J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 6651--6670

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Mental Health Assessment of College Students Based on Social Sentiment Analysis and Multi-Branch Neural Networks

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*

ABSTRACT
Aiming at the complexity of mental health assessment for students in colleges and universities, this paper proposes an innovative framework that integrates social sentiment analysis and multi-branch neural networks. A multilevel mental health assessment system is constructed through cross-modal feature interaction CNN+BiGRU with heterogeneous graph structure modeling. In the model design, image feature extraction is pre-trained by five-branch CNN structure ViT, text features are fused by dynamic word embedding with multi-scale convolution, and a virtual node and metapath-driven heterogeneous graph neural network H-GNN is introduced to strengthen the global relationship modeling. Experiments show that the model achieves 89.7% and 91.2% accuracy on Twitter-15 and Twitter-17 datasets, respectively, and the F1 values are improved by 3.24% and 2.32% from the optimal baseline BICCM. In the actual college mental health monitoring, the model successfully captured the time-series fluctuations of depression index and anxiety level, and found that the rational-perceptual dimension was highly correlated with the examination cycle, with 0.69 during the midterm examination and 0.68 during the final examination. Through the ten-fold cross-validation comparison experiments, the model significantly outperforms the cutting-edge models, such as MIMN-BERT, EF-NET and so on on the weighted average index, with an average accuracy rate of 99.02% and F1 value of 98.08%. The study shows that the framework provides a highly accurate and interpretable technical solution for mental health risk early warning, which is especially suitable for dynamic monitoring scenarios in universities.

Keywords: social sentiment analysis; multi-branch neural network; heterogeneous graph modeling; mental health assessment

✉ Corresponding author.
E-mail address: muzi2000718@163.com (L. Li).
Received 01 March 2024; Revised 25 May 2024; Accepted 10 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-364

1. Introduction

Mental health of college students is one of the most important issues that are widely concerned in today's society. In modern society, college students face pressures from various aspects such as academics, interpersonal relationships, and future employment, which may negatively affect their mental health [1-3]. Therefore, it is especially crucial to emphasize the mental health problems of college students and provide necessary support and assistance. When assessing the mental health of college students, traditional mental health assessment methods rely on interviews, psychological tests, and behavioral observations [4]. Although these methods have some application value in individual diagnosis, their limitations are becoming increasingly apparent as mental health problems become more complex and widespread [5-6]. For example, the assessment cycle is long and costly, which makes it difficult to meet the demand for large-scale screening, and is limited by the assessor's professionalism and subjective judgment, which makes it unable to effectively respond to individualized needs [7-9]. In addition, traditional health assessment methods are often static, making it difficult to provide real-time feedback on mental health status [10-11]. Therefore, colleges and universities should actively establish an accurate and practical digital assessment model of college students' mental health to provide more effective intervention and support strategies for schools and related organizations by comprehensively understanding the mental health status of college students [12-14].

With the wide application of the Internet and mobile devices, social media-based communication among college student groups has become the main vehicle for them to record their lives, express their views, and share their communication [15-16]. The information contained in these social data can truly and accurately reflect the emotional status and psychological state of college students, which can provide effective data support for college students' mental health education, and provide new data guarantee for constructing psychological early warning mechanism for college students [17-20]. College student groups in the network social publishing content is not limited to the text expression form, images and network expressions as supplementary information often appear together with the main text [21-22]. In view of the complexity of the emotional expression of college student groups, the introduction of multi-branch neural networks to extract students' key information and establish a prediction model, so as to quickly and accurately realize the prediction and intervention of college students' mental health, which is expected to provide new ideas and methods for mental health assessment [23-25].

In this paper, we propose an innovative framework that integrates social sentiment analysis and multi-branch neural networks to construct a multi-level mental health assessment system through cross-modal feature interaction and heterogeneous graph structure modeling. It mainly focuses on three core models. The cross-modal sentiment classification model based on feature fusion adopts a multi-branch convolutional neural network (CNN) to extract hierarchical features of image and text respectively, and combines with bidirectional gated recurrent unit (BiGRU) to realize bidirectional inter-modal information interaction, which breaks through the limitations of the traditional splicing and fusion. The multimodal sentiment analysis algorithm based on heterogeneous graph neural network introduces virtual nodes and meta-path mechanism, constructs a global information fusion center, and captures cross-modal semantic associations through hierarchical node aggregation. The mental health assessment model for college students further integrates the technical advantages of the previous two, and designs a three-branch deep learning architecture (CNN, Random Forest, and Fully Connected Network) to process positive, negative, and neutral psychological features respectively, and ultimately realizes the prediction of depression risk through feature fusion.

2. Model construction based on social sentiment analysis with multi-branch neural networks

2.1. Cross-modal sentiment classification model based on feature fusion

Utilizing the advantages of CNN neural network model for feature extraction, this paper designs two deep convolutional neural networks with multiple branches for extracting multilevel features of images and text respectively to accomplish the feature extraction task.

2.1.1 CNN-based feature extraction. To realize cross-modal sentiment classification, this section first proposes a multi-branch CNN-based feature extraction framework. Through the local connectivity and weight sharing properties, CNN can efficiently capture the local semantic features of image and text. In the following section, the specific designs of image and text feature extraction are described separately.

1) Image Feature Extraction. CNN is a structure of deep neural network, which can be regarded as a kind of locally connected network, and its biggest features are: local connectivity and weight sharing compared with fully connected network. That is, for a point p in a two-dimensional grid, the closer the point p is to the point, the greater the influence on it; according to the statistical properties of the natural image, the weights of one localization can also be used for the analysis of another localization. The weight sharing here is the convolutional kernel parameter sharing, two convolutional kernels with unequal weights can extract two kinds of features on a two-dimensional mesh.CNN has achieved great success in various computer vision tasks in recent years, and it is feasible to use CNN to extract image features in the task of image emotion classification. The image convolutional neural network in this paper adds multiple intermediate branches to the CNN.Zhou's visualization experiments show that in a convolutional neural network, as the number of layers increases, the activation acceptance domains of the pooling layers of different convolutional layers become more meaningful and easier to be recognized by humans. Based on this, our multi-branch image convolutional neural network adopts the design of multiple branches in extracting image features in order to obtain different levels of image features from low to high, not only the overall image features, for the next feature fusion model selection and feature fusion, where the design of the branches is taken as an example of branch one, as shown in Fig. 1.

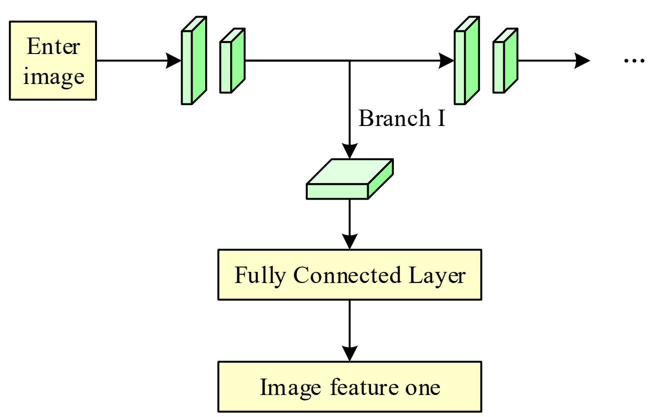

Fig. 1. Branch design of image convolutional neural network

Each of these branches contains a 128-dimensional convolutional layer of size 1×1, a fully connected layer for dimensional transformation.

The image convolutional neural network used to extract image features in the paper has multiple branches, which are used to extract image features at many different levels, where the number of branches can be set from 1 to 5, which is set to 5 in this paper.The pre-processed image P is fed into the pre-trained image convolutional neural network (denoted as PicCNN) to obtain the image features $\{p_i\}_{i=1}^{m}$, where m denotes the number of branches, i.e.:

$$p_i = PicCNN_i\left(P\right) \tag{1}$$

2) Text Feature Extraction. Convolutional Neural Networks and Recurrent Neural Networks have been successfully applied in the field of Natural Language Processing. The advantage of Convolutional Neural Networks is that they can easily extract local features in the sentence by using filters and reduce the influence of irrelevant words on the final result. Convolutional neural networks with multiple branches can also extract key features in different ranges. And the size of the convolutional kernel can be changed arbitrarily according to the need, which is more flexible. In this paper, the convolutional neural network is used for text feature extraction, and the same multi-branch design is used, the branch design of the text multi-branch convolutional neural network is shown in Figure 2.

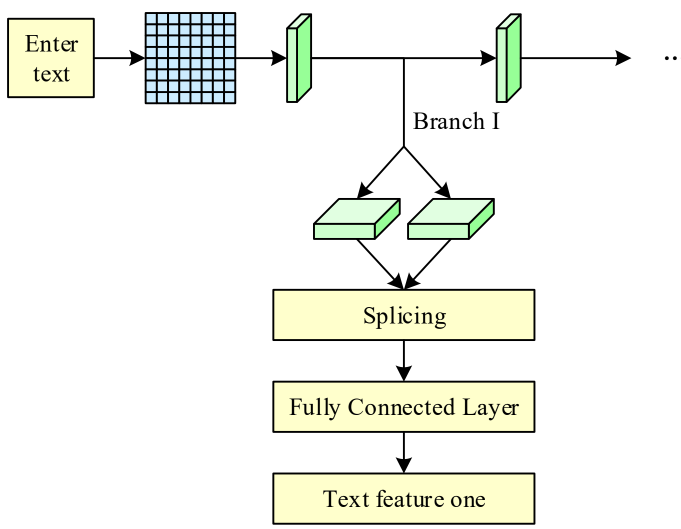

Fig. 2. Branch design of text multi-branch convolutional neural network

The text multibranch convolutional neural network has the same number of layers as the image convolutional neural network, and differs from the image convolutional neural network in the branch design and the size of the convolutional kernel in each layer.

Specifically, our method text indexes the text with a preprocessed lexicon to obtain $\{T_i\}_{i=1}^{h}$ (h denotes the length of the text after indexing), and then inputs $\{T_i\}_{i=1}^{h}$ to the Embedding module to be converted into a word vector representation $\{e_i\}_{i=1}^{h}$, and subsequently feed the word vector representation into a pre-trained textual multi-branch convolutional neural network (denoted by TextCNN), i.e., $\{e_i\}_{i=1}^{h}$ into our TextCNN to obtain the textual feature $\{t_i\}_{i=1}^{m}$ (m denotes the number of branches), i.e.:

$$t_j = TextCNN_j(\{e_i\}_{i=1}^{h}) \tag{2}$$

2.1.2 BiGRU-based feature fusion. After completing the hierarchical extraction of image and text features, the semantic associations between different modalities still need to be further mined. To this end, this section introduces a bidirectional feature fusion mechanism based on BiGRU, which realizes the deep interaction of cross-modal information through temporal modeling.

In the cross-modal sentiment classification task, the current mainstream feature fusion methods only perform simple splicing, addition, subtraction, multiplication and division operations on the features of different modalities, which are difficult to closely link the features of different modalities, and a BiGRU-based fusion method is designed by taking advantage of the long time memory of RNN. That is, on the basis of already extracted image and text features, the information of image and text features is interacted and further extracted by a BiGRU, and the fused features are obtained by combining the fusion mechanism in two directions, from low-level features to high-level features and from high-level features to low-level features.

Specifically, our method first splices the image features and text features output from different branches, and then inputs them into a BiGRU with the same length and number of branches in a back-and-forth order, the BiGRU can be regarded as consisting of two unidirectional GRUs, one for forward propagation and one for backward propagation, and the neuron structure of the BiGRU is shown in Fig. 3.

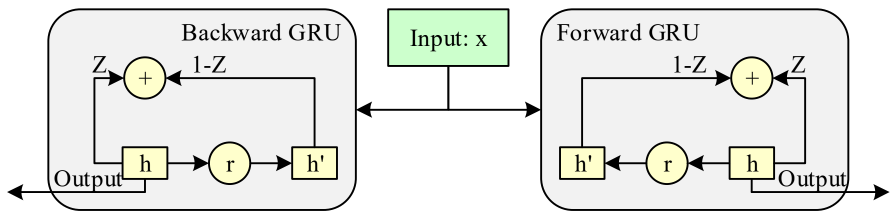

Fig. 3. BiGRU neuron structure

which at moment t contains the reset gate r_t, the update gate z_t, and the hidden state h_t, which is updated as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{3}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{4}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{5}$$

That is, the update gate z is used to control the extent to which state information from the previous moment affects the current state, and the reset gate r is used to control the extent to which state information from the previous moment is ignored.

The output of BiGRU is computed from two unidirectional GRUs as follows:

$$\hat{y}_t = g(H \cdot [\overrightarrow{h_t}, \overleftarrow{h_t}] + c) \tag{6}$$

We splice image features $\{p_i\}_{i=1}^m$ and text features $\{t_i\}_{i=1}^m$ in sequential order and input them into the BiGRU fusion model, to obtain the bottom-up features containing the relationship from the low-level features to the high-level features (the forward output of BiGRU) and the top-down features (backward output of BiGRU), and finally the two are spliced together to obtain the fused feature vector y, i.e.

$$y_{last} = \overrightarrow{BiGRU}\left(\{concat(p_i, t_i)\}_{i=1}^m\right) \tag{7}$$

$$y_{first} = \overleftarrow{BiGRU}\left(\{concat(p_i, t_i)\}_{i=1}^m\right) \tag{8}$$

$$y = concat\left(y_{last}, y_{first}\right) \tag{9}$$

2.2. *Multimodal Sentiment Analysis Algorithm Based on Heterogeneous Graph Neural Networks*

Although BiGRU can effectively integrate bimodal features, a more scalable relationship modeling framework needs to be constructed in the face of heterogeneous data from multiple sources, such as psychological questionnaires and social behaviors. Based on this this section proposes a multimodal sentiment analysis algorithm based on heterogeneous graph neural network, which realizes global information integration through virtual node and meta-path mechanism.

2.2.1. Graph initialization. Based on the graph attention mechanism, it is also necessary to convert each piece of data into a heterogeneous node. Since different data contain different modal types, the number of nodes contained is also different, and a heterogeneous graph $G=(V,E)$ is constructed for all the data, where V is the set of nodes and E is the set of edges. Each node $v \in V$ has a type $\phi_v \in \Gamma_v$ and each edge $e \in E$ has a type $\phi_e \in \Gamma_e$. Γ_v and Γ_e are the set of node types and the set of edge types, respectively. Each node v also has a feature vector $x_v \in R^d$, where d is the feature dimension.

2.2.2. Virtual Node Construction. In GNNs, the concept of virtual nodes is often used to enhance the information transfer capability of the network. A virtual node can be defined as v^*, which is designed to provide a global information fusion center. Each modal counterpart is connected to this virtual node through a special type of edge that can be labeled ϕ_{e^*}. The existence of virtual nodes allows the network to transmit information without being restricted to a single modality, thus creating connections between different modalities and effectively integrating information from multiple sources.

The feature vector x_{v^*} of a virtual node can be initialized by different strategies, in this paper, by averaging the node feature vectors of all modalities to obtain a comprehensive representation as a feature fusion strategy. This design enables the virtual node to act as a global information hub and provide a reference feature description for other nodes in the network.

2.2.3. Figure Embedding. In heterogeneous graphs, meta-path is an important concept to characterize the complex relationships between different types of nodes and edges. A metapath P is defined as an ordered combination of a series of edge types e, which can be expressed as:

$$P=\phi_{e_1} \rightarrow \phi_{e_2} \rightarrow \cdots \rightarrow \phi_{e_k} \tag{10}$$

Here ϕ_{e_i} denotes a type of edge, and a metapath P expresses a specific type of path between a start node and a termination node. Meta-paths are able to capture the complex semantic relationships between nodes in heterogeneous graphs, which helps to understand the interactions and patterns between different entities.

In order to efficiently process the heterogeneous information in the graph and compress this information into low-dimensional feature representations, in this paper, we define an H-GNNs as H, with the aim of mapping the feature vector x_v of each node v to a low-dimensional embedding vector $h_v \in R^k$, where k denotes the dimension of the embedding space.

Specifically, in this paper, the embedding h_v of node v is set as a compressed representation of its original feature vector, which is used to capture the structural information of node v as well as the relational information with its neighboring nodes. The process of generating the embedding vectors requires the computation of neighborhood aggregation and feature propagation using a message-passing mechanism in this paper, where nodes receive and send information, and the network iteratively updates the node's embedding according to well-defined rules. This process can be formalized as the following equation:

$$h_v^{(l+1)} = f\left(h_v^{(l)}, \left\{h_u^{(l)} : u \in \mathrm{N}(v)\right\}\right) \tag{11}$$

Where $h_v^{(l)}$ and $h_v^{(l+1)}$ denote the embedding vectors of node v in the l th and $l+1$ th layers, respectively, f is a learnable aggregation function, and $\mathrm{N}(v)$ denotes the set of neighbor nodes of node v. Through iterative message passing and feature aggregation, the embedding vectors of each node will be optimized during the training process to more accurately reflect the characteristics of the nodes and the relationships between nodes.

2.2.4. Hierarchical node aggregation based on meta-paths. In heterogeneous graphs, different types of nodes and edges carry rich and diverse information, so how to effectively perform information aggregation and transformation is a key issue in the study of heterogeneous graph neural networks. Each layer in the heterogeneous graph neural network is responsible for information aggregation and transformation, for the l th layer in the network H, the embedding update of node v can be expressed by the following equation:

$$h_v^{(l)} = f^{(l)}\left(x_v, \left\{h_u^{(l-1)} : u \in N_v\right\}\right) \tag{12}$$

Where $h_v^{(l)}$ is the embedding representation of node v at layer l, $f^{(l)}$ is the information aggregation and transformation function at layer l, which is responsible for combining the node's own features with its neighbor's embedding representations and outputting updated embeddings, and N_v denotes the set of neighboring nodes of node v, which also includes the virtual node v^*, this is to model and learn the self-connectivity property of nodes.

Since metapaths can not only help capture the heterogeneity among nodes, but also enhance the connection semantics among nodes, this paper uses metapaths to define a more detailed set of neighboring nodes to guide the information aggregation process. For a given meta-path P, node v is defined as follows with respect to the set of neighbor nodes N_v^P under this meta-path:

$$N_v^P = \left\{u : u \in N_v, \phi_{(u,v)} = P\right\} \tag{13}$$

Here $\phi_{(u,v)}$ is the type of meta-path between node u to node v, and N_v^P is the set of neighbors of node v under the meta-path P. In this way, in this paper, we set different meta-paths for each type of nodes and edges in the heterogeneous graph in order to achieve selective aggregation of information.

In the process of information aggregation, nodes will synthesize their own attributes as well as information from different types of neighbors, and through the cascading GNN layers, the representation of each node will obtain broader neighborhood information layer by layer. In practice, the $f^{(l)}$ function uses an attention mechanism to weight the influence of neighboring nodes.

In addition, in order to enable the model to be trained smoothly and achieve optimization, this paper also introduces some regularization means, such as dropout and weight decay, to control the complexity of the model and prevent the occurrence of overfitting phenomenon.

2.3. Mental Health Assessment Model for Higher Education Students

Based on the aforementioned cross-modal fusion and heterogeneous graph modeling techniques, this section further constructs a three-branch deep learning model for mental health assessment of college students, which combines the multimodal analysis capability with the need for psychological feature classification, and which fully combines the multibranch CNN feature extraction capability with the idea of heterogeneous graph relational modeling to achieve accurate prediction of depression risk.

In order to predict students' depression status through the data from the questionnaire, the article used a three-branch deep learning model as shown in Figure 4. Each branch is designed to process a specific type of psychometric data from the students' questionnaire: positive trait data, negative trait data, and neutral trait data. The model structure is briefly described below.

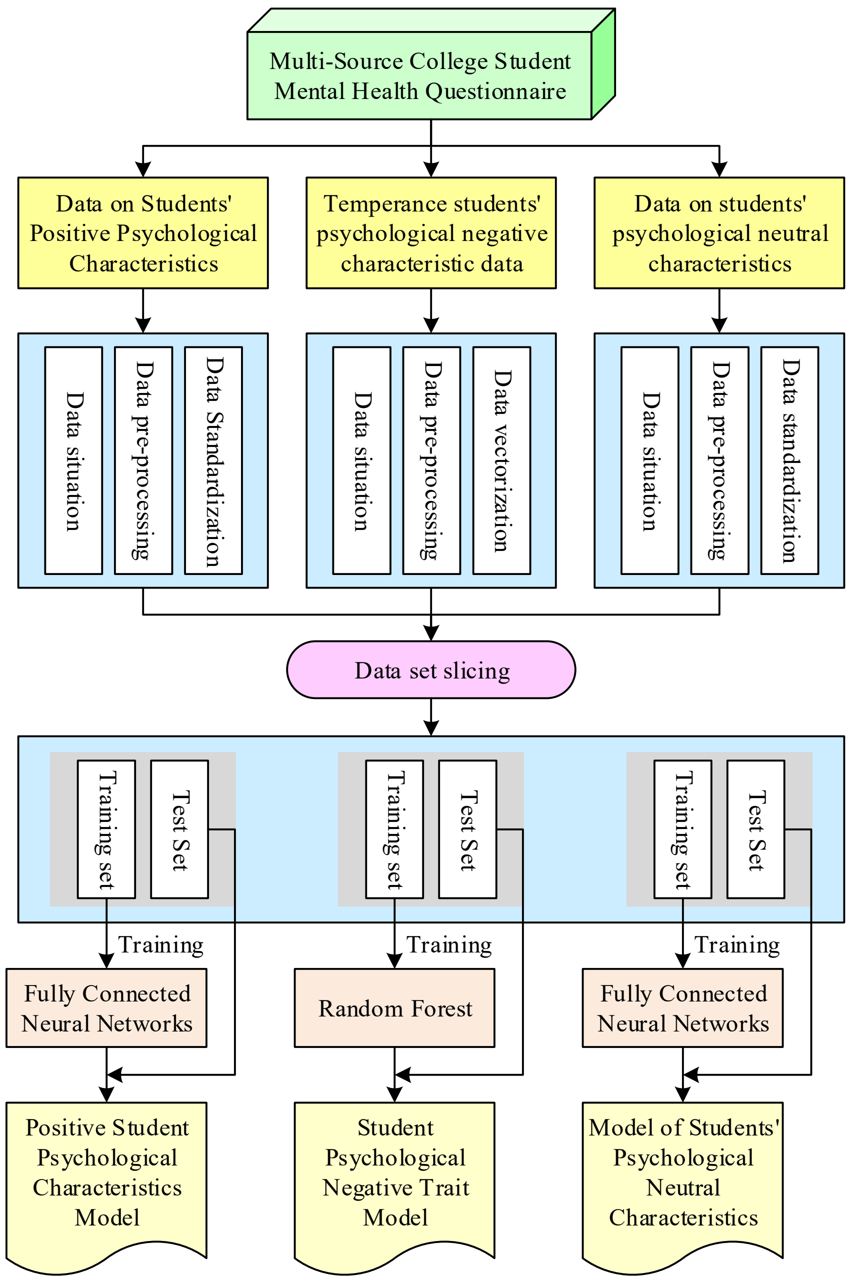

Fig. 4. Multi-branch deep learning psychological assessment model framework

First, we constructed the previously mentioned multimodal neural network branch incorporating CNN and RNN to process student positive trait class data. This class of data includes participants' intelligibility, responsibility, daring, independence, and self-discipline. According to the research, high intelligence increases pressure and expectation, high responsibility leads to self-blame, high daring faces setbacks, high independence may feel lonely, and high self-discipline increases stress, and these positive psychological traits have potential negative effects on students' depressive conditions, and the constructed neural network can capture the nonlinear patterns in these positive psychological data to provide some of the informative features for the model to predict students' depressive conditions.

Second, we constructed a random forest model to deal with the data in the category of psychological assessment of students' negative traits. This category of data includes students' fantasizing, worldliness, dominance, sensitivity, skepticism, apprehension, and nervousness scores. These traits are data in the area of negative psychological traits of individual students and are more associated with depression than positive psychological traits. Higher fantasizing and worldliness, greater dominance, sensitivity, skepticism, apprehension, and nervousness may be associated with an increased risk of depression. Lack of social support, high mood swings, excessive control over others, anxiety, sensitivity, suspicion, apprehension, and tension may contribute to depressed mood. This information depicts in depth the negative psychological states and behavioral manifestations of students, and since these attributes and depressiveness are used as negative psychological indicators, they have a high predictive value. Random forest, as a nonparametric integrated learning model, can effectively handle the complex interactions between various features and analyze the psychological assessment data in depth.

Finally, we construct a new branch of fully-connected neural network to deal with neutral feature data in students' psychology. This type of data includes students' congeniality, stability, arousal, and openness scores. These traits relate to an individual's introverted and extroverted tendencies and acceptance of novel experiences. The relationship between them and depression may be relatively weak, and similar to the CNN that handles data from the positive traits category, this CNN also captures the non-linear relationships between these traits and further aids the primary traits to improve the accuracy of the predictions.

Below is an algorithmic representation of the model:

Algorithm: three-branch deep learning psychological assessment model

Input: students' positive trait data X_1, negative (negative) trait data X_2, neutral trait data X_3

Output: predicted depressive status Y

Steps:

1) Initialize the model parameters.
2) Perform the following steps for each training instance:

(1) Input the positive feature data X_1 into the Fully Connected Neural Network (FCN) branch to obtain the feature characterization result Y_1; input the negative feature data X_2 into the Random Forest Model branch to obtain the feature characterization result Y_2; (2) Input the neutral feature data X_3 into the Fully Connected Neural Network (FCN) branch to obtain the feature characterization result Y_3; (3) Splicing and integrating the prediction results Y_1, Y_2, and Y_3 from the three branches to generate the combined prediction result Y.

3) Use cross-entropy loss to calculate the gap between the prediction result Y and the actual value.
4) Use stochastic gradient descent to update the parameters of the model.
5) Repeat steps two through four until the model converges, i.e., the loss value no longer decreases significantly or reaches 100 training rounds.
6) Output the predicted depression status Y.

3. Multimodal Sentiment Analysis Experiment

In order to verify the effectiveness of the cross-modal sentiment classification model, heterogeneous graph neural network and mental health assessment framework proposed in Chapter 2, this chapter conducts multimodal sentiment analysis and mental health prediction experiments based on publicly available datasets and self-constructed data, and verifies the superiority of the models through comparative analysis and empirical research.

3.1. Data set and basic experimental setup

3.1.1. Data sets. In this chapter, multimodal aspect-level sentiment analysis experiments will be conducted based on the publicly available Twitter-15 and Twitter-17 datasets and the datasets constructed in this paper. Firstly, each dataset is divided into three parts: training set, validation set and test set.The statistics of Twitter-15 dataset and Twitter-17 dataset are shown in Table 1.

Table 1. Twitter-15 and Twitter-17 datasets statistics

Dataset	Twitter-15			Twitter-17		
	Train	Dev	Test	Train	Dev	Test
Positive	638	425	517	1723	472	582
Neutral	2034	816	684	1882	466	409
Negative	412	108	98	162	112	214
Total	3084	1349	1299	3767	1050	1205

3.1.2. Experimental environment and parameter settings. In this study, the PyTorch framework is used as the experimental environment. A pre-trained BERT with 12 layers and 828 hidden dimensions was used as a feature extractor for text modality in the experiments. Using pre-trained ViT model was used as a feature extractor for image modality. First, the images were resized to 224 × 224 and then they were divided into 14 × 14 blocks of 16 × 16 pixels. In the grouped coordinate convolutional model, the text is divided into 2 groups and the images are divided into 4 groups with a convolutional kernel size of 1×1. The parameters are set as follows: the Batch size is 16 the learning rate is adjusted to 2e-5, the Dropout is 0.1, the period is set to 30, the early stop 5, and the optimizer is Adam.

3.1.3. Evaluation indicators. Macro-averaged F1 values and accuracy ACC were used as evaluation metrics to measure the performance of the model. True case TP is the number of samples that predict positive cases into positive cases, false positive case FP is the number of samples that predict negative cases into positive cases, true negative case TN is the number of samples that predict negative cases into negative cases, and false negative case FN is the number of samples that predict positive cases into negative cases. From the primary metrics, secondary metrics are derived: the check accuracy rate P, the recall rate R, and the F1 value, as follows:

The check accuracy rate is the ratio of the number of correctly identified by the system to the total number of entities of the class identified. The formula for the check accuracy rate is as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{14}$$

The recall rate is the ratio of the number of correctly recognized by the system to the total number of entities of that type in the real data. The formula for recall rate is as follows:

$$R = \frac{TP}{TP + FN} \times 100\% \tag{15}$$

P value and R value in the case of positive and negative sample category imbalance will appear contradictory situation, then it is necessary to consider a combination of the F1 value is commonly used to judge the whole, F1 value is more concerned about the low value, therefore, F1 value is high only when the rate of checking the accuracy and the recall rate are high. The F1 formula is as follows:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{16}$$

Accuracy is an intuitive performance metric that is the ratio of the number of samples predicted correctly by the model to the total number of samples. The formula for calculating it is as follows:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \tag{17}$$

3.2. Comparative Experiments and Model Performance Analysis

After completing the setting of the experimental environment and evaluation metrics, this section further selects multi-class baseline models for comparison experiments to verify the performance advantages of this paper's method in cross-modal sentiment classification tasks.

3.2.1. Baseline Comparison Model. In this section, in order to validate the effectiveness of this paper's multimodal fusion model that incorporates CNNs and RNNs as well as heterogeneous graphical neural networks based on heterogeneous graphical neural networks, the following baseline methods are chosen as a benchmark for comparison in this chapter.

1) Visual-only approaches. Resnet- Aspect: visual and aspect features are extracted using ResNet and BERT respectively. Features from both modalities are merged for predicting aspectual emotional polarity.

Faster R-CNN-Aspect: similar to Resnet- Aspect, visual features are extracted by Faster R-CNN.

2) Text-only methods. RAM: Captures the important information of the aspect in the text using the attention mechanism and GRU model.

MGAN: Combines aspects interacting with the context using both coarse- and fine-grained attention mechanisms, and then connects the two to predict sentiment polarity.

BERT: A text-based pre-trained language model that uses a multi-layer Transformer encoder to obtain word vector representations.

BERT-BL: an additional BERT layer added to the basic BERT model for fusing different modalities.

3) Multi-modal based approaches. Res-MGAN: combines image features extracted by ResNet-152 and text features extracted by MGAN, and then simply connects these two features.

MIMN: employs two interacting memory networks that fuse given aspects with textual and visual information, capable of learning cross-modal and endomodal interaction effects.

ESAFN: introduces an entity-sensitive attention and fusion network that uses ResNet-152 and BiLSTM to extract image and text features, respectively, and uses a gating mechanism to eliminate noisy visual information.

TomBERT: ResNet and BERT are applied to extract features from image and text modalities, then another self-attention layer is added to capture inter-modal interactions, followed by multimodal information fusion via another BERT.

ViLBERT: a pre-trained visual-verbal model that directly connects specific aspects to input sentences to predict affective polarity.

Topic-MABSA: augmenting target information with topics, then capturing multimodal interactions through a multi-head attention mechanism, and ultimately fusing textual and visual information through a gated cell design.

MIMN-BERT: Replacing the textual feature extraction part of the MIMN model with a pre-trained BERT model to generate aspects and text embeddings for further feature fusion.

EF-NET: Uses Multihead Attention (MHA)-based network and ResNet-152 to process text and images, respectively, and captures interactions between multimodal inputs using MHA.

IFNRA: extracts features from text and images using BERT and a pre-trained bottom-up attention model, and proposes an iterative attention decoder that learns aspect-specific affective features step-by-step.

BICCM: Text and image features are extracted using BERT and ResNet, and a bi-directional complementary correlation strategy is used to generate image-sensitive text features and text-sensitive image features. Finally, the features of both kinds are fused.

3.2.2. Experimental results and analysis. Figure 5 shows the comparison of the results of different models on the public datasets Twitter-15 and Twitter-17. It can be clearly seen that the model in this paper outperforms all the baseline models in terms of accuracy and F1.

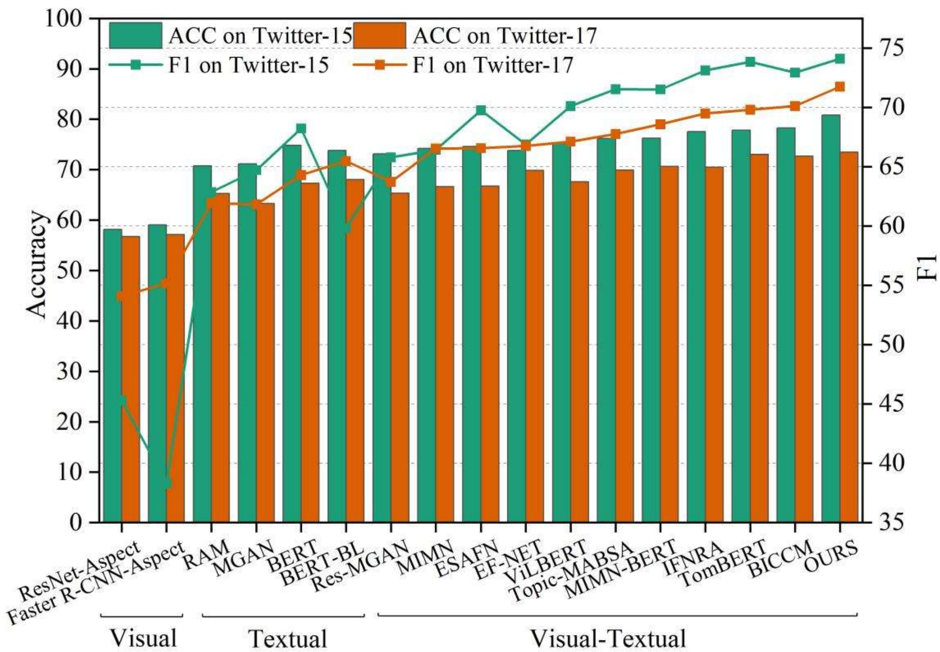

Fig. 5. Different model performances comparison on public datasets

As can be seen in Fig. 5, in the unimodal approach, the performance of the visual modal model is lower than that of the textual modal model. For example, on the Twitter-15 dataset, the BERT method outperforms the ResNet-Aspect method by 28.63% in terms of accuracy. Even the RAM method, which does not use a pre-trained model, outperforms the ResNet-Aspect method with a 21.66% improvement in accuracy. This indicates that textual modality analysis is crucial in multimodal aspect-level sentiment analysis tasks. In textual modality, BERT and BERT-BL outperform the RAM method due to BERT's ability to learn rich linguistic representations and strong contextual understanding as a result of BERT's training on large-scale corpora. However, the significant improvement in accuracy and F1 values of the multimodal approaches proves the equally important contribution and value of non-textual modalities in multimodal sentiment analysis. Res-MGAN, MIMN, and ESAFN, which lacked the use of pre-trained models, performed poorly on ACC and F1.EF-NET outperforms the above methods, suggesting that attentional mechanisms are indispensable in aspect-based multimodal sentiment analysis.Topic-MABSA makes innovations in introducing thematic augmentation of information and gated fusion design, but it still struggles to capture the difference in the space between aspectual words and the graphic content, but it still falls short in capturing the subtle spatial relationship between aspect words and graphic content, whereas this paper's method outperforms it by using multi-hop interactive attention to further capture the association between aspect words and graphic features. The pre-trained ViLBERT model performs worse than TomBERT, probably because the dataset used for pre-training of ViLBERT is much smaller than that of BERT. Compared to IFNRA, TomBERT and BICCM, the methods in this chapter improved 4.24%, 3.89% and 3.28% on ACC and 1.35%, 0.36% and 1.60% on F1 on the Twitter-15 dataset, respectively. Similarly, on the Twitter-17 dataset, ACC improved by 4.23%, 0.64% and 1.11% and F1 improved by 3.24%, 2.78% and 2.32% respectively. Although the above methods use BERT as a text feature extractor, as well as a bottom-up approach, ResNet pre-trained model as an image feature extractor, the method in this paper outperforms them by using a pre-trained ViT model for image feature extraction.ViT overcomes the traditional CNN and ResNet by introducing the self-attention mechanism of the Transformer model's limitation of local receptive fields in image feature extraction, capturing the global contextual relationships in images and better understanding their semantics and structure. In summary, the method proposed in this paper is able to effectively utilize the positional information within the modality, while strengthening the degree of association between aspectual words and graphic text, proving the effectiveness of the model.

3.3. Empirical Research on Mental Health Assessment of Students in Higher Education Institutions

After verifying the generalizability of the model based on the experimental results of the public dataset, this section further combines the mental health assessment needs of college students with an empirical study of social network opinion data to explore the correlation between psychological characteristics and online sentiment trends.

3.3.1. Analysis of Sentiment Trends in Social Network Public Opinion. The experimental dataset comes from the media information crawled from microblogs, and gets monitored on different media, using the method of topic extraction and feature screening, and realizes the topic crawling analysis of the risk prediction of network public opinion through data cleaning and Chinese lexicon. The F1 test analysis method is used to analyze the trend of a certain network public opinion sentiment on the social network about college students in the dataset, and this study is carried out on the information of blog posts and comments in the social network respectively, and the trend of network public opinion sentiment of blog posts and comments is shown in Figure 6.

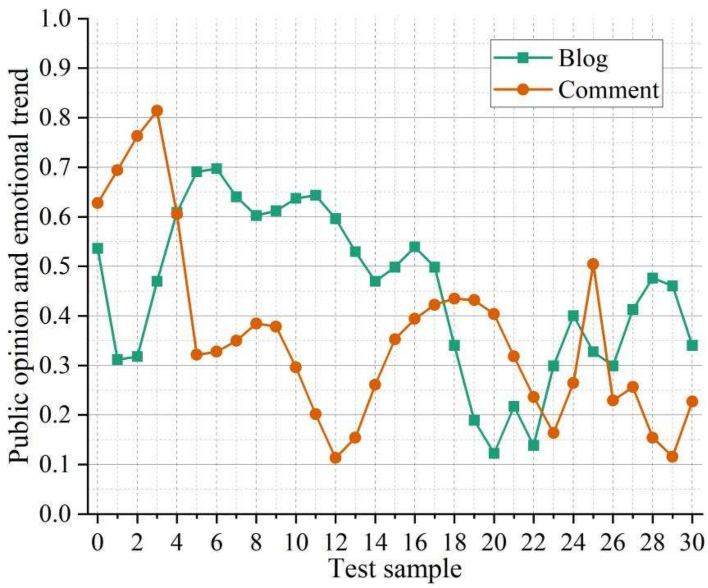

Fig. 6. Online sentiment trends of blog posts and comments

Analyzing Figure 6, we learned that through the analysis of the evolution of public opinion risk, the overall sentiment value is low, and the average sentiment value is between 0.2 and 0.4. By analyzing the public opinion information in the time period of the peak and trough of the sentiment value.

3.3.2. Sentiment Analysis of Public Emergencies. Using Python to call the SnowNLP library, the public event of "Fang Datong's death" from March 1 to March 10, 2025 was used as the network public opinion sentiment analysis. The sentiment value of the public opinion text at each stage is calculated and divided into three sentiment categories: "positive", "neutral" and "negative", and the number of texts corresponding to the three emotions in each stage is counted, and the "emotion-stage" change curve is drawn as shown in Figure 7.

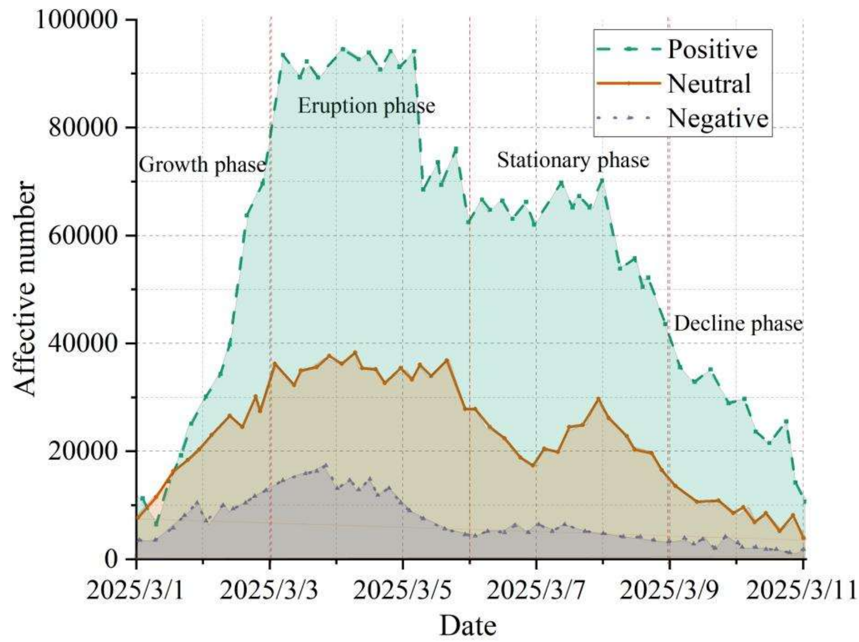

Fig. 7. Emotion category curves at different stages

In the periods of growth, outbreak, stabilization and decline of public opinion, positive sentiments predominate, accounting for 57.81%, 72.56%, 76.88% and 80.06% of the number of sentiments in each period, respectively. As time passes, negative emotions decline rapidly, and the proportion of neutral comments (objective discussion) and positive comments (remembrance and homage) increases, indicating that emotions are gradually calming down and shifting to positive reflection. This phenomenon suggests that mental health education needs to pay attention to the stage-by-stage intervention of crisis events, channeling negative emotions at an early stage and guiding positive value transfer at a later stage.

4. Application and Validation of Mental Health Assessment System for College Students

Chapter 3 verifies the validity of the model on the public dataset through multimodal sentiment analysis experiments. In order to further apply the model to practical scenarios, this chapter combines the specific needs of mental health assessment of college students with systematic applied research and empirical analysis to verify the performance and practical value of the model in real scenarios.

4.1. Correlation analysis of the dimensions of the scales

Validating the intrinsic correlation between the dimensions of the psychological scales provides a basis for the screening of the model input features and ensures the validity of the feature fusion supported by the subsequent experiments based on psychological theories. A total of three assessment scales were used in this chapter, with a total of six dimensions, namely, internal and external tendency, sensory intuition, rationality and sensibility, judgmental perception, depression and anxiety, and the correlation coefficient matrix heat map between the dimensions is shown in Figure 8.

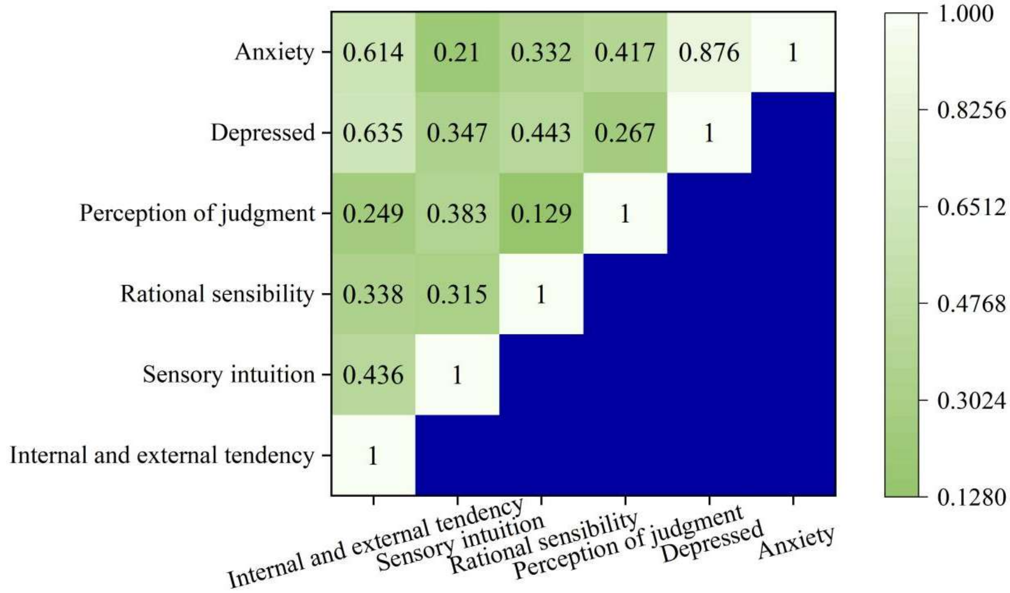

Fig. 8. Heat Map of each Dimension Correlation Coefficient Matrix

From the correlation coefficient matrix heat map of each scale dimension, it can be found that there is a strong correlation between internal and external tendency, depression and anxiety, respectively 0.635, 0.614 and 0.876, all of which reach more than 0.5, and the degree of correlation between depression and anxiety is higher, close to about 0.8. Therefore, for this topic, when constructing the early warning model of students' psychological state, in addition to the common concern of depression symptoms, it is also necessary to study the internal and external tendency dimension of their personality and the dimension of anxiety symptoms.

4.2. Dynamic Monitoring and Trend Analysis of Mental Health of Higher Education Students

On the basis of completing the correlation analysis of the dimensions of each scale, this section further takes students of a university as the research object, and analyzes the time-series change trend of mental health status in conjunction with the dynamic monitoring data to reveal the correlation between psychological characteristics and external events.

Based on the multimodal fusion model proposed in Chapter 3, the time-series changes in students' mental health status are tracked in real time, and the impact of external events (e.g., examination cycle) on psychological indicators is analyzed to verify the sensitivity and practicality of the model in dynamic monitoring. The mental health assessment system for college students based on social sentiment analysis and multi-branch neural network constructed in the article is applied in practice. Taking students of a university as the object of applied research, the system can help students assess their mental health status, and its persistent use can help students identify potential mental health problems.

Figure 9 shows the trend of total scores and factors over time for the period 2024.2.25-2024.7.1 for students in this university.

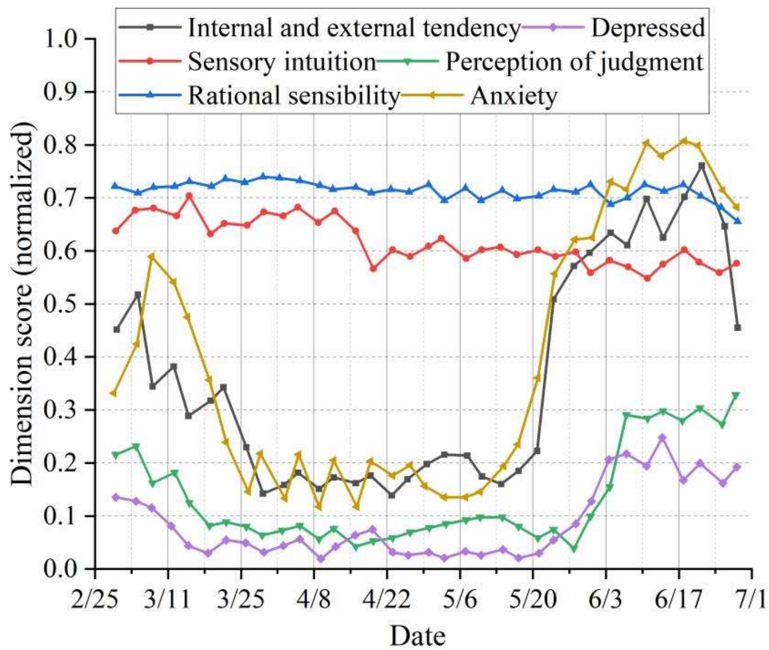

Fig. 9. The variation trend of each factor with time

Based on the mental health monitoring data of a university from February 25 to July 1, 2024, this study launched a dynamic analysis of six psychological dimensions (internal and external tendencies, feeling intuition, rational sensibility, judgment perception, depression, anxiety). As shown in Figure 9, the scores of each dimension showed differential fluctuations within the standardized interval of 0.1-1.0, with the black curve for internal and external tendencies, the mean value of which was 0.48±0.07, the red curve for sensory intuition, the mean value of which was 0.53±0.05, and the purple curve for the level of anxiety at a medium-high level, while the yellow curve for depression and the green curve for judgment perception scores were relatively low, and the blue curve for rationality and sensibility was significantly affected by the academic cycle. Sensibility is significantly affected by the academic cycle, showing regular oscillations.

In terms of time series, the system captured multidimensional linkage fluctuations from mid-March to late April, especially the depression index peaked at 0.81 on May 6), while the anxiety level showed a bimodal phenomenon with a peak of 0.79 on March 25th and June 17th. It is worth noting that the rational-perceptual dimension is highly correlated with the examination cycle, with two peaks of 0.69 on April 8 and 0.68 on May 20, both of which correspond to the midterm and final examination phases in the school calendar. The judgmental-perceptual dimension, on the other hand, exhibits negative fluctuation characteristics, with scores plummeting from 0.73 to 0.45 after the June 3 exam week.

4.3. Comparative experiments and performance validation based on multimodal models

In order to verify the comprehensive performance of the models in practical applications, in this section, the cutting-edge models in the field of multimodal sentiment analysis are selected to conduct comparative experiments with MIMN-BERT, EF-NET, IFNRA, and BICCM, and to analyze in depth the advantages and disadvantages of each model in mental health assessment tasks from the indexes of accuracy rate, F1 value, and so on.

4.3.1. Experimental program. The strong association rule feature data obtained from feature association analysis is selected for experimentation, 70% of the data is selected for training the model, 30% of the data is used for testing the model, and the experimental evaluation indexes are weighted average, and the weight is the proportion of the number of samples of each category in the total samples. In the improved particle swarm optimization algorithm the number of particles is set to 5, the maximum number of iterations of the particle swarm max_iter is set to 50, the maximum value of the inertia weight is 0.8, the minimum value is 0.3, and the variation rate of the particles is 0.6, and the cognitive coefficient c1 and the social coefficient c2 are set to be random numbers in the interval [1.38212, 2], and the speed range of the particles is set as $-3 \leqslant v \leqslant 3$, the range of particle's position is set as $-1 \leqslant p \leqslant 1$, and the initial values of particle's velocity and position are random numbers in the interval [0,1]. In order to avoid the impact of inconsistent data ranges of different feature data on the model performance, it is necessary to standardize the feature data before the experiment, and the maximum-minimum standardization method is selected to standardize the feature data, so that the range of each feature data falls within the [0,1] interval. The maximum-minimum standardization method is

$$d_i^* = \frac{d_i - \min(D)}{\max(D) - \min(D)} \tag{18}$$

Where di is the feature data, D is the feature data column, d_i^* is the normalized feature data. After the preliminary experiments, it was found that the PDNN neural network model containing two hidden layers and the number of nodes in the hidden layers were all 5 worked best, and the highest accuracy and F1 score were obtained. Therefore, the number of hidden layers of the model in the experiment is set to two layers, and the number of nodes in each layer is set to 5, where the activation function of the input layer and hidden layer adopts the tanh function, the activation function of the output layer adopts the sigmoid function, and the loss function of the model adopts the cross-entropy loss function.

4.3.2. Comparison of Experimental Results and Analysis of Model Strengths. In order to avoid the dependence of the model on dataset division, different test and training sets are chosen for each experiment. In the specific process, the dataset will be divided into 10 portions evenly in the order from front to back, assuming that each portion is labeled as 1, 2, 3, ..., 9, 10, respectively, then the 1st experiment will select 1 to 7 portions as the training set, and the rest as the test set. The 2nd experiment will pick 2 to 8 copies as the training set and the rest as the test set. And so on until the 10th division is taken and then the average of each evaluation metric is taken. The study compares the experimental results of the four best performing models in the multimodal sentiment analysis experiments MIMN-BERT, EF-NET, IFNRA and BICCM on each evaluation metric as shown in Figure 10.

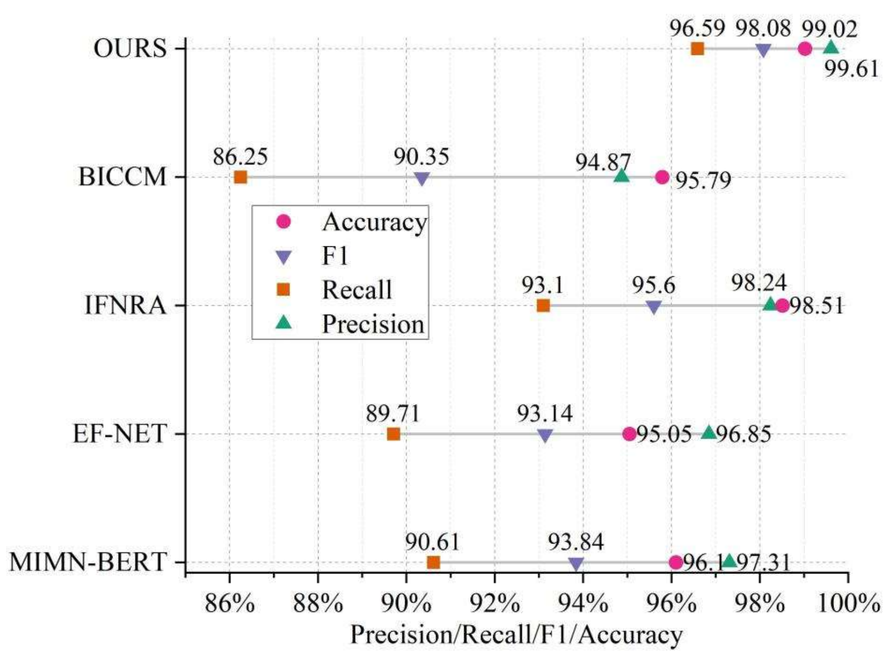

Fig. 10. Comparison of model experiment results

As can be seen from Fig. 10, the model of this paper is still optimal compared to the neural network models of other intelligent algorithms, and achieves optimal results in four evaluation indexes, including Precision, Recall, F1-score, and Accuracy, with mean values of 99.61%, 96.59%, 98.08%, and 99.02%, respectively, which are higher than those of the other four models. It shows that the model of this paper can predict students' mental health status more effectively according to their daily behaviors, which is of practical significance to help schools find students with psychological problems in time and carry out relevant interventions and guidance to make them return to normal life and study.

5. Conclusion

In this paper, we propose a multimodal fusion and heterogeneous graph neural network-based mental health assessment framework for college students, which effectively integrates social opinion, psychological scales and behavioral data through multi-branch CNN feature extraction, BiGRU cross-modal interaction, and virtual node-driven global graph modeling.

In the experiments on public datasets, the accuracy of the model on Twitter-15 and Twitter-17 is improved by 3.28% and 1.11%, and the F1 value is improved by 3.24% and 2.32%, respectively, compared with the existing optimal model, BICCM, which verifies the effectiveness of the ViT global feature extraction and the semantic association of meta-paths.

Practical application shows that the model can accurately identify key psychological risk indicators: depression index shows significant fluctuation in the middle and late semester, with a peak value of 0.81 on May 6, anxiety level shows a bimodal feature by unexpected events, with a peak value of 0.79 on March 25 and June 17, and rational-emotional dimensions are strongly correlated with the examination cycle.

In the comparison experiments, the average performance of the model under ten-fold cross-validation with 99.61% accuracy and 98.08% F1 value is significantly better than the models such as 97.31% for MIMN-BERT and 96.85% for EF-NET, which proves the adaptability of the three-branch architecture (CNN, Random Forest, and Fully-connected Networks) to the heterogeneous feature classification.

Acknowledgements

This work was supported by the 2023 National Social Science Fund Special Project on Ideological and Political Education in Higher Education, titled "Enhancing the Effectiveness of 'Situation and Policy' Course Teaching by Skillfully Utilizing the 'Grand Ideological and Political Course' Approach" (23VSZ062); 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] Onuoha, C. U., & Idemudia, E. S. (2020). Academic demands and mental health among undergraduate students in full-time employment: the moderating role of perceived social support. Journal of Psychology in Africa, 30(2), 89-95.

[2] Peltz, J. S., Bodenlos, J. S., Kingery, J. N., & Rogge, R. D. (2021). The role of financial strain in college students' work hours, sleep, and mental health. Journal of American college health, 69(6), 577-584.

[3] Campbell, F., Blank, L., Cantrell, A., Baxter, S., Blackmore, C., Dixon, J., & Goyder, E. (2022). Factors that influence mental health of university and college students in the UK: a systematic review. BMC public health, 22(1), 1778.

[4] Zhang, X. (2022). Problems and countermeasures of college students' mental health education. Journal of healthcare engineering, 2022(1), 6430855.

[5] Liu, Q. (2025). 56 effect of observation and experimental intervention of psychological education on college students with mental disorders. Schizophrenia Bulletin, 51(Supplement_1), S30-S30.

[6] Sontag-Padilla, L., Dunbar, M. S., Ye, F., Kase, C., Fein, R., Abelson, S., ... & Stein, B. D. (2018). Strengthening college students' mental health knowledge, awareness, and helping behaviors: the impact of active minds, a peer mental health organization. Journal of the American Academy of Child & Adolescent Psychiatry, 57(7), 500-507.

[7] Nepal, S., Liu, W., Pillai, A., Wang, W., Vojdanovski, V., Huckins, J. F., ... & Campbell, A. T. (2024). Capturing the college experience: a four-year mobile sensing study of mental health, resilience and behavior of college students during the pandemic. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, 8(1), 1-37.

[8] Pace, K., Silk, K., Nazione, S., Fournier, L., & Collins-Eaglin, J. (2018). Promoting mental health help-seeking behavior among first-year college students. Health communication, 33(2), 102-110.

[9] Bantjes, J., Hunt, X., & Stein, D. J. (2022). Public health approaches to promoting university students' mental health: a global perspective. Current psychiatry reports, 24(12), 809-818.

[10] Grégoire, S., Lachance, L., Bouffard, T., & Dionne, F. (2018). The use of acceptance and commitment therapy to promote mental health and school engagement in university students: A multisite randomized controlled trial. Behavior therapy, 49(3), 360-372.

[11] Wang, T. (2022). Research on fuzzy comprehensive evaluation index system of mental health education for college students. Journal of Healthcare Engineering, 2022(1), 7106926.

[12] 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.

[13] Cornish, P. A., Berry, G., Benton, S., Barros-Gomes, P., Johnson, D., Ginsburg, R., ... & Romano, V. (2017). Meeting the mental health needs of today's college student: Reinventing services through Stepped Care 2.0. Psychological services, 14(4), 428.

[14] Li, C., Yang, M., Zhang, Y., & Lai, K. W. (2022). An intelligent mental health identification method for college students: A mixed-method study. International Journal of Environmental Research and Public Health, 19(22), 14976.

[15] Boukhechba, M., Daros, A. R., Fua, K., Chow, P. I., Teachman, B. A., & Barnes, L. E. (2018). DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. Smart Health, 9, 192-203.

[16] Jung, H., Park, H. A., & Song, T. M. (2017). Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals. Journal of medical internet research, 19(7), e259.

[17] Moeller, R. W., Seehuus, M., & Peisch, V. (2020). Emotional intelligence, belongingness, and mental health in college students. Frontiers in psychology, 11, 499794.

[18] Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Romine, W., Sheth, A., Monadjemi, A. H., ... & Hitzler, P. (2020). Multimodal mental health analysis in social media. Plos one, 15(4), e0226248.

[19] Makwana, B., & Gawas, T. (2025). Research on Sentimental Analysis on Mental Health Using Social Media. International Journal of Scientific Research and Technology, 3(02).

[20] Saha, K., Yousuf, A., Boyd, R. L., Pennebaker, J. W., & De Choudhury, M. (2022). Social media discussions predict mental health consultations on college campuses. Scientific reports, 12(1), 123.

[21] Lu, Z. (2022). Analysis model of college students' mental health based on online community topic mining and emotion analysis in novel coronavirus epidemic situation. Frontiers in Public Health, 10, 1000313.

[22] Benrouba, F., & Boudour, R. (2023). Emotional sentiment analysis of social media content for mental health safety. Social Network Analysis and Mining, 13(1), 17.

[23] Li, P., & Liang, F. (2022). An assessment and analysis model of psychological health of college students based on convolutional neural networks. Computational Intelligence and Neuroscience, 2022(1), 7586918.

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

[25] Pei, J. (2022). Prediction and analysis of contemporary college students’ mental health based on neural network. Computational Intelligence and Neuroscience, 2022(1), 7284197.