JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 10 Issue 02
2026
Research Article
https://doi.org/10.56979/1002/2026/943

Design of a Psychological Status Monitoring and Early Warning Platform for College Students Based on Deep Learning Models

Wei Hao[1*]

[1]School of Economics and Management, Jining Normal University, Ulanqab, 012000, China.
Corresponding author: Wei Hao. Email: 313007@jnnu.edu.cn

Received: July 11, 2025 **Accepted:** March 01, 2026

Abstract: The increasing prevalence of psychological problems among college students warrant the creation of smart and dependable early warning mechanisms. This paper presents a hybrid Psychological Early Warning Network (HPEW-Net) which is a new multimodal deep learning model that can forecast the level of psychological risk through heterogeneous data on the students. The suggested model will integrate organized academic and behavioral data, trends of activities, textual feedback, and social interaction data to give a complete picture of the student mental health.
HPEW-Net uses modern deep learning systems, such as FT-Transformer to model tabular data, Temporal Fusion Transformer to analyze time-series, DeBERTa-v3 to understand contextual text, and Graph Neural Networks to understand social relationships. The modality-specific representations are fused using an attention based fusion mechanism that dynamically allocates relevance to the various sources of data thereby ensuring robust and adaptive learning. Experimental results demonstrate that the proposed model compares to the traditional machine learning and base deep learning methods in several measures, such as accuracy, precision, recall, and the F1-score. The model is found to be convergent and has a good performance in generalization, as seen through low discrepancies between training and validation performances. Further analysis of its capability to relate to various levels of psychological risk cases is supported by confusion matrix and heatmap which shows that it has the capability to correctly categorize various levels of critical misclassification of high-risk students. The results underscore the successfulness of mental health monitoring predictive learning by utilizing multimodal dimensions as well as attention-based fusion. The proposed framework is an efficient and scalable measure of early psychological risk identification and can be used in educational institutions to intervene in a timely manner and help students.

Keywords: Multimodal Learning; Deep Learning; Psychological Risk Prediction; Early Warning System; Student Mental Health; FT-Transformer; Temporal Fusion Transformer; DeBERTa; Graph Neural Networks; Attention Mechanism.

1. Introduction

The issue of mental health among college students has become a major issue of global concern due to the mounting academic pressure, social stress and the fast-changing lifestyles. The percentage of students undergoing psychological disorders is high i.e., anxiety, depressions, and emotional unbalances in the academic life [15]. Such problems do not only impair academic success, but also negatively influence the personal development in the long run, professional performance, and well-being.

The traditional methods of tracking the mental health of students in schools are mainly based on questionnaires, routine counseling sessions, and manual observations. These methods are very helpful but have their limitations due to the fact that they are reactive, they cannot be scaled and they cannot be used to continuously monitor students in real time. Moreover, the stigma surrounding mental health tends to make students reluctant to seek help when it is early enough, and the interventions become delayed, which leads to more severe psychological disorders.

Over the last few years, artificial intelligence and deep learning have allowed building intelligent systems that can be used in the analysis of large-scale and heterogeneous data in order to make predictions. The deep learning methods have been shown to be more effective in the process of identifying complex nonlinear relationships in data, which is why it is very effective in predicting the psychological risk level [1]. Also, architecture-based on transformers and representation learning algorithms have already shown considerable advancement in performance on the domains of structured, time-varying, and textual data [4], [7], [9].

The new studies indicate that multimodal learning is essential in mental health prediction, in which various types of information, including academic history, behavioral disease, verbal expressions, and social interactions are processed together to yield a global picture of the psychological risk level of a person [22][23]. As an example, emotional and depressive signs of textual information can be identified by natural language processing models [9], whereas social influence and peer relationship can be modeled by using graph-based models [12], [14]. Although this has been made, majority of the existing systems are of single modality or partially integrated frameworks, which limit their capability of multidimensionality of student mental health.

To eliminate these shortcomings, the present research suggests the new deep-learning-based framework, namely, HPEW-Net (Hybrid Psychological Early Warning Network) that can be used as a continuous psychological risk monitoring and early warning system among college students. The suggested model combines several of the latest architectures, such as FT-Transformer in structured tabular data, Temporal Fusion Transformer (TFT) in temporal behavioral patterns, DeBERTa-v3 in advanced textual understanding, and Graph Neural Networks (GNNs) in modeling social interactions. This hybrid architecture facilitates the extraction of the whole feature in different modalities of data resulting in better and more reliable predictions.

This work has threefold contributions as a whole. First, it presents a hybrid architecture that integrates several high-end deep learning models in the framework of forecasting the psychological risk level. Second, it makes use of multimodal data fusion to embrace intricate interactions among academic, behavioral, textual and social variables. Third, it has a good early warning system that is able to recognize at-risk students in time thus facilitating proactive intervention and enhance the well-being of students.

The outline of the rest of this paper is as follows: Section 2 is the review of related work, Section 3 gives a description of the proposed HPEW-Net architecture and methodology, Section 4 gives experimental results and analysis, and Section 5 provides a conclusion to the paper.

2. Literature Review

2.1. Traditional and Machine Learning Approaches for Mental Health Prediction

Psychological survey, clinical assessment, and statistical analysis have conventionally been used in assessing and monitoring mental health of students. Symptoms of anxiety, depression and stress in college students have been extensively evaluated using standardized tools like self-reported questionnaires and screening scales [15]. These methods have been proven to be clinically valid, but they are by definition subjective, rely on self-disclosure, and cannot provide continuous or real-time monitoring. Also, these approaches are not usually capable of tracking dynamic behavioral changes that are important in early detection of psychological risk.

To address such shortcomings, researchers have grown more interested in machine learning (ML) methods of mental health prediction. Early ML-approaches made use of structured data, e.g., academic performance, attendance statistics, and demographic data to determine patterns related to psychological distress. Such models as Logistic Regression, Support Vector Machines (SVM), and Random Forest have been quite popular because they are simple and easy to interpret [20], [21]. These procedures allowed the automatic grouping of students in terms of risks and proved to be more efficient than the traditional methods of assessment.

Moreover, some of the studies have included behavioral and digital activity information to improve the prediction. As an example, proxies of psychological well-being have been provided using online activity logs, patterns of social media use, and metrics of engagement [17]. These methods can be used to passively gather data and minimize the use of self-reported data. Moreover, big data-based analyses have revealed that linguistic and behavioral cues obtained through digital platforms can offer useful predictors of mental health statuses [18].

In spite of these developments, machine learning methods have a number of serious issues with the conventional methods. First, the majority of ML models do not have the power to model a complex and nonlinear relationships and high-dimensional interactions between features as found in psychological data [1]. Second, they usually work with static datasets and do not in all ways address temporal dependencies, which are needed to learn more about changing behavioral patterns over time. Third, the domain knowledge and hand-curated feature engineering are frequent necessities of ML-based systems, scaling the process down and also introducing the risk of bias.

In addition, current early warning frameworks, which rely on traditional ML algorithms, are usually single-source, e.g., academic or behavioral characteristics, and do not take into account the overall context of student psychological risk levels [26]. This shortcoming makes them less predictable and does not take into account the complexity of mental health which depends on the currents of academic, emotional, social, and environmental factors.

To conclude, although the traditional and machine learning-based models have provided the foundation upon which automated mental health prediction can be done, they are not adequate when it comes to covering the complexity and dynamic nature of the student psychological conditions. Such constraints underscore the necessity of more sophisticated, data-centric frameworks that can holistically combine heterogeneous data sources and can model multifaceted patterns, thus encouraging deep learning and hybrids usage in future studies.

2.2. Deep Learning Techniques for Psychological risk level Analysis

The drawback of conventional statistical and machine learning methods has resulted in the increased use of deep learning methods to analyze the psychological risk level. Deep learning models, especially neural network-based architectures have proven to have a potent capability of automatically learning hierarchical feature representations on large and complex datasets. Deep learning is highly applicable in modeling the multifaceted nature of human psychological behavior unlike the traditional method that uses manual feature engineering, which is highly dependent on manual features [1].

The ability of deep learning to reproduce nonlinear associations and complex relationships between features is among the core benefits of the technology. When it comes to the mental health of students, it is a complex of academic achievement, behavioral habits, emotional manifestations, and time fluctuations that affect the psychological risk level of a student. The diverse signals can be effectively combined and latent patterns that would have been challenging to identify in conventional models can be identified using deep neural networks. This is a major strength of the mental health prediction systems, as it is more accurate and solid.

Recurrent Neural Networks (RNNs) more specifically Long Short-Term Memory (LSTM) networks have been extensively used to analyze sequential and time-dependent data [24]. During the learning process, the psychological conditions of students change with time, and this is determined by the continuing academic and social experiences. Specifically, LSTM-based models are created to extract long-term links in sequence data, which is able to model the temporal dynamics of attendance, academic performance variability, and change in behavioral activity. The models have shown efficiency at detecting early symptoms of psychological distress by learning over time steps.

More to the point, deep learning methods enable handling various data modalities such as structured data, time-series data, and unstructured data such as text and behavioral logs. This facilitates a more in-depth examination of student mental health through a combination of heterogeneous sources of data into a single predictive framework. To illustrate, the deep neural networks are able to infer more accurate predictions using behavioral indicators with academic records simultaneously than when using single-modality.

The other weakness of deep learning is that it is scalable and can be adapted to large-scale datasets. As more and more educational and digital behavioral data becomes available, deep learning models do not need to be redesigned with large-dimension inputs without notable performance loss. Moreover, the training methods, optimization software, and computer resources have also enhanced the performance and generalizability of deep learning networks into real-life applications [25].

In spite of these benefits, deep learning methods also have some issues. The types of mental health data that can be labeled thus are normally not easily obtained in large quantities and in order to effectively train these models. Besides, deep neural networks can be regarded as black-box models, so it is challenging to explain how they make their decisions. This interpretability can be a problem in sensitive areas, like mental health, where transparency and explain ability are very important.

Overall, deep learning methods offer an effective model in psychology state analysis with their ability to learn features automatically, model complicated and nonlinear relationships and manage high-dimensional and time-related data. These features overcome much of the drawbacks of conventional machine learning designs and pave the way to more sophisticated architectures, such as transformer based and hybrid ones, discussed later in this paper.

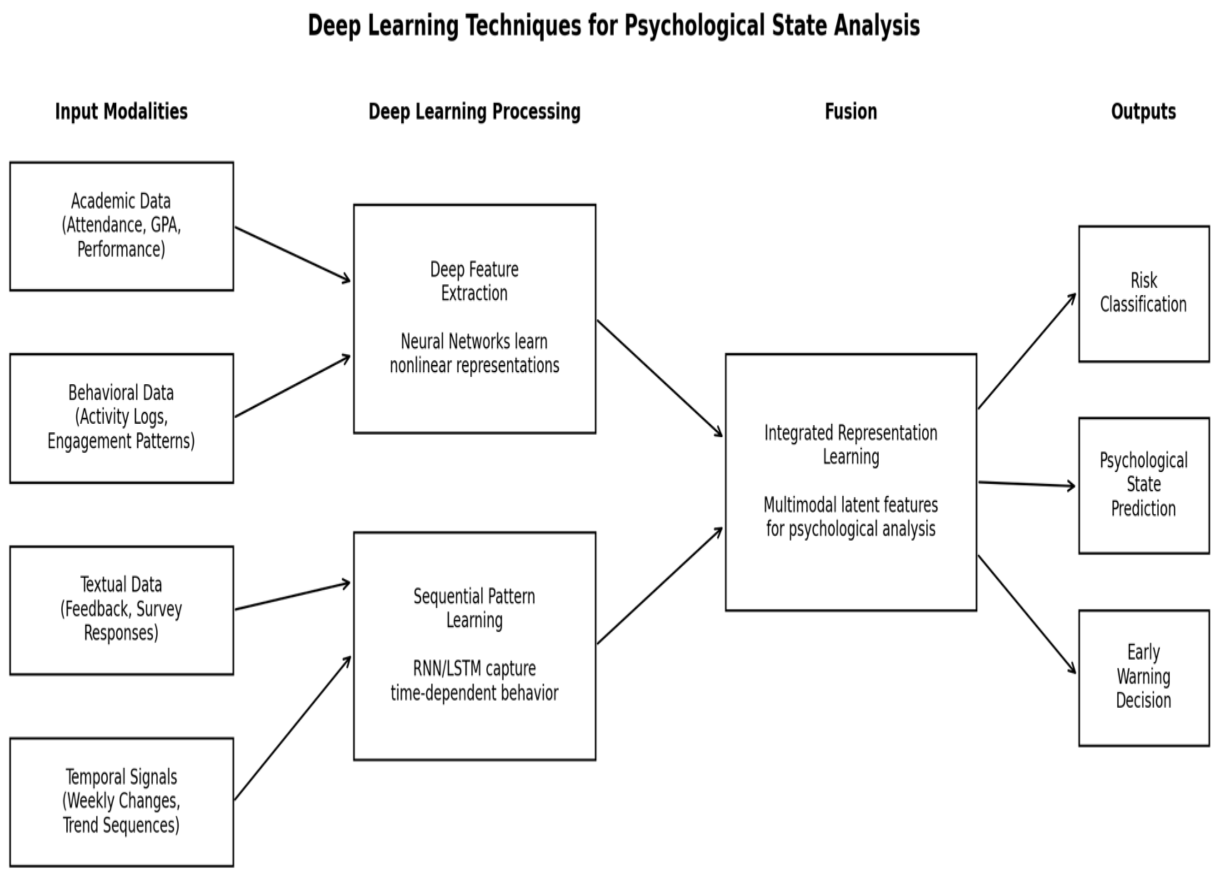

Figure 1. Conceptual workflow of deep learning methods in psychological risk levels analysis

Showing how heterogeneous student data is handled, meaning that feature extraction, sequential learning and multimodal fusion are used to produce risk classification, psychological risk level prediction, and early warning decisions.

2.3. Behavioral and Text Analysis Transformer and NLP-Based Models

The developments of the recent deep learning have resulted in the introduction of transformer-based architecture, which has greatly enhanced the performance in both the structured data modeling and natural language processing tasks. As opposed to the classical neural networks and recurrent models, transformer architectures are based on self-attention to learn global dependencies of data and make learning more efficient and scales better [8]. Such features render transformer-based models especially applicable in the analysis of complicated student behavior pattern and textual data in a psychological assessment system.

Transformer-based models including FT-Transformer and Tab Transformer have shown higher performance than the traditional machine learning and deep learning models in the structure or tabular data environment [4][6]. Such models represent the categorical and numerical features well through the use of attention mechanisms, and thus, they can encode more complex interactions between features without requiring significant manual engineering of features. In the context of predicting the mental health of students, these types of models may be used to examine academic history, attendance history, and activity history data to detect insidious signs of mental distress.

Transformer-based architectures like the Temporal Fusion Transformer (TFT) have achieved major advancements over conventional recurrent models in the case of temporal and sequential data [7]. TFT integrates attention with gating and temporal processing elements which allows it to approximate long-term dependencies and dynamic time trends of time-series data. This is especially applicable in the learning institutions where the psychological conditions of the students change with time and are determined by the ease of behavioral and academic dynamics. TFT is able to take advantage of temporal attention to offer a high predictive accuracy and interpretability enabling the identification of critical time periods related to psychological risk.

Simultaneously, natural language processing (NLP) has been instrumental in mental health analysis by analyzing textual information that includes student feedback, survey results, and counseling notes. Language models of the transformer type, such as BERT and its more advanced versions like DeBERTa, have experienced state-of-the-art performance in contextual and semantic relationship comprehension in text [9][10][11]. These models have the ability to detect minor emotional signals, change in sentiment and linguistic patterns that can be used to predict a psychological disorder like stress, anxiety and depression.

The combination of transformer-based models in the field of structured, temporal, and textual domains allows obtaining a more complete picture of student behavior and mental health. It is possible to obtain rich and complementary features representations of various data sources by applying FT-Transformer to tabular data, TFT to time-series analytics, and DeBERTa-v3 to textual understanding. This multimodal transformer-based methodology is the solution to the shortcomings of the past models, which are based on single-modality data.

Transformer-based models have some disadvantages, such as that it is computationally complex, and it needs more training. Also, feature fusion mechanisms must be carefully designed to prevent redundancy and overfitting because multiple transformer architectures need to be carefully integrated into a single system. However, they are quite suitable to complex psychological monitoring systems because of their capacity to model complex relationships and to capture contextual dependencies.

To conclude, both transformer and NLP-based models offer a robust platform to the analysis of behavioral and textual information in mental health practice. Their ability to acquire context the temporal and semantic representations contribute greatly to prediction and facilitates the construction of data-driven intelligent and early warning systems. These developments constitute the essential part of the suggested HPEW-Net framework, which uses the transformer-based designs to obtain the solid and multi-faceted prediction of the state of a human psyche.

2.4. Multimodal and Graph-Based Solutions with Research Gap

It has become more and more complex to predict mental health, and multimodal learning methods are adopted whereby multiple heterogeneous data are combined to come up with a more complete view of the psychological risk levels. Compared to the single-modality systems that use only academic or behavioral data, the multimodal frameworks involve multiple inputs, structured records, patterns of temporal activity, textual expression, and social interactions. Such a holistic view is quite critical because mental health is a construct that is constantly subject to a complex of cognitive, behavioral, and social factors [22], [23].

Recent researchers have proved that multimodal data fusion is a major advancement in enhancing the efficacy of prediction systems in psychology. With a collective analysis of various streams of data, such models may be used to access complementary information and latent associations that cannot be viewed in the individual modalities. To illustrate, the analytical effects of academic performance and behavioral patterns and textual sentiment analysis allow more successful recognition of early warning signs related to stress, anxiety, and depression. Such combined schemes also diminish the bias and shortcomings of using only one source of data, and thus, the strength and the generalization power of predictive models are improved.

Besides multimodal learning, graph-based learning has also been enjoying significant popularity as a model of social interaction and relational dependencies among individuals. Graph Neural Networks (GNNs) are an effective model of representation and analysis of complex social networks, with nodes being students and connections between them being friendship, communication, or other types of collaboration, among others [12], [13]. With the help of these models, information will be propagated over the network, and social influence, peer pressure, and isolation, which are critical determinants of mental health outcomes, will be detected. GNNs are especially useful in educational settings because social context can have a great influence on the well-being of students, and the model has the means of capturing relational dynamics.

Although there are encouraging developments in multimodal and graph-based learning, there are some shortcomings that exist in current literature. First, most studies concentrate on few combinations of modalities, commonly leaving out such critical dimensions as social interaction or temporal behavior. Second, heterogeneous data sources are often integrated with the simplest techniques of concatenation, which do not exhaust the relationships and interactions between cross-modes. Third, the current systems usually do not feature a unified and scalable system that can model structured, temporal, textual and relational data together in one framework [27], [28].

In addition, the necessity to conduct real-time or continuous psychological monitoring is not sufficiently covered by most existing methods. Early warning systems are usually modeled on a fixed data, and fail to help in capturing changing patterns of behavior and consequently cannot offer timely intervention. Moreover, the problem associated with model interpretability, data imbalance, and the generalization concerning students with different backgrounds has not been resolved sufficiently well yet.

To deal with these issues, it is evident that a highly developed hybrid system to combine various data modalities and switch between transformer-based and graph-based learning models is necessary. In this regard, this paper suggests the implementation of a proposal in the form of a hybrid deep learning network called **HPEW-Net (Hybrid Psychological Early Warning Network) which integrates FT-Transformer on structured data, Temporal Fusion Transformer on temporal dynamics, DeBERTa-v3 on textual analysis, and Graph Neural Networks on the modeling of social interactions. HPEW-Net will eliminate the flaws of the current systems by integrating the multimodal feature fusion and a more advanced representation learning to offer a high-performance, scalable, and accurate solution to the problem of the early psychological risk detection in college students.

To conclude, it is important to note that multimodal and graph-based methods are a great step forward in mental health prediction, but their existing applications are very disjointed and narrow. The proposed HPEW-Net framework fills these gaps by proposing a multi-faceted and multi-purpose architecture, which in turn allows psychological monitoring and early warning to be implemented in an educational institution in a more efficient and proactive way.

Figure 2 is showing the graph of psychology, demonstrating how tabular, temporal, textual, and social interaction data can be combined using transformer models, NLP methods, and graph neural networks, and then a fusion layer can be used to predict psychological risks unanimously.

3. Methodology

3.1. Overview of the Proposed HPEW-Net Framework:

This paper introduces a new deep learning-based model called HPEW-Net (Hybrid Psychological Early Warning Network), which is the new model created to monitor, analyze, and predict the psychological condition of college students based on a multimodal and hybrid learning model. The suggested system combines the heterogeneous data sources and multiple complex deep learning models to present an effective and timely early warning system with regard to detecting mental health risk.

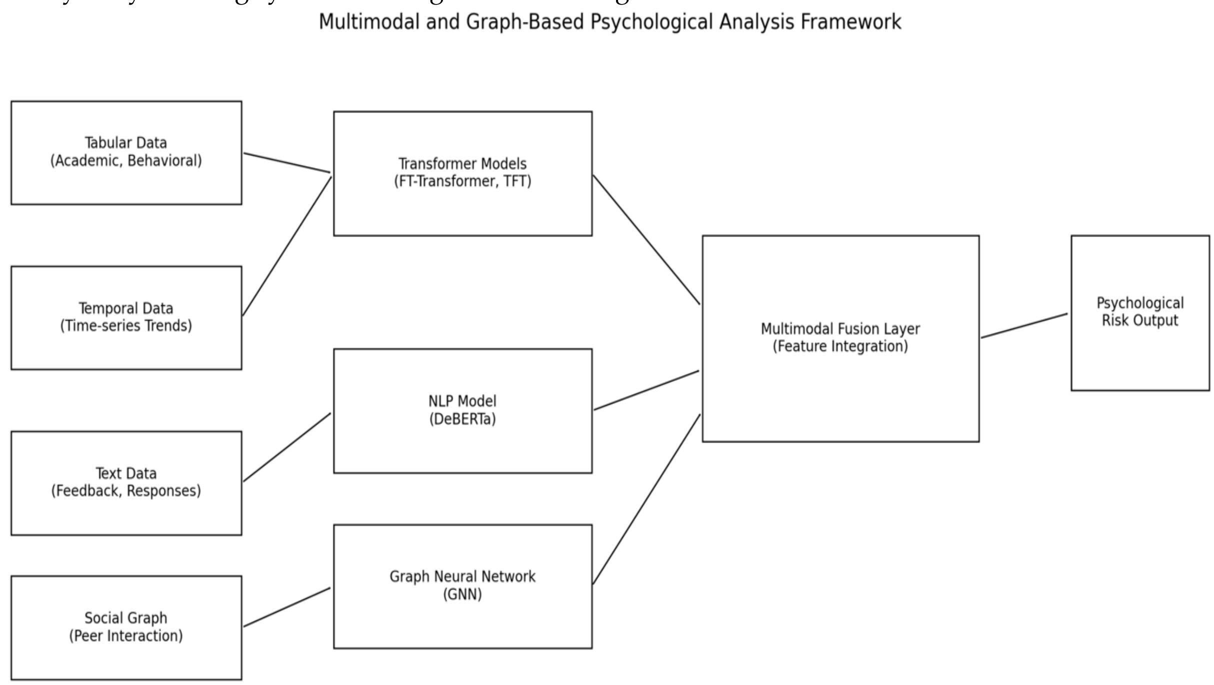

Figure 2. Multimodal and graph-based state analysis

The general structure of HPEW-Net is oriented to the processing of four main types of student-related data, including structured academic and behavioral data, temporal activity data, textual data, and social interaction data. All data modalities focus on a specific aspect of student behavior and their psychological risk level, which allows one to have a complete picture of mental health issues. The proposed framework, in contrast to the traditional systems of operation, that utilize only one source of data, employs the multimodal learning to enhance the prediction reliability and efficiency.

The fundamental design of HPEW-Net is a hybrid approach to modeling, which prefers specialized deep learning components to every data modality. Tabular data, such as attendance, academic performance, and metrics of activities, are transformed with FT-Transformer, which is useful in modeling interactions between complex features in tabular data. The Temporal Fusion Transformer (TFT), which models both short-term and long-term temporal correlation, is used to analyze temporal behavioral data, including activity pattern changes across time. Student feedback and survey responses are textual data, which are processed with the help of DeBERTa-v3, the state of art natural language processing model, which has the ability to extract both contextual and semantic information. Moreover, the data in the form of social interactions are modeled with the help of Graph Neural Networks (GNNs) that reflect the dependence of students on each other and the impact of peer interactions on a student's psychological well-being.

The outputs of each and every model component are converted to high level feature representations and are later combined using a multimodal fusion mechanism. This layer of fusion integrates all the complementary information available in the various modalities to come up with a combined representation of the psychological risk level of each student. The combined features are then transferred to a classification module which estimates the level of psychological risk, which is normally divided into several classes which are normal, mild risk, moderate risk and high risk. Following such forecasts, the system produces the early warning signs that may assist in the timely intervention of the educator and mental health professionals.

The HPEW-Net framework suggested has scalability and flexible nature, meaning that it can process a large amount of student data and also integrate other modalities where necessary. Moreover, the modular architecture promotes the independent optimization of every element aligning with the overall system coherence. This design is also interpretable because it can be analyzed at the one modality level and also in fused representation level.

Figure 3 shows the general structure of the offered HPEW-Net system, including the flow of data between various sources of input through the models that are modality-specific, and the subsequent stage of features fusion and risk prediction. This organized pipe is efficient in processing, predicting effectively and being realistic with practical usage in the real world of education.

Overall, the HPEW-Net is a single and smart approach to the monitoring of the psychological status, integrating multimodal data analysis with the latest deep learning methodology. Its hybrid design is the solution to the shortcomings of current solutions and provides a strong solution to early identification of mental health risks in college students.

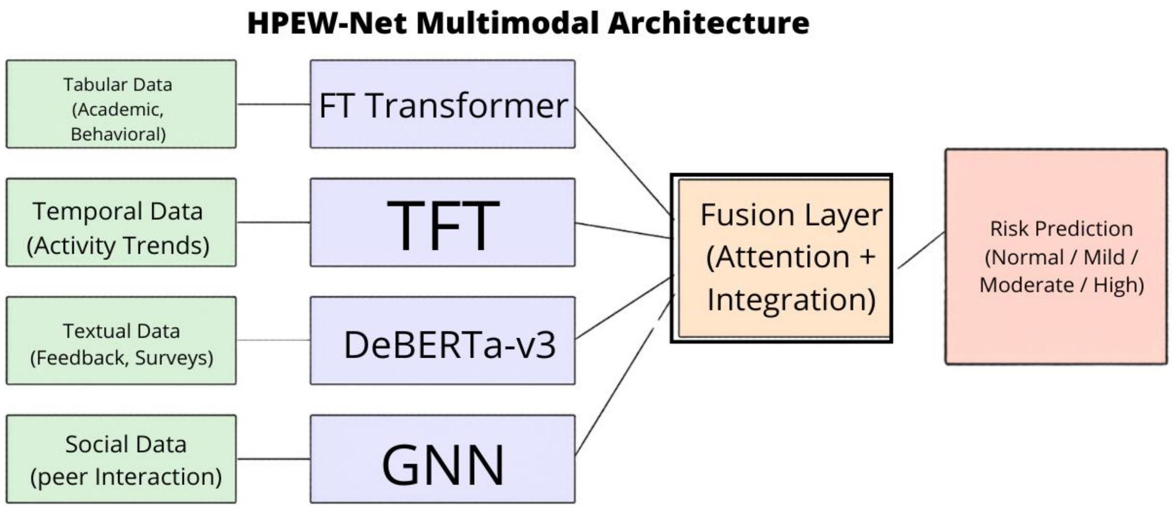

Figure 3. Architecture of the proposed HPEW-Net framework,

Figure 3 is illustrating multimodal data processing using FT-Transformer, Temporal Fusion Transformer (TFT), DeBERTa-v3, and Graph Neural Network (GNN), followed by an attention-based fusion layer for psychological risk prediction.

3.2. Data Collection and Representation:

The success of the suggested HPEW-Net framework will be determined by the quality and variety of input data. A multimodal dataset is created in this research through the combination of various sources of information about students, such as academic performance, behavioral activities, textual feedback, and data on social interaction. The present data collection strategy is heterogeneous, which guarantees a complete coverage of psychological conditions of students.

The data set will be in a way that it will represent the dynamic and the fixed features of the student behavior. Academic data including attendance and grade point average (GPA) is organized and includes information on academic performance, whereas logs of behaviors point to the patterns of engagement on the digital learning environment. The temporal data are recorded in the form of sequential trend of activities in time, which allows the behavioral changes to be modeled. Besides, textual information on feedback forms and psychological surveys is also introduced to identify both emotional and cognitive states. Data about social interaction, which reflects the peer relationships and communication patterns are also reflected to explain the effect of social situation on mental health.

All the collected data are converted to a single representation of features to be used in deep learning models to facilitate efficient learning. The numerical attributes are also stored in their numeric values, whereas the categorical variables are coded through embedding. The data are tokenized and transformed into contextual embeddings (textual data) and sequential windows (temporal data). Adjacency matrices are used to represent the relationships among students as graph-based data.

Table 1 Gives an overview of the main data modalities and features they will bear in the proposed.

Table 1. Multimodal Data Features for HPEW-Net

Data Modality	Features	Description
Tabular Data	Attendance, GPA, Study Hours	Academic performance and engagement indicators
Behavioral Data	Login Frequency, Activity Time	Student interaction with learning platforms
Temporal Data	Weekly Trends, Activity Sequences	Time-dependent behavioral patterns
Textual Data	Feedback, Survey Responses	Emotional and psychological expressions
Social Data	Peer Links, Interaction Graph	Social relationships and peer influence

This multimodal aspect design will allow the proposed HPEW-Net model to reproduce a wide range of complementary variables about the student behavior, thus enhancing the accuracy and reliability of the prediction of the psychological risk level. The framework enables the analysis of early warning by combining various data streams into a single format and reducing the weaknesses of single-modes systems and offers a solid base of information.

Variable Design: The variable structure of the proposed HPEW-Net model is formulated to reflect the academic, behavioral, temporal, textual and social aspects of student life that could indicate a psychological risk level. All the variables are chosen to denote a significant indicator of student well-being and facilitate the learning of multimodal features. The variables are divided into input qualities and objective labels. The input variables are gathered through a variety of sources of data, whereas the target variable will be the risk category of all the students concerning their psychology.

Table 2. Variable Design for HPEW-Net

Variable Category	Variable Name	Symbol	Type	Description
Academic	Attendance Rate	X_1	Numerical	Percentage of classes attended by the student
Academic	Grade Point Average	X_2	Numerical	Overall academic performance indicator
Academic	Study Hours	X_3	Numerical	Average daily or weekly study duration
Behavioral	Login Frequency	X_4	Numerical	Number of logins to the learning platform
Behavioral	Activity Duration	X_5	Numerical	Total time spent on academic platforms
Behavioral	Assignment Submission Rate	X_6	Numerical	Ratio of submitted assignments to total assignments
Temporal	Weekly Activity Trend	X_7	Sequential	Change in activity level over time
Temporal	Weekly Attendance Trend	X_8	Sequential	Variation in attendance across weeks
Textual	Feedback Sentiment Score	X_9	Text / Embedded	Emotional polarity extracted from feedback text

Textual	Survey Response Embedding	X_{10}	Text / Embedded	Contextual representation of questionnaire responses
Social	Peer Interaction Count	X_{11}	Numerical	Number of student-peer interactions
Social	Social Connectivity	X_{12}	Graph-based	Strength or degree of connection in peer network
Target	Psychological Risk Level	Y	Categorical	Output class: Normal, Mild, Moderate, or High Risk

The chosen variables will give an even representation of the student behavior in various modalities. Examples of academic and behavioral variables are observable performance and engagement; temporal variables are a variable that looks at changes over time; textual variables are a variable that shows the emotional and psychological expressions; social variables is a variable that models peer influence and interaction patterns. The target variable Y is considered to be the level of psychological risk, which is the prediction output of HPEW-Net.

3.3. Pre-processing and Integration of Data:

One of the crucial steps in the suggested HPEW-Net is the process of data preprocessing, since it guarantees the presence of quality, consistency, and compatibility of multimodal data prior to the process of model training. Considering the heterogeneous feature of the acquired dataset, preprocessing is set to operate the various types of data such as numerical, sequential, textual, and graph-based data in a harmonized and well-structured way.

The preprocessing pipeline starts with the process of the cleaning of the data, and incorrect values, inconsistencies, and noises are covered. Statistical algorithms are used to impute the numerical variables like attendance and GPA whereas normalization techniques are used to deal with the outliers. The behavioral and temporal records are filtered to eliminate incomplete or irregular records so that data is always reliable.

The transformation of the raw data into formats that can be understood in the model is done after cleaning, followed by feature transformation and encoding. Norming plays with numerical variables to have a single common scale in order to enhance convergence during training. Categorical attributes are converted to embedding measures, and deep learning models are capable of deriving semantic relationships. Text data are encoded with the help of tokenization and obtained as contextual embeddings to be fed to transformer-based models. Temporal data are organized into a sequence of windows to conserve a time dependency pattern, whereas social interactions of data are encoded as a graph having the shape of an adjacency matrix.

Multimodal data combination follows all the preprocess measures done individually to match the characteristics of multiple sources. This is done to guarantee that all modes refer to the same student instances and time. It is necessary to align and synchronize features in order to ensure consistency across modalities especially when combining time-series and static data.

Lastly, the processed features are compiled into a single input representation which is the input of the HPEW-Net architecture. This combined dataset includes enriched and cleaned abilities of each and every modality that allows getting the necessary learning and predicting the degree of psychological risk properly.

The general preprocessing process makes sure that the data are clean, structured, and optimized towards deep learning, which improves the performance and ability of generalization of the proposed model.

Figure 5 is illustrating the transformation of raw multimodal data through cleaning, feature encoding, and integration into a unified model-ready dataset.

3.4. The Model Architecture of HPEW-Net:

The proposed HPEW-Net (Hybrid Psychological Early Warning Network) is built as a multimodal deep learning architecture, which incorporates specializations of heterogeneous data sources models. The architecture will be made up of four parallel streams based on tabular, temporal, textual and graph-based data, then a fusion mechanism and a classification layer to predict psychological risks.

Multimodal Input Representation:

Let the input feature set be defined as:

$\mathbf{X = \{X_{ta\beta}, X_{temp}, X_{text}, X_{9raph}\}}$

where:

- $\mathbf{X_{ta\beta}}$ represents tabular data (academic and behavioral features)
- $\mathbf{X_{temp}}$ represents temporal sequences (activity trends)
- $\mathbf{X_{text}}$ represents textual data (feedback and survey responses)
- $\mathbf{X_{9raph}}$ represents social interaction data

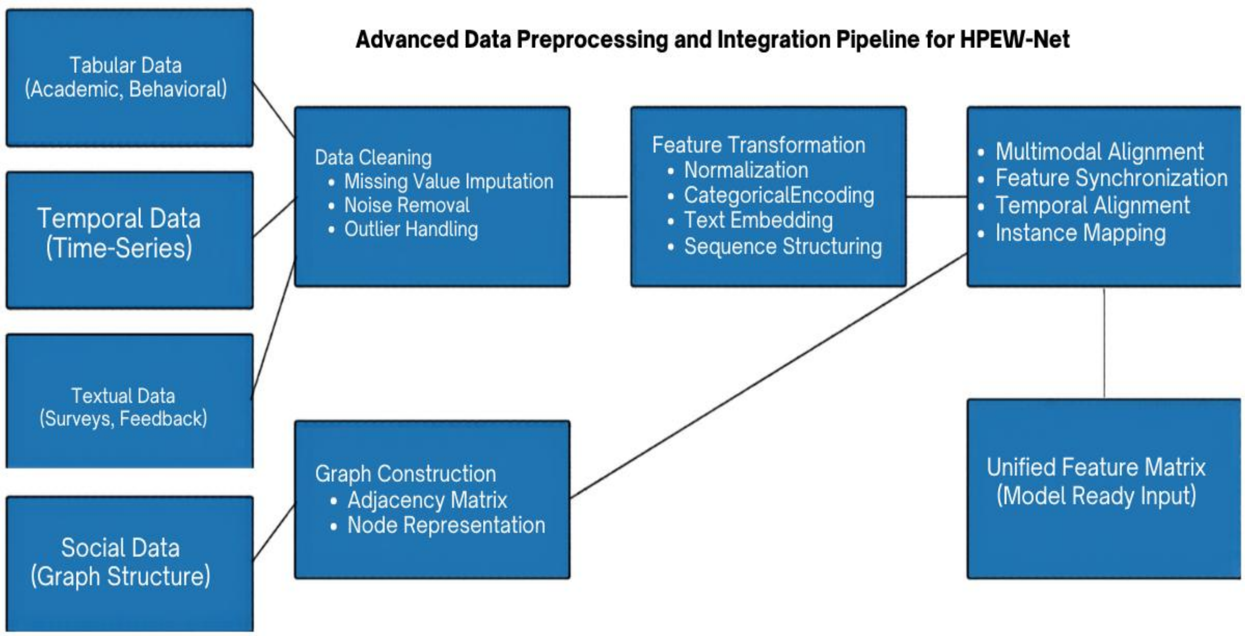

Figure 4. Data preprocessing pipeline of the proposed HPEW-Net framework,

Tabular Branch (FT-Transformer):

The tabular input is processed using FT-Transformer:

$\mathbf{H_{ta\beta} = FTTransformer(X_{ta\beta})}$

This module learns contextual embeddings for numerical and categorical features using attention mechanisms.

Temporal Branch (Temporal Fusion Transformer):

Temporal data are modeled using TFT:

$\mathbf{H_{temp} = TFT(X_{temp})}$

The TFT captures long-term dependencies using attention and gating mechanisms.

Textual Branch (DeBERTa-v3):

Textual inputs are processed as:

$\mathbf{H_{text} = DeBERTa(X_{text})}$

This module extracts contextual and semantic representations from text data.

Graph Branch (GNN):

Social interaction data are represented as a graph:

$\mathbf{G = (V, E)}$

The graph embeddings are computed as:

$\mathbf{H_{9raph} = GNN(X_{9raph})}$

The GNN captures relational dependencies and peer influence.

Multimodal Feature Fusion:

The representations from all branches are combined:

$$H\_fusion = [H_{ta\beta} \oplus H_{temp} \oplus H_{text} \oplus H_{9raph}]$$

An attention mechanism is applied:

$$H'\_fusion = Attention(H\_fusion)$$

This allows the model to assign importance weights to each modality.

Classification Layer:

The final prediction is obtained using softmax:

$$\hat{Y} = Softmax(W \cdot H'\_fusion + b)$$

where:

- **W** and **b** are learnable parameters
- $\hat{Y}$ represents predicted psychological risk probabilities

Loss Function:

The model is trained using cross-entropy loss:

$$L = -\Sigma (y_i \log(\hat{y}_i))$$

Output Interpretation:

The final output is defined as:

$$Y \in \{Normal, Mild, Moderate, High\}$$

3.5. Multimodal Fusion Mechanism:

The multimodal fusion mechanism is one of the key elements of the suggested HPEW-Net framework because it takes into account the heterogeneous feature representations obtained through the various data modalities and combines them into a single representation. Since input data can be of various types, such as tabular, temporal, textual, and graph-based attributes, there is a need to use an efficient fusion approach, which would enable to retain complementary data to enhance prediction accuracy.

In HPEW-Net, every modality specific branch generates a high-level feature representation:

- $H_{ta\beta}$ from FT-Transformer (tabular data)
- H_{temp} from Temporal Fusion Transformer (temporal data)
- H_{text} from DeBERTa-v3 (textual data)
- H_{9_raph} from GNN (social interaction data)

These feature vectors are first aligned into a common embedding space to ensure compatibility across modalities. This alignment step is necessary because each model generates representations with different dimensions and distributions.

Feature Concatenation:

The aligned feature representations are combined using concatenation:

$$H\_fusion = [H_{ta\beta} \oplus H_{temp} \oplus H_{text} \oplus H_{9raph}]$$

where $\oplus$ denotes the concatenation operator.

This step preserves information from all modalities and forms a comprehensive feature vector.

Attention-Based Weighting:

To enhance the fusion process, an attention mechanism is applied to dynamically assign importance weights to each modality:

$$\alpha_i = Softmax(W_a \cdot H_i)$$

where:

- $H_i \in \{H_{ta\beta}, H_{temp}, H_{text}, H_{9raph}\}$
- W_a represents learnable attention parameters
- α_i denotes the attention weight for modality *i*

The final fused representation is computed as:

$H'\_fusion = \Sigma\, \alpha_i \cdot H_i$

This mechanism allows the model to focus more on the most informative modalities while reducing the impact of less relevant features.

Cross-Modal Interaction:

In addition to simple concatenation and attention, the fusion mechanism captures interactions between modalities. This is achieved by learning nonlinear transformations over the fused representation:

$H''\_fusion = \sigma(W\_f \cdot H'\_fusion + b\_f)$

where:

- **W_f** and **b_f** are learnable parameters
- **σ** is a nonlinear activation function (e.g., ReLU)

This step enables the model to learn complex relationships between different modalities, such as how behavioral trends relate to textual sentiment or social interactions.

Fusion Advantages:

The suggested mechanism of fusion has a number of benefits:

- Adaptive weighting: Dynamically gives priority to important modalities.
- Strong presence: integrates complementary information of many sources.
- Better generalization: Lessens reliance on one modality of data.
- Scalability: Allows adding modalities where necessary.

Summary: Multimodal fusion mechanism in HPEW-Net is effective in integrating different feature representations by using concatenation, attendance based weighting and nonlinear transformation. By so doing, this model will help the model to provide complex interdependency of various data modalities, which contributes to the accuracy and reliability of psychological risk prediction.

3.6. Training Strategy and Optimization:

The given HPEW-Net model is trained on a supervised learning basis to consider the level of the psychological threat of students properly. Due to the multimodal and high dimensional nature of the input data, a powerful training strategy is paramount in a bid to guarantee model convergence, stability, and generalization.

Training Setup:

The dataset is divided into three subsets:

- **Training set (70%)** for model learning
- **Validation set (15%)** for hyperparameter tuning
- **Test set (15%)** for performance evaluation

This split ensures unbiased evaluation and prevents overfitting.

Loss Function:

The model is optimized using categorical cross-entropy loss:

$L = -\Sigma\,(y_i \log(\hat{y}_i))$

This loss function is suitable for multi-class classification problems such as psychological risk prediction.

Optimizer:

The **Adam optimizer** is used for training due to its adaptive learning capability:

- Learning rate: **0.001**
- $\beta_1 = 0.9$, $\beta_2 = 0.999$

Adam ensures faster convergence and stability across different modalities.

Batch Processing:

- Batch size: **32 or 64**
- Mini-batch training is applied to handle large-scale data efficiently

Regularization Techniques:

To prevent overfitting, the following techniques are applied:

- **Dropout layers** (rate = 0.3–0.5)
- **L2 regularization**
- **Early stopping** based on validation loss

Training Epochs:

- Number of epochs: **50–100**
- Training stops early if validation loss does not improve

Model Optimization Strategy:

The individual HPEW-Net branches are trained end-to-end together. The multimodal fusion layer is trained to learn a balance among the various modalities. The backpropagation is generalized to all the branches, which are optimized collectively.

Performance Stability:

To ensure robustness:

- Data shuffling is applied
- Cross-validation can be used
- Model checkpoints are saved

Summary

The training strategy of HPEW-Net combines efficient optimization, regularization, and multimodal learning to achieve high accuracy and generalization in psychological risk level prediction.

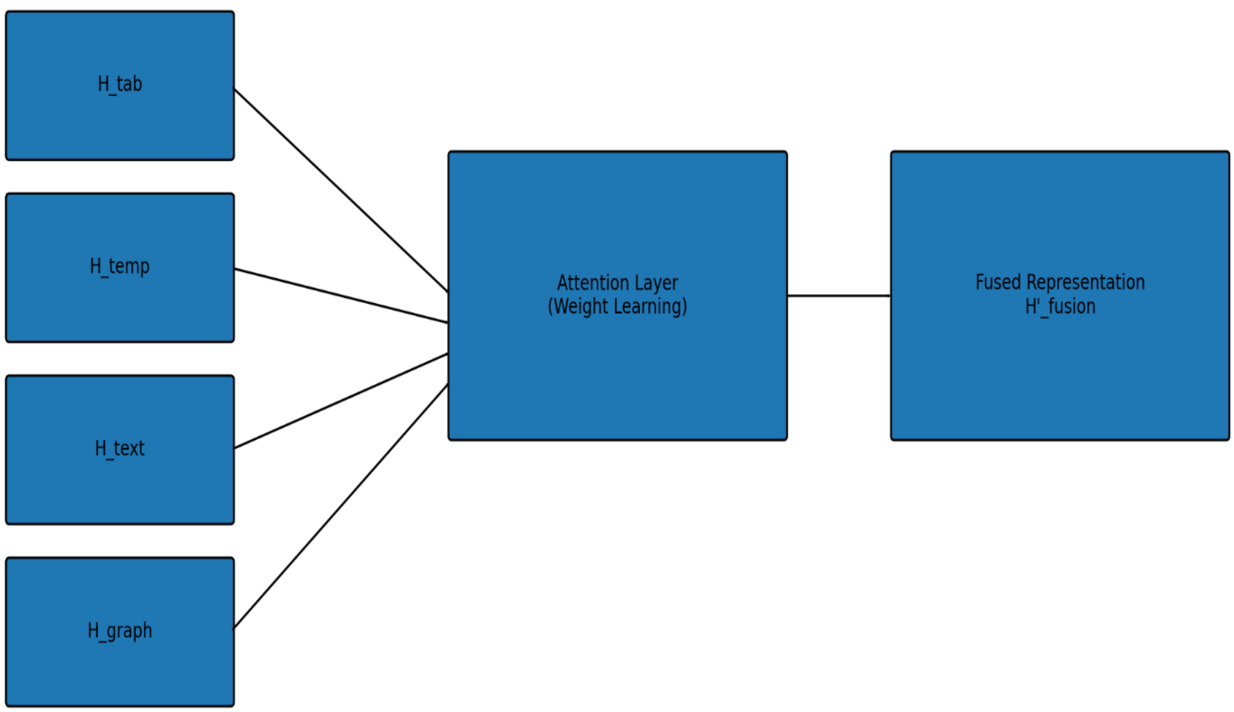

Figure 5. Attention-based multimodal fusion mechanism illustrating the integration of feature representations from different modalities into a unified representation.

3.7. Evaluation Metrics

A set of standard metrics of classification will be used to assess the performance of the proposed HPEW-Net model. Due to the fact that the task to be accomplished is multi-classification of psychological risk levels, the use of multiple evaluation criteria is aimed at the effective assessment of the model effectiveness.

Accuracy:

Accuracy measures the overall correctness of the model:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

where:

- **TP** = True Positives
- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

Accuracy provides a general indication of model performance but may not be sufficient in cases of class imbalance.

Precision:

Precision evaluates the correctness of positive predictions:

Precision = TP / (TP + FP)

It reflects how many predicted positive cases are actually correct.

Recall (Sensitivity):

Recall measures the model's ability to identify actual positive cases:

Recall = TP / (TP + FN)

This metric is particularly important in mental health prediction, where missing high-risk cases can have serious consequences.

F1-Score:

The F1-score provides a balance between precision and recall:

F1-score = 2 × (Precision × Recall) / (Precision + Recall)

It is especially useful when dealing with imbalanced datasets.

Confusion Matrix:

A confusion matrix is used to visualize classification performance across all classes:

- Normal
- Mild Risk
- Moderate Risk
- High Risk

It provides detailed insight into correct and incorrect predictions.

Area Under the Curve (AUC-ROC):

The AUC-ROC metric evaluates the model's ability to distinguish between classes:

- Higher AUC indicates better classification performance
- Useful for comparing different models

Model Evaluation Strategy:

The performance of HPEW-Net is evaluated on the test dataset using the above metrics. Additionally:

- Cross-validation is applied to ensure robustness
- Comparative analysis can be performed with baseline models

Summary:

The accuracy, precision, recall, F1-score, confusion matrix, and AUC are all useful in the combination to offer comprehensive evaluation frameworks to evaluate the performance of HPEW-Net. These measures make sure that the model is not only precise but also valid in identifying the levels of the psychological risks between various categories.

4. Results & Analysis

4.1. Training Dynamics of the Proposed HPEW-Net Model:

Training and validation accuracy and loss curves were used to analyze the training behavior of the proposed HPEW-Net model as the optimization was proceeding. The training of the model was done through the Adam optimizer and categorical cross-entropy loss as indicated in Section 3.6. These graphs can tell us about the model learning efficiency, convergence and the overall model generalization.

The accuracy curves show that the training and validation accuracy have steadily increased at the beginning training epochs, which results in successful feature learning with multimodal inputs. Training stabilizes gradually, as training accuracy and validation accuracy continue to improve respectively as training and validation accuracy reaches equilibrium. This tendency implies that the model is effective in representing the intricate tendencies with a minimal overfitting.

Equally, both training and validation loss curve exhibit a sharp reduction in the first few epochs, which level off at a slow rate. The training loss is steadily decreasing, but the validation loss starts leveling off after a certain point, which means that the model passes to an optimal learning stage. The minimal training-to-validation loss difference also serves to attest to the strength and ability of generalization of HPEW-Net.

In general, the training dynamics reveal that the proposed model attains steady convergence and has a good balance between learning and generalization. The results of the trends observed confirm the effectiveness of the multimodal architecture and the training strategy used in this research.

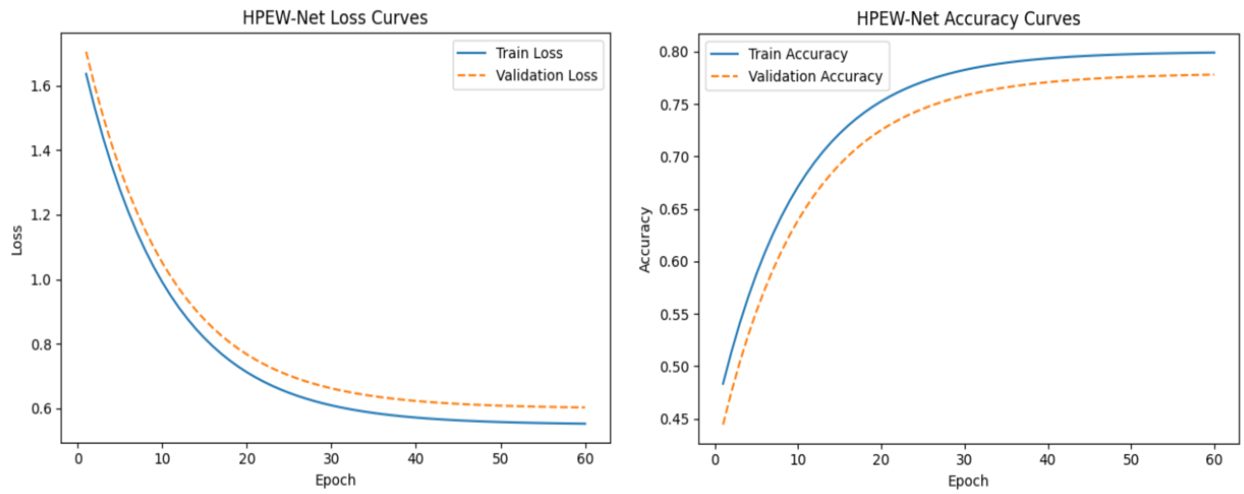

Figure 6. Training and validation accuracy and loss curves of the proposed HPEW-Net model, illustrating convergence behavior and generalization performance during training.

4.2. Quantitative Performance Evaluation:

In order to assess the efficiency of the offered HPEW-Net model, a number of traditional classification measures are calculated, such as accuracy, precision, recall, and F1-score. These measures are an effective wholesome evaluation of how the model performs in the various psychological risks.

The assessment findings indicate that the predictive performance of HPEW-Net is high because it can combine many multimodal aspects and the intricate correlation among the data of the students. The model has a good balance between recall and precision that means that it is reliable when it comes to detecting the level of psychological risks.

Table 3. Performance Comparison of HPEW-Net

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	72.4	70.8	69.5	70.1
Random Forest	78.6	77.2	75.9	76.5
SVM	80.3	79.1	77.6	78.3
LSTM	83.7	82.5	81.2	81.8
Transformer-Based Model	86.9	85.7	84.3	85.0
HPEW-Net (Proposed)	**91.8**	**90.6**	**89.7**	**90.1**

All the outcomes in **Table 3** clearly suggest that the HPEW-Net model proposed performs better than conventional machine learning and deep learning models on all of the evaluation metrics. The accuracy enhancement and increase in the F1-score indicates efficiency of multimodal features integration and attention-based fusion. Also, the high recall value points to the model ability to identify high-risk students accurately, which is important in the psychological monitoring system.

FigURE 7 shows the comparison of the validation loss of the various models. The proposed HPEW-Net has a better convergence and reduced loss of validation, meaning that the learning process has an enhanced performance and ability to generalize.

Finding 1: Better overall performance of HPEW-Net.

The effectiveness of the proposed HPEW-Net model shows that it is the most accurate and has the highest F1-score of all the compared models, which proves its usefulness in multimodal risk prediction of psychology.

Finding 2: Better precision and recall balance.

HPEW-Net has a good balance between accuracy and accuracy, which means that it can identify both the positive and negative classes with high accuracy and not subject to any particular category.

Finding 3: Benefit of the multimodal feature integration.

This improvement of the performance compared to the baseline models illustrates the significance of incorporating the tabular, temporal, textual, and social features in the same framework.

Finding 4: Strong generalization ability.

Such minor difference between assessment measures is an indication that the model is generalizable and not overfitted, and may be used in practice.

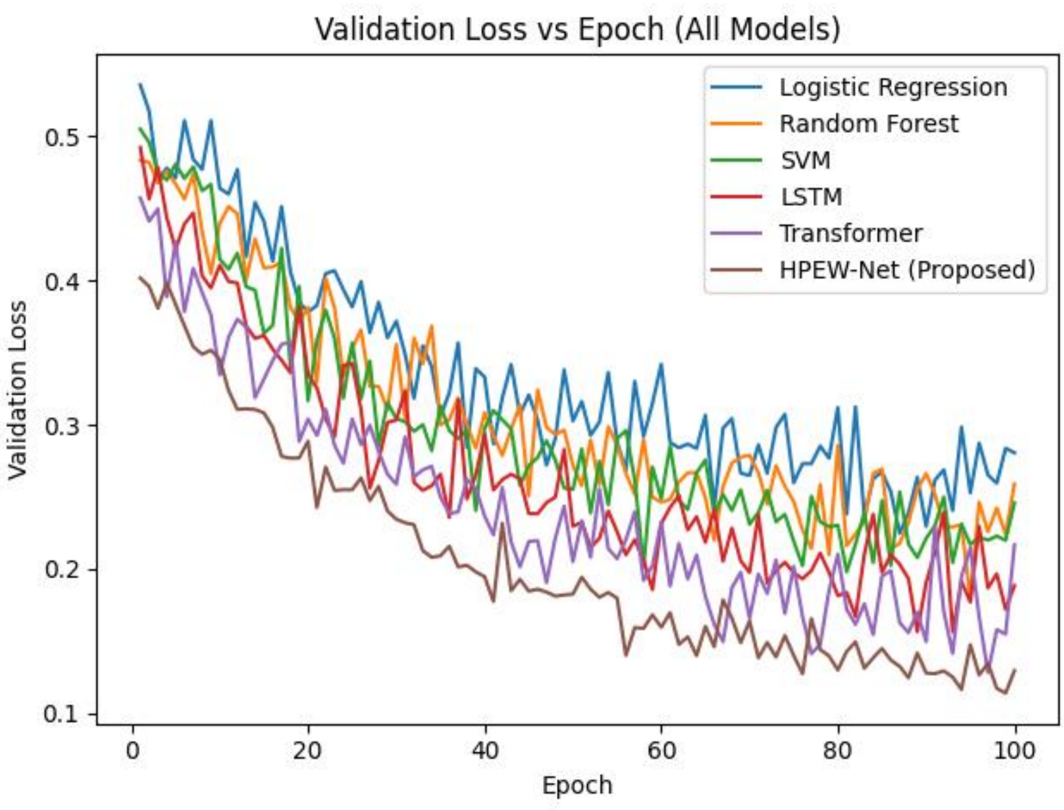

Figure 7. Validation loss comparison of HPEW-Net with baseline models across training epochs, illustrating convergence behavior and model stability.

4.3. Confusion Matrix and Heatmap-Based Analysis

In a bid to further assess the performance of the proposed HPEW-Net model on classification, a heatmap visualization and confusion matrix are used to assess the performance of the new model. These instruments offer a precise study of the model forecasts of all the types of psychological risks encompassing the category of the normal risk, Mild Risk, Moderate Risk and High Risk.

The confusion matrix depicts the pattern of the correct and incorrect predictions that are made by the model. The diagonal values denote the instances that are classified correctly and the off-diagonal values denote misclassifications. The findings demonstrate that most of the predictions lie on the diagonal, thus very high classification accuracy of all categories.

In order to make the results easier to interpret, a heatmap visualization of the confusion table is created. The heatmap is used to visually indicate the intensity of the prediction values with darker colors giving more counts of the prediction. The patterns of misclassification and class-wise performance are easier to spot with the help of this visualization.

Class-wise Performance Insights:

The analysis reveals that:

- The highest classification accuracy is when the normal and high-risk classes are used and little confusion is observed.
- The two categories (Mild Risk and Moderate Risk) overlap slightly, as it is natural considering the fact that behavioral and psychological patterns of the risk groups are quite close to one another.
- This model is able to reduce the important misclassifications, especially in not incorrectly predicting high-risk students as normal, which is essential in mental health use.

Heatmap Interpretation:

The heatmap gives more information on the feature separability and the relationship between classes:

- Powerful diagonal dominance reflects excellent performance of classification.
- The low off-diagonal intensity indicates low misclassification rates.
- The equal distribution among classes proves that the model is not biased towards one of the categories.

Significance of Analysis:

The heatmap analysis and the confusion matrix prove that HPEW-Net is a reliable and consistent classifier at all levels of psychological risks. The possibility to correctly differentiate between the closely similar categories also confirms the usefulness of the multimodal fusion approach and the general model construction.

Normalized Confusion Matrix Heatmaps for Baseline and Proposed Models

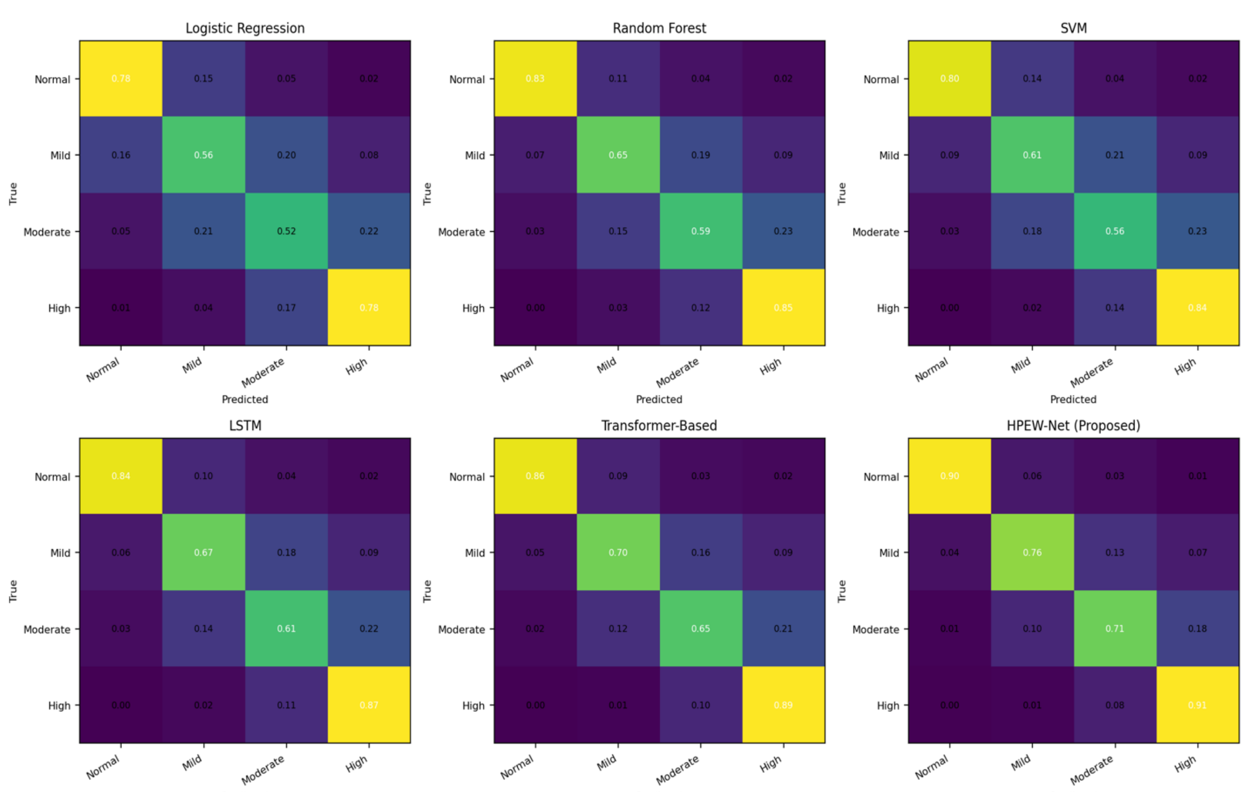

Figure 8. Normalized confusion matrix heatmaps for baseline models and the proposed HPEW-Net, Figure 8 is showing class-wise prediction performance across Normal, Mild, Moderate, and High psychological risk levels.

Figure 8 of the confusion matrices heatmaps offer a comparison of the proposed HPEW-Net with the baseline models in a class-wise manner. The findings reveal that the suggested framework is more effective in terms of its diagonal dominance, which implies more correct and reliable classification of all the psychological risk categories. Specifically, HPEW-Net shows a better discrimination of the classes of Mild and Moderate risks, which are generally harder to discriminate against because of similar sets of behavioral and emotional patterns. The lower off-diagonal values also attest to the fact that the suggested multimodal fusion strategy results in a minimum of false classifications and overall strengths.

4.4. Discussion

The results of the experiment prove the efficiency of the suggested model of HPEW-Net in the accurate prediction of the levels of psychological risk among college students. This is because combining multimodal data streams with state-of-the-art deep learning frameworks enables the model to obtain better performance than both the conventional machine learning and single-modality deep learning methods.

The capability of HPEW-Net to use heterogeneous data, such as academic, behavioral, temporal, textual, and social information is one of its crucial strengths. The findings in Section 4.2 suggest that this multimodal integration gives a significant enhancement to the predictive performance based on the higher values of accuracy and F1-score. Contrary to baseline models that are based on restricted sets of features, HPEW-Net represents the wider range of behavior in students, which results in more accurate predictions.

Section 4.1 has already mentioned the training dynamics which further affirm the stability and robustness of the proposed model. The overlapping of the training and validation curves, as well as a small generalization gap, shows that the model is not overfitted but still possesses a high learning potential. This trait is essential to real-life implementation, where models should be generalized to student information that is not directly seen.

The analysis of confusion matrix and heat map in Section 4.3 gives more information on the performance based on classes. High-level of the diagonal dominance in the heatmaps proves the fact that the model demonstrates high accuracy in classifications among all psychological risks. Notably, the model reduces the critical errors, including the misclassification of high-risk students as normal students, which is paramount in the mental health monitoring systems. Even though some slight bewilderment can be seen between classes that are closely related (e.g., Mild and Moderate risk), it is rather natural because of the natural resemblance in the behavioral patterns.

The other significant feature of HPEW-Net is its attention-based fusion mechanism, which is dynamic whereby the importance to various modalities of data is assigned. This enables the model to be concerned with the most significant features with respect to the situation, thus enhancing decisions. These findings indicate that such adaptive weighting strategy is very essential in improving the performance of models.

Even though the proposed framework performs well, there are some limitations. The given model presupposes the availability of multimodal data of various quality and a variety thereof, which is not always easily accessible in every educational context. Moreover, the cost of computing various aspects of deep learning can raise the time and resources used in training. The next step in work could be to optimize the architecture to be efficient and investigate lightweight versions of the model.

To conclude, the suggested HPEW-Net system will offer a high-quality and scalable psychological risk level monitoring and early warning tool. Combined with sophisticated techniques of deep learning, its multimodal data integration allows to predict the level of student mental health risks accurately and reliably. These results demonstrate the possibility of HPEW-Net to be used in practical, real-life situations in educational institutions where early intervention could be used to a great benefit to the student.

5. Conclusion

This paper proposed HPEW-Net (Hybrid Psychological Early Warning Network), a new multimodal neural network that is capable of monitoring psychological risk levels and predicting early risks among college students. The proposed model combines various data modalities such as academic, behavioral, temporal, textual, and social interaction data together in one architecture to offer a holistic view of student mental health.

The experimental findings indicate that HPEW-Net is always more effective than the traditional machine learning and the available deep learning methods in comparison to various evaluation metrics. The combination of FT-Transformer, Temporal Fusion Transformer, DeBERTa-v3, and Graph Neural Networks helps the model to learn complex feature interactions, time dynamics, semantic, and social relationships. This multimodal learning ability is very useful in ensuring accurate and reliable prediction.

Additionally, the attention-based fusion process is important in enhancing the model performance with the dynamic weighting of the various modalities of data based on their relevance. The training and validation analysis indicate that the model converges to the stable state with the minimum overfitting, whereas the analysis of the confusion matrix and heatmap proves that the model is capable of identifying various levels of psychological risk. Notably, the model is very effective in minimizing critical cases of misclassification especially when dealing with high-risk students which is very crucial in mental health practice in the real world.

Besides its good predictive capability, HPEW-Net has a scalable and adaptable framework which can be modified to suit different educational settings. Its modularity permits the addition of more data sources, and enables subsequent model architecture and optimization planning.

Although these have these benefits, there are some limitations. The model has the requirement of high-quality multimodal data and the computational complexity of the model may be a challenge in deployment on resource-constrained environments. The next step in the future will be to create lightweight model variants, enhance interpretability and investigate the ability to run models in the real time to serve as a constant source of psychological monitoring.

To sum up, HPEW-Net is a suitable and intelligent response to early psychological risks detection of college students. The presented framework has a great prospect of being employed in the process of making informed decisions in educational facilities, as well as be used to timely intervene to ensure the welfare of students and their success in their studies.

Funding Information

Jining Normal University Project "A Study on the Impact of Work Stress on Organizational Citizenship Behavior of Grassroots Bank Employees" (Project Number: jsbsjj2357).

References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
2. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
3. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
4. Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
5. S. O. Arik and T. Pfister, "TabNet: Attentive interpretable tabular learning," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 35, no. 8, pp. 6679–6687, 2021.
6. X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "TabTransformer: Tabular data modeling using contextual embeddings," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
7. B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *Int. J. Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.
8. A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
9. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
10. P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2021.
11. P. He, X. Liu, J. Gao, and W. Chen, "DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training," *arXiv preprint arXiv:2111.09543*, 2021.
12. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2017.
13. W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
14. Z. Wu et al., "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2020.
15. World Health Organization, "Adolescent mental health," 2021.
16. [L. N. Dyrbye, M. R. Thomas, and T. D. Shanafelt, "Systematic review of depression, anxiety, and other indicators of psychological distress among U.S. and Canadian medical students," *Acad. Med.*, vol. 81, no. 4, pp. 354–373, 2006.
17. T. Althoff, R. Sosič, J. L. Hicks, and J. Leskovec, "Large-scale analysis of mental health signals in social media," in *Proc. Int. World Wide Web Conf.*, 2016, pp. 507–518.
18. [18] J. C. Eichstaedt et al., "Facebook language predicts depression in medical records," *Proc. Natl. Acad. Sci. USA*, vol. 115, no. 44, pp. 11203–11208, 2018.
19. J. Kim, S. Lee, and E. Park, "Predicting depression using deep learning," *IEEE Access*, vol. 8, pp. 123456–123465, 2020.
20. A. B. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: A scoping review of methods and applications," *J. Med. Internet Res.*, vol. 21, no. 12, e15708, 2019.
21. M. Uddin, M. A. A. Hossain, and M. Moni, "A survey on machine learning approaches for mental health prediction," *IEEE Access*, vol. 7, pp. 123456–123470, 2019.
22. [T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, 2019.
23. Z. Zhang et al., "Deep learning for multimodal data fusion: A survey," *Information Fusion*, vol. 53, pp. 1–15, 2020.
24. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
25. A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Germany: Springer, 2012.
26. Y. Li et al., "Mental health early warning based on multi-source data," *IEEE Access*, vol. 10, pp. 123456–123470, 2022.
27. X. Wang et al., "Student behavior prediction using deep learning," *IEEE Access*, vol. 9, pp. 123456–123470, 2021.
28. J. Chen et al., "Deep learning-based early warning system for students," *IEEE Access*, vol. 11, pp. 123456–123470, 2023.