

<https://doi.org/10.70517/ijhsa463308>

# The application of affective computing models in student psychoeducation and the optimal design of intervention pathways

Xuebo Hu<sup>1,\*</sup>

<sup>1</sup> Mental Health Education Center, Zhengzhou Shengda University, Zhengzhou, Henan, 451191, China

Corresponding authors: (e-mail: 100218@shengda.edu.cn).

**Abstract** In order to more intelligently and accurately analyze the mental health status of students through their emotions, and to promote the good development of students' mental health. In this paper, we propose a student mental health analysis method based on multimodal social emotion classification, using the multi-head self-attention mechanism in the BERT model to extract and train the textual features of the data, and then using the VGG16 model as a pre-training model to obtain the image features of the data, and then fusing the two features into a multimodal feature through the fully connected layer. The features are inputted into GRU layer and fully connected layer to get the subjective emotion of the student, and finally the dynamic matching image is emotionally categorized to get the side emotion of the student, and then the model of this paper is applied to the recommendation of psychological services for students. The student multimodal sentiment computation model improves its main accuracy by 2.86% compared to the better performing model in the comparison experiments. Psychological interventions for students based on the student multimodal emotion computational model led to significant improvements in students' total mental health, obsessive-compulsive symptoms, and other psychological factors, realizing the combination of psychological education work for college students with new technologies, advancing the progress of intelligent psychological work mechanisms, and enhancing the scientificity and relevance of psychological intervention mechanisms.

**Index Terms** Emotion classification, Multimodal fusion, BERT model, VGG16 model, Psychological intervention

## I. Introduction

Currently, students are facing a complex developmental environment, and statistics show that academic stress, interpersonal relationships, and self-exploration are the main sources of stress for students [1]. However, most students tend to hide negative emotions and tend to rely on self-growth rather than professional treatment to relieve psychological stress [2]. Therefore, accurate mental health testing to detect mental health problems among college students through scientific methods is extremely important to improve the effectiveness of students' psychological education [3]. The current mental health detection methods in colleges and universities mainly include self-assessment (self-scales testing or questionnaires) and other-assessment (offline questioning), but both methods have many limitations when applied to campus mental health education [4], [5]. Unlike traditional testing methods, affective computing technology can capture sensitive sources of human emotions (e.g., respiratory system, nervous system, speech text, etc.) and its own algorithms to process the emotions, and ultimately provide feedback to the user through "anthropomorphic" methods [6]-[8]. This approach not only saves the cost of the school, but also greatly increases the accuracy of the test, bringing new development space for the field of mental health education.

The concept of affective computing first originated in the Massachusetts Institute of Technology (MIT) Media Laboratory in the United States with the book "Affective Computing" by Prof. Picard, R. W published in 1995 [9]. Affective computing is a computational approach that relates to, is inspired by, and can have an impact on emotions [10]-[12]. Researchers in this field are currently exploring how to utilize intelligent products to mimic natural humans for emotional expression, and to conduct emotional experiments using the established emotional models, and then design or improve emotional robots based on the experimental data [13]. Some researchers have also begun to collect large amounts of data and information, dissecting the way algorithms are optimized, and incorporating Artificial Intelligence (AI) to achieve a large number of research results [14]. For example, literature [15] outlines affective computing models applied to address challenges associated with mood and cognitive disorders in later life (e.g., depression and Alzheimer's disease), solving the problem of misdiagnosis and underdiagnosis that may result from the results of traditional methods of detection. Literature [16] states that affective computing combines

hardware and software to detect and recognize a person's emotional state, and that affective computing relies on objective data information rather than subjective human judgments, so that the detector is able to keenly capture subtle physiological signal fluctuations in the subject, which can help to reveal the user's hidden psychological changes. Literature [17] proposes a taxonomy to organize the field of affective computing, which is an area that involves designing emotionally intelligent machines, and organizes existing research into three main areas: emotion generation, emotion understanding, and applications. In his study, literature [18] states that affective computing aims to enable computers to recognize, interpret, and simulate human emotions in a variety of ways including body movement analysis. Literature [19] used probabilistic programming as a modeling approach for affective computing to represent psychological theories of emotion as generative models and to enable flexible, modular and effective data learning.

In the field of psychoeducation, affective computing technology can provide subjects with more specialized and objective treatment plans, and its main research includes three aspects, including physiological signal acquisition and processing, physiological signal analysis, and personality-based interactive feedback [20], [21]. The integration of affective computing technology and psychoeducation has the following advantages. First, accuracy. Affective computing relies on objective data information rather than human subjective judgment, so the tester is able to keenly capture the subtle physiological signal fluctuations of the subject, which helps to reveal the hidden psychological changes of the user as well as the minor symptoms that may be overlooked in manual testing [22], [23]. Secondly, the emotional test feedback can be based on the test results of each subject, combined with the family and personal information of the subject students, their growth environment, cultural background, personality characteristics and other factors in the process of conducting the test in the most suitable way for the subject to provide feedback to complete the entire testing process, reducing the degree of students' aversion to [24], [25]. However, the related research of affective computing technology in the field of psychoeducation is still relatively small, and the related research is more in the theoretical aspect, which has certain limitations.

In this paper, a multimodal data fusion based method is designed to fuse the features of different modalities in order to realize the analysis of students' multimodal emotions. In this paper, BERT model and VGG16 model are first introduced to extract the text and image features of the data respectively. Due to the large difference of different modal features, this paper applies the fully connected layer with high scalability and flexibility to integrate the two features, and inputs the fused feature vectors obtained from the integration into the GRU layer and the fully connected layer, to predict the students' subjective emotions. Then the side emotions are obtained through the VGG16 model to supplement the analysis of the students' emotions. The model in this paper can recommend appropriate psychological services for students through the results of emotion recognition, and then based on the model, we designed a psychological intervention path for students, and combined with the questionnaire survey to verify the effect of the model on the psychological intervention of students.

## II. Multimodal Affective Computing Model for Students

### II. A. Overall structure

Based on the student population will friend circle and microblogging data, this study proposes a sentiment analysis method that is divided into three main parts. The first is a subject sentiment analysis model that integrates text, image and network expression features. The second is to convert the network emoji into text, and then to get the side sentiment by calculating the text similarity [26]. The last is to categorize the sentiment of the dynamic accompanying images to get the side sentiment 2. So the student's sentiment analysis task is defined as a combination of sentiments from the subjective sentiment space and the side sentiment space, and the model architecture is shown in Fig. 1.

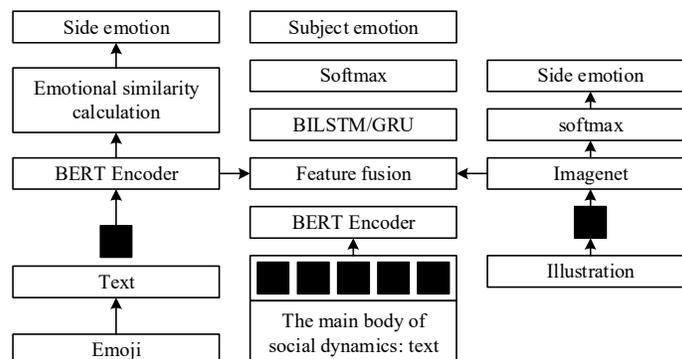


Figure 1: Model structure diagram

## II. B. Text Feature Extraction

BERT [27] can learn both local and global features in the text through the multi-head self-attention mechanism [28], and the contextual environment information of the text is taken into account. In this study, we use BERT pre-training model to extract the text features and obtain the corresponding text vectors. For the text defined as  $T$  in the dynamics posted by student groups in their friend circles or microblogs,  $T$  may contain multiple sentences  $S_i$ .

$$T = [S_1, S_2, S_3, \dots, S_M] \quad (1)$$

For each sentence  $S_i$  consists of a sequence of words:

$$S = [w_1, w_2, w_3, \dots, w_N] \quad (2)$$

BERT encodes the text vector as a text encoding layer BERT Encoder to get the text vector  $H$ :

$$H = BERTEncoder(T) = [h_1, h_2, h_3, \dots, h_k] \quad (3)$$

## II. C. Image Feature Extraction

This study uses the VGG16 model trained on the ImageNet dataset as a pre-trained model to acquire image features. For a given set of paired images  $P$  in the dynamic data posted by the students, which consists of a combination of 1 to  $J$  images.

$$P = [p_1, p_2, p_3, \dots, p_j] \quad (4)$$

After the VGG16 network, the feature vector  $H^p$  of the image can be obtained at the last pooling layer POOL, and the feature vector  $H_j^p$  of the response can be obtained for each image  $p_j$  in the set  $P$ .

$$H_j^p = POOL(p_j) = [h_1^p, h_2^p, h_3^p, \dots, h_k^p] \quad (5)$$

## II. D. Subject Sentiment Calculation

After extracting text and image features, the text feature vector  $H$ , the image feature vector  $H^p$  and the web expression text vector  $H^e$  are feature spliced and fused to obtain the fused vector  $H^c$ . Since the web expression itself is a small image determined based on the specified text, the web expression can be converted to a textual representation, and then the textual feature vector  $H$  and the web expression feature vector  $H^e$  are vectorially spliced to obtain the spliced feature vector  $H^{ce}$  by the splicing function.

$$H^{ce} = concatenate(H \parallel H^e) \quad (6)$$

Since the expressions of text and image are very different, this study fuses the text feature vector  $H^{ce}$  and image feature vector  $H^p$  by adding a fully connected layer, and then obtains the final fused feature vector  $H^c$  of multimodal data after the fully connected layer. In Eq. (7)  $\theta$  is the fully connected layer parameter.

$$H^c = FC(H^{ce}, H^p; \theta) \quad (7)$$

The resulting fused feature vector  $H^c$  is passed through the GRU layer and the fully connected layer to finally obtain the predicted subject sentiment. Formally, for the input at position  $t$  as  $I_t$ , the vector of the hidden layer at the current position is denoted as  $h_t^g$ , and the vector of the hidden layer at the previous position is denoted as  $h_{t-1}^g$ . The  $\sigma$  is a Sigmoid function for controlling the update gating unit denoted as  $U$ , where the learning parameter is  $U_w$ . For the definition of update gating see equation (8).

$$U_t = \sigma(U_w \cdot [I_t, h_{t-1}^g]) \quad (8)$$

The gating unit used to control the reset is denoted  $R$ , where the learning parameter is  $R_w$ :

$$R_t = \sigma(R_w \cdot [I_t, h_{t-1}^g]) \quad (9)$$

$\tanh$  is the activation function, and  $H_w$  is computed for the hidden layer output  $h_t^g$  at the current position as shown in equation (10).

$$h_t^g = (1 - U_t) \times h_{t-1}^g + U_t \times \tanh \left( H_w \cdot \left[ R_t \times h_{t-1}^g, I_t \right] \right) \quad (10)$$

$Out_w$  is the learning parameter of the output layer, and the final current position output  $y_i$  is computed as shown in equation (11).

$$y_i = \sigma(Out_w \cdot h_t^g) \quad (11)$$

After passing through the GRU layer and then the fully connected layer, the predicted subject sentiment  $\tilde{M}$  can finally be obtained through the SoftMax activation function (see equation (12)).

$$\tilde{M} = SoftMax(GRU(H^e)) \quad (12)$$

## II. E. Side Emotion Calculation

Web Emoji as a special symbol has the form of an image but essentially defines the meaning by specifying it through text. Similarly, as a text, the text vector  $H^e$  of the web emoji is obtained after converting the web emoji into text input and passing through the text encoding layer BERT Encoder. If there are multiple web expressions in the data, the text vectors obtained from all web expressions are spliced together before sentiment classification. The obtained text vectors are passed through the fully connected layer and then the side sentiment is calculated directly. After extracting the image features, the features of the matching images in the dynamic can be obtained. Side sentiment classification of images can be done directly through VGG16 model [29].

## III. Classification effects of a multimodal affective computing model for students

### III. A. Experimental comparison results of model classification effects

In order to verify the effectiveness of the fusion strategy of multi-head attention mechanism, the following nine models are selected for comparison and verification on Sina Weibo dataset. In the text feature extraction part, the text sentiment classification model RBCF-Net is used. VGG-14, ResNet-50, Inception-ResNet-v2 and the improved Inception-ResNet-v2 models are used for feature extraction on image data. In the data fusion part, Direct Splicing and Transformer, which are commonly used in existing research methods, are chosen to compare with the multi-head attention mechanism used in this paper. The final experimental results are shown in Table 1.

The text feature extraction algorithm and the image feature extraction algorithm used in this paper achieved the best results in the single-modal sentiment classification task, respectively.

From the experimental results of (3), (4) and (5), it can be concluded that the image feature vectors extracted by VGG-14 and ResNet-50 contribute less to the accuracy of the graphic and text fusion sentiment classification task, and will even be lower than the accuracy of the image in a single modality.

From the comparison of the results of experiments (1), (2), and (8), it can be seen that the accuracy of the graphic fusion sentiment classification algorithm based on the multi-attention mechanism reaches 89.67%. Compared with the emotion recognition algorithm based on single modality text and image data, it is improved. This indicates that data of different modalities can play a complementary role to each other. Image features and text features can play an enhancing role to each other. Therefore, combining the fusion features of image and text modalities can achieve better classification results.

Table 1: Comparison of different emotional classification models

Model	Precision rate/%	Recall rate/%	Macro-F1/%	Accuracy rate/%
RBCF-Net (1)	71.56	65.43	68.33	78.37
IRCF-Net (2)	76.67	74.71	75.78	81.54
RBCF-Net+VGG-14 (3)	73.61	73.97	67.04	70.31
RBCF-Net+ResNet-50 (4)	76.81	75.04	75.85	82.53
RBCF-Net+Inception-ResNet-v2 (5)	77.01	75.2	76.13	84.88
RBCF-Net+IRCF-Net(CONCAT) (6)	78.86	77.35	78.12	85.72
RBCF-Net+IRCF-Net(Transformer) (7)	81.02	83.01	82.11	86.81
This model (8)	83.97	85.26	84.49	89.67

From the comparison of the experimental results of (6), (7), and (8), it can be seen that compared with the (6) method that takes direct splicing of feature vectors of different modalities, or the (7) method that feeds the sequences into the Transformer network for fusion after splicing, this paper is based on the multi-attention mechanism that can efficiently capture correlations between graphic and textual modal features, and it can weight

the attentionality of the features in different modalities to realize the adaptive fusion of features. The accuracy is improved by 3.95% and 2.86%, respectively. And the Macro-F1 value of the fusion method used in this paper is better than the other two methods, and the best results are achieved.

Considering the usage requirements and scope of the system, this paper focuses on model training and evaluation on graphic datasets containing Chinese text in previous research work. In order to further validate the effectiveness and robustness of this paper's approach and to enhance the experimental persuasiveness, in addition to experiments on the microblogging dataset, comparative experiments with other graphic-text fusion sentiment categorization models are also conducted on some other publicly available datasets. When conducting experimental comparisons, in order to be able to make direct comparisons with cutting-edge technologies, it is necessary to ensure the fairness and reliability of the comparison results by conducting experiments and evaluations on the same datasets. Therefore, this paper chooses two publicly available English graphic datasets collected by MCRLab Lab on Twitter that are more widely used in cutting-edge technologies: the MVSA-Single dataset and the MVSA-Multi dataset.

The proposed method in this paper is experimentally compared with the seven models, and the results are shown in Table 2.

From the analysis of the experimental results, it can be seen that models (1) and (2) have the worst performance among all the methods, which is because models (1) and (2) do not take advantage of the correlation between different modalities, although they take into account the information of two modalities. Models (3) ~ (8) all consider the interactions between the graphic modalities. Among them, model (3) has the worst performance, which is due to the fact that the method only considers the unidirectional relationship of image to text. The unidirectional relationship of image to text is mainly applied in some application scenarios where text is the main focus and image is the secondary focus. Model (4) considers the bidirectional relationship between image and text, which is an improvement compared to model (3). Model (5) utilizes Attention mechanism for feature fusion between image and text to solve the problem of modal alignment. Model (6) establishes the relationship between image and text from multiple perspectives. Model (7) has a greater improvement in the experimental results, which is due to the fact that model (7) uses a single-modal feature extraction algorithm with more excellent performance. In this paper, not only the unimodal feature extraction algorithm is optimized. In the modal fusion part, the correlation between different modalities and modality-unique information are considered. The recognition accuracy of this paper's method on both datasets reaches more than 80%, indicating that the graphic fusion sentiment classification algorithm designed in this paper achieves superior performance compared with other models.

Table 2: Compare the results with his model

Model	MVSA-Single		MVSA-Multi	
	Acc/%	Macro-F1/%	Acc/%	Macro-F1/%
CNN-Multi (1)	61.13	58.34	66.31	64.2
DNN-LR (2)	61.49	61.01	67.78	66.29
MultiSentiNet (3)	69.94	69.57	68.89	66.09
CoMN (4)	70.53	69.98	69.01	68.84
VistaNet (5)	72.42	71.91	70.9	69.13
MVAN (6)	73.02	72.93	72.39	72.29
DR-Transformer (7)	76.78	73.47	75.11	73.22
This model (8)	80.51	75.57	80.77	74.33

### III. B. Visualization of Attention Weights for Graphic Fusion Features

Figure 2 shows the visual view of the attention of the text part, and it can be found that the words "grandpa", "golden", "corn", "happy", "smile" and "gratitude" in the text data have attracted special attention, so the algorithm judges the text data as positive emotions. From the visualization results, it can be seen that analyzing the correlation between different modalities can help the model improve its accuracy.

## IV. Multimodal affective computing in student psychoeducation

Multimodal feature fusion for affective computing is a method of collecting multimodal data features and fusing the information to help researchers understand the process of affective changes in the test subjects. Currently, multimodal affective computing involves personal information, behavioral, physiological and psychological dimensions. With the development of data mining technology and artificial intelligence fusion application, the

collection of multimodal data to implement affective computing has become the future development direction of mental health assessment and psychological problem identification.

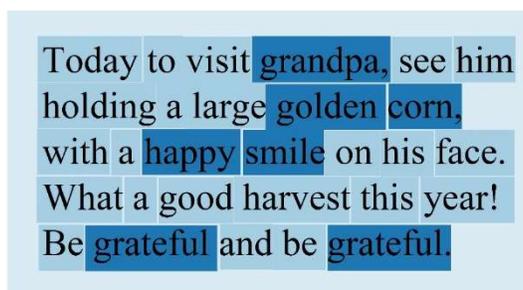


Figure 2: Visual attention weight visualization

#### **IV. A. Multimodal Affective Data Processing**

After data collection, the data first needs to be cleaned, noise reduction, dimensionality reduction, labeling and classification. Due to the presence of noisy data in multimodal data which is not beneficial for machine learning, Principal Component Analysis, Random Forest, and Bayes can be used to remove the noise, followed by linear or nonlinear dimensionality reduction of the data in order to improve the robustness when modeling multimodal data. Labeling of data is an important means of classifying and labeling the cleaned data using manual labor, thus helping machine learning methods to build corresponding sentiment computation models from the manually labeled learning labels. Here the emotions of older adults are categorized in terms of multiple dimensions such as normal, mildly depressed, severely depressed, mildly anxious, severely anxious, hostile, sad, horrible, and paranoid.

#### **IV. B. Multimodal Affective Data Fusion**

Data fusion is the process of further analyzing and fusing processed multimodal data with machine learning models. For personal information, physiological information and psychological information, the information platform of comprehensive service of students' psychological education mostly records relevant information in the form of text, which is a special kind of sequence data, and if we can capture the contextual correlation information of the text, we can help to explore the emotional tendency embedded in it. Psychology shows that there is a correlation between behavioral characteristics, physiological characteristics and psychological traits, and the collection of multimodal comprehensive features to build a psychological model is better than the modeling effect of a single feature, which can more comprehensively identify the psychological characteristics of students. Therefore, for students, multimodal data can be collected for psychological modeling of virtual doppelgangers, such as personal data, physiological data, and behavioral data.

#### **IV. C. Recommending psychological services in combination with emotion recognition results**

After obtaining affective prediction results using machine learning affective computational models, the need to understand the triggering causes of a particular affective state is crucial for providing accurate psychological services. Based on value theory, emotion attribution includes individual, family, and social factors. After understanding the triggers of negative emotions through the method of emotion attribution, corresponding psychological services can be implemented. Currently, existing psychological education services for students ignore individualized psychological needs and only provide limited psychological comfort services, such as chatting to relieve boredom, recreational activities, and dealing with disputes, which, however, cannot satisfy the different needs of different students with different psychological problems. On the other hand, the psychological service resources of third-party service institutions and fragments have not been effectively integrated into the construction of the supply of personalized psychological service resources. After grasping the emotional state of students and emotional attribution, deep learning algorithms and service-oriented architecture can be used to realize personalized psychological service recommendation.

### **V. Optimized design of psychoeducational intervention pathways for students**

#### **V. A. Subjects of the survey**

Adopting stratified random sampling, the survey sampling was conducted in high schools in several districts of province A. A total of 1600 questionnaires were distributed, invalid questionnaires were excluded, and finally 1560 valid questionnaires were obtained, with a total of 30 classes. Among the recovered valid questionnaires, there were

730 in the intervention group and 830 in the control group, (the age of the students was 15-20 years old, and the intervention group used the emotion calculation model based on the multimodal fusion of this paper to carry out the psychoeducational interventions for the students, and the control group used the traditional psychological intervention model to carry out the psychoeducational interventions for the students.

### **V. B. Survey instruments**

The psychological well-being of the subjects was assessed using the Chinese Mental Health Inventory for High School Students (MMHI-60), which consists of 60 items, including 10 subscales, including obsessive-compulsive (6 items), paranoia (6 items), hostility (6 items), interpersonal sensitivity (6 items), depression (6 items), anxiety (6 items), feelings of stress in study (6 items), mood volatility (6 items) and psychological disequilibrium (6 items). There are 10 subscales including psychological imbalance (6 entries). The MMHI-60 is scored on a 5-point scale (1=never, 2=mild, 3=moderate, 4=severe, and 5=severe), with higher total scale scores indicating lower levels of psychological well-being and higher levels of psychological crisis among the subjects. The factor scores of the scale represent the subjects' problematic situation in each aspect of mental health, and the higher the score, the more serious the psychological problems in that aspect and the higher the degree of psychological crisis. The test reliability of the MMHI-60 scores was examined using internal consistency reliability (alpha coefficient and synthetic reliability), and the results showed that the alpha coefficient and synthetic reliability of the MMHI-60 total score were 0.95 and 0.98, respectively, and the alpha coefficients and synthetic reliabilities of each factor (except for the obsessive-compulsive factor) were more satisfactory. In addition, a validated factor analysis based on the factor structure of the MMHI-60 showed that the fitting indexes of the first-order 10-factor model of the MMHI-60 were greater than 0.90 for the NNFI and CFI, and less than 0.06 for the RMSEA, which satisfied the relevant criteria. This indicates that the structural validity of MMHI-60 is better supported in this study.

### **V. C. Investigation process**

To ensure that the survey process was strictly standardized, after uniform training, psychology teaching and researchers from various regions, states, and cities in Guizhou Province went to secondary schools in their respective areas from March 2022 to September 2022 to conduct the field questionnaire survey. The questionnaires were distributed by class, administered in groups, with unified instructions, filled out anonymously, and collected on the spot. The questionnaires were finally entered into the computer by the group for management, and SPSS25.0 statistical software was used for descriptive statistical analysis, as well as alpha coefficients and difference tests. Due to the large sample of this survey, small data differences can also cause significant differences and lead to misjudgment, so the effect size index is used to test the actual significance of data differences. The significance test was conducted using the  $d$  value as the effect size, which was categorized into small ( $d < 0.2$ ), medium ( $0.2 < d < 0.7$ ) and large ( $d > 0.7$ ). Validated factor analysis and synthetic reliability were calculated using Mplus 7.0 statistical software.

### **V. D. Methods of psycho-pedagogical intervention for students**

The multimodal fusion of emotion calculation model of this paper was chosen for the psychoeducational intervention of students.

The intervention includes health education on adolescent physiological hygiene knowledge, health education on adolescent psychological knowledge, and the promotion of healthy lifestyles. Physiological health knowledge mainly includes the definition of puberty, recognizing the first and second sex signs, and how to do a good job of health care during puberty. Mental health knowledge mainly includes correctly recognizing and coping with psychological problems during puberty, how to deal with interpersonal interactions and conflicts, and how to correctly deal with relationships with the opposite sex. Healthy behavioral lifestyle mainly includes how to do reasonable diet, appropriate amount of exercise, and maintain good sleep.

All the subjects in the intervention school will be gathered together, and the pediatric experts, psychological teachers and graduate students of Chongqing Medical University will conduct unified training on the intervention curriculum based on the multimodal fusion of emotional calculation model, and the effect test will be conducted before and after the training, and two educator trainings will be carried out in March 2022 and September 2022 respectively, so that the students will have the knowledge of the training related to the intervention curriculum based on the multimodal fusion of emotional calculation model in place. Knowledge in place.

### **V. E. Results of psychoeducational interventions for students**

#### **V. E. 1) Comparison of balance between intervention and control groups**

The results of the pre-intervention balance comparison between the intervention and control groups are shown in Table 3, where the differences between the intervention and control groups in terms of whether they lived in school,

whether they stayed behind, their parental relationship, their relationship with their fathers, their family's economic status, and the composition of their parenting styles were all statistically significant (all p-values < 0.05).

Table 3: Pre-intervention equilibrium comparison

General feature		Intervention group (n=730)	Control group (n=830)	$\chi^2$	<i>P</i>
Gender	Man	372	407	0.85	0.33
	Female	357	423		
The only child	Yes	120	160	1.85	0.15
	No	610	667		
Whether to stay in school	Yes	40	336	245.15	<0.01
	No	680	491		
Whether or not to stay	Yes	115	255	46.51	<0.01
	No	612	577		
Parental relationship	Good	598	730	13.65	0.00
	Medium	110	79		
	Bad	20	21		
Relationship with father	Good	577	712	11.03	0.00
	Medium	120	96		
	Bad	29	22		
Relationship with mother	Good	633	730	0.55	0.76
	Medium	71	70		
	Bad	25	26		
Academic performance	Good	210	229	0.33	0.84
	Medium	372	405		
	Bad	152	179		
Family economy	Good	277	280	20.85	<0.01
	Medium	375	401		
	Bad	71	146		
Breeding style	Democracy	477	596	15.86	0.00
	Dictatorship	155	122		
	Doting	53	75		
	Neglect	32	22		

Table 4: Experimental group psychological health problem detection rate

Problem indicator	Preintervention/%	After intervention/%	$\chi^2$	<i>P</i>
Obsessive-compulsive Disorder	55.4	45.8	23.55	<0.01
Paranoia	38.6	36.4	60.54	<0.01
Antagonism	37.5	33.2	73.45	<0.01
Interpersonal relation	43.8	38.1	69.65	<0.01
Depression	40.3	38.3	73.85	<0.01
Anxiety	42.5	40.2	69.54	<0.01
Learning pressure	48.3	45.3	64.56	<0.01
Maladjustment	37.8	39.5	36.54	<0.01
Mood inequality	52.6	49.1	58.12	<0.01
Psychological inequality	32.8	30.2	55.26	<0.01
Total mental health	39.8	36.7	88.65	<0.01

#### V. E. 2) Comparison of mental health status before and after the intervention

Comparison of the detection rates of mental health problems before and after the intervention and control groups are shown in Tables 4 and 5, respectively. The detection rates of total mental health, obsessive-compulsive symptoms, paranoia, hostility, interpersonal tension and sensitivity, depression, anxiety, study stress, emotional imbalance and psychological imbalance problems of the students in the intervention group were lower than those of the pre-intervention period, and the detection rate of maladjustment after the intervention period (39.5%) was

higher than that of the pre-intervention period (37.8%), and the differences were statistically significant ( $p$ -value  $< 0.05$  in all cases). The detection rates of total mental health, obsessive-compulsive symptoms, paranoia, interpersonal tension and sensitivity, depression, anxiety, maladjustment, emotional imbalance and psychological imbalance problems of the control group students were higher than those of the pre-intervention period, and the detection rate of obsessive-compulsive symptoms was lower (54.3%) than that of the pre-intervention period (55.8%), and the differences were statistically significant (all  $p$ -values  $< 0.05$ ).

Table 5: Control group psychological health problem detection rate

Problem indicator	Preintervention/%	After intervention/%	$\chi^2$	$P$
Obsessive-compulsive Disorder	55.8	54.3	56.65	$<0.01$
Paranoia	37.2	41.8	80.02	$<0.01$
Antagonism	31.3	41.4	56.54	$<0.01$
Interpersonal relation	39.2	43.2	66.35	$<0.01$
Depression	33.5	44.35	63.21	$<0.01$
Anxiety	38.4	51.6	62.56	$<0.01$
Learning pressure	38.9	52.4	64.56	$<0.01$
Maladjustment	31.6	45.6	73.54	$<0.01$
Mood inequality	49.6	57.5	74.12	$<0.01$
Psychological inequality	22.9	32.5	43.56	$<0.01$
Total mental health	31.2	43.5	104.56	$<0.01$

### V. E. 3) Comparison of mental health status after intervention

Logistic regression analyses of psychoeducational interventions on the detection of students' mental health problems based on this paper's multimodal fusion of affective computational models are shown in Table 6, with total mental health, obsessive-compulsive symptoms, paranoia, hostility, interpersonal tensions and sensitivities, depression, anxiety, academic stress, maladjustment, emotional imbalance, and whether or not mental imbalance was detected after the interventions as the dependent variable (yes=1, no=0), respectively and whether intervention (Yes=1, No=0) was the independent variable, controlling for confounders that constituted statistically significant differences between the intervention and control groups, and binary logistic regression analyses were performed. The results showed that the post-intervention detection rates of total mental health, obsessive-compulsive symptoms, paranoia, hostility, interpersonal tension and sensitivity, depression, anxiety, and academic stress in the intervention group were 0.65, 0.75, 0.74, 0.62, 0.66, 0.75, 0.64, and 0.73 times higher than those in the control group, respectively, for the adolescent students ( $p$ -value  $< 0.05$  for all).

Table 6: Logistic regression analysis

Dependent variable	$\beta$	Standard error	Wald $\chi^2$	$P$ value	OR value
Obsessive-compulsive Disorder	-0.33	0.11	7.68	0.01	0.75
Paranoia	-0.27	0.11	5.62	0.01	0.74
Antagonism	-0.58	0.11	16.54	0.00	0.62
Interpersonal relation	-0.44	0.12	10.32	0.00	0.66
Depression	-0.35	0.12	7.54	0.00	0.75
Anxiety	-0.37	0.13	10.36	0.01	0.64
Learning pressure	-0.33	0.12	7.45	0.00	0.73
Maladjustment	-0.14	0.13	1.56	0.23	0.85
Mood inequality	-0.23	0.11	3.65	0.05	0.82
Psychological inequality	-0.21	0.13	2.56	0.12	0.82
Total mental health	-0.45	0.13	12.89	0.00	0.65

## VI. Conclusion

In this paper, we design a method for analyzing and calculating the expression of students' emotional state from the perspective of subjective and lateral emotions, and then apply the method to the optimization of students' psychoeducation and students' psychological intervention path.

The multi-head attention mechanism added in the model effectively captures the correlation between image features and text features, and weights the attention to different forms of features, which makes the model's emotion

classification accuracy improve by 2.86% compared with the better-performing model. And the model in this paper also achieved good results on the public datasets MVSA-Single and MVSA-Multi.

Except for the maladjustment factor, the students in the intervention group showed significant improvement in all psychological state factors after the psychoeducational intervention based on this paper's multimodal fusion of affective computing model. The interventions in this paper were significantly effective in improving students' total mental health, obsessive-compulsive symptoms, paranoia, hostility, interpersonal tension and sensitivity, depression, anxiety, and academic stress, with P values less than 0.05, and were not significant in improving the maladjustment, emotional imbalance, and psychological imbalance factors, with P values of 0.85, 0.82, and 0.82, respectively, and the reasons for the non-significant improvement in the above three factors may be Due to the limited understanding and learning ability of some students, some adverse factors in the family, school and social environment affect the intervention effect of this paper's model, but on the whole this paper's model can improve students' mental health well and make up for the shortcomings of traditional mental health education.

## References

- [1] Bernaras Iturrioz, E., Insúa Cerretani, P., & Bully Garay, P. (2018). Prevalence and severity of psychological problems in university students. *British Journal of Guidance & Counselling*, 46(4), 418-428.
- [2] Cuijpers, P., Smit, F., Aalten, P., Batelaan, N., Klein, A., Salemink, E., ... & Karyotaki, E. (2021). The associations of common psychological problems with mental disorders among college students. *Frontiers in Psychiatry*, 12, 573637.
- [3] Storrie, K., Ahern, K., & Tuckett, A. (2010). A systematic review: students with mental health problems—a growing problem. *International journal of nursing practice*, 16(1), 1-6.
- [4] Pedrelli, P., Nyer, M., Yeung, A., Zulauf, C., & Wilens, T. (2015). College students: mental health problems and treatment considerations. *Academic psychiatry*, 39(5), 503-511.
- [5] Layte, R. (2012). The association between income inequality and mental health: testing status anxiety, social capital, and neo-materialist explanations. *European Sociological Review*, 28(4), 498-511.
- [6] Bakkialakshmi, V. S., & Sudalaimuthu, T. (2021, December). A survey on affective computing for psychological emotion recognition. In 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 480-486). IEEE.
- [7] Gaggioli, A., Riva, G., Peters, D., & Calvo, R. A. (2017). Positive technology, computing, and design: shaping a future in which technology promotes psychological well-being. In *Emotions and affect in human factors and human-computer interaction* (pp. 477-502). Academic press.
- [8] Esposito, A., Esposito, A. M., & Vogel, C. (2015). Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognition Letters*, 66, 41-51.
- [9] Picard, R. W. (2000). *Affective computing*. MIT press.
- [10] Guo, R., Guo, H., Wang, L., Chen, M., Yang, D., & Li, B. (2024). Development and application of emotion recognition technology—a systematic literature review. *BMC psychology*, 12(1), 95.
- [11] Marin-Morales, J., Linares, C., Guixeres, J., & Alcañiz, M. (2020). Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors*, 20(18), 5163.
- [12] Desideri, L., Ottaviani, C., Malavasi, M., Di Marzio, R., & Bonifacci, P. (2019). Emotional processes in human-robot interaction during brief cognitive testing. *Computers in Human Behavior*, 90, 331-342.
- [13] Russo, S., Lorusso, L., D'Onofrio, G., Ciccone, F., Tritto, M., Nocco, S., ... & Giuliani, F. (2023). Assessing feasibility of cognitive impairment testing using social robotic technology augmented with affective computing and emotional state detection systems. *Biomimetics*, 8(6), 475.
- [14] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83, 19-52.
- [15] Smith, E., Storch, E. A., Vahia, I., Wong, S. T., Lavretsky, H., Cummings, J. L., & Eyre, H. A. (2021). Affective computing for late-life mood and cognitive disorders. *Frontiers in psychiatry*, 12, 782183.
- [16] Greene, S., Thapliyal, H., & Caban-Holt, A. (2016). A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, 5(4), 44-56.
- [17] Schwark, J. D. (2015). Toward a taxonomy of affective computing. *International Journal of Human-Computer Interaction*, 31(11), 761-768.
- [18] Zacharatos, H., Gatzoulis, C., & Chrysanthou, Y. L. (2014). Automatic emotion recognition based on body movement analysis: a survey. *IEEE computer graphics and applications*, 34(6), 35-45.
- [19] Ong, D. C., Soh, H., Zaki, J., & Goodman, N. D. (2019). Applying probabilistic programming to affective computing. *IEEE Transactions on Affective Computing*, 12(2), 306-317.
- [20] Williamson, B. (2017). Moulding student emotions through computational psychology: Affective learning technologies and algorithmic governance. *Educational Media International*, 54(4), 267-288.
- [21] Liu, J., & Wang, H. (2022). Analysis of educational mental health and emotion based on deep learning and computational intelligence optimization. *Frontiers in Psychology*, 13, 898609.
- [22] Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology*, 47(6), 1304-1323.
- [23] Ren, D. (2024). Research on the Application of Information Grain-Based Affective Computing Model in College Students' Psychoeducation. *Contemp. Readings L. & Soc. Just.*, 16, 381.
- [24] Huangfu, B., & Cheng, W. (2025). Cognitive computing method based on decoding psychological emotional states. *International Journal of Cognitive Computing in Engineering*, 6, 32-43.
- [25] Akbiyik, C. (2010). Can affective computing lead to more effective use of ICT in Education. *Revista de Educación*, 352(4), 181-185.
- [26] Marwah Bani Saad, Lidia Jackowska Strumillo & Wojciech Bieniecki. (2025). Hybrid ANN-Based and Text Similarity Method for Automatic Short-Answer Grading in Polish. *Applied Sciences*, 15(3), 1605-1605.



- [27] Min Zou & ZhongPing Wang. (2025). A REVIEW OF THE APPLICATION OF BERT MODEL IN TEXT CATEGORIZATION. *World Journal of Information Technology*,3(2),
- [28] Chun Chun Jia, Hongwen He, Jiaming Zhou, Kunang Li, Jianwei Li & Zhongbao Wei. (2024). A performance degradation prediction model for PEMFC based on bi-directional long short-term memory and multi-head self-attention mechanism. *International Journal of Hydrogen Energy*,60,133-146.
- [29] 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.