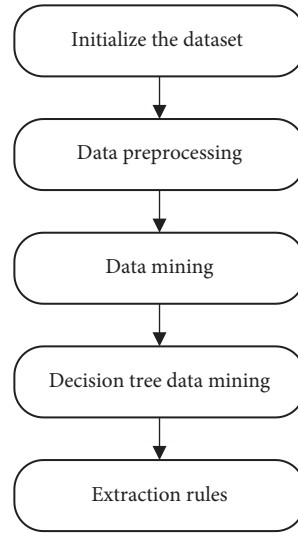
FIGURE 1: Overall structure of the system.

FIGURE 2: Complete data mining process based on decision tree.

ascending order, and the minimum and maximum values in the interval are assigned to MIN and MAX, respectively, to obtain the value sequence of the attribute value which is $\{A_1, A_2, \ldots, A_n\}$.

(2) Suppose A_i is the equal division point. If there are N equal division points in the interval $[\text{MIN}, \text{MAX}]$, calculate in order, and the formula for the i-th equal division point is $A_i = \text{MIN} + i * (\text{MAX} - \text{MIN})$, where $i = 1, 2, \ldots, N$.

(3) Calculate the information gain values in the interval values $[\text{MIN}, A_i]$ and $[A_i, \text{MAX}]$, respectively, and there is a total of $N - 1$ dividing points for comparison.

(4) Suppose the segmentation threshold of continuity attribute A is the segmentation point with the largest information gain rate A_k, divide the data set based on this, and set the attribute value in $[\text{MIN}, A_k]$ and $[A_k, \text{MAX}]$.

According to the above discretization method, the information gain rate of all attributes in the candidate attribute set is solved, and the test attribute is the attribute with the largest information gain rate. Use all possible values in the sample set to divide the sample to obtain several subsample sets. Use the same method to continue to divide all the subsample sets. Until it cannot be divided, the decision tree

is generated. When using the C4.5 algorithm to generate a decision tree, the selected target class affects the determination of each class. When evaluating the decision tree, information entropy needs to be used to obtain

$$S = -\sum_i (P_i * \log_2 P_i). \qquad (3)$$

Information gain is the effective reduction of information entropy. According to information entropy, the variable level used for classification can be determined. If there are two classes, class A and class B, in the training set S, and they, respectively, include x and y records belonging to class A and class B, then the formula for determining the total amount of information for a certain record in the training set S is

$$\text{Info}(s) = \text{Info}(s_a, S_b) = -\left(\frac{x}{x+y}\right) \cdot \log\frac{x}{x+y}$$
$$+ \frac{y}{x+y} \cdot \log\frac{y}{x+y}. \qquad (4)$$

Suppose variable C is the root node of the decision tree, and the subclasses of training set S are divided into $\{S_1, S_2, \ldots, S_k\}$; then $S_i (i = 1, 2, \ldots, k)$ includes X_i, Y_i records that belong to class A and class B, respectively. Get the information volume formula classified in all subcategories:

$$\text{Info}(C, S) = \sum_{i=1}^{k} \frac{x_i + y_i}{x + y} \text{Info}(S_{iA}, S_{iB}). \quad (5)$$

Suppose variable C is the classification node of the decision tree, and its information increment value is the largest among all variable information increment values. Then the information increment formula of variable C is

$$\text{Gain}(C) = \text{Info}(S) - \text{Info}(A, S). \quad (6)$$

Then the information gain function is defined as

$$\text{Gain}(D, C) = \text{Info}(S) - \text{Info}(D, S),$$

$$\text{Info}(S) = -(P_1 * \log P_1 + \cdots + P_k * \log P_k), \quad (7)$$

$$\text{Info}(D, S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} \text{Info}(S_i).$$

3.1.2. Improved Decision Tree. From the basic principle of the C4.5 algorithm, it can be known that the selection of attributes when generating a decision tree is based on the principle of information theory. Since the calculation of the information gain rate formula involves multiple logarithmic function operations, this requires that the library function is called multiple times during calculation, which greatly increases the calculation time. In response to this problem, an improved method for calculating the information gain rate is proposed; that is, the mathematical Taylor formula and McLaughlin's formula are used to simplify the calculation of the information gain rate of the C4.5 algorithm, which greatly reduces the calculation of the algorithm. The improved C4.5 algorithm is named TAM-C4.5 algorithm.

Since the derivative of $\ln(x)$ at $x = 0$ is meaningless, and the commonly used probability value range in the calculation formula of information gain rate is between $[0, 1]$, this paper chooses the Maclaurin formula of $\ln(x + 1)$ which improves the calculation formula of information gain rate in traditional C4.5, as the formula

$$\ln(x + 1) \approx x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots + (-1)^{n-1}x^n,$$

$$\ln(x) \approx (x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 \quad (8)$$

$$+ \cdots + (-1)^{n-1}(x - 1)^n.$$

When $x \in (0, 1)$, the formula will become as follows:

$$\ln(x) \approx (x - 1) - \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3. \quad (9)$$

Through the above approximate simplification, it is possible to convert logarithmic operations into non-logarithmic operations, and the abovementioned conversion characteristics can be used to eliminate complex logarithmic operations in the information gain rate formula, so as to simplify the calculation formula and improve the efficiency of tree construction.

The conversion formula of category information entropy is as follows:

$$\text{Info}'(S) = -\sum_{i=1}^{m} \frac{S_i}{S}\log_2\frac{S_i}{S}. \quad (10)$$

Similarly, the conversion formula of conditional information entropy and split information entropy is as follows:

$$\text{Info}'_A(S) = -\frac{1}{\ln(2s)}\sum_{j=1}^{k}\sum_{i=1}^{m}\left[\frac{s_{ij}(s_{ij} - s_j)(11s_j^2 + 2s_{ij}^2 - 7s_{ij}s_j)}{6s_j^2}\right],$$

$$\text{Info}'(A) = -\frac{1}{\ln(2s)}\sum_{j=1}^{k}\left[\frac{s_j(s_j - s)(11s^2 + 2s_j^2 - 7s_js)}{6s^2}\right]. \quad (11)$$

Therefore, the formula for calculating the information gain rate after conversion is as follows:

$$\text{Gain} - \text{Ratio}'(A) = \frac{\text{Info}'(S) - \text{Info}'_A(S)}{\text{Info}'(A)}. \quad (12)$$

Analyzing the improved calculation formula shows that the category information entropy is the same every time the information gain rate value of the condition attribute is calculated. Since $1/\ln(2s)$ is omitted in each part of the simplified formula in this article, in order to ensure the classification accuracy of the algorithm, this article is calculating the category conditional entropy, the improved formula is used to try to ensure that the order of the information gain rate of each condition attribute is not changed, and the classification accuracy is not affected. The traditional C4.5 algorithm needs to call a function to perform a large number of logarithmic function operations. The improved algorithm proposed in this paper only needs a simple four-mixed operation, which eliminates the frequently called logarithmic operation in the information gain rate calculation formula, and the system operation speed is greatly improved.

3.2. Improved ANN Algorithm

3.2.1. ANN for Prediction. An ANN is a structure and calculation model that imitates a biological neural network. It is usually used to estimate or approximate a function. The neural network is mainly composed of the input layer, hidden layer, and output layer. In practice, each neuron in the network's input layer represents a feature, and the number of neurons in the output layer is one. If the Softmax classifier as shown in Figure 3 is used, the number of output neurons is two, a multiclassification problem. The number of hidden layers and hidden layer neurons is manually set. Figure 4 shows a basic neural network model.

The MP neuron model receives input signals from other neurons, and the connections are transmitted. Then the total input received by the neuron is compared with the neuron's threshold, and after the activation function is processed, the neuron's output is generated.
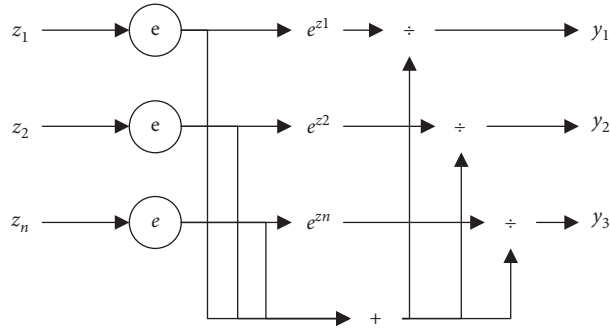
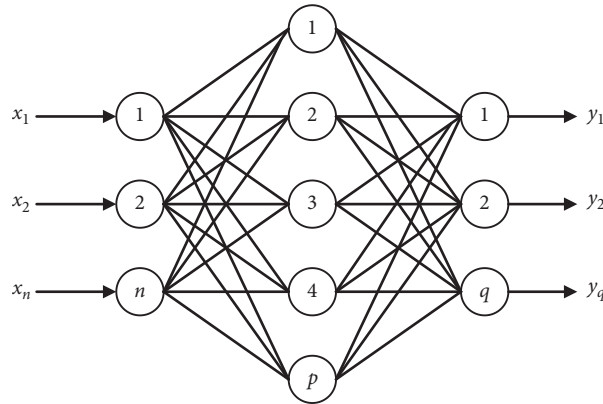FIGURE 3: The structure of Softmax classifier.

FIGURE 4: The structure of ANN.

Ideally, the form of the activation function should be a step function (that is, the modified linear unit ReLU). There are two output types: 0 or 1. 0 means that the neuron is not excited, and 1 means that the neuron is excited. But the step function is not smooth and discontinuous, so the activation function usually chooses the sigmoid function.

It can be seen from the function image that the value range of the function is $(0, 1)$. That is, the function value falls between 0 and 1. The property of the sigmoid function is that it can compress the input value that changes in a larger range into the $(0, 1)$ interval, so it is also called a squashing function. By connecting many neural units according to a certain level, a neural network is obtained.

3.2.2. The Cross-Entropy Cost Function to Optimize Backpropagation. When using this method, the real output y should be transformed by the Softmax function first and then calculated in the substitution cost function. The calculation formula of the softmax function and classification cross-entropy is as follows:

$$y_{\text{out}}^i = \text{softmax}\left(y_k^i\right) = \frac{e^{y_k^i}}{\sum_{q=1}^k e^{y_k^i}},$$

$$J(\theta) = -\sum_{i=1}^m \sum_{k=1}^k y_k^i \ln y_{\text{out}}^i + \frac{\lambda}{Z_m} \sum_{i=1}^s \sum_{j=1}^{sl+1} \theta_{j.i}^{(i)^2}.$$

(13)

When we improve the classification loss, we need to optimize the gradient of a step size through gradient descent. At this time, we require loss to give the partial derivative of each weight matrix and then apply the chain rule. Then the first step in this process is to derive the softmax.

Calculate the error for the output y that the Softmax function has not transformed; that is, find the derivative of a. According to the chain rule, it can be transformed into an error function. The derivative of the output after the transformation of the softmax function is multiplied by the derivative of the output after the transformation of the SoftMax function, and the calculation formula for the derivative of a is as follows:

$$\frac{\partial J(\theta)}{\partial a} = \frac{\partial J(\theta)}{\partial s} \cdot \frac{\partial s}{\partial a}. \tag{14}$$

Among them, the formula for the loss function to derive the softmax layer is

$$\frac{\partial J(\theta)}{\partial s} = -\sum_{k=1}^K y_k \cdot \frac{1}{S_k},$$

$$S_k = \frac{e^{a_k^3}}{e^{a_k^1} + e^{a_k^2} + e^{a_k^3} + \cdots + e^{a_k^n}},$$

(15)

where y is the expected output value of the k-th neuron and s_K is the output of the k-th neuron transformed by the softmax function, namely, y_{out}.

3.2.3. Improved ANN Algorithm.

Based on ANN's three-layer perceptron and nonlinear optimization capability, it can approximate any nonlinear function. However, ANN also has areas that need to be improved. (1) There are many times of training, and the convergence speed is relatively slow. The ANN algorithm also needs a lot of learning and training to converge for simple and common problems. For more complex problems, the training time will be longer. Using the gradient descent method to minimize the loss function will inevitably cause jagged images, resulting in low algorithm efficiency. In order to better solve this problem and improve the convergence speed of the ANN network, the momentum term can be appropriately increased, the error function can be improved, the learning rate can be adaptively changed, and the steepness factor can be used. (2) It is easy to fall into the local optimum, and the global optimum cannot be guaranteed. Because the back-propagation algorithm uses the gradient descent method, the weight space is a parabola with a minimum point, and there are multiple minimum points. Therefore, different training starting points will result in no optimal solution.

At present, there are many researches on ANN improvement. The more commonly used optimization algorithms are additional momentum algorithm, variable algorithm, adaptive learning rate method, RPROP method, conjugate gradient algorithm, Newton algorithm, Levenberg–Marquardt algorithm, etc. LM algorithm has the fastest convergence speed and the best robustness among the above algorithms.

In Newtonian algorithms, if Hessian matrix is not a positive definite matrix, then the Newtonian direction may point to a local optimum. This phenomenon can be solved by adding a positive definite matrix to Hessian matrix. LM algorithm is a combination of gradient descent method and Gauss–Newton method. It also has local convergence and global characteristics of gradient descent method. LM algorithm is also more stable and converges faster than gradient method. In ANN, the loss function ist

$$E(x) = \frac{1}{2} \sum_{i=1}^{n} [d(i) - y(i)]^2. \tag{16}$$

The updated weight is calculated as follows:

$$x^{(k+1)} = x^{(k)} + \Delta x. \tag{17}$$

With Newton's algorithm, the amount of change is calculated as follows:

$$\Delta x = -\frac{\nabla E(x)}{\nabla^2 E(x)}. \tag{18}$$

Approximate them so that all Hessian matrices are invertible as follows:

$$\nabla E(x) = J^T(x)e(x). \tag{19}$$

LM algorithm is able to improve the Gauss–Newton method, and the improved weight and threshold adjustment rule is calculated as follows:

$$\Delta x = -\left[J^T(x)J(x) + \mu I\right]^{-1} J(x)E(x). \tag{20}$$

4. Experiments and Discussion

4.1. Data Set.

This experiment selects Student-Life and Reach Out online forum post data as the data source. The Student-Life data set is a data set researched by Dartmouth College. It records 49 students' psychological perception data for ten consecutive weeks, including academic data, online psychological test data, and questionnaire survey data. Reach Out online forum posts include user posting information, posting time, likes and views, etc. Select 5 million data points from each of the two data sets for a total of 10 million data sets. The data set is equally divided into 10 equal parts, of which six parts are used for neural network model training and four parts are used for experimental testing.

4.2. Evaluation Metric.

To show the superiority of our proposed method, four evaluation metrics are used to evaluate each algorithm, named MSE, MAE, RMSE, and MAPE.

The MAE of normalized data is given by the following:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |P_i - T_i|. \tag{21}$$

The MSE of prediction is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (P_i - T_i)^2. \tag{22}$$

The equation of MAPE is as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{P_i - T_i}{T_i}\right|. \tag{23}$$

The RMSE is described as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - T_i)^2}. \tag{24}$$

4.3. Comparison with Other Methods.

Select the following experimental indicators for verification and analysis. Feature fusion accuracy rate: the accuracy of feature fusion is directly related to the accuracy of the system evaluation results, so this indicator is selected for analysis. System recall rate: the recall rate generally refers to the recall rate, which refers to the ratio of the information obtained from the system query to the total amount of system information. The system recall rate is tested to verify the performance of the system in this article. Running time: time consumption is usually an important indicator of system performance. The comparison results of feature fusion accuracy are shown in Figure 5.

According to the comparison curve of the feature fusion accuracy rate of different systems in Figure 5, the system's
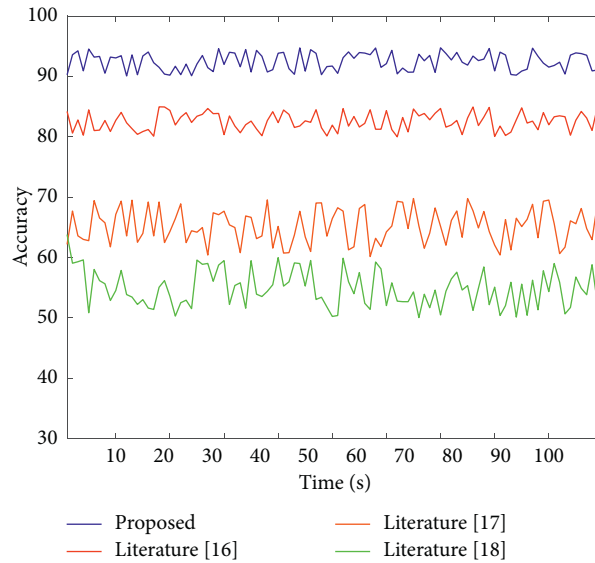
FIGURE 5: The comparison results of feature fusion accuracy.

accuracy rate in this paper is always the highest, with an average value of about 90%. Among the other three results, [16] has higher feature fusion accuracy, the highest rate can reach 80%, and the accuracy rate of [17] and [18] is low. It can be seen that this paper is based on a neural network for multifeature fusion, fully utilizes the ability of a neural network to process data in parallel, and obtains the feature fusion effect with higher accuracy. The higher the system recall rate, the higher the recall rate, and the superior system performance can be obtained. Table 1 and Figure 6 show the comparison results of the recall rate between the system in this paper and the system in [16], [17], and [18].

By analyzing the table data, it can be found that the system in this paper has a high data recall rate. Under different test data volume conditions, the recall rate is high, and it is much higher than the results of other documents. The system in this paper has certain advantages. By analyzing Figure 6, we can see that the data processing time of the system in this paper has been lower than the other two systems. As the amount of data increases, the data processing time of the three systems has changed. The data processing time of comparison system 1 and comparison system 2 has increased sharply with the increase in the amount of information, and the fluctuation range is large, and the stability is poor. The system's data in this article increase in processing time is small, the curve is smoother, and the stability is good. After the data volume reaches 5×10 GB, the data processing time gradually stabilizes, proving that the system in this paper has high data processing efficiency and strong stability.

Table 2 shows the values of each metric obtained in 10 experiments for each algorithm. To further demonstrate the superiority of our proposed algorithm, the histogram is used to compare the average of ten experiments. It is not difficult to find that the error of our proposed algorithm is much smaller than the other two methods. In addition, the box plot shown in Figure 7 is used here to compare the robustness of various algorithms, and we use MSE as the comparison standard. The information entropy method combined with neural networks has far better immunity to extreme and outlier data than other methods.

4.4. Evaluation on Joint Optimization. As mentioned before, the proposed joint optimization model consisted of the improved decision network and the improved ANN. To verify the effectiveness of this joint model, the comparative experiment based on the single improved model is conducted, and the result is shown in Figure 8. IDT is improved decision tree. IANN is improved artificial neural network.

It can be seen from the figure that the joint optimization model can obtain the best performance, with the lowest error index and the highest accuracy index. Its performance is better than any single improved model, which verifies the correctness and reliability of the joint optimization model proposed in this paper. At the same time, it can be seen that, compared to the IDT model, the performance of IANN is better, but it is still lower than the joint optimization model.

In addition, this paper uses the corresponding algorithm to improve the original decision tree and ANN. In order to verify the effectiveness of this improvement measure, a comparative experiment was also carried out, and the experimental results are shown in Figures 9 and 10. DT is original decision tree.

It can be seen that, compared with the initial decision tree algorithm and ANN algorithm, after introducing the information entropy improvement algorithm and the LM algorithm to improve the two, respectively, the performance can be effectively improved. The error index shows a downward trend, while the accuracy rate shows an upward trend, which proves the effectiveness and correctness of these two improvement measures.

TABLE 1: Comparison results of recall rate.

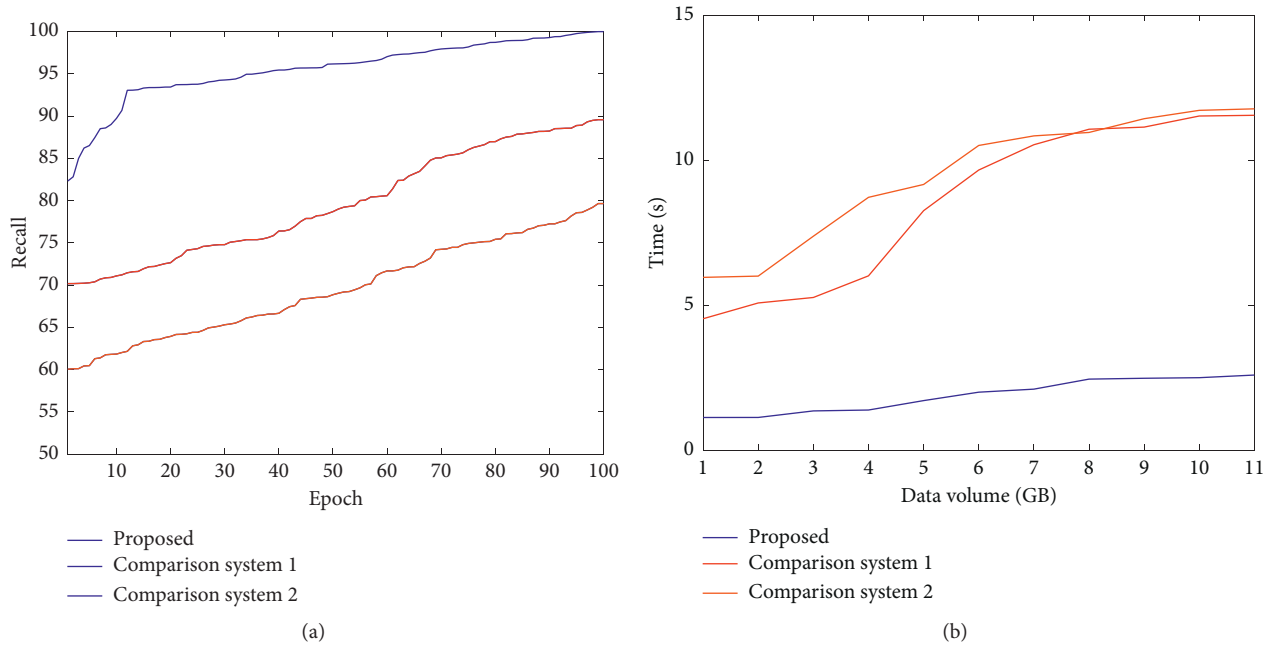| Test numbers	Proposed (%)	Literature [16] (%)	Literature [17] (%)	Literature [18] (%)
100	95	85	62	59
200	96	83	63	58
300	93	81	67	59
400	94	82	65	54

(a)

(b)

FIGURE 6: Comparison of three systems. (a) Recall. (b) Time consumption.

TABLE 2: Comparison of different metrics of different method.

No. 1	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$4.86E + 01$	$1.51E + 02$	$5.36E + 01$	$3.81E + 03$
MSE	$4.71E + 04$	$8.50E + 04$	$8.98E + 03$	$1.64E + 07$
RMSE	$2.17E + 02$	$2.92E + 02$	$9.48E + 01$	$4.05E + 03$
MAPE	$2.93E - 02$	$4.10E - 02$	$1.63E - 02$	$1.76E + 00$
No. 2	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$3.49E + 01$	$3.22E + 02$	$4.02E + 02$	$3.29E + 03$
MSE	$2.71E + 04$	$8.75E + 05$	$4.31E + 06$	$1.44E + 07$
RMSE	$1.65E + 02$	$9.36E + 02$	$2.07E + 03$	$3.80E + 03$
MAPE	$3.20E - 03$	$5.75E - 02$	$2.59E - 02$	$1.59E + 00$
No. 3	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$1.43E + 02$	$1.50E + 03$	$1.83E + 03$	$2.89E + 03$
MSE	$2.02E + 05$	$3.25E + 07$	$4.21E + 07$	$3.84E + 07$
RMSE	$4.49E + 02$	$5.70E + 03$	$6.49E + 03$	$6.20E + 03$
MAPE	$2.22E - 02$	$3.85E - 02$	$4.20E - 02$	$6.59E - 01$
No. 4	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$2.06E + 01$	$1.32E + 02$	$7.63E + 01$	$2.08E + 03$
MSE	$7.36E + 03$	$7.08E + 04$	$2.14E + 04$	$5.93E + 06$
RMSE	$8.58E + 01$	$2.66E + 02$	$1.46E + 02$	$2.44E + 03$
MAPE	$2.10E - 02$	$4.44E - 02$	$3.31E - 02$	$1.01E + 00$
No. 5	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$3.55E + 01$	$1.80E + 02$	$2.23E + 02$	$1.96E + 03$
MSE	$2.98E + 04$	$2.00E + 05$	$7.67E + 05$	$6.48E + 06$
RMSE	$1.73E + 02$	$4.47E + 02$	$8.76E + 02$	$2.55E + 03$
MAPE	$2.80E - 03$	$2.52E - 02$	$2.05E - 02$	$5.35E - 01$

TABLE 2: Continued.

| No. 6	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$5.61E + 01$	$6.66E + 02$	$4.40E + 02$	$1.86E + 03$
MSE	$2.03E + 04$	$3.65E + 06$	$1.31E + 06$	$6.13E + 06$
RMSE	$1.43E + 02$	$1.91E + 03$	$1.14E + 03$	$2.48E + 03$
MAPE	$1.80E - 03$	$3.30E - 02$	$2.00E - 02$	$7.11E - 01$
No. 7	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$1.85E + 01$	$2.79E + 02$	$2.53E + 02$	$1.56E + 03$
MSE	$4.57E + 03$	$6.75E + 05$	$8.77E + 05$	$4.37E + 06$
RMSE	$6.76E + 01$	$8.22E + 02$	$9.36E + 02$	$2.09E + 03$
MAPE	$2.10E - 03$	$2.92E - 02$	$1.96E - 02$	$6.45E - 01$
No. 8	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$1.59E + 01$	$1.48E + 02$	$1.12E + 02$	$2.52E + 03$
MSE	$2.71E + 03$	$1.35E + 05$	$7.61E + 04$	$3.26E + 07$
RMSE	$5.21E + 01$	$3.67E + 02$	$2.76E + 02$	$5.71E + 03$
MAPE	$4.30E - 03$	$3.54E - 02$	$2.25E - 02$	$8.38E - 01$
No. 9	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$4.02E + 01$	$8.19E + 02$	$9.23E + 02$	$3.56E + 03$
MSE	$2.48E + 04$	$2.42E + 07$	$2.88E + 07$	$2.10E + 07$
RMSE	$1.58E + 02$	$4.92E + 03$	$5.37E + 03$	$4.58E + 03$
MAPE	$2.50E - 03$	$2.24E - 02$	$2.08E - 02$	$1.04E + 00$
No. 10	Proposed	Literature [16]	Literature [17]	Literature [18]
MAE	$2.56E + 02$	$3.80E + 02$	$2.53E + 02$	$3.82E + 03$
MSE	$2.47E + 06$	$1.35E + 06$	$6.03E + 05$	$1.79E + 07$
RMSE	$1.57E + 03$	$1.16E + 03$	$7.76E + 02$	$4.24E + 03$
MAPE	$2.97E - 02$	$6.51E - 02$	$2.53E - 02$	$1.56E + 00$
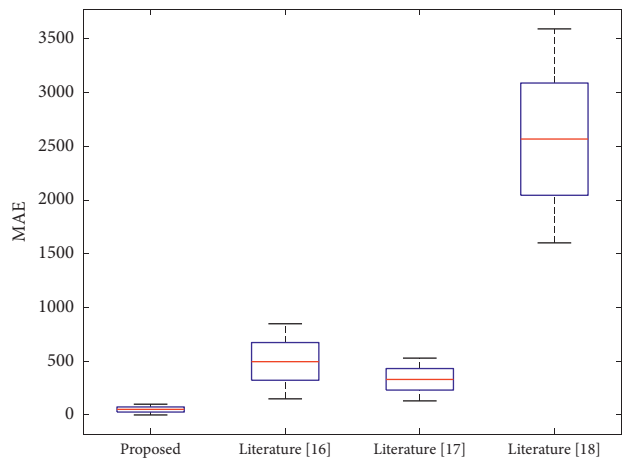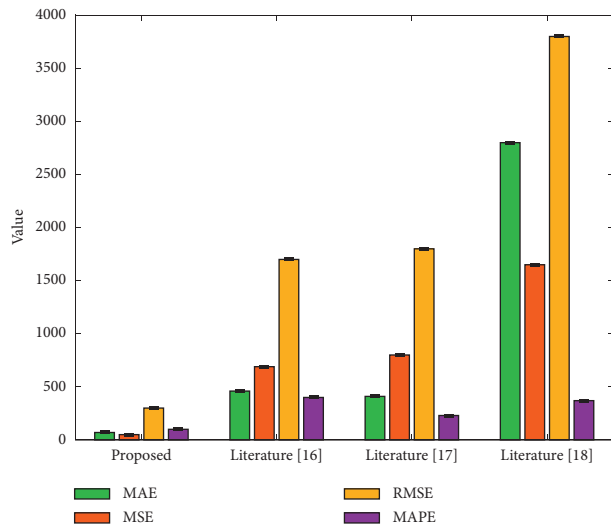
FIGURE 7: Comparison of different evaluation metrics with different methods. (a) Four metrics. (b) Box plot.
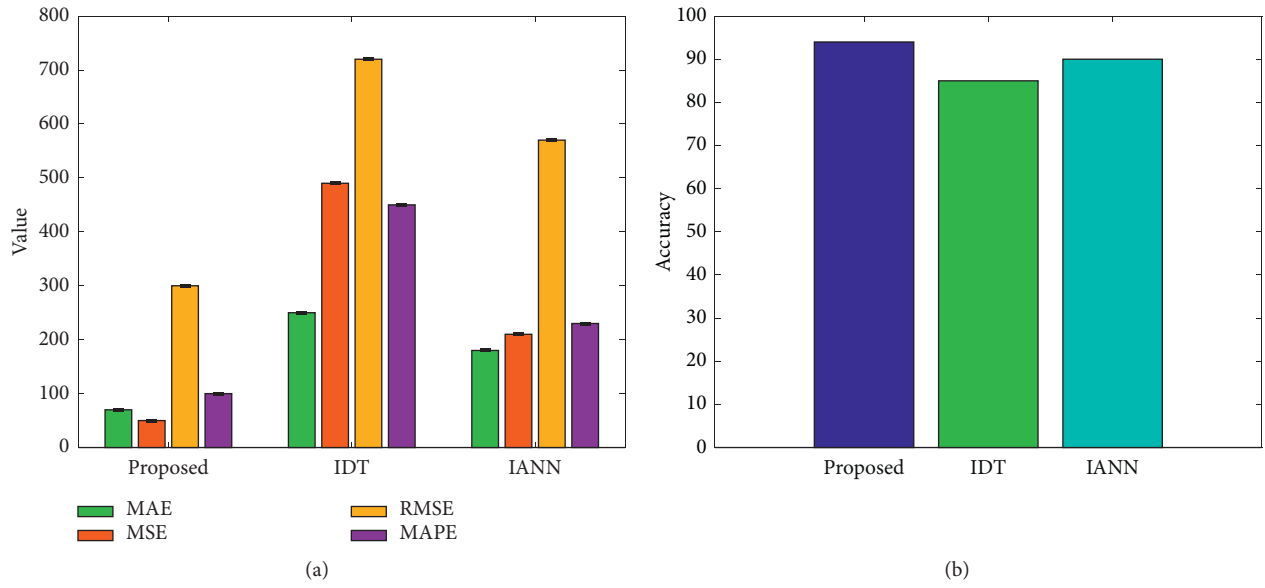
(a)

(b)

Figure 8: Comparison of joint optimization and single improve network. (a) Four metrics. (b) Accuracy.
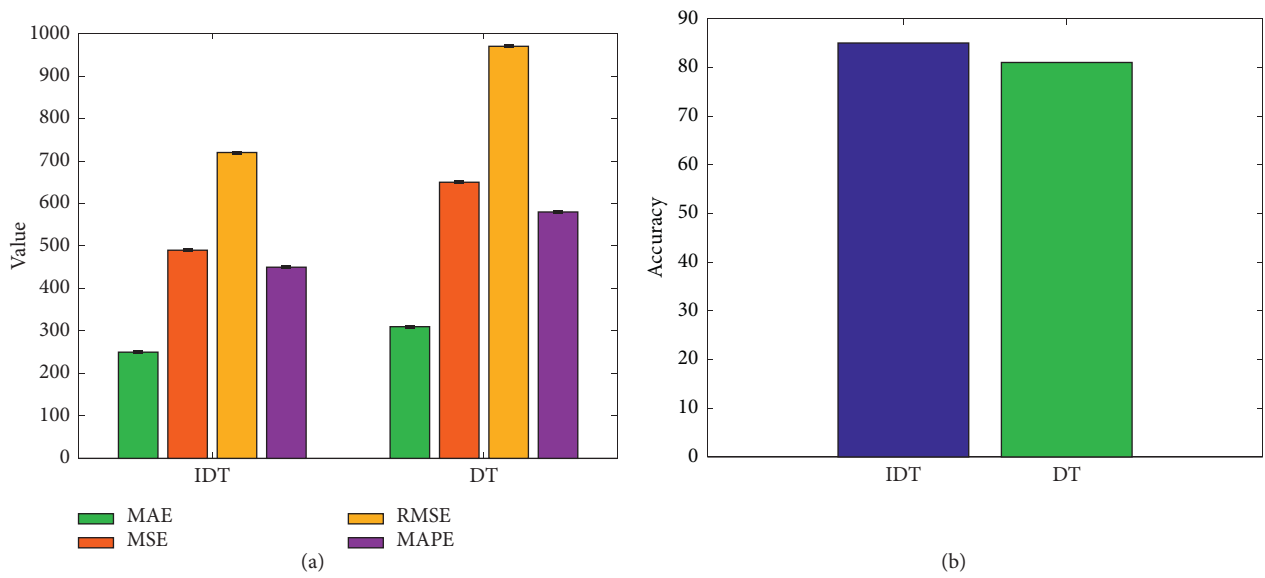

(a)

(b)

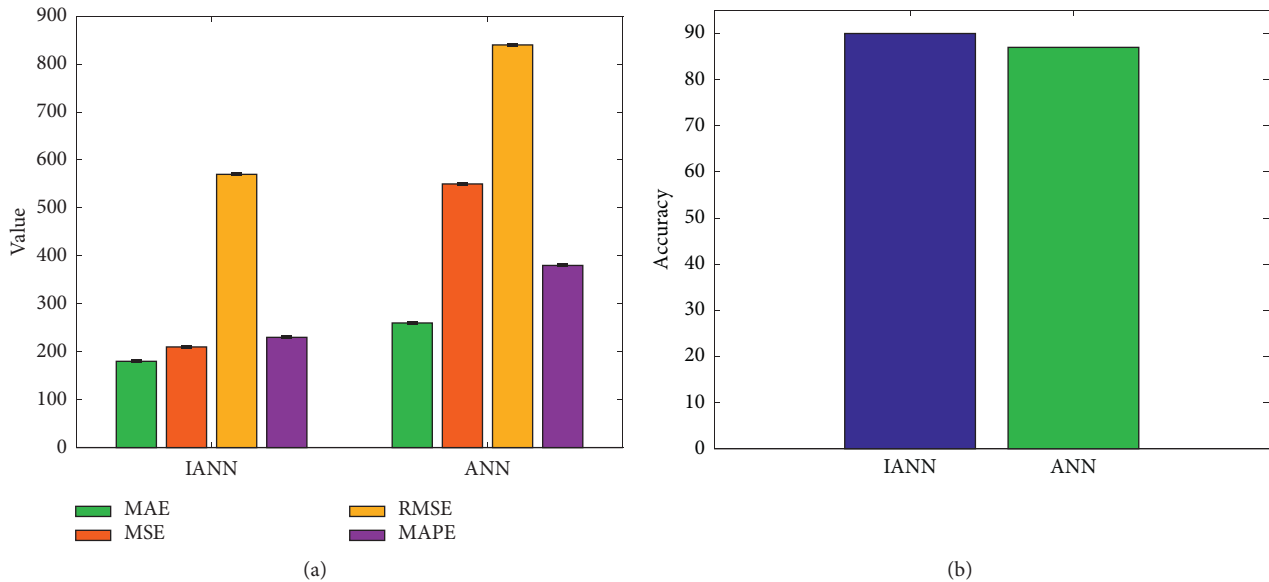Figure 9: Comparison of IDT and DT. (a) Four metrics. (b) Accuracy.

Figure 10: Comparison of IANN and ANN. (a) Four metrics. (b) Accuracy.

5. Conclusions and Outlook

In order to improve the effectiveness of mental health intelligence evaluation, this paper constructs a mental health intelligent evaluation system based on the joint optimization model, which consists of the improved decision tree algorithm and the improved ANN. The proposed model can fully help users understand their mental health, solve their psychological problems, and enhance self-awareness. In order to make the system of this article more functional and have better development, it is also necessary to improve the analysis of the system evaluation results, improve the system data loss during a power failure, increase the data backup and recording functions of the system, and solve the problem of waste of storage space—more perfect the function of this text system.

This article has conducted in-depth research on students' mental state perception based on network behavior data and has achieved certain results. However, there is still a lot of work that is not perfect and needs further discussion and research. The future research work will be from the following aspects: (1) the data source is not wide enough. This article only uses the network behavior data in the student's on-campus behavior data, which does not fully represent the student's behavior data. Therefore, the next step is to focus on adding other on-campus network behavior data to the network behavior data. Behavioral data, such as library borrowing records and meal card consumption flow records, are used to construct a perceptual model of students' mental state. (2) There is not enough feature dimensions. The next step will be to explore new feature dimension construction methods. The currently used feature dimension data is only constructed with two indicators of regularity and dependence. In the future, more intermediate variables based on network behavior data will be explored. The existence continuously enriches the feature dimension of the sample data set. (3) The scope of model consideration is not large enough. The model used in this article is only two classification models in machine learning, and the current more popular deep learning model is not used. The next step will be to use the network structure model of deep learning to conduct model experiments and compare the performance of the authors' model.

Data Availability

This experiment selects Student-Life and Reach Out online forum post data as the data source. The Student-Life data set is a data set researched by Dartmouth College. The data is open and available.

Disclosure

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. Grossiord, T. N. Buckley, L. A. Cernusak et al., "Plant responses to rising vapor pressure deficit," *New Phytologist*, vol. 226, no. 6, pp. 1550–1566, 2020.

[2] C. Ma, Z. Yang, R. Xia et al., "Rising water pressure from global crop production-A 26-yr multiscale analysis," *Resources, Conservation and Recycling*, vol. 172, Article ID 105665, 2021.

[3] M. T. Schmitt, S. D. Neufeld, C. M. L. Mackay, and O.D. Steenbergen, "The perils of explaining climate inaction in terms of psychological barriers," *Journal of Social Issues*, vol. 76, no. 1, pp. 123–135, 2020.

[4] M. Spinelli, F. Lionetti, M. Pastore, and M. Fasolo, "Parents' stress and Children's psychological problems in families facing the COVID-19 outbreak in Italy," *Frontiers in Psychology*, vol. 11, 2020.

[5] W. Li, Y. Yang, Z. H. Liu et al., "Progression of mental health services during the COVID-19 outbreak in China," *International Journal of Biological Sciences*, vol. 16, no. 10, pp. 1732–1738, 2020.

[6] Y. Kotera, M. V. Laethem, and R. Ohshima, "Cross-cultural comparison of mental health between Japanese and Dutch workers: relationships with mental health shame, self-compassion, work engagement and motivation," *Cross Cultural & Strategic Management*, vol. 27, no. 3, pp. 511–530, 2020.

[7] J. Pei, K. Zhong, and J. Li, "ECNN: evaluating a cluster-neural network model for city innovation capability," *Neural Computing and Applications*, 2021.

[8] Y.-T. Xiang, Y.-J. Zhao, Z.-H. Liu et al., "The COVID-19 outbreak and psychiatric hospitals in China: managing challenges through mental health service reform," *International Journal of Biological Sciences*, vol. 16, no. 10, pp. 1741–1744, 2020.

[9] C. Holingue, L. G. Kalb, A. Klein, and J. B. Beasley, "Experiences with the mental health service system of family caregivers of individuals with an intellectual/developmental disability referred to START," *Intellectual and Developmental Disabilities*, vol. 58, no. 5, pp. 379–392, 2020.

[10] M. A. Furst, N. Bagheri, and L. Salvador-Carulla, "An eco-systems approach to mental health services research," *BJPsych International*, vol. 18, no. 1, pp. 23–25, 2021.

[11] T. Wand, S. H. Buchanan, K. Derrick, and M. Harris, "Are current mental health assessment formats consistent with contemporary thinking and practice?" *International Journal of Mental Health Nursing*, vol. 29, no. 2, pp. 171–176, 2020.

[12] P. A. Areán, K. H. Ly, and G. Andersson, "Mobile technology for mental health assessment," *Dialogues in Clinical Neuroscience*, vol. 18, no. 2, p. 163, 2016.

[13] R. Pocobello, T. Sehity, L. Negrogno, C. Minervini, M. Guida, and C. Venerito, "Comparison of a co-produced mental health service to traditional services: a co-produced mixed-methods cross-sectional study," *International Journal of Mental Health Nursing*, vol. 29, no. 3, pp. 460–475, 2020.

[14] A. Haque, "Mental health concepts in Southeast Asia: diagnostic considerations and treatment implications," *Psychology Health & Medicine*, vol. 15, no. 2, pp. 127–134, 2010.

[15] C. Guo, G. Tomson, C. Keller, and F. Soderqvist, "Prevalence and correlates of positive mental health in Chinese adolescents," *BMC Public Health*, vol. 18, no. 1, pp. 1–11, 2018.

[16] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318, PA, USA, July 2002.

[17] J. Soet and T. Sevig, "Mental health issues facing a diverse sample of college students: results from the college student mental health survey," *NASPA Journal*, vol. 43, no. 3, pp. 410–431, 2006.

[18] E. J. R. David, S. Okazaki, and A. Saw, "Bicultural self-efficacy among college students: initial scale development and mental health correlates," *Journal of Counseling Psychology*, vol. 56, no. 2, pp. 211–226, 2009.

[19] K. E. Miller, P. Omidian, A. S. Quraishy et al., "The Afghan Symptom Checklist: a culturally grounded approach to mental health assessment in a conflict zone," *American Journal of Orthopsychiatry*, vol. 76, no. 4, pp. 423–433, 2006.

[20] Q. Yang, Y. Su, C. Chi et al., "Ultrathin graphene-based membrane with precise molecular sieving and ultrafast solvent permeation," *Nature Materials*, vol. 16, no. 12, pp. 1198–1202, 2017.

[21] A. Macaskill, "The mental health of university students in the United Kingdom," *British Journal of Guidance and Counselling*, vol. 41, no. 4, pp. 426–441, 2013.

[22] K. Zhong, Y. Wang, J. Pei, S. Tang, and Z. Han, "Super efficiency SBM-DEA and neural network for performance evaluation," *Information Processing & Management*, vol. 58, no. 6, Article ID 102728, 2021.

[23] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[24] L. He, L. Bai, D. D. Dionysiou et al., "Applications of computational chemistry, artificial intelligence, and machine learning in aquatic chemistry research," *Chemical Engineering Journal*, vol. 426, Article ID 131810, 2021.

[25] N. Artrith, K. T. Butler, F.-X. Coudert et al., "Best practices in machine learning for chemistry," *Nature Chemistry*, vol. 13, no. 6, pp. 505–508, 2021.

[26] J. Pei, "Solving the problem of charging and discharging of electric vehicles based on particle swarm algorithm," in *Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 534–538, Dalian, China, September 2019.

[27] M. O'Brien, C. A. Smith, E. R. Sokol et al., "ecocomDP: a flexible data design pattern for ecological community survey data," *Ecological Informatics*, vol. 64, Article ID 101374, 2021.

[28] Y. Zhu, M. Kim, X. Zhu, D. Kaufer, and G. Wu, "Long range early diagnosis of Alzheimer's disease using longitudinal MR imaging data," *Medical Image Analysis*, vol. 67, Article ID 101825, 2021.

[29] D. P. Cobb, H. Jashami, and D. S. Hurwitz, "Bicyclists' behavioral and physiological responses to varying roadway conditions and bicycle infrastructure," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 80, pp. 172–188, 2021.

[30] X. Liu, F. Zhang, Z. Hou et al., "Self-supervised learning: generative or contrastive," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[31] X. Wang, S. Zhang, Z. Qing et al., "Self-supervised learning for semi-supervised temporal action proposal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1905–1914, New Orleans, LA, USA, June 2021.

[32] J. Pei, J. Li, B. Zhou, M. Gao, and Q. Dat, "A recommendation algorithm about choosing travel means for urban residents in intelligent traffic system," in *Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE*, pp. 2553–2556, Chongqing, China, March 2021.

[33] A. Rives, J. Meier, T. Sercu et al., "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," *Proceedings of the National Academy of Sciences*, vol. 118, no. 15, 2021.

[34] M. Abspoel, D. Escudero, and N. Volgushev, "Secure training of decision trees with continuous attributes," *Proceedings on Privacy Enhancing Technologies*, vol. 2021, no. 1, pp. 167–187, 2021.

[35] P. S. Karunadsa, U. D. Annakkage, and B. A. Macdonald, "Dynamic security control using secure regions derived from a decision tree technique," in *Proceedings of the Power Engineering Society Summer Meeting*, IEEE, Seattle, WA, USA, July 2000.

[36] H. Zhang, L. Jiang, and L. Yu, "Attribute and instance weighted naive Bayes," *Pattern Recognition*, vol. 111, Article ID 107674, 2021.

[37] J. Fan, J. Zheng, L. Wu, and F. Zhang, "Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models," *Agricultural Water Management*, vol. 245, Article ID 106547, 2021.

[38] R. Xiao, X. Cui, H. Qiao et al., "Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net," *Biomedical Signal Processing and Control*, vol. 66, Article ID 102362, 2021.

[39] K. K. Sharma and A. Seal, "Spectral embedded generalized mean based k-nearest neighbors clustering with s-distance," *Expert Systems with Applications*, vol. 169, Article ID 114326, 2021.

[40] L. Sun, X. Qin, W. Ding, J. Xu, and S. Zhang, "Density peaks clustering based on k-nearest neighbors and self-recommendation," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 7, pp. 1913–1938, 2021.

[41] S. Georganos, T. Grippa, A. Niang Gadiaga et al., "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto International*, vol. 36, no. 2, pp. 121–136, 2021.

[42] V. Brandstätter and K. Bernecker, "Persistence and disengagement in personal goal pursuit," *Annual Review of Psychology*, vol. 73, 2021.

[43] S. I. Popkov, "Research of interactive and traditional tasks aimed at studying information technologies among students," *Psychological-Educational Studies*, vol. 13, no. 2, pp. 17–39, 2021.

[44] L. A. C. Ribeiro, T. Bresolin, D. R. G. J. Magalhães, D. R. Casagrande, M. A. C. Danes, and J. R. R. Dórea, "Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data," *Journal of Animal Science*, vol. 99, no. 9, 2021.

[45] W. Tian, Y. Li, J. Zhou et al., "Implantable and biodegradable micro-supercapacitor based on a superassembled three-dimensional network Zn@PPy hybrid electrode," *ACS Applied Materials & Interfaces*, vol. 13, no. 7, pp. 8285–8293, 2021.

[46] W. Zhu, Y. Ai, F. Fang, and H. Liao, "Application of thermosonication in red pitaya juice processing: impacts on native microbiota and quality properties during storage," *Foods*, vol. 10, no. 5, 2021.

[47] J. H. Carpenter, "Forty-year natural history study of Bahalana geracei Carpenter, 1981, an anchialine cave-dwelling isopod (Crustacea, Isopoda, Cirolanidae) from San Salvador Island, Bahamas: reproduction, growth, longevity, and population structure," *Subterranean Biology*, vol. 37, pp. 105–156, 2021.

[48] J. D. Margolis and J. P. Walsh, "Misery loves companies: rethinking social initiatives by business," *Administrative science quarterly*, vol. 48, no. 2, pp. 268–305, 2003.