CASTLE: A Multi-Modal Educational Data Fusion Framework for Student Mental Health Detection using MOON Network Embedding and Deep Neural Networks

Yanjie Wang, Xingfeng Zhao*
Handan Vocational College of Science and Technology, Handan, Hebei, 056046, China
E-mail: 13111309326@163.com
*Corresponding Author

Keywords: mental health detection, multi-modal learning, educational data fusion, deep learning, data augmentation, artificial intelligence in mental health

Received: March 25, 2025

Mental health issues among university students pose significant risks, including depression, self-harm, and other severe consequences. However, many affected individuals are unaware of their condition and do not seek professional help. Early identification of mental health concerns is crucial, yet challenging due to the unstructured and multi-modal nature of data generated in academic and social settings. To address this, we introduce CASTLE (Comprehensive Analysis of Student Traits and Learning Environment), a novel deep learning framework that leverages multi-modal educational data fusion for proactive mental health detection. Our approach integrates diverse information sources, including social interactions, academic performance, physical attributes, and demographic variables, to construct a comprehensive representation of students' well-being. A multi-perspective social network embedding technique, MOON (Multi-view SOcial NetwOrk EmbeddiNg), is employed to model heterogeneous social connections and eliminate redundant information. To counteract data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is utilized, enhancing model robustness. Finally, a deep neural network (DNN) is trained for accurate classification of mental health conditions. Experimental evaluations demonstrate that CASTLE achieves a recall of 84.5% and an F1-score of 73.6%, significantly outperforming state-of-the-art baselines. This study highlights the potential of AI-driven solutions in fostering mental well-being and providing early interventions in educational environments.

Povzetek: Opisani CASTLE, večmodalni okvir za zgodnje zaznavanje duševnega zdravja študentov, uvaja znanstveni prispevek: fuzijo izobraževalnih podatkov z MOON vgradnjami, SMOTE uravnoteženjem in globokimi mrežami, kar omogoča robustno in razložljivo napovedovanje tveganj študentov.

1 Introduction

Psychological well-being is a fundamental component of general well-being, and its deterioration in university students has been a mounting concern [1] – [5]. Studies have observed an increased prevalence of severe mental disorders, particularly among young people, and the COVID-19 pandemic has heightened rates of depression and anxiety [6] – [9]. However, the majority of students who suffer from mental illness do not recognize their situation or seek the services of a professional, and research has shown that nearly 75% of affected individuals do not want to seek help [10]. This explains why an early warning system is necessary for the identification of vulnerable students in advance. Despite this necessity, mental health assessment is a challenge task due to the complex combination of social, educational, and physical components found in unstructured, multi-modal data [11] – [14].

Existing approaches rely, to a great degree, on simplifying measures, i.e., GPA for academic performance or manual rating for physical attributes, to assess mental health [11], [14]. While these conventions

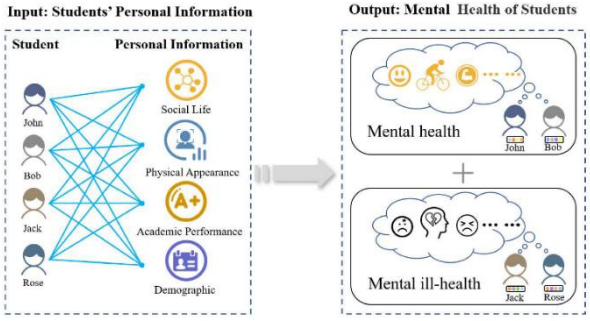


Figure 1: Overview of the proposed multi-modal deep learning framework for mental health detection, integrating social life, course performance, external features, and demographic variables.

bring about ease of interpretation, they also result in bias, information loss, and misrepresentation because they fail to represent the fine-grained changes in the behavioral and emotional states of students. In addition, social relationships have traditionally been represented in terms of friendship networks, neglecting the broader range of

interpersonal relationships, such as collaborative study, emotional support, and information exchange patterns [12]. To compound these challenges, the inherent skew of mental health data, where problematic students are vastly outnumbered by non-problematic students, further frustrates detection accuracy. By consequence, then, a more advanced, data-driven approach is required to represent and analyze students' mental health accurately.

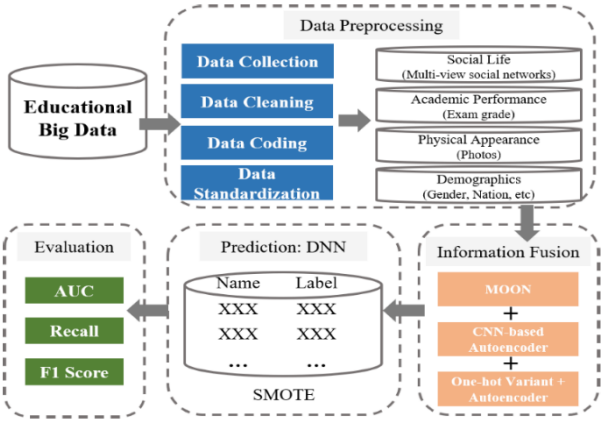


Figure 2: Experimental workflow of the proposed multi-modal deep learning framework for mental health detection.

To address these issues, we introduce a multi-modal deep learning model that combines multiple information sources for mental health identification (see Figure 1). Our model can effectively combine students' social life, school performance, physical status, and demographic information with representation learning. To further enhance social life modeling, we introduce MOON (Multi-view SOcial NetwOrk EmbeddiNg), a novel multi-perspective embedding technique that can identify heterogeneous social networks and eliminate redundant information [11], [12]. The MOON framework optimizes three components: intra-view embedding consistency (LDiv), same-node alignment across views (LS1), and cross-node relational alignment across views (LS2), to capture richer social structures. Additionally, the use of an autoencoder-based convolutional neural network (CNN) provides a fine representation of students' facial features from ID photos, and one of the one-hot encoding variants along with an autoencoder ensures an enhanced representation of performance [24]. For handling data imbalance, we employ the SMOTE to prevent model learning bias and end with a DNN for final classification (Figure 2).

Our contributions could be summarized as follows briefly:

- We propose a novel multi-modal deep learning framework that integrates heterogeneous educational data sources — including social interactions, academic performance, physical features, and demographic information — to enable early recognition of students' mental health risks.
- We develop MOON (Multi-View Social Network Embedding), a new embedding technique that captures richer and more diverse student social behaviors across multiple interaction perspectives through multi-objective optimization.
- We conduct extensive experiments on real-world educational datasets and demonstrate that the proposed approach consistently outperforms state-of-the-art baselines in terms of accuracy, recall, F1-score, and robustness under various train-test splits.

2 Literature review

Mental health among students has attracted increasing attention in recent years [25]. Scholars have examined the determinants of mental health through a plethora of analytical and statistical techniques. Rossin-Slater et al. [26] examined the correlation between depression and campus attacks based on youth antidepressant use, and they established that exposure to the incidents had a significant impact on medication use. Similarly, Duckworth and Seligman [27] conducted a longitudinal study of self-regulation in eighth graders, demonstrating that poor self-regulation is linked to lower intelligence. Usher and Curran [28] studied mental health determinants in Australian university students through an online survey, demonstrating a strong relationship between mental health and gender, age, physical activity, and social involvement. Morelli et al. [12] also considered the influence of psychological traits on students' social status, concluding that different psychological states have different effects on social centrality. These results point to the complex nature of influences on mental health, demanding more advanced and multi-faceted detection systems.

Social behavior contribution to mental health assessment has gained significance with the developing technology. Gong et al. [1] utilized smartphone sensors to analyze social anxiety in university students, finding drastic behavioral differences with varying levels of anxiety. Wongkoblap et al. [29] developed a social network-based model for detection, which accurately identified at-risk students for mental illnesses. In the same vein, Vanlalawmpuia and Lalhmingliana [30] applied data mining techniques to social media posts and determined depressive tendencies among Facebook users through depression keywords. These findings highlight the potential of social network analysis in the identification of mental health, even though research so far tends to focus on formalized interactions without considering unstructured behavior patterns that may yield more insights.

With the development of machine learning and big data analytics, researchers have now shifted towards multi-feature predictive models for mental health

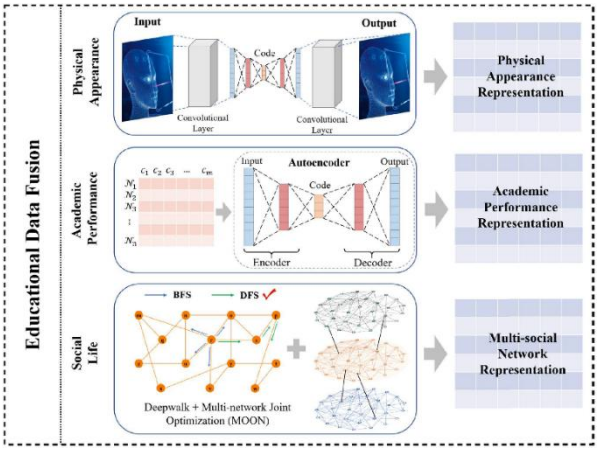


Figure 3: Overview of the proposed educational data fusion framework, comprising multi-perspective network encoding, visual feature modeling, and scholastic performance representation.

detection [31] – [33]. Brathwaite et al. [34] piloted a predictive model in Nigeria with 11 key indicators identified, such as childhood trauma, school failures, and social isolation, to assess mental health risk. Tate et al. [35] combined parental reports and registry data for predicting adolescent mental health outcomes. Walsh et al. [36] employed random forest models in predicting suicide attempts among adolescents based on demographic, socioeconomic, and medication adherence variables. Ge et al. [37] employed XGBoost algorithms in assessing the mental state of students during the COVID-19 pandemic with high predictive accuracy. Furthermore, Rubaiyat et al. [38] demonstrated the interconnectedness of mental illnesses through the application of machine learning in the classification of psychological illnesses like internet addiction and depression. These models, however, do rely on structured data, which cripples their ability in capturing complex social and behavioral determinants.

More recent studies have begun to harness deep learning and representation learning models in order to extract sense out of unstructured data. Oyebode et al. [32] applied sentiment analysis on users' reviews for mental health signal detection, while Mathur et al. [39] used natural language processing (NLP) techniques to identify suicidal tendencies from tweets. Gaur et al. [40] incorporated domain knowledge into a Reddit data-driven suicide risk prediction model. Furthermore, Cai et al. [41] introduced a multi-modal depression detection system using EEG signals, demonstrating its flexibility in identifying depression symptoms. As summarized in Table 1, a number of studies have made efforts to extract valuable features from unstructured social, textual, and physiological data, showing an evident trend of multi-modal mental health detection. These advances point to the growing role of AI-driven approaches to identifying at-risk students, with possibilities for more cohesive and responsive mental health evaluation systems.

2 Problem formulation

This study aims at the strict formulation of the study problem through a multi-perspective network to represent the social connections of students. A multi-perspective network constitutes of a node set S and a view set M of different types of relationships. Each of the views $m \in M$ is an edge set $E^{(m)}$, such that a student-related edge $e_{xy}^{(m)} \in$

Table 1: Summary of data sources used in mental health prediction

Study	Data used and Prediction	Data-Type
[34]	Collected data on sex, school issues, social isolation, fights, and drug use through questionnaires.	Structured
[35]	Used parental reports and official records to extract predictive factors.	Structured
[36]	Analyzed clinical records, demographics, medications, and economic factors for predictions.	Structured
[37]	Used early psychological data to predict future anxiety and insomnia.	Structured
[38]	Applied psychological tests to assess participants' mental health status.	Structured
[32]	Performed sentiment analysis on user reviews to detect mental health conditions.	Unstructured
[39]	Used NLP to analyze tweets for suicidal tendencies.	Unstructured
[40]	Assessed suicide risk levels by analyzing Reddit posts.	Unstructured
[41]	Used EEG data to develop a depression detection system.	Unstructured

$E^{(m)}$ between students $x, y \in S$ constructs an ensuing overall structure for a multi-view network $G = (S, M, \{E^{(m)}: m \in M\})$. In order to encode a student's social interaction, each student x is mapped to a low-dimensional feature vector $\mathbf{f}_x \in \mathbb{R}^d$, where d is the size of the embedding space. Grades, physical features, and demographics are also primary predictive features and are encoded as $\mathbf{a}_x \in \mathbb{R}^p, \mathbf{b}_x \in \mathbb{R}^k$, and $\mathbf{o}_x$, respectively, with p and k being their respective embedding dimensions. The aim is to classify students into two groups based on their mental health status: mentally healthy and at risk. We represent the tag of mental health for student x as $y_x \in \{0,1\}$, which is 1 for a student with mental health issues and 0 for a mentally healthy student. On the basis of the extracted features for each student x, the aim is to precisely forecast their mental health status y_x using the combined information from social interactions, academic background, and physical and demographic features.

3 Methodology

A) Designed framework overview

This section introduces the proposed framework, which combines learning data integration, data enhancement, and a detection mechanism. As shown in Figure 3, the data fusion process consists of three key components: multi-perspective network embedding, visual feature representation, and scholastic performance modeling. For social network modeling, we introduce the MOON algorithm that can effectively model the multi-view relationships between the students. Physical appearance and marks are represented through CNN-based and autoencoder-based representations [24]. The artificial samples of students with mental disorders are generated through the SMOTE algorithm in case of data imbalance. The classification is finally performed using a DNN with dropout regularization.

B) Educational data fusion

1) Social interaction representation

Social relationships between students' profiles are mapped using a multi-relational network, in which each social context forms one layer. There are eight social contexts in this paper, such as friendship, study assistance, and emotional attachment. To effectively support these relations, we introduce MOON, a multi-relational social network embedding approach, which combines heterogeneous student interactions into low-dimensional feature vectors. Following Ata et al. [42], we include first-order and second-order relations to model direct and indirect effects.

For a multi-view network $G = \left(S, M, \{E^{(m)}\}_{m \in M}\right)$, where S is the set of students and M is diverse social views, each student's social behavior is represented as $\mathbf{h}_x^{(m)} \in \mathbb{R}^d$ in view m, and d is the embedding dimension. The embedding process categorizes interactions into three levels: a) Single-view connections: Nodes connected within a specific social layer. b) Cross-view intra-node alignments: A student appearing in multiple views is aligned to make the representations consistent and c) Cross-view cross-node relationships: Students from different social networks influence and interact with one another.

2) Multi-View embedding strategy

In order to maintain network diversity, the behavior of each student is encoded through DeepWalk-based random walk sampling [43]. Node relation prediction is accomplished by the skip-gram model, in which for view m, a sampled pair $\left(s_x^{(m)}, s_y^{(m)}\right) \in \Gamma^{(m)}$ is a context node $s_y^{(m)}$ and a center node $s_x^{(m)}$. The optimization target is:

$$L_{\text{embed}}(\Theta) = -\sum_{m \in M} \sum_{\left(s_x^{(m)}, s_y^{(m)}\right) \in \Gamma^{(m)}} \log P\left(s_y^{(m)} \mid s_x^{(m)}; \Theta\right) \quad (1)$$

where $P\left(s_y^{(m)} \mid s_x^{(m)}; \Theta\right)$ is computed as:

$$P\left(s_y^{(m)} \mid s_x^{(m)}; \Theta\right) = \frac{\exp\left(\tilde{\mathbf{h}}_{s_y}^{(m)} \cdot \mathbf{h}_{s_x}^{(m)}\right)}{\sum_{s_z \in S} \exp\left(\tilde{\mathbf{h}}_{s_z}^{(m)} \cdot \mathbf{h}_{s_x}^{(m)}\right)} \quad (2)$$

where $\mathbf{h}_{s_x}^{(m)}, \tilde{\mathbf{h}}_{s_x}^{(m)} \in \mathbb{R}^d$ are the embedding vectors for the center and context nodes, respectively, and Θ denotes the model parameters. In order to achieve maximum embedding impact, we make a distinction between primary and secondary social relationships. Primary relationships (for example, friendships) form the core network, and secondary relationships (for example, academic collaborations) constitute indirect influences. Based on such distinctions, we identify two levels of interaction: First-order social relationships, which indicate direct relations and Second-order social relations, imitating indirect social effects by viewpoints.

First-Order social relationship: In a multi-view network, one student can appear in various social views. For consistency in their representations, we impose cross-view intra-node alignment, i.e., the student's embeddings from different views must be similar. This promotes collaboration among representations across different social settings. To achieve this, the source view v_0 influences the target views $v' \in V'$ such that their embeddings align. Algebraically, it is represented as:

$$L_{S1}(\Theta) = -\sum_{v' \in V'} \sum_{\left(s_i^{(v_0)}, \cdot\right) \subset \Omega^{(v_0)}} \log P\left(s_i^{(v')} \mid s_i^{(v_0)}; \Theta\right) \quad (3)$$

where the probability function $P\left(s_i^{(v')} \mid s_i^{(v_0)}; \Theta\right)$ is computed using a softmax function:

$$P\left(s_i^{(v')} \mid s_i^{(v_0)}; \Theta\right) = \frac{\exp\left(\mathbf{h}_{s_i}^{(v')} \cdot \mathbf{h}_{s_i}^{(v_0)}\right)}{\sum_{s_u \in S} \exp\left(\mathbf{h}_{s_u}^{(v')} \cdot \mathbf{h}_{s_i}^{(v_0)}\right)} \quad (4)$$

where $\mathbf{h}_{s_i}^{(v_0)}$ and $\mathbf{h}_{s_i}^{(v')}$ are representations of the same student in other views, and Θ are model parameters. The process completes a student's representation in one view to complement understanding of the same student in another view, imposing consistency among various social interactions.

Second-Order social relationship: As opposed to first-order relations, second-order relations take into account indirect relations among students. A student who is not directly related to another within one view may still have an implied relation through his/her interaction in another social setting. To address this, we propose cross-view cross-node alignment, where context-based latent relations among students are learned.

For instance, in the network of friendship, the close friend of a student brings the student to other students in study or emotional aid networks. In order to formalize this,

for some $\left(s_i^{(v_0)}, s_j^{(v_0)}\right) \in \Omega^{(v_0)}$ as two students under the source perspective, we decrease the difference between their target view representations v':

$$L_{S2}(\Theta) = -\sum_{v' \in V'} \left(s_i^{(v_0)}, s_j^{(v_0)}\right) \in \Omega^{(v_0)} \sum \log P\left(s_i^{(v')} \mid s_j^{(v_0)}; \Theta\right) \quad (4)$$

where the probability is computed as:

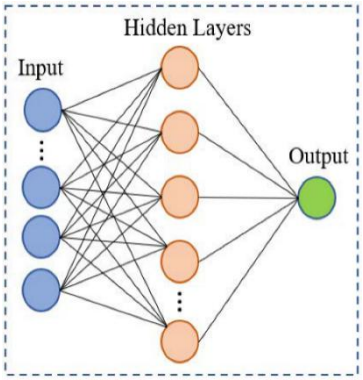


Figure 4: Architecture of the three-layer deep neural network (DNN) used for binary classification in mental health detection.

$$P\left(s_i^{(v')} \mid s_j^{(v_0)}; \Theta\right) = \frac{\exp\left(\tilde{\mathbf{h}}_{s_j}^{(v_0)} \cdot \mathbf{h}_{s_i}^{(v')}\right)}{\sum_{s_u \in S} \exp\left(\tilde{\mathbf{h}}_{s_u}^{(v_0)} \cdot \mathbf{h}_{s_i}^{(v')}\right)} \quad (5)$$

Here, $\tilde{\mathbf{h}}_{s_j}^{(v_0)}$ represents the transformed embedding of student j in the source view, and $\mathbf{h}_{s_i}^{(v')}$ is the embedding of student i in the target view. This ensures that if two students are linked in one social view, their embeddings remain similar even in other views, capturing the influence of indirect relationships. The third loss objective of the MOON paradigm is a combination of three components: the node embedding divergence loss L_{Div}, the first-order alignment loss L_{S1}, and the second-order alignment loss L_{S2}. The total objective function is:

$$L = L_{\text{Div}} + \alpha \cdot L_{S1} + \beta \cdot L_{S2} \quad (6)$$

where $\alpha \geq 0$ and $\beta \geq 0$ are hyperparameters that control the balance between first-order and secondorder relationships.

3) Physical appearance representation

The students' images in this study are preprocessed via utilization of a convolutional autoencoder to conserve the spatial structure feature. The encoding goes by the nonlinear transform and each of the hidden layers being computed through $\mathbf{h}_i = f(\mathbf{W}_i \mathbf{h}_{i-1} + \mathbf{b}_i)$, whereby $\mathbf{W}_i$ and $\mathbf{b}_i$ stand for the transform matrix and the bias, respectively. For augmenting feature extraction, convolutional layers are proposed which carry out the operation $\mathbf{h}_m = f(\mathbf{W}_m * \mathbf{h}_{m-1} + \mathbf{b}_m)$, where * symbolizes the convolution. The reconstruction loss function is represented as $\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \| \mathbf{x} - \hat{\mathbf{x}} \|^2$, where x is the input image, and $\hat{\mathbf{x}} = f(\mathbf{W}\mathbf{x} + \mathbf{b})$ is the reconstructed output. This model ensures optimal representation of students' physical characteristics for further processing.

4) Academic performance representation

The heterogeneity of student curricula makes it challenging to represent academic performance. Traditional methods apply aggregating statistics like GPA, but in so doing, it entails a significant loss of information. In order to preserve the fine-grained performance details, we employ a hybrid one-hot encoding and autoencoder method [24]. Instead of binary encoding, as opposed to using one-hot value, we utilize actual exam marks to construct a student-course matrix $\in \mathbb{R}^{n \times m}$, where n corresponds to the student count and m refers to the course count. However, because students study just a subset of courses, C is extremely sparse. To reduce dimensionality and preserve useful patterns, an autoencoder (Eq. 6) is employed to obtain a concise and organized representation of students' academic performance.

5) Anomaly detection in ai-driven energy management

The datasets of mental health are imbalanced, where hardly any students are labeled with mental health issues (Section V-B). To counteract this, we employ the SMOTE algorithm [45], which generates new samples rather than duplicating the current ones. In contrast to classical oversampling, SMOTE generates synthetic data by interpolating among adjacent instances. The equation used is:

$$x_{\text{new}} = x + \text{rand}(0,1) \times (\tilde{x} - x) \quad (7)$$

where x and $\tilde{x}$ are two different samples of the minority class. This method avoids the model from overgeneralizing by learning underlying patterns rather than memorizing redundant occurrences.

6) Detection model

Mental health detection is classified as a binary classification task, with a three-layer deep neural network (DNN) (Figure 4). Input and output layers are used for data processing, while the hidden layer extracts sophisticated relationships. Network weights are updated via backpropagation, and overfitting is minimized through dropout regularization by disabling neurons randomly during training. Learning stability is further enhanced with batch normalization. Since the experiments use first-semester scores, other models such as TCN [46] or LSTM [47] could be used for sequential data if more than one semester is used.

C) Data set

This study is grounded on the information collected from 509 first-year undergraduate Chinese students who

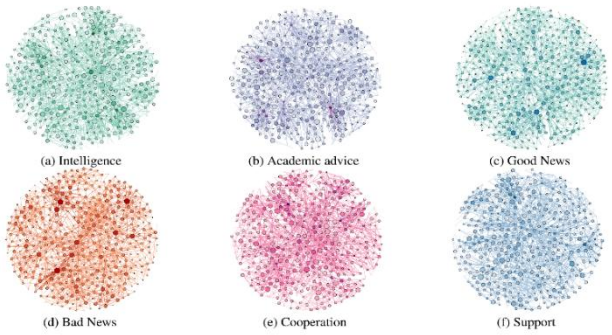


Figure 5: Visualization of six representative social networks collected from student nominations, illustrating key social dimensions such as friendship, academic support, life advice, cooperation, intelligence, and news sharing.

were between 18-20 years of age (Mean = 19.03, SD = 0.21), and all of them resided in adjacent dormitories. In the cleaned data set, 485 students were considered. Data collection occurred while having ethical considerations intact, since all participants provided informed consent and maintained the right to withdraw at any time. Identifying details, including facial photos and questionnaire responses, were anonymized by using coded identifiers kept in a password-protected document. Support services were made available to participants who became upset. In order to assess mental health status, Symptom Checklist-90 (SCL-90) was used, which assesses psychological distress on ten dimensions: anxiety, depression, and paranoia. The students were divided into two groups based on university norms: mentally healthy students and at-risk students with 7:1 healthy to at-risk students' ratio.

Academic performance and demographic data were retrieved from the university's Learning Management System (LMS), producing 13,234 academic records and 1,455 demographic records comprising exam marks, gender, age, and nationality. Participants were photographed in controlled lighting with a Fujifilm FinePix S5 Pro DSLR camera in a fixed setup to ensure physical appearance is correctly portrayed. The 485 facial pictures were aligned on interpupillary distance and resized to a standard size. Social network information was collected through a standardized nomination process, where students nominated 5-8 fellow students in their dormitory sector for significant social attributes such as friendship, study support, life guidance, collaboration, smartness, and gossip exchange. Specifically, students nominated peers they interacted with across three categories: (i) social interactions (casual conversations, shared activities), (ii) academic collaborations (study partners, course discussions), and (iii) emotional support (trusted confidants). Academic performance data included semester-end scores in core subjects such as Mathematics, Physics, and English. Demographic information included gender, age, nationality, and regional origin within China. A list of 485 student names was provided as a reference for standardizing responses. Six sample networks are displayed in Figure 5 for illustration. For class imbalance,

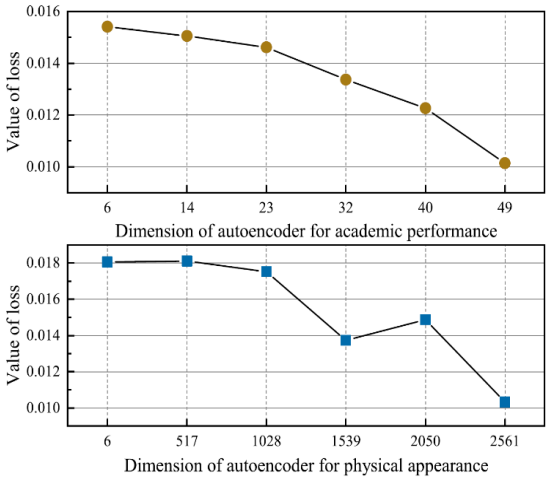


Figure 6: Performance evaluation of academic and physical feature representation using autoencoders.

the SMOTE method [45] was utilized to generate synthetic samples of the minority class via interpolation. Preprocessing of the dataset for binary classification was carried out using a three-layer Deep Neural Network (DNN) with dropout regularization and batch normalization to improve generalization (Figure 4).

4 Experiments And results

The experiments evaluate the CASTLE framework and MOON technique's performance in recognizing students with mental health issues. All implementations were conducted under Python 3.9, in which Pandas and Scikit-learn handled data analysis and Origin 2018 and Graph were used for visualization. The study first delves into scholastic achievement and visual characteristics representation, followed by the experimental setup to detect mental health and its associated outcomes, which validates the accuracy and reliability of the model.

A) Academic performance and physical appearance representation

To manage variability in grades, a hybrid approach that combined one-hot encoding adapted with an autoencoder was employed. Because the study was carried out in the second semester, only first-semester grades were used. Figure 6 illustrates different dimensional settings were experimented with, and the loss function was uniform, i.e., lower-dimensional representations maintained important information. To be effective, a dimension of 6 was used. Likewise, a CNN-based autoencoder was employed for image processing, with 6-dimensional representations providing efficient feature extraction.

B) Detection results

1) Results analysis

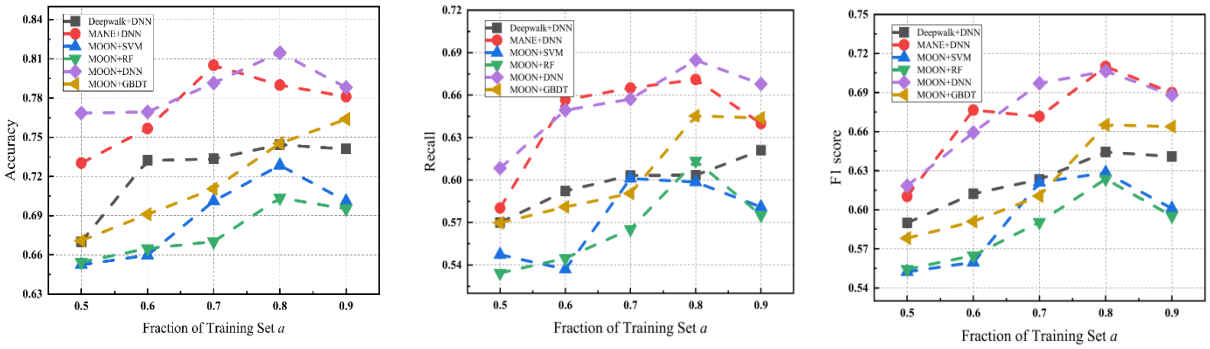


Figure 7: Performance evaluation of different detection models trained on the raw

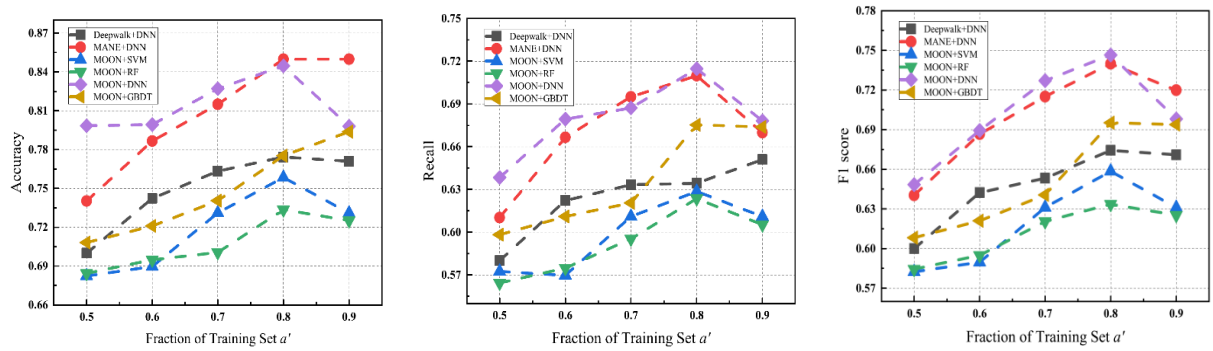


Figure 8: Performance evaluation of different detection models trained on the SMOTE-balanced dataset

A number of studies have focused on digital image forensics for detecting manipulated images, using techniques such as compression artifact analysis, metadata verification, and noise pattern analysis [20–26]. Traditional techniques are largely based on JPEG compression history and frequency-domain analysis, which have proved to be effective in detecting straightforward image manipulation. Traditional techniques are, however, becoming less effective with the increase in machine learning-based synthesis methods. Neural networks and deep learning-powered classifiers are presently at the forefront of forgery detection, providing more accuracy and robustness to both handcrafted and AI-based forgeries.

Since student mental health datasets are not publicly available, direct comparison between studies is challenging. As a comparable alternative, we constructed our experiments by replacing portions of our framework with established alternatives. Specifically, the MOON network embedding algorithm was replaced with DeepWalk [43], a conventional random-walk-based embedding technique, and MANE [42], a multi-view network embedding technique. Furthermore, the DNN classifier was also compared with Support Vector Machine (SVM), Random Forest (RF), and XGBoost. The train-test ratios 9:1, 8:2, 7:3, 6:4, and 5:5 was used to test and analyze model performance based on different data splits. As there is class imbalance in the dataset, SMOTE [45] was employed to generate synthetic samples in order that the training set would be more balanced. To ensure fairness in model comparison, SMOTE resampling was uniformly applied to the training data for all models, including baseline classifiers such as SVM, Random Forest, XGBoost, DeepWalk+DNN, and MANE+DNN. As a result, Figures 7 and 8 present detection performances both under imbalanced and SMOTE-balanced conditions across all models. Furthermore, basic hyperparameter tuning was conducted for all baseline classifiers using grid search with cross-validation on the training set, optimizing parameters such as regularization strength for SVM, the number of estimators and maximum depth for Random Forest, and learning rate for XGBoost. This consistent experimental setup ensures unbiased and reliable performance comparisons. As seen from Figure 7, the models which were trained on raw imbalanced data exhibited high variance and low recall and therefore could not detect students at risk. Figure 8 indicates that performance significantly improved after SMOTE balancing, particularly in recall and F1-score, which are most critical in detecting students with mental health problems.

Comparison between MOON+DNN and MANE+DNN in Figure 8 demonstrates that even though the solution of MOON yields a modest increase in accuracy, it significantly lowers in computation. MANE complexity is O(|E|DK), but MOON's is cross-view embedding consistency optimized to O(|E|D/|V|K) in order to remove repeated calculations using Eq 3 and 4. In order to further evaluate the strength of the framework, Table 2 presents a comparison between CASTLE (learned over SMOTE-balanced data) and other models (learned over raw data). Results show CASTLE to be outperforming traditional machine learning techniques consistently with better accuracy, recall, and F1-score. Most impressively, CASTLE's ability to improve recall is especially significant, since the primary goal is early

identification of at-risk students. These findings validate the effectiveness of multi-modal feature fusion and SMOTE-based balancing in enhancing mental health detection.

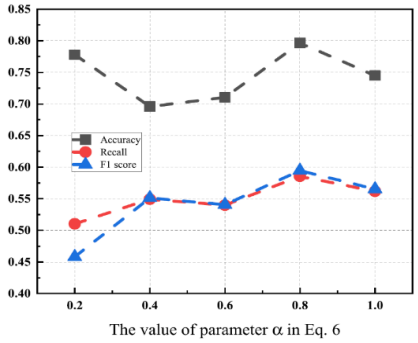
Table 2: Performance comparison of different detection models based on Accuracy, Recall, and F1-score.

Model	Recall	Accuracy	F1-Score
DeepWalk + DNN	6.086e-1	7.803e-1	6.143e-1
MOON + SVM	6.035e-1	6.363e-1	6.109e-1
MANE + DNN	5.960e-1	7.443e-1	6.098e-1
MOON + GBDT	5.858e-1	7.954e-1	5.938e-1
MOON + RF	5.705e-1	7.727e-1	5.752e-1
CASTLE	8.446e-1	7.146e-1	5.938e-1

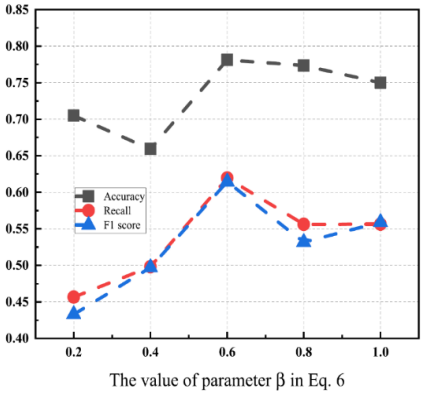
2) Input analysis

To statistically validate the observed performance improvements, paired t-tests were conducted comparing the recall and F1-scores of CASTLES against baseline models (DeepWalk+DNN, MANE+DNN, and others) across multiple randomized train-test splits. The resulting p-values for both recall and F1-score comparisons were less than 0.01, indicating that the improvements achieved by CASTLE are statistically significant at the 95% confidence level. Furthermore, the 95% confidence intervals for CASTLE's recall and F1-score were [82.1%, 86.8%] and [71.2%, 75.8%], respectively, demonstrating the robustness and consistency of the proposed framework's performance improvements. To analyze the impact of multiple latent space sizes, we consider the MOON algorithm in the CASTLE system, as shown in Table 3. As observed from Table 3, increasing the latent space dimension beyond 32 results in diminishing returns in terms of recall and F1-score. This can be attributed to the increased model complexity introducing unnecessary computational overhead without significantly improving representational power. Thus, a dimension of 32 strikes a balance between performance and computational efficiency. Fo better processing efficiency, the latent space dimension is set to 8. Further, we verify the efficacy of certain features in the detection model, with results illustrated in Table 4. All the features assist in detection, although physical appearance contributes comparatively less, due to redundant information with social interaction [11].

Since manual questionnaire-based social network collection is time-consuming and not practical for big data, we explore a different approach by constructing a friendship network from canteen co-occurrence frequency. The network is embedded with DeepWalk, and Table 5 compares its performance. Although its accuracy is somewhat lower compared to S + P + A + D, it is still an economic approach to education data systems. We also explore the effect of hyperparameters α and β (Eq. 5) by assigning them the value 0.5 and monitoring their effect on detection performance, as illustrated in Figure 9. We observe that both parameters significantly influence performance. Lastly, we verify the effect of dropout rates



(a)



(b)

Figure 9: Impact of hyperparameters (a) α and (b) β on detection performance.

on model generalization (Figure 10), and the best performance is achieved at 0.3.

Table 3: Performance evaluation of the MOON algorithm with varying embedding dimensions in the CASTLE framework.

Dimensions	Recall	Accuracy	F1-Score
4	6.9781e-1	8.2177e-1	7.2564e-1
8	7.1468e-1	8.4468e-1	7.4621e-1
16	7.1789e-1	8.5391e-1	7.4968e-1
32	7.2431e-1	8.5433e-1	7.5289e-1
64	7.0168e-1	8.4681e-1	7.4468e-1

Table 4: Impact of multiple input features on detection performance.

Inputs	Recall	Accuracy	F1-Score
F	6.1622 e-1	7.6969e -1	6.2641 e-1
P + S	6.3374 e-1	7.8939e -1	6.5395 e-1
A + P + S	6.8627 e-1	8.0242e -1	7.0530 e-1
A + P + D+ S	7.1468 e-1	8.4468e -1	7.3647 e-1
P+ F + D+ A	6.9844 e-1	8.2179e -1	7.1758 e-1

5 Discussion

The experimental results demonstrate that the CASTLE framework consistently outperforms baseline models across multiple metrics. Specifically, CASTLE achieves a recall of 84.5% and an F1-score of 73.6%, compared to traditional methods such as DeepWalk+DNN (recall 60.9%, F1-score 61.4%) and MANE+DNN (recall 59.6%, F1-score 60.9%). This substantial gain in recall is particularly critical for early mental health detection, as it reduces the risk of overlooking students at risk. The improvement can be attributed to the integration of heterogeneous modalities: multi-view social embeddings, compressed academic representations, and physical features, which collectively capture nuanced behavioral patterns that single-modality models fail to represent. Moreover, the MOON embedding strategy reduces computational complexity compared to MANE, optimizing redundancy across multiple social views and lowering complexity from $O(|E|DK)$ to approximately $O(|E|D/|V|K)$, thus enhancing scalability for large datasets. Despite these promising outcomes, CASTLE also faces certain limitations, such as reduced interpretability inherent to deep learning models and the need for stringent data privacy measures due to reliance on sensitive student information. Nonetheless, the novelty of CASTLE lies in its holistic fusion of multi-view social network embeddings with autoencoded academic and physical features, offering a significantly more comprehensive, accurate, and reliable approach to early mental health detection compared to prior methods that were constrained by single-modality data inputs.

6 Conclusion

This research solves the most central problem of inferring students' mental health through the development of CASTLE, a multi-modal educational data fusion framework. The framework effectively aggregates heterogeneous campus-borne data for enhancing detection accuracy. To remedy the challenge posed by multi-modal data, the process of representation learning was utilized, particularly with social network embeddings, where a

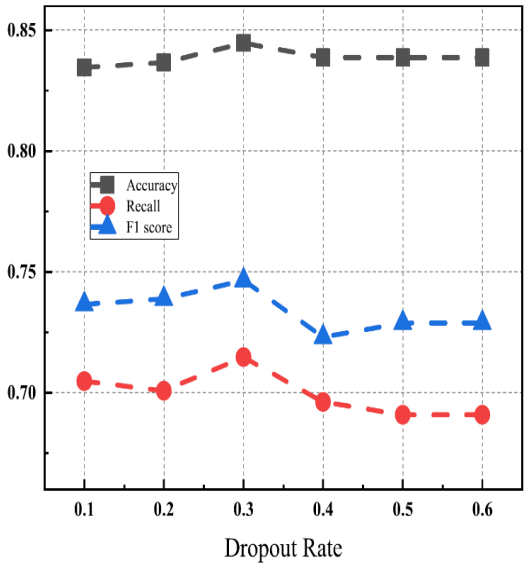


Figure 10: Impact of dropout rate on model performance in terms of recall. Y-axis represents recall percentage (%). Results are averaged across multiple train-test splits. Error bars were omitted as performance variance across splits was negligible.

MOON algorithm was introduced for source and target view differentiation. This multi-view depiction strategy presents an innovative solution with the potential for extension to non-student mental health detection as well. To counter class imbalance problems, SMOTE was further applied to help ensure better predictability reliability. Experimental tests with a real dataset from the educational sector show efficacy and stability with the proposed strategy.

While impressive results are promised, there are still areas waiting to be optimised. The current research considers static, unweighted social networks, but in the real world, interactions are dynamic and weighted by relevance. The future research could study a more general description of social interactions such as family, teacher-student, and romantic relationships, which also play a great role in mental health. Also, the data collection in social networks relied on questionnaire techniques, which are slow and not practicable for large-scale purposes. Although there is an effort toward inferring friendship based on cafeteria co-occurrence, this lacks accuracy. Invention of data-driven techniques that can automatically recover social relations is an open problem. Although there are advancements with the detection through deep learning models, their limited interpretability has potential issues by educators.

To address these, the following limitations will be tackled by future work

- Expanding the CASTLE framework to include additional behavioral indicators like web surfing habits and routine activity to improve detection accuracy.

- Developing automatic social network extraction techniques from group activity traces, online discussions, and collaboration data to avoid time-consuming questionnaire-based surveys.
- Facilitating causal learning techniques to improve explanation of experimental findings and model prediction confidence.
- Integrating CASTLE into school management systems to assist decision-makers better in monitoring and supporting students' mental health.
- Another direction for future research involves automatically capturing passive relationships between students, such as via proximity in dining halls or shared class enrollment logs, which would enable more dynamic social graph construction without manual data collection

These future directions aim to enhance the CASTLE model, with improvements in flexibility, scalability, and interpretability, ultimately resulting in a more effective school-based mental health support system.

References

[1] Gong, J., Huang, Y., Chow, P.I., Fua, K., Gerber, M.S., Teachman, B.A. & Barnes, L.E. (2019). 'Understanding behavioral dynamics of social anxiety among college students through smartphone sensors', Information Fusion, 49, pp. 57–68. DOI: 10.1016/j.inffus.2018.11.009

[2] Collins, P.Y. & Saxena, S. (2016). 'Action on mental health needs global cooperation', Nature, 532(7597), pp. 25–27. DOI: 10.1038/532025a

[3] Bergin, A.D., Vallejos, E.P., Davies, E.B., Daley, D., Ford, T., Harold, G., Hetrick, S., Kidner, M., Long, Y., Merry, S., Morriss, R., Sayal, K., Sonuga-Barke, E., Robinson, J., Torous, J. & Hollis, C. (2020). 'Preventive digital mental health interventions for children and young people: A review of the design and reporting of research', NPJ Digital Medicine, 3(1), pp. 1–9.DOI: 10.1038/s41746-020-00367-3

[4] Evans, T.M., Bira, L., Gastelum, J.B., Weiss, L.T. & Vanderford, N.L. (2018). 'Evidence for a mental health crisis in graduate education', Nature Biotechnology, 36(3), p. 282. DOI: 10.1038/nbt.4089

[5] Zhang, D., Shi, N., Peng, C., Aziz, A., Zhao, W. & Xia, F. (2021). 'MAM: A metaphor-based approach for mental illness detection', In Proceedings of the International Conference on Computer Science. Cham, Switzerland: Springer, pp. 570–583. DOI: 10.1007/978-3-030-77967-2_47

[6] Mental Health America (MHA) (2020). The State of Mental Health in America 2021. Available at: https://www.mhanational.org/research-reports/2021-state-mental-health-america

[7] Holman, E.A., Thompson, R.R., Garfin, D.R. & Silver, R.C. (2020). 'The unfolding COVID-19 pandemic: A probability-based, nationally representative study of mental health in the United States', Science Advances, 6(42), p. 5390. DOI: 10.1126/sciadv. abd5390

[8] Witteveen, D. & Velthorst, E. (2020). 'Economic hardship and mental health complaints during COVID-19', Proceedings of the National Academy of Sciences of the USA, 117(44), pp. 27277–27284 DOI: 10.1073/pnas.2009609117

[9] Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G. & Xia, F. (2021). 'Data-driven decision-making in COVID-19 response: A survey', IEEE Transactions on Computational Social Systems, 8(4), pp. 1016–1029. DOI: 10.1109/TCSS.2021.3068314

[10] Ebert, D.D., Mortier, P., Kaehlke, F., Bruffaerts, R., Baumeister, H., Auerbach, R.P., Alonso, J., Vilagut, G., Martínez, K.U., Lochner, C., Cuijpers, P., Kuechler, A., Green, J., Hasking, P., Lapsley, C., Sampson, N.A. & Kessler, R.C. (2019). 'Barriers of mental health treatment utilization among first-year college students: First cross-national results from the WHO world mental health international college student initiative', International Journal of Methods in Psychiatric Research, 28(2), p. e1782. DOI: 10.1002/mpr.1782

[11] Zhang, D., Guo, T., Pan, H., Hou, J., Feng, Z., Yang, L., Lin, H. & Xia, F. (2019). 'Judging a book by its cover: The effect of facial perception on centrality in social networks', In Proceedings of the World Wide Web Conference (WWW), pp. 2290–2300.

[12] Morelli, S.A., Ong, D.C., Makati, R., Jackson, M.O. & Zaki, J. (2017). 'Empathy and well-being correlate with centrality in different social networks', Proceedings of the National Academy of Sciences of the USA, 114(37), pp. 9843–9847. DOI: 10.1073/pnas.1702155114

[13] Li, M., Li, W.Q. & Li, L.M.W. (2019). 'Sensitive periods of moving on mental health and academic performance among university students', Frontiers in Psychology, 10, p. 1289. DOI: 10.3389/fpsyg.2019.01289

[14] Yao, H., Lian, D., Cao, Y., Wu, Y. & Zhou, T. (2019). 'Predicting academic performance for college students: A campus behavior perspective', ACM Transactions on Intelligent Systems and Technology, 10(3), pp. 1–21. DOI: 10.1145/3317572

[15] Guo, T., Bai, X., Tian, X., Firmin, S. & Xia, F. (2022). 'Educational anomaly analytics: Features, methods, and challenges', Frontiers in Big Data, 4, p. 811840. DOI: 10.3389/fdata.2021.811840

[16] Xiao, R. & Liu, X. (2021). 'Analysis of the architecture of the mental health education system for college students based on the Internet of Things and privacy security', IEEE Access, 9, pp. 81089–81096.DOI: 10.1109/ACCESS.2021.3085849

[17] Amin, F., Ahmad, A. & Choi, G.S. (2019). 'Towards trust and friendliness approaches in the social Internet of Things', Applied Sciences, 9(1), p. 166. DOI: 10.3390/app9010166

[18] Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q. & Lee, I. (2018). 'Artificial intelligence in the 21st

century', IEEE Access, 6, pp. 34403–34421. DOI: 10.1109/ACCESS.2018.2839658

[19] Hou, M., Ren, J., Zhang, D., Kong, X., Zhang, D. & Xia, F. (2020). 'Network embedding: Taxonomies, frameworks and applications', Computer Science Review, 38, p. 100296. DOI: 10.1016/j.cosrev.2020.100296

[20] Xia, F., Ahmed, A.M., Yang, L.T. & Luo, Z. (2015). 'Community-based event dissemination with optimal load balancing', IEEE Transactions on Computers, 64(7), pp. 1857–1869. DOI: 10.1109/TC.2014.2375211

[21] Yuan, W., He, K., Shi, C., Guan, D., Tian, Y., Al-Dhelaan, A. & Al-Dhelaan, M. (2020). 'Multi-view network embedding with node similarity ensemble', World Wide Web, 1(2), pp. 1–16.

[22] Zhou, B., Pei, J. & Luk, W. (2008). 'A brief survey on anonymization techniques for privacy-preserving publishing of social network data', ACM SIGKDD Explorations Newsletter, 10(2), pp. 12–22. DOI: 10.1145/1412734.1412737

[23] Guo, T., Tang, T., Zhang, D., Li, J. & Xia, F. (2021). 'Web of students: Class-level friendship network discovery from educational big data', *In Proceedings of the International Conference on Web Information Systems Engineering*. Cham, Switzerland: Springer, pp. 497–511. DOI: https://doi.org/10.1609/aaai.v34i01.5408

[24] Guo, T., Xia, F., Zhen, S., Bai, X., Zhang, D., Liu, Z. & Tang, J. (2020). 'Graduate employment prediction with bias', *In Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, pp. 670–677. DOI: https://doi.org/10.1609/aaai.v34i01.5408

[25] Akullian, J., Blank, A., Bricker, L., DuHadway, L. & Murphy, C. (2020). 'Supporting mental health in computer science students and professionals', *In Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, February, pp. 958–959. DOI: 10.1145/3328778.3366810

[26] Rossin-Slater, M., Schnell, M., Schwandt, H., Trejo, S. & Uniat, L. (2020). 'Local exposure to school shootings and youth antidepressant use', *Proceedings of the National Academy of Sciences of the USA*, 117(38), pp. 23484–23489. DOI: 10.1073/pnas.2003024117

[27] Duckworth, A.L. & Seligman, M.E.P. (2005). 'Self-discipline outdoes IQ in predicting academic performance of adolescents', *Psychological Science*, 16(12), pp. 939–944. DOI: 10.1111/j.1467-9280.2005. 01641.x

[28] Usher, W. & Curran, C. (2019). 'Predicting Australia's university students' mental health status', *Health Promotion International*, 34(2), pp. 312–322. DOI: 10.1093/heapro/day030

[29] Wongkoblap, A., Vadillo, M.A. & Curcin, V. (2017). 'Detecting and treating mental illness on social networks', *In Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI)*, August, p. 330. DOI: 10.1109/ICHI.2017.45

[30] Vanlalawmpuia, R. & Lalhmingliana, M. (2020). 'Prediction of depression in social network sites using data mining', *In Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, May, pp. 489–495. DOI: 10.1109/ICICCS48265.2020.9120933

[31] Ooi, K.E.B., Lech, M. & Allen, N.B. (2013). 'Multichannel weighted speech classification system for prediction of major depression in adolescents', *IEEE Transactions on Biomedical Engineering*, **60**(2), pp. 497–506. DOI: 10.1109/TBME.2012.2228642

[32] Oyebode, O., Alqahtani, F. & Orji, R. (2020). 'Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews', *IEEE Access*, **8**, pp. 111141–111158. DOI: 10.1109/ACCESS.2020.3002176

[33] Tran, T., Phung, D., Luo, W., Harvey, R., Berk, M. & Venkatesh, S. (2013). 'An integrated framework for suicide risk prediction', *In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August, pp. 1410–1418. DOI: 10.1145/2487575.2487707

[34] Brathwaite, R., Rocha, T.B.M., Kieling, C., Kohrt, B.A., Mondelli, V., Adewuya, A.O. & Fisher, H.L. (2020). 'Predicting the risk of future depression among school-attending adolescents in Nigeria using a model developed in Brazil', *Psychiatry Research*, **294**, p. 113511. DOI: 10.1016/j.psychres.2020.113511

[35] Tate, A.E., McCabe, R.C., Larsson, H., Lundström, S., Lichtenstein, P. & Kuja-Halkola, R. (2020). 'Predicting mental health problems in adolescence using machine learning techniques', *PLoS ONE*, **15**(4), p. e0230389. DOI: 10.1371/journal.pone.0230389

[36] Walsh, C.G., Ribeiro, J.D. & Franklin, J.C. (2018). 'Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning', *Journal of Child Psychology and Psychiatry*, **59**(12), pp. 1261–1270. DOI: 10.1111/jcpp.12916

[37] Ge, F., Zhang, L.W.D. & Mu, H. (2020). 'Predicting psychological state among Chinese undergraduate students in the COVID-19 epidemic: A longitudinal study using machine learning', *Neuropsychiatric Disease and Treatment*, **16**, p. 2111. DOI: 10.2147/NDT.S262736

[38] Rubaiyat, N., Apsara, A.I., Chaki, D., Arif, H., Israt, L., Kabir, L. & Alam, M.G.R. (2019). 'Classification of depression, internet addiction and prediction of self-esteem among university students', *In Proceedings of the 22nd International Conference on Computer and Information Technology (ICCIT)*, December, pp. 1–6.

[39] Mathur, P., Sawhney, R. & Shah, R.R. (2020). 'Suicide risk assessment via temporal psycholinguistic modeling', *In Proceedings of the AAAI Conference on Artificial Intelligence*, **34**(10). Palo Alto, CA, USA: AAAI Press, pp. 13873–13874. DOI: 10.1609/aaai. v34i10.7245

[40] Gaur, M., Alambo, A., Sain, J.P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., Sheth, A., Welton, R. & Pathak, J. (2019). 'Knowledge-aware assessment of severity of suicide risk for early intervention', *In Proceedings of the World Wide Web Conference*

(WWW), pp. 514–525. DOI: 10.1145/3308558.3313695

[41] Cai, H., Qu, Z., Li, Z., Zhang, Y., Hu, X. & Hu, B. (2020). 'Feature-level fusion approaches based on multimodal EEG data for depression recognition', *Information Fusion*, **59**, pp. 127–138. DOI: 10.1016/j.inffus.2020.03.001

[42] Ata, S.K., Fang, Y., Wu, M., Shi, J., Kwoh, C.K. & Li, X. (2021). 'Multi-view collaborative network embedding', *ACM Transactions on Knowledge Discovery from Data*, **15**(3), pp. 1–18. DOI: 10.1145/3442384

[43] Perozzi, B., Al-Rfou, R. & Skiena, S. (2014). 'DeepWalk: Online learning of social representations', *In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August, pp. 701–710. DOI: 10.1145/2623330.2623732

[44] Xia, F., Liu, J., Nie, H., Fu, Y., Wan, L. & Kong, X. (2020). 'Random walks: A review of algorithms and applications', *IEEE Transactions on Emerging Topics in Computational Intelligence*, **4**(2), pp. 95–107. DOI: 10.1109/TETCI.2019.2952908

[45] Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. (2002). 'SMOTE: Synthetic minority over-sampling technique', *Journal of Artificial Intelligence Research*, **16**(1), pp. 321–357. DOI: 10.1613/jair.953

[46] Bai, S., Kolter, J.Z. & Koltun, V. (2018). 'An empirical evaluation of generic convolutional and recurrent networks for sequence modeling', *arXiv preprint arXiv:1803.01271*. DOI: 10.48550/arXiv.1803.01271

[47] Yu, Y., Si, X., Hu, C. & Jianxun, Z. (2019). 'A review of recurrent neural networks: LSTM cells and network architectures', *Neural Computation*, **31**(7), pp. 1235–1270. DOI: 10.1162/neco_a_01199