International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6259-6270

https://doi.org/10.70517/ijhsa463537

Research on Mental Health Assessment Methods for College Students Based on Sentiment Analysis Technique and Multilayer Perceptual Machine

Liping Li[1,2,*]
[1] College of Marxism, Suqian University, Suqian, Jiangsu, 223800, China
[2] College of Marxism, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, 211106, China
Corresponding authors: (e-mail: muzi2000718@163.com).

Abstract Mental health problems of students in colleges and universities are becoming more and more obvious, and traditional assessment methods have limitations such as poor timeliness and narrow coverage. This study constructs a mental health assessment method for college students based on sentiment analysis technology and multilayer perceptual machine. The method mines the psychological state features implied in students' web content data from two dimensions, text sentiment computing and image sentiment computing, respectively, through a multimodal fusion computing framework, and adopts a multilayer perceptual machine model for mental health level assessment. In the textual sentiment analysis part, the constructive study builds a three-layer neural network of word embedding layer-bidirectional long and short-term memory layer-dense connectivity layer to capture the textual contextual sentiment information; and in the image sentiment analysis part, a convolutional neural network based on VGG16 is used to accurately recognize the emotional tendency in the images through fine-tuning strategies. Aiming at the sample imbalance problem, the study introduces a cost-sensitive method to optimize the training process. The experimental results show that the evaluation method performs well on the sentiment classification MSD dataset, with an accuracy of 92.3% for text sentiment classification and 96.2% for image sentiment classification, both of which are better than the traditional CNN and BiLSTM models; and on the public dataset Yelp, the average accuracy reaches 67.08%, which is an improvement of more than 5% compared with other algorithms. The multilayer perceptual machine model is used for the student mental health assessment task with an accuracy of 92%, showing better generalization performance. The study shows that the multimodal fusion sentiment analysis technique combined with the multilayer perceptual machine model can effectively realize the automatic assessment of students' mental health status in colleges and universities, providing a scientific basis for psychological intervention.

Index Terms Multimodal fusion, Sentiment analysis, Multilayer perceptual machine, Mental health assessment, Bidirectional long and short-term memory network, Convolutional neural network

I. Introduction

With the development of society and the improvement of living standards, people's attention to mental health is increasing [1]. Higher education students not only have to bear the pressure from study and life, but also have to deal with employment pressure, which is a common cause of various psychological disorders, and as one of the important groups in the future of the society, the assessment of students' psychological health is particularly important [2]-[4]. Traditional psychological assessment methods require specialized skills and are expensive, often making it difficult to cover all school students [5], [6]. With the development of artificial intelligence, big data and other technologies, mental health assessment methods based on sentiment analysis techniques and multilayer perceptual machines are getting more and more attention and importance [7], [8].

Sentiment analysis technology, also known as emotion recognition technology, is a technology that can automatically recognize emotional information in text, speech or images [9], [10]. It mainly analyzes the input data for sentiment classification and emotional tendency by using algorithms such as machine learning and natural language processing, and has been widely used in the fields of social media, market research, and public opinion monitoring, providing an effective means to understand and analyze a large amount of emotional data [11]-[14]. In the mental health assessment of students in colleges and universities, through the use of sentiment analysis technology, we can understand the emotional and psychological state of students more objectively and scientifically, and provide more accurate assessment and intervention [15], [16]. Multilayer perceptron is a deep learning algorithm based on the structure of neural network, which is a fully-connected neural network consisting of multiple neurons stacked on top of each other, where each layer is connected with multiple neurons [17]-[19]. It can be used to train models by back propagation algorithm to realize classification tasks, regression tasks, etc [20]. And in the mental health assessment of college students, the application of multilayer perceptual machine needs to be combined with multidimensional data collection and intelligent analysis technology, so as to form a dynamic assessment and intervention system [21]-[23].

Higher education students are in a critical period of psychological development, facing multiple pressures such as academics, employment, and interpersonal relationships, and mental health problems occur frequently. Traditional mental health assessment methods mainly rely on self-assessment scales and manual interviews, which have limitations such as long assessment period, narrow coverage, and strong subjectivity, making it difficult to meet the demand for real-time monitoring of the mental health of large-scale student groups in colleges and universities. Online social networking platforms have become an important channel for college students to express their emotions and share their lives, and students often reveal their real psychological status and emotional attitudes when posting text and pictures. Sentiment analysis technology can mine individual psychological characteristics from these data, providing a new technical way for mental health assessment. Deep learning methods have achieved remarkable results in the field of sentiment analysis due to their powerful feature learning capabilities. Among them, bidirectional long and short-term memory networks can effectively capture contextual semantic information when processing text sequence data, and convolutional neural networks have advantages in image feature representation. However, single-modal sentiment analysis is often difficult to comprehensively reflect an individual's psychological state, and multimodal fusion computing can more accurately identify an individual's emotional tendency by integrating data information from different modalities. In addition, there is a general sample imbalance in students' mental health data, and the sample of serious psychological problems is relatively sparse, which poses a challenge to the training and evaluation of the model. As a classical deep learning model, the multilayer perceptron, although simple in structure, has a powerful nonlinear feature learning capability that makes it suitable for application to complex mental health assessment tasks. Based on the above background, the study proposes a mental health assessment method for college students based on sentiment analysis techniques and multilayer perceptual machines. A multimodal fusion computing framework for mental health assessment is constructed, including four parts: data cleaning preprocessing, text sentiment computing, image sentiment computing and mental health assessment model generation. In the text emotion calculation module, a word embedding model trained based on the microblogging corpus and a bidirectional long and short-term memory network are used to deeply excavate the semantic features and emotional tendencies of the text; in the image emotion calculation module, based on the VGG16 network, the implicit emotional information in the image is captured through the transfer learning strategy and network fine-tuning. For the sample imbalance problem, the study compares the effects of various methods, such as oversampling, undersampling, integrated sampling and cost-sensitivity. In terms of feature selection and extraction, principal component analysis is used to reduce the dimensionality and mitigate the effect of “dimensionality catastrophe”. Finally, a mental health assessment model based on multilayer perceptual machine was constructed, and the network parameters and training strategy were optimized to achieve accurate assessment of students' mental health status. This study comprehensively evaluates the effectiveness of the proposed method through experimental validation on the self-constructed MSD dataset and the public dataset Yelp, and provides a new technical path for monitoring the mental health of college students.

II. Construction of mental health assessment model based on multimodal fusion computing

The model design is based on the concept of systematic approach, analyzing and determining the assessment level, designing the automatic assessment framework, selecting the assessment strategy, implementing the model assessment, evaluating the model assessment effect, correcting the model parameters and applying them. In order to realize the fusion of multimodal data information, the model is designed from the following four parts: data cleaning and preprocessing, text-based emotion calculation, image-based emotion calculation, and mental health assessment model generation.

The mental health assessment model of multimodal fusion computing [24] is a process of mining the real emotions implied behind students' modal data such as text, images, and expressions, and synthesizing the students' psychological changes over a period of time to achieve fast and accurate identification of students' mental health levels. Based on this, an automatic mental health assessment model for college students based on multimodal data fusion is constructed, and the model framework is shown in Figure 1.

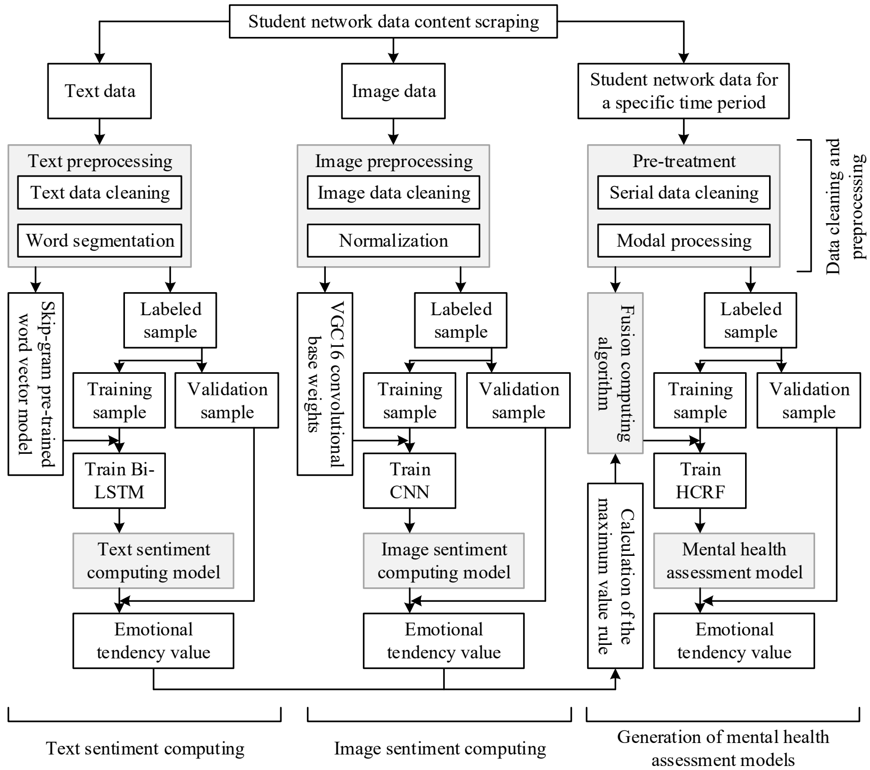

Figure 1: The college students' mental health automatic evaluation mode

II. A. Data Cleaning and Preprocessing

The collected data on students' web content could not be directly used for processing and analyzing psychological characteristics, and these raw data needed to be cleaned as well as pre-processed. Firstly, the data obtained from the depression self-assessment questionnaire was cleaned, i.e., subjects with zero or perfect scores on the questionnaire and those who filled in the questionnaire for less than 3 minutes were removed. Secondly, the data obtained by means of web crawlers were cleaned, i.e., subjects with a number of web content data below a threshold were removed. Finally, the multimodal data were preprocessed and converted into model-recognizable symbols before computer processing. For example, operations such as removing irrelevant symbols from text data and converting fonts. Format conversion, resizing, normalization and normalization are performed on image data.

II. B. Text-based Sentiment Calculation

Textual information is the basic information for human beings to convey their emotions and express their thoughts, and it is an important form of external expression of an individual's psychological state. Therefore, mining the psychological state and emotional attitudes of individuals when publishing content plays an important role in accurately identifying mental health conditions. Text is typical sequential data, if the sentence context information can be captured, the emotional tendency of the text can be well mined based on semantic understanding, therefore, this study establishes a three-layer neural network framework of word embedding layer - bidirectional long and short-term memory (Bi-LSTM) layer - dense connectivity layer, which crosses the contextual intervals and learns the implicit sentence sentiment information.

There are two types of text vectorization: one-hot coding and distributed representation. The distributed representation “word embedding” can map words into dense and low-dimensional vectors and ensure that semantically similar words are closer to each other in the vector space, which is more suitable to handle the task of sentiment analysis than the one-hot coding representation. Since both Weibo text and talking text are characterized by colloquialism and short and concise expression, the Skip-gram pre-trained word embedding model based on the Weibo corpus is used as the word embedding layer of the model, so that the vector representation of the talking text is more accurate and relevant.

LSTM is well known for its unique gating structure and memory units that can avoid long-term dependency and gradient vanishing problems, but information can only be propagated unidirectionally when learning text sequence features. In order to deeply understand the semantics of sayings and learn effective sentiment feature representations, the paper designs a bidirectional long and short-term memory network layer to fully grasp the contextual information of the saying text. The o_t -speaking feature representation at t moment needs to acquire the forward hidden state h_t and the backward hidden state $h't$ at t moment, where $\oplus$ denotes the integration of the two in a splicing way. The forward hidden state h_t at moment t is computed from the input $x_t, t-1$ moment of hidden state information at moment t. The backward hidden state h at moment t is computed from the input $x_t, t+1$ moment hidden state information at moment t, where the f function is an LSTM nonlinear function, W, U, W', U' denote the weights of the function, and b, b' denotes the bias of the function. This is shown in the following equation:

$$o_t = h_t \oplus h't \tag{1}$$

$$h_t = f(W \times x_t + U \times h_{t-1} + b) \tag{2}$$

$$h't = f(W' \times x_t + U' \times h_{t+1} + b') \tag{3}$$

II. C.Image-based sentiment computation

Image information is an important supplement to textual information, and college students often attach images to their speech to share their life status, enhance their emotional expression, and even use images to express psychological states that cannot be depicted and described by language, so accurately recognizing the implicit emotions in image modality is conducive to assessing the mental health of an individual more accurately on the basis of text.

VGG16 network [25] is a convolutional neural network with a 16-layer structure, which is notably effective in image feature representation and generalization ability. In this study, VGG16 is used as the baseline model, and the CNN model is constructed by fine-tuning the strategy in order to capture the hidden emotional tendencies behind the images.

Image-based emotion computation is a complex visual problem, and the only way for CNN to accurately compute students' mental emotions when posting images is to learn a large number of parameters and effective features. Transfer learning technique is utilized to solve the scale dilemma of labeled data, and the VGG16 convolutional basis trained based on ImageNet large-scale dataset is used as a pre-trained model for image-based emotion computation to learn the generalized feature representations of the images, and the original dense connectivity layer setup is changed to adapt to the task of image-based emotion computation.

The lower convolutional layers in the convolutional basis learn localized generic features of the image, while the higher top layers learn more abstract, specialized feature representations. In this study, the weight of the fifth convolutional block of the VGG is released to learn the image emotion representation when training the CNN, both to make the network structure more suitable for computing image emotions and to avoid the risk of overfitting. After the convolutional base learns the emotion representation of the image, the emotion tendency value of the image can be obtained through dense connected layer integration and softmax [26] classification, which is calculated as shown in the following equation:

$$S(V_i) = \frac{\exp(V_i)}{\sum_{j=1}^{n} \exp(V_j)} \tag{4}$$

where V_i refers to the i th category element of the output vector of the dense connectivity layer, $S(V_i)$ is the i th category sentiment probability value of the image, and n represents the positive and negative sentiment categories.

II. D.Methods of assessing the mental health of students in higher education institutions

II. D. 1) Feature Selection and Feature Extraction

The student data samples obtained from the student mental health assessment model are characterized by high dimensionality and large scale, and many of them have strong correlation in the physical sense and belong to redundant features, all of which will affect the training effect of the deep learning model, so feature selection and data processing will be carried out, and feature selection will be based on the following 4 aspects.

1) High missing value percentage features.

2) Covariance features. In this paper, Pearson coefficient is used to calculate the correlation between features:

$$\rho_{X,Y} = \frac{c_{ov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2 \sum_{i=1}^{N}(Y_i - \overline{Y})^2}} \tag{5}$$

where, N is the number of samples. X_i and Y_i denote the sample points of the 2 features. $\overline{X}$ and $\overline{Y}$ denote the sample mean of the 2 features, the closer the value is to ±1, the higher the correlation between the 2 features.

3) Low importance features. In this paper, we use the lightweight gradient boosting machine algorithm (LGBM) to calculate the importance of features, and sequentially sort them to delete the remaining features after the cumulative importance reaches 0.99.

4) Single-valued features. Features with only one value have no basis for differentiation in the model and can be deleted directly.

In this paper, we use Principal Component Analysis (PCA) method based on Singular Value Decomposition (SVD) for feature extraction. This method retains the main features of the data, and the missing information does not affect the classification effect of the model, but rather improves the computational speed and mitigates the impact of the "dimensionality catastrophe".

At this point, the original dataset for training has been constructed, and the following deep learning model is introduced.

II. D. 2) Multi-layer perception machines

The multilayer perceptron [27], also known as deep neural network (DNN), is one of the simplest neural network structures, consisting of input, hidden and output layers, each of which consists of multiple neurons. Although its structure is simple, it can learn deeper nonlinear features in the data and is suitable for application to large nonlinear systems such as student mental health assessment models.

The output of the implicit layer is:

$$H = f(XW_h + b_h) \tag{6}$$

The output layer output is:

$$O = f(HW_0 + b_0) \tag{7}$$

where, X is the input. W_h and b_h are the implicit layer weights and thresholds, respectively. W_o and b_o are the output layer weights and thresholds respectively, they are vectors, the dimension is determined by the number of neurons in the layer, and f is the activation function, and the commonly used activation functions are the sigmoid function, the tanh function, and the ReLU function.

The gradient backpropagation uses RAdam adaptive learning rate optimizer to achieve fast and accurate convergence, and cross entropy is used as the loss function:

$$f_{loss}(x,i) = -\ln\left(\frac{\exp(x[i])}{\sum_j \exp(x[j])}\right) = -x[i] + \ln\left(\sum_j \exp(x[j])\right) \tag{8}$$

where, x denotes the unsoftmaxed output with dimension equal to the number of categories. i denotes the category. $x[i]$ denotes the output of this sample corresponding to the category. $x[j]$ denotes the output corresponding to each category in x.

The Gaussian Error Linear Units (GELUs) function is chosen for the activation function of each hidden layer and is defined as follows:

$$f_{GELU}(x) = xP(X \le x) = \frac{x}{2}\left[1 + e_{rf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{9}$$

Among them:

$$e_{rf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}dt = \frac{2}{\sqrt{\pi}}\int_{0}^{x} e^{-t^2}dt \tag{10}$$

Its can be approximated as:

$$f_{GELU}(x)=\frac{1}{2}x\left(1+\tanh\left[\sqrt{\frac{2}{\pi}}(x+0.044715x^{3})\right]\right) \tag{11}$$

It introduces stochastic regularization while maintaining the nonlinear mapping capability to better avoid overfitting and gradient vanishing problems.

Due to the large number of neurons in the network, a portion of neurons are randomly deleted for each hidden layer using the dropout method to prevent overfitting The final improved multilayer perceptron structure is shown in Fig. 2.

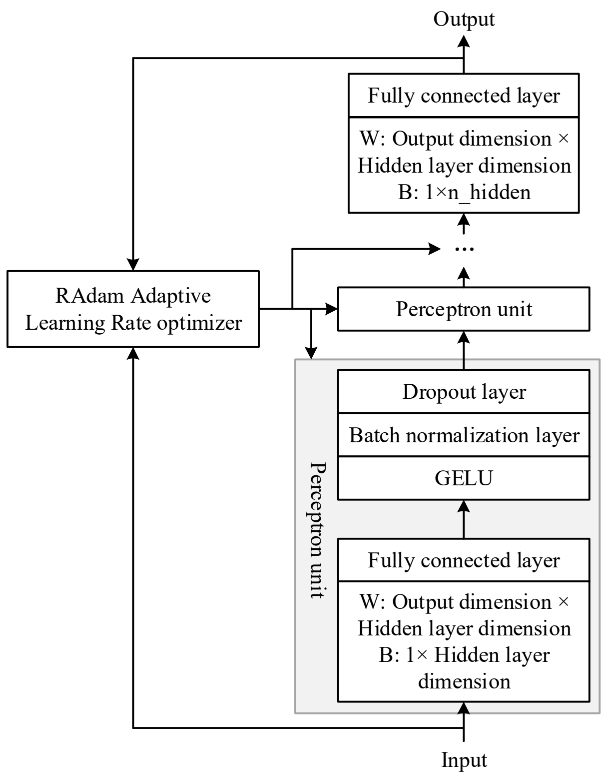

Figure 2: Improved multi-layer perceptron structure

II. D. 3) Unbalanced sample training methods

Sample imbalance can have a great impact on the training effect of deep learning models. The student mental health assessment model has a typical sample imbalance characteristic for its data due to its high level of security measures and very few serious accidents, so the sample imbalance problem must be solved.

There are four main unbalanced sample training methods in the field of deep learning as follows.

1) Oversampling: balancing the number of samples in each category by increasing the number of samples in a few categories, of which there are mainly 2 types of methods: data reselection and data expansion.

2) Undersampling: to achieve sample balance by reducing the number of samples of most classes.

3) Comprehensive sampling: for the undersampling method, effective information may be missing due to the deletion of the original data samples. Due to the above problems, comprehensive sampling methods that combine the two have emerged, and the following 2 methods are introduced in this paper.

① SMOTETomek (ST) method: first use SMOTE oversampling method to expand the minority class samples, and then use Tomek Link undersampling method to reduce the boundary samples, and finally realize the sample balance.

② SMOTEENN(SE) method: similar to the above method, first use SMOTE oversampling method, then use ENN undersampling method to realize sample balance.

4) Cost-sensitive (SC) method: this method starts from the loss function during model training, assigns different weights to the minority class samples and the majority class samples, and increases the misclassification penalties

of the minority class samples, so that the model is more inclined to classify the minority class correctly during the training, and thus improves the sample imbalance problem.

For the multilayer perceptual machine using cross entropy as the loss function, the weights of the samples of different classes in its dataset are calculated according to the following equation:

$$\delta[i] = \frac{n_s}{n_c c_s} \tag{12}$$

The cross-entropy loss function for adding a cost-sensitive method can be expressed in the following form:

$$f_{loss}(x,i) = \delta[i]\left\{-x[i] + \log\left[\sum_j \exp(x[j])\right]\right\} \tag{13}$$

In order to compare with existing methods of risk assessment using support vector machine models, this paper will also simulate the support vector machine (SVM) algorithm and also introduce a cost-sensitive training method.

The linear support vector machine algorithm is essentially a constrained optimization problem for solving a geometrically maximally spaced separating hyperplane $\omega x + b = 0$, which can be formulated as follows:

$$\begin{cases} \min\limits_{\omega,b,\xi_i} \frac{1}{2}\|\omega\|^2 + C\sum\limits_{i=1}^{m}\xi_i \\ s.t. y_i(\omega x_i + b) \ge 1 - \xi_i \end{cases} \tag{14}$$

where, the sample point is (x_i, y_i). The ξ_i is a "slack variable", $\xi_i = \max[0, 1 - y_i(\omega x_i + b)]$, which is used to characterize the extent to which the sample does not satisfy the constraints. The SVM-based cost-sensitive training method is realized by changing the value of c, which can be achieved by replacing the value of C with the sample weights during training. In contrast, the nonlinear support vector machine is used to partition the samples by mapping x_i to a higher dimensional space using a suitable mapping function (kernel function) $\Phi(\cdot)$, i.e., in the above constrained optimization problem, x_i is converted to $\Phi(x_i)$.

III. Multi-layer perceptual machine model construction

III. A. Data pre-processing

Deep learning networks are very sensitive to the distribution of data, and it is difficult to guarantee the quality of the data obtained through us. In order to speed up the training and eliminate the negative impact of the data, the data is standardized. That is, data with different ranges of distribution are scaled to the same range, which generally requires standardization to a standard normal distribution with a mean of 0 and a standard deviation of 1. The mean value of 0 makes the data centralized, and the standard deviation of 1 makes the data present the same unit vector in all dimensions, that is, the original size range of the distribution of different data scaled to a uniform scale, to prevent the model gradient explosion or gradient dispersion.

In this paper, we choose the maximum-minimum standardization method to standardize the data, using the maximum-minimum boundary value of the data to scale the training set and test set data to [0, 1].

$$x = \frac{[x - \min(x)]}{[\max(x) - \min(x)]} \tag{15}$$

III. B. Determination of network parameters

According to the results of correlation analysis, the data with higher correlation coefficients with students' mental health scores, cumulative absolute velocity, root-mean-square acceleration and other parameters are used as inputs to the multilayer perceptron, and macroscopic students' mental health scores are used as the outputs to establish a multilayer perceptron deep learning network in the tensorflow framework. Through the continuous adjustment of hyperparameters. The activation function between the input layer and the hidden layer and between the hidden layer and the hidden layer is a segmented linear function relu, and the hidden layer and the output layer is a sigmoid.

III. C. Training and validation of the network

After determining the network parameters, the training data are input into the multilayer perceptron deep learning network model, the number of training times is 1500, and the model training results are shown in Figure 3. The model error decreases with the increase of training times, and after 600 epochs of iteration, the loss curves of the training set and the test set stabilize, and the training error and the test error are basically stabilized at about 0.22, and the model reaches the optimum. From Fig. 3, we can see that the loss function curves of the training set and the test set basically overlap, indicating that the multilayer perceptron model does not produce overfitting phenomenon.

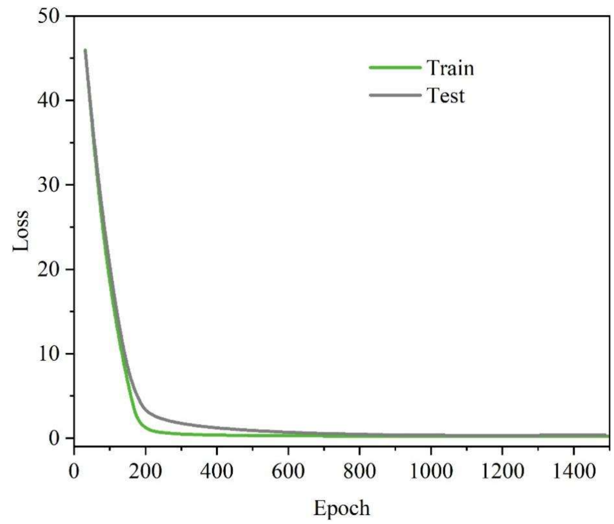

Figure 3: Loss function curve

Figures 4 and 5 show the prediction results of the model training set and test set, in which there are 50 training samples and 25 test samples. The mental health scores of college students predicted by the multilayer perceptron deep learning network are basically consistent with the mental health scores of students in the actual field survey. By analyzing the prediction results, the accuracy of the model training set is obtained as 92%, and the accuracy of the test set is 92%, and the prediction effect of the model is good.

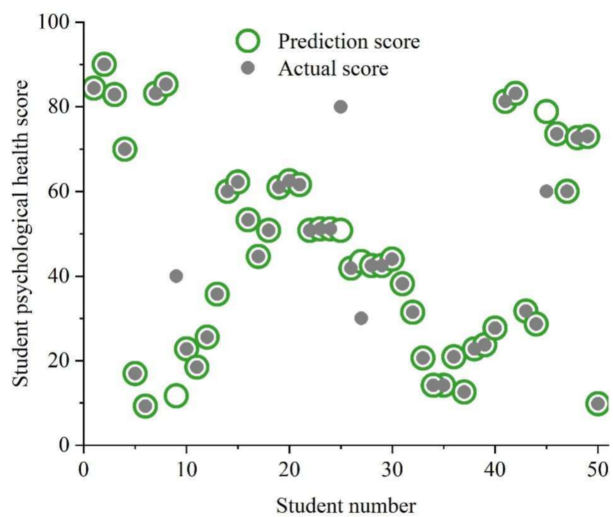

Figure 4: The prediction value of the training set is compared to the actual value

III. D. Performance comparison of different models

In order to minimize the effect of the randomness of the data on the results of the test, the entire test was repeated and executed several times. The test was repeated three times, and the number of repetitions for each time was 30, 50, and 70 times, respectively. During the test, stratified sampling of the data was performed to ensure the consistency of the distribution of the training data set and the test data set. The final ratio of the number of minority classes to the number of majority classes achieved during repeated oversampling was 1.0.

Common evaluation metrics for class-imbalanced datasets include recall, F1-measure, and AUC. The AUC metric used in this experiment is often used as a classification metric for analyzing class-imbalanced datasets in binary classification problems.

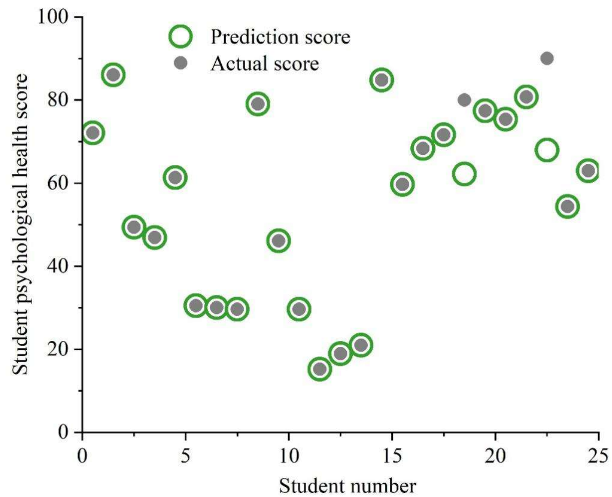

Figure 5: Comparison of predicted values of test sets

In order to compare the effects of different classification models on the missile test interference assessment dataset, several common classification models were implemented to perform comparative tests under the same conditions. The five classification models used in the test are as follows:

1) Multilayer Perceptron Machine model, which is the model used in this paper and is denoted as mlp in the test.
2) Random forest model, denoted as rf.
3) Support Vector Machine, denoted as svm.
4) Plain Bayesian model, denoted as nb.
5) A 2-stage random forest based model with random forest as the classification model, denoted as rf_imb.

Figure 6 shows the AUC box plots of the different classification models for the case of 70 repeated runs. The mlp classifier obtains the best classification performance in terms of median, and its median is nearly 53% better than the comparison method. Meanwhile, the other four classification models all show a large degree of consistency, with median values around 0.55 on the one hand, and significantly larger fluctuations in classification performance on the other.

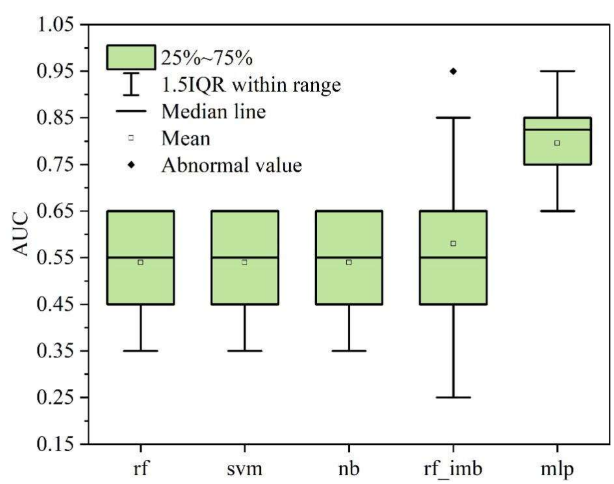

Figure 6: Repeat the number of times 70 times

Figures 7 and 8 show the number of repetitions of 30 and 50 runs, respectively. The mlp classifier obtained the best classification performance in terms of median in all cases with different number of repetitions for different trials. And its classifiers likewise all performed poorly, with their median being around 0.55 for the case of 30 repetitions of runs, which indicates the lack of practical maneuverability of these classifiers. Therefore, it can be concluded that the mlp classifier has better generalization and classification performance and is suitable for the task of assessing the mental health of college students.

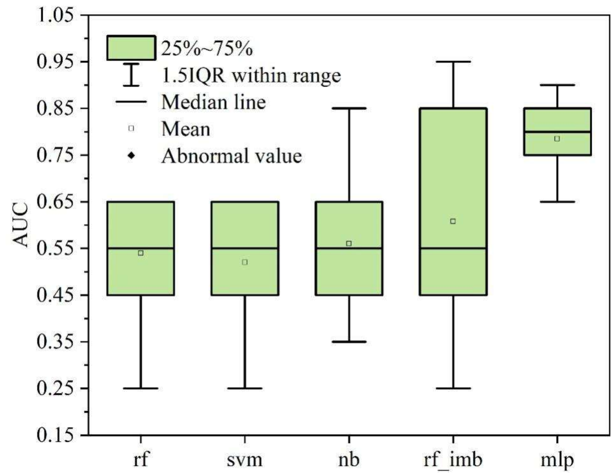

Figure 7: Repeat the number of times 30 times

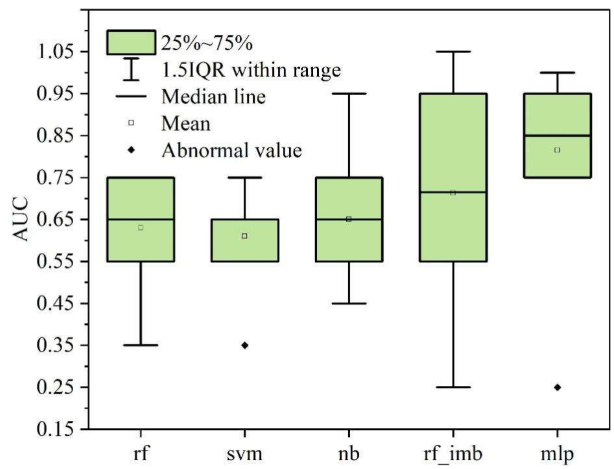

Figure 8: Repeat the number of times 50 times

IV. Performance analysis of the model on sentiment classification

IV. A. Experimental setup

In this study, the friend circle and microblogging dynamics of 400 college students were collected to construct the sentiment dataset MSD as the experimental dataset. The dataset has a total of 5600 samples. Among them, emotions are divided into 3 types: positive (2000 articles), negative (1600 articles), and neutral (2000 articles).

IV. B. Analysis of results

For the emotion classification experiments, the experimental results are shown in Table 1, which indicate that the method proposed in this study can effectively classify the side emotions of two different modal data, namely, web expressions and matching images, individually. The sentiment classification accuracy of the models in this paper are all above 90%, which is better than the comparison models.

Table 1: The performance of different models in the MSD data set

	Model	Acc/%	Recall/%	F1/%
Web expression	CNN	78.1	75.6	76.5
	BiLSTM	88.1	85.1	87.3
	OURS	92.3	92.1	95.1
Mapping	CNN	75.3	71.2	75.7
	BiLSTM	65.1	62.2	65.4
	OURS	96.2	84.6	86.5

To further demonstrate the effectiveness of the proposed method, further tests are conducted on the public dataset Yelp, and the experimental results are shown in Table 2. It contains five different U.S. cities in the Yelp dataset, where a and m denote the two hyperparameters, average pooling and maximum pooling, respectively, of the algorithms involved in the experiment. The experimental results show that the method proposed in this paper improves the accuracy of the algorithm on the public dataset Yelp by 5% and above compared to other algorithms.

Table 2: Experimental results on the open data set yelp

Model	BO/%	CH/%	LA/%	NY/%	SF/%	Average accuracy/%
TFN-aVGG	46.28	43.71	43.89	43.81	42.86	44.11
TFN-mVGG	48.28	47.11	46.68	46.68	47.6	47.27
BiGRU-aVGG	51.255	51.26	49.02	49.48	48.66	49.94
BiGRU-mVGG	53.89	53.48	52.06	52.23	51.38	52.61
HAN-aVGG	55.15	54.84	53.1	52.9	52.03	53.60
HAN-mVGG	56.71	57.07	55.05	54.67	53.71	55.44
VistaNet	63.78	65.74	62.06	61.06	60.17	62.56
OURS	70.14	68.31	65.44	68.36	63.14	67.08

V. Conclusion

The mental health assessment method of multimodal fusion computing combined with multilayer perceptual machine significantly improves the accuracy and automation level of mental health assessment for college students. The experimental results of sentiment analysis show that on the MSD dataset, the method achieves 92.3% accuracy, 92.1% recall, and 95.1% F1 value for sentiment classification of web emojis; and 96.2% accuracy for sentiment classification of accompanying images, which are all better than traditional CNN and BiLSTM models. On the public dataset Yelp, the average accuracy of the method reaches 67.08%, which is more than 5% higher than the comparison algorithm, verifying the generalization ability of the method. When the multilayer perceptual machine model is applied to the task of students' mental health assessment, the accuracy of both the training and test sets reaches 92%, and the performance is stable in repeated experiments, with the median AUC improved by nearly 53% over the comparison method, which reflects a significant classification performance advantage. The data imbalance processing and feature selection strategies effectively solve the problem of uneven distribution of student mental health data, so that the model error is stable at about 0.22 after 600 rounds of iterations, avoiding the overfitting phenomenon. The multimodal fusion framework makes full use of the complementary advantages of the two modal data, text and image, to make the mental health assessment more comprehensive and accurate. The method realizes automatic analysis of students' online behavior data and real-time monitoring of mental health, which provides a scientific basis for mental health intervention in colleges and universities and has good application prospects.

Acknowledgements

This work was supported by the 2025 Major Project of Philosophy and Social Sciences Research in Jiangsu Universities, titled "Research on the Generation Logic and Teaching Effectiveness Evaluation of the Digital Narrative System of Ideological and Political Courses in Universities", and a phased outcome of the 2021 Major Project of Philosophy and Social Sciences Research in Jiangsu Universities, titled "Research on the Construction Path of the 'Grand Ideological and Political Course' in the New Era" (2021SJZDA084).

References

[1] Keller, A. S., Leikauf, J. E., Holt-Gosselin, B., Staveland, B. R., & Williams, L. M. (2019). Paying attention to attention in depression. Translational psychiatry, 9(1), 279.
[2] Katz, D. S., & Davison, K. (2014). Community college student mental health: A comparative analysis. Community College Review, 42(4), 307-326.
[3] Fusar-Poli, P., Davies, C., & Bonoldi, I. (2018). A case of a college student presenting with mild mental health problems. JAMA psychiatry, 75(12), 1298-1299.
[4] Bravo, A. J., Villarosa-Hurlocker, M. C., & Pearson, M. R. (2018). College student mental health: An evaluation of the DSM–5 self-rated Level 1 cross-cutting symptom measure. Psychological Assessment, 30(10), 1382.
[5] El-Den, S., Moles, R., Choong, H. J., & O'Reilly, C. (2020). Mental Health First Aid training and assessment among university students: a systematic review. Journal of the American Pharmacists Association, 60(5), e81-e95.
[6] Hubbard, K., Reohr, P., Tolcher, L., & Downs, A. (2018). Stress, mental health symptoms, and help-seeking in college students. Psi Chi Journal of Psychological Research, 23(4), 293-305.
[7] Newson, J. J., Hunter, D., & Thiagarajan, T. C. (2020). The heterogeneity of mental health assessment. Frontiers in psychiatry, 11, 76.

[8] De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. Psychological bulletin, 141(4), 858.

[9] Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., & Wang, R. (2024). A review of key technologies for emotion analysis using multimodal information. Cognitive Computation, 16(4), 1504-1530.

[10] Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017, September). Emotion detection: a technology review. In Proceedings of the XVIII international conference on human computer interaction (pp. 1-8).

[11] Hakak, N. M., Mohd, M., Kirmani, M., & Mohd, M. (2017, July). Emotion analysis: A survey. In 2017 international conference on computer, communications and electronics (COMPTELIX) (pp. 397-402). IEEE.

[12] Colombo, D., Fernández-Álvarez, J., Garcia Palacios, A., Cipresso, P., Botella, C., & Riva, G. (2019). New technologies for the understanding, assessment, and intervention of emotion regulation. Frontiers in psychology, 10, 1261.

[13] Bettis, A. H., Burke, T. A., Nesi, J., & Liu, R. T. (2022). Digital technologies for emotion-regulation assessment and intervention: A conceptual review. Clinical Psychological Science, 10(1), 3-26.

[14] Katirai, A. (2024). Ethical considerations in emotion recognition technologies: a review of the literature. AI and Ethics, 4(4), 927-948.

[15] Shanthi, N., Stonier, A. A., Sherine, A., Devaraju, T., Abinash, S., Ajay, R., ... & Ganji, V. (2022). An integrated approach for mental health assessment using emotion analysis and scales. Healthcare Technology Letters.

[16] Fei, Z., Yang, E., Li, D. D. U., Butler, S., Ijomah, W., Li, X., & Zhou, H. (2020). Deep convolution network based emotion analysis towards mental health care. Neurocomputing, 388, 212-227.

[17] Riedmiller, M., & Lernen, A. (2014). Multi layer perceptron. Machine learning lab special lecture, University of Freiburg, 24.

[18] Al Bataineh, A., Kaur, D., & Jalali, S. M. J. (2022). Multi-layer perceptron training optimization using nature inspired computing. IEEE Access, 10, 36963-36977.

[19] He, X., & Chen, Y. (2021). Modifications of the multi-layer perceptron for hyperspectral image classification. Remote Sensing, 13(17), 3547.

[20] Malalur, S. S., Manry, M. T., & Jesudhas, P. (2015). Multiple optimal learning factors for the multi-layer perceptron. Neurocomputing, 149, 1490-1501.

[21] Tran, A., Tran, L., Geghre, N., Darmon, D., Rampal, M., Brandone, D., ... & Avillach, P. (2017). Health assessment of French university students and risk factors associated with mental health disorders. PloS one, 12(11), e0188187.

[22] Li, K., & Yu, W. (2021). A mental health assessment model of college students using intelligent technology. Wireless Communications and Mobile Computing, 2021(1), 7485796.

[23] Shi, Q., Cai, N., & Jiao, W. (2022). Monitoring and evaluating college Students' mental health based on big data analysis. American journal of health behavior, 46(2), 164-176.

[24] Jianjun Peng,Yunhao Cui,Zhidan Zhong & Yi An. (2023). Ore Rock Fragmentation Calculation Based on Multi-Modal Fusion of Point Clouds and Images. Applied Sciences,13(23),

[25] Asif Shahriar Arnob,Ashfakul Karim Kausik,Zohirul Islam,Raiyan Khan & Adib Bin Rashid. (2025). Comparative result analysis of cauliflower disease classification based on deep learning approach VGG16, inception v3, ResNet, and a custom CNN model. Hybrid Advances,10,100440-100440.

[26] Irwan Budi Santoso,Shoffin Nahwa Utama & Supriyono. (2025). Meta-learning based softmax average of convolutional neural networks using multi-layer perceptron for brain tumour classification. Array,26,100398-100398.

[27] Fabricio A Chiappini,Mirta R Alcaraz & Liliana Forzani. (2025). A bootstrap-assisted methodology for the estimation of prediction uncertainty in multilayer perceptron-based calibration.. Analytica chimica acta,1353,343954.