

# Development of a Multimodal Hierarchical Early Warning System for College Students' Mental Health Based on Federated Learning and Its Application in Ideological and Political Intervention

Shangfei Lu

School of Education, Guangxi Vocational Normal University, Nanning, 530007, Guangxi, China  
lushangfei583@163.com

**Abstract:** As mental health issues in colleges and universities become increasingly complex and hidden, traditional methods of mental health monitoring that rely on questionnaires and manual interviews have significant shortcomings in terms of data coverage, dynamic identification capabilities, and privacy protection, making it difficult to achieve continuous perception and accurate early warning of college students' mental state. To this end, this paper proposes a multimodal hierarchical early warning system for the mental health of college students in Guangxi based on federated learning, and explores its collaborative application mechanism in ideological and political intervention. A multimodal psychological feature modeling method is constructed, integrating learning behavior data, campus life behavior data, online interaction data, and psychological assessment data. Through feature representation learning and a multimodal fusion model, psychological characteristics such as students' emotional fluctuations, behavioral abnormalities, and learning pressure are extracted. An intelligent matching mechanism for ideological and political resources based on student psychological profiles is constructed to realize the transformation of psychological early warning results into personalized ideological and political intervention strategies. Experimental results show that the multimodal fusion model achieves an accuracy of 0.913 and an F1-score of 0.894 in the mental health identification task, which is about 10%–14% higher than the single-behavioral data model. In terms of collaborative training performance, the federated learning model achieves an accuracy of 0.912, reduces the single-round communication volume to 7.3 MB, and shortens the total training time to 20.1 min, balancing model performance with data privacy protection requirements.

**Keywords:** College Student Mental Health Early Warning; Ideological and Political Intervention in Universities; Multimodal Feature Modeling; Federated Learning; Intelligent Risk Classification Model

## 1. Introduction

Traditional mental health management models in universities mainly rely on psychological assessment questionnaires, counselor observation, and individual interviews to identify psychological risks. However, these methods typically have limitations such as single data sources, long monitoring cycles, and insufficient dynamic perception capabilities, making it difficult to continuously track and provide timely early warnings of students' mental states. At the same time, with the continuous development of the information technology environment in universities, students have generated a large amount of behavioral data on learning platforms, campus life systems, and online interactive platforms. These data to some extent reflect students' learning pressure, emotional changes, and social status, providing a new data foundation for identifying mental health risks. Therefore, how to integrate multi-source behavioral data to achieve intelligent recognition of college students' psychological states has become an important research direction for the governance of mental health in universities.

Based on the above background, this article takes universities in Guangxi as the research scenario, constructs a multimodal grading warning system for college students' mental health based on federated learning, and explores the application mechanism of mental health warning results in ideological and political intervention, realizing the construction of students' psychological profiles, intelligent matching of ideological and political resources, and optimization of intervention feedback. By conducting systematic experiments to verify the performance of the model in identifying psychological risks and enhancing the efficiency of collaborative training, a feasible technical implementation path is provided

for monitoring mental health and promoting ideological and political education in universities.

## 2. Related Works

In recent years, with the rapid development of artificial intelligence and big data technology, researchers have begun to explore the use of machine learning and deep learning methods for early identification and intervention of mental health risks in adolescents and college students, in order to achieve more accurate and actionable mental health management. Zhang proposed a hybrid model that integrates Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) network to address mental health issues such as depression and anxiety in adolescents, achieving early risk warning. The model is trained on electronic health records and brain imaging data of over 50000 adolescents, combined with spatial feature extraction and time series analysis, achieving a prediction accuracy of 95% and an AUC of 97% [1]. Perochon et al. developed and evaluated a digital screening application for autism, which analyzed children's behavioral characteristics through computer vision and machine learning, and tested 475 children aged 17-36 months. The results showed that the algorithm had high diagnostic accuracy (AUC=0.90), sensitivity of 87.8%, and specificity of 80.8% [2]. Kirkbride et al. explored the impact of social determinants on mental health, emphasizing intergenerational health disparities and risks for vulnerable groups caused by structural inequality. Based on high-quality evidence, they proposed a prevention framework and intervention strategies, including universal, selective, and targeted primary prevention, as well as secondary and tertiary interventions to promote rehabilitation [3]. O'Connor et al. reviewed the global issue of premature death caused by comorbidities of suicide and mental illness, and proposed strategies to translate research evidence into priority actions. Through international expert review, 12 key risk factors and mechanisms have been identified. They also proposed 18 actionable solutions, covering principles such as integrating psychological and physical health, prioritizing prevention, and coordinating intervention [4]. Uhlhaas pointed out that the age range of 12 to 25 is a high-risk period for mental illness, emphasizing the role of risk and protective factors, as well as brain development, in the formation of adolescent psychopathology. They proposed the need to develop new methods for early diagnosis and intervention, combined with new service models and scientific research practice exchanges, to construct a transformative early intervention paradigm, in order to improve the mental health level of adolescents and promote the prevention of severe mental disorders [5]. Garg reviewed research on social media based mental health analysis, with a focus on exploring methods for detecting stress, depression, and suicide risk using machine learning and deep learning models. He constructed a classification system for mental health research, sorted out feature extraction, model progress, public datasets, and future research directions, and summarized 92 relevant literature [6]. However, existing research generally suffers from issues such as data silos, privacy protection challenges, and insufficient model generalization ability, which limits the application of mental health prediction and intervention methods across institutions and real-world scenarios.

## 3. Methods

### 3.1 Multimodal Feature Modeling of Psychological Health among College Students in Guangxi

#### 3.1.1 Multi-source behavior data collection and structured processing

In the process of constructing a psychological health warning system for college students in Guangxi, multimodal data is an important source of information for characterizing students' psychological states. The system collects multi-source behavior data through the campus information platform, learning management system, psychological assessment system, and campus network behavior platform, mainly including learning behavior data ( $D^{learn}$ ), campus life behavior data ( $D^{life}$ ), network interaction behavior data ( $D^{net}$ ), and psychological assessment data ( $D^{psy}$ ). Integrating data from different sources into a multimodal dataset through a unified data interface:

$$D=D^{learn},D^{life},D^{net},D^{psy} \quad (1)$$

To improve data quality, it is necessary to preprocess the raw data. Firstly, identify abnormal data points through anomaly detection methods, such as using standard deviation method for outlier filtering:

$$x_i' = \begin{cases} x_i, & |x_i - \mu| \leq 3\sigma \\ \mu, & |x_i - \mu| > 3\sigma \end{cases} \quad (2)$$

$(x_i)$  is the raw data,  $(\mu)$  and  $(\sigma)$  are the sample mean and standard deviation, respectively. Subsequently, normalize the data of different dimensions to unify them into the  $([0,1])$  interval:

$$x_i^{\text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

After preprocessing, a unified student behavior feature vector is formed:

$$X_i = [x_1, x_2, \dots, x_n] \quad (4)$$

$(X_i)$  represents the multimodal behavioral feature vector of the  $(i)$  th student, providing an input basis for subsequent mental health risk identification models.

### 3.1.2 Extraction and Representation Learning of Mental Health Risk Characteristics

In this study, multimodal behavioral data was mapped to a unified psychological feature space, and the transformation from behavioral data to psychological state features was achieved through embedding functions  $(f(\cdot))$ :

$$z_i = f(X_i; \theta) \quad (5)$$

$(X_i)$  is the student behavior feature vector,  $(\theta)$  is the model parameter, and  $(z_i)$  is the representation of student psychological characteristics.

In specific implementation, multi-layer neural networks can be used for feature extraction. If the network has  $(L)$  layers, then the output of the  $(l)$  th layer is:

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}) \quad (6)$$

$(h^{(l)})$  is the hidden vector of the  $(l)$  th layer,  $(W^{(l)})$  is the weight matrix,  $(b^{(l)})$  is the bias vector, and  $(\sigma(\cdot))$  is the activation function. By mapping layer by layer, potential psychological risk features such as changes in learning stress, abnormal social behavior, and emotional fluctuations can be gradually extracted.

The final representation vector of students' psychological characteristics is formed:

$$Z_i = [z_{i1}, z_{i2}, \dots, z_{ik}] \quad (7)$$

$(Z_i)$  represents the characteristic vector of students' psychological state, which can comprehensively reflect the correlation between learning behavior, social behavior, and emotional state, providing important basis for subsequent mental health risk assessment.

### 3.1.3 Construction of multimodal feature fusion and psychological state recognition model

If the learning behavior characteristics, life behavior characteristics, and network behavior characteristics are respectively  $(Z^{\text{learn}})$ ,  $(Z^{\text{life}})$ , and  $(Z^{\text{net}})$ , then the multimodal fusion features can be represented as:

$$Z^{\text{fusion}} = W_1 Z^{\text{learn}} + W_2 Z^{\text{life}} + W_3 Z^{\text{net}} + W_4 Z^{\text{psy}} \quad (8)$$

$(W_1, W_2, W_3, W_4)$  are weight coefficients for different modalities, used to reflect the importance of different data sources in identifying mental health.

Construct a psychological state recognition model based on fused features, and output the probability of student psychological risk through a classification function  $(g(\cdot))$ :

$$P(y_i = 1 | Z_i^{\text{fusion}}) = \sigma(W Z_i^{\text{fusion}} + b) \quad (9)$$

$(y_i)$  represents whether students have mental health risks, and  $(\sigma(\cdot))$  is the Sigmoid function.

To further achieve the classification of psychological risk levels, multi-level risk thresholds can be set:

$$R_i = \begin{cases} 0, & P_i < \tau_1 \\ 1, & \tau_1 \leq P_i < \tau_2 \\ 2, & P_i \geq \tau_2 \end{cases} \quad (10)$$

$(R_i)$  represents the level of psychological risk, corresponding to normal state, attention state, and

high-risk state respectively.

### 3.2 Construction of a Psychological Health Grading Early Warning Model Based on Federated Learning

#### 3.2.1 Design of Distributed Collaborative Training Mechanism for Federated Learning

Assuming there are (K) universities participating in model training within the Guangxi region, the local dataset of the (k) th university is represented as  $(D_k)$ , with a scale of  $(n_k)$ .

Under the federated learning framework, universities only train mental health recognition model parameters locally without uploading raw data. Let the global model parameters be  $(w)$ , and the local model optimization objective function be:

$$\min_w F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} L(w; x_i, y_i) \quad (11)$$

$(L(\cdot))$  is the loss function,  $(x_i)$  is the multimodal behavior feature of students, and  $(y_i)$  is the psychological health label.

During each round of federated training, each university updates the model parameters locally

$$w_k^{t+1} = w^t - \eta \square F_k(w^t) \quad (12)$$

$(w^t)$  is the global model parameter for the (t) th round, and  $(\eta)$  is the learning rate.

Subsequently, each university only uploads model parameter updates to the central server for aggregation. Using the classic FedAvg parameter aggregation algorithm to construct a global mental health prediction model:

$$w^{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1} \quad (13)$$

$n = \sum_{k=1}^K n_k$  represents the total sample size of all universities.

#### 3.2.2 Mental Health Risk Grading Assessment Model

Let the multimodal psychological feature vector of students be  $Z_i = [z_{i1}, z_{i2}, \dots, z_{im}]$ . The psychological risk prediction model calculates the probability of student psychological risk  $P_i = g(Z_i; w)$  through the function  $(g(\cdot))$ . Among them,  $(w)$  is the global model parameter obtained by federated learning.

In practical applications, logistic regression or neural networks can be used to output the probability of student psychological risk:

$$P_i = \sigma(w^T Z_i + b) \quad (14)$$

$(\sigma(\cdot))$  is the Sigmoid function  $\sigma(x) = \frac{1}{1+e^{-x}}$ . After obtaining the probability of psychological risk, the classification of mental health levels is achieved by setting a risk threshold. The setting of thresholds  $(\tau_1)$  and  $(\tau_2)$  enables the construction of a three-level psychological risk assessment model, defined as:

$$R_i = \begin{cases} 0, & P_i < \tau_1 \\ 1, & \tau_1 \leq P_i < \tau_2 \\ 2, & P_i \geq \tau_2 \end{cases} \quad (15)$$

$(R_i=0)$  indicates a normal psychological state;  $(R_i=1)$  indicates that students need mild attention;  $(R_i=2)$  represents high-risk warning.

To improve the reliability of the evaluation, a multi-index comprehensive risk function can also be introduced:

$$S_i = \alpha_1 P_i + \alpha_2 B_i + \alpha_3 E_i \quad (16)$$

$(P_i)$  is the predicted probability of the model;  $(B_i)$  is an indicator of abnormal behavior;  $(E_i)$  is an indicator of emotional fluctuations;  $(\alpha_1, \alpha_2, \alpha_3)$  are weight coefficients. The stability and explanatory power of mental health risk assessment can be further improved through comprehensive scoring  $(S_i)$ , thereby providing reliable basis for mental health grading warning.

### 3.2.3 Dynamic warning and continuous monitoring mechanism

The psychological state of college students has obvious dynamic changes, so it is necessary to construct a continuous monitoring mechanism for psychological state based on time series data. If the psychological characteristic vector of students at time (t) is  $(Z'_t)$ , then the probability of psychological risk is  $P'_t=g(Z'_t;w)$ . This paper constructs student psychological risk change sequences  $P'_1, P'_2, \dots, P'_T$  through continuous time windows. To identify the trend of psychological risk changes, a sliding time window can be used to calculate the risk change  $\Delta P'_t=P'_t-P'_{t-1}$ . When the risk change rate exceeds the preset threshold ( $\gamma$ ), the system can trigger the abnormal psychological state detection  $\Delta P'_t>\gamma$ .

If the results of multiple consecutive tests exceed the threshold, it is determined to be a state of potential psychological risk increase. Simultaneously, an automatic warning mechanism is constructed by combining the graded risk thresholds ( $\tau_1, \tau_2$ ). When the probability of student psychological risk meets  $P'_t \geq \tau_2$ , the system automatically generates a high-risk warning signal and pushes it to the psychological counseling and ideological and political management platform; When  $\tau_1 \leq P'_t < \tau_2$ , the system marks it as a key focus object and enters continuous tracking and monitoring. By combining time series dynamic monitoring with automatic warning mechanisms, it is possible to continuously monitor the mental health status of college students in Guangxi, identify risk trends, and trigger graded warnings, thereby providing timely data support for subsequent ideological and political interventions and psychological counseling.

## 3.3 Application Mechanism of Ideological and Political Intervention Driven by Psychological Health Warning

### 3.3.1 Accurate identification and portrait construction of psychological risk groups

On the basis of the output risk results of the federated learning mental health warning model, it is necessary to further construct a psychological portrait model for college students to achieve accurate identification and characterization of psychological risk groups. Let the multimodal psychological feature vector of students be  $(Z_i=[z_{i1}, z_{i2}, \dots, z_{im}])$ , and the probability of psychological risk be  $(P_i)$ . By integrating feature clustering and risk level classification, student groups can be classified into different psychological state categories, and their classification functions can be expressed as  $(C_i=f(Z_i, P_i))$ . Among them,  $(C_i)$  represents the category of psychological risk group to which the student belongs. In practical implementation, clustering methods can be used for pattern recognition of student behavior characteristics. By minimizing the intra class distance function  $J=\sum_{k=1}^K \sum_{x_i \in C_k} |x_i - \mu_k|^2$ , psychological behavior patterns can be grouped. Among them,  $(C_k)$  is the (k) th student group, and  $(\mu_k)$  is the group center vector.

On this basis, this article constructs a student psychological portrait vector  $(U_i=[P_i, B_i, S_i, E_i])$ . Among them,  $(P_i)$  represents the probability of psychological risk,  $(B_i)$  represents the indicator of behavioral abnormalities,  $(S_i)$  represents the learning stress index, and  $(E_i)$  represents the indicator of emotional fluctuations. Through multidimensional psychological characterization, different psychological risk patterns such as learning stress, social isolation, and emotional fluctuations can be identified, providing accurate data basis for the formulation of subsequent ideological and political intervention strategies.

### 3.3.2 Intelligent matching of ideological and political resources and generation of intervention strategies

After identifying the psychological risk characteristics of students, it is necessary to construct an intelligent matching mechanism for ideological and political education resources, to achieve the coordinated application of psychological health intervention and ideological and political education resources. The ideological and political resource library is given by  $(R=r_1, r_2, \dots, r_n)$ . Among them,  $(r_j)$  represents different types of ideological and political education resources, such as psychological counseling courses, themed educational activities, and value guidance content.

The system calculates the matching degree between the student's psychological profile vector  $(U_i)$  and ideological and political resources, and the matching function can be expressed as  $M_{ij}=sim(U_i, r_j)$ . Among them,  $(sim(\cdot))$  represents the similarity function, such as cosine similarity:

$$\text{sim}(a,b)=\frac{a \cdot b}{|a||b|} \tag{17}$$

When the matching degree satisfies  $M_{ij} \geq \theta$ , the system can automatically recommend corresponding ideological and political intervention resources.

In addition, an intervention strategy generation mechanism can be constructed by combining rule models. For example, when the student's psychological risk level ( $R_i \geq 2$ ), the system prioritizes matching psychological counseling with individual conversation resources; When ( $R_i = 1$ ), priority is given to recommending mental health courses and themed educational activities. By combining intelligent matching with rule-based reasoning, personalized ideological and political intervention strategies can be formed, thereby improving the synergistic effect of mental health education and ideological and political education.

### 3.3.3 Closed loop feedback and continuous optimization mechanism for ideological and political intervention

In order to continuously improve the implementation effectiveness of psychological intervention and ideological and political education, it is necessary to establish a mechanism for feedback on intervention effects and strategy optimization. If the psychological risk probabilities of students before and after intervention are ( $P_i^{\text{before}}$ ) and ( $P_i^{\text{after}}$ ), respectively, then the intervention effect can be expressed by the risk change amount as follows:

$$\Delta P_i = P_i^{\text{before}} - P_i^{\text{after}} \quad (18)$$

When ( $\Delta P_i > 0$ ), it indicates that intervention has a positive effect on reducing psychological risk.

At the same time, a comprehensive intervention effect evaluation function can be constructed by combining the behavior change index ( $B_i$ ) and the emotional stability index ( $E_i$ ):

$$F_i = \alpha \Delta P_i + \beta \Delta B_i + \gamma \Delta E_i \quad (19)$$

( $\alpha, \beta, \gamma$ ) are weight coefficients.

The system continuously optimizes ideological and political intervention strategies based on the evaluation results of intervention effectiveness. When ( $F_i$ ) falls below the preset threshold ( $\lambda$ ), the system can readjust the resource matching scheme or upgrade the intervention level, thus forming a closed-loop management process of "psychological risk warning - ideological and political intervention implementation - effectiveness evaluation - strategy optimization".

## 4. Results and Discussion

### 4.1 Experimental Data and Operating Environment Configuration

The experimental data is sourced from campus information systems and psychological assessment platforms of multiple universities in Guangxi. This article constructs a multimodal dataset that includes learning behavior data, campus behavior data, network interaction data, and psychological scale data. To ensure the operability of the experiment, the data is distributed and divided by universities to simulate real federated learning training scenarios.

The experimental data scale is set as shown in Table 1:

*Table 1. Experimental Data*

Data Type	Number of Features	Sample Size
Learning Behavior Data	12	4200
Campus Life Data	10	4200
Online Interaction Data	8	4200
Psychological Assessment Data	6	4200

Each student sample contains 36 feature dimensions. The data is divided into 5 local data nodes based on university nodes, with approximately 840 samples per node.

The experimental environment adopts Python and PyTorch framework to implement the federated learning model. The experiment deployed a global model aggregation module on the server-side and local model training modules on each university node. The main parameter settings are shown in Table 2:

Table 2. Parameter Settings

Parameter	Value
Number of Federated Nodes	5
Local Training Rounds	5
Federated Communication Rounds	50
Learning Rate	0.001
Batch Size	32

#### 4.2 Performance Experiment

Based on the multimodal feature fusion model, a psychological state recognition task is constructed and compared with a single modal model. The experiment uses learning behavior data, psychological assessment data, and multimodal fusion data to train the model and compare the recognition performance under different feature combinations.

The evaluation indicators include accuracy, precision, recall, and F1 score.

To evaluate the collaborative training effectiveness of federated learning frameworks in cross university data scenarios, the federated learning model was compared with traditional centralized training models, and the following indicators were calculated: global model convergence rounds, model prediction accuracy, communication data volume per round, and total model training time. By analyzing the convergence speed and communication overhead of the model, the training efficiency and collaborative ability of federated learning in mental health prediction tasks can be evaluated.

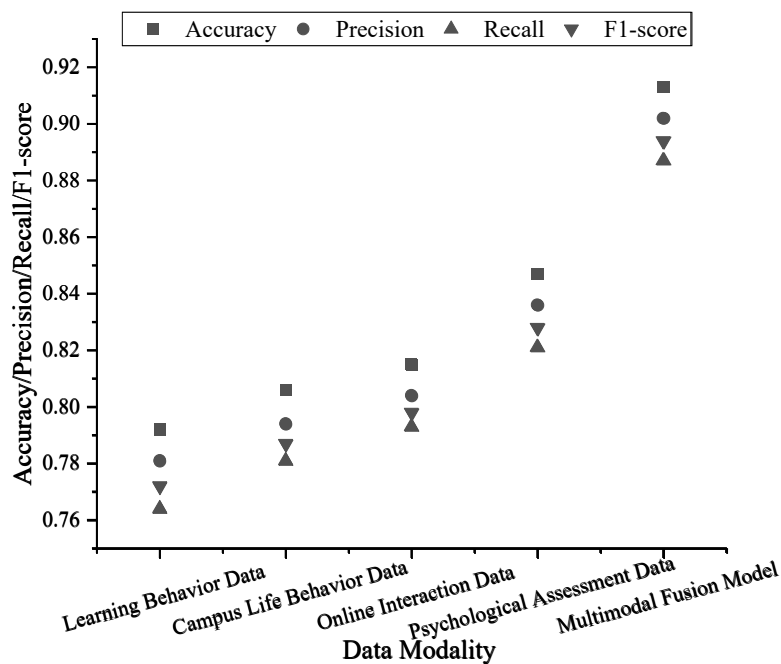


Figure 1. Performance comparison of mental health recognition models under different data modalities

From Figure 1, it can be seen that there are significant differences in the performance of different data modalities in the task of identifying the mental health of college students. When only using learning behavior data, the accuracy of the model is 0.792, and the F1 score is 0.772, indicating that although a single learning behavior feature can reflect a certain degree of learning pressure and behavioral abnormalities, its characterization of students' psychological state is still relatively limited. After further adopting the multimodal fusion model, the model performance was significantly improved, with accuracy increased to 0.913, precision reached 0.902, recall reached 0.887, and F1 score reached 0.894. Compared with a single data modality, the multimodal fusion model achieves a maximum accuracy improvement of about 12.1%, indicating that by integrating multiple sources of data such as learning behavior, campus life, online interaction, and psychological assessment, students' psychological state characteristics can be more comprehensively characterized.

Table 3. Comparison of Model Training Performance under Different Collaborative Training Modes

Training Mode	Model Accuracy	Convergence Rounds	Communication per Round (MB)	Total Training Time (min)
Single-School Local Training	0.842	31	0	18.6
Centralized Training	0.901	24	0	26.4
Federated Learning (FedAvg)	0.908	27	12.5	21.8
Federated Learning + Feature Compression	0.912	26	7.3	20.1

From Table 3, it can be seen that there are significant differences in model performance and training efficiency among different collaborative training modes. The single school local training mode, relying solely on data from a single university for model learning, has a model accuracy of 0.842, which is significantly lower than other collaborative training methods. At the same time, the convergence rounds reach 31, indicating that the model's generalization ability is weak in the case of limited data size. The federated learning (FedAvg) model achieves cross university collaborative training through parameter aggregation without sharing raw data. The model accuracy reaches 0.908, which is close to the centralized training effect. At the same time, the total training time is reduced to 21.8 minutes, demonstrating good collaborative modeling ability.

## 5. Conclusion

This article focuses on the needs of monitoring the mental health of college students and promoting ideological and political education. It proposes a multimodal grading warning system for the mental health of college students in Guangxi based on federated learning, and constructs an overall technical framework from multimodal psychological feature modeling, federated learning collaborative training to psychological warning driven ideological and political intervention. By integrating learning behavior data, campus life data, network interaction data, and psychological assessment data, a multimodal psychological feature representation model is constructed to achieve comprehensive characterization of students' psychological states. There are still certain limitations to this study, such as the relatively limited scale and source range of experimental data, which may have an impact on the model's generalization ability in larger scale university scenarios. Future research can further expand the scale of data from multiple universities and integrate more unstructured psychological data to continuously improve the intelligence level and application effectiveness of mental health warning systems.

## Acknowledgement

This work was supported by the 2025 Research Project on Ideological and Political Education Theory and Practice for College Students in Guangxi Universities (Grant No. 2025SZ071).

## References

- [1] Zhang Z. Early warning model of adolescent mental health based on big data and machine learning [J]. *Soft computing*, 2024, 28(1): 811-828.
- [2] Perochon S, Di Martino J M, Carpenter K L H, et al. Early detection of autism using digital behavioral phenotyping[J]. *Nature Medicine*, 2023, 29(10): 2489-2497.
- [3] Kirkbride J B, Anglin D M, Colman I, et al. The social determinants of mental health and disorder: evidence, prevention and recommendations[J]. *World psychiatry*, 2024, 23(1): 58-90.
- [4] O'Connor R C, Worthman C M, Abanga M, et al. Gone Too Soon: priorities for action to prevent premature mortality associated with mental illness and mental distress[J]. *The Lancet Psychiatry*, 2023, 10(6): 452-464.
- [5] Uhlaas P J, Davey C G, Mehta U M, et al. Towards a youth mental health paradigm: a perspective and roadmap[J]. *Molecular psychiatry*, 2023, 28(8): 3171-3181.
- [6] Garg M. Mental Health Analysis in Social Media Posts: A Survey: M. Garg[J]. *Archives of Computational Methods in Engineering*, 2023, 30(3): 1819-1842.