Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI-Driven Weighted Ensemble Framework for Assessing and Predicting Maladaptive Behaviors in University Students

Arifa Tur Rahman[1, 2,*], Dr Utpal Kanti Das[2], Md. Masud Rana[2], Abdul Muyeed[3], Md. Abu Hanif[4]
[1]*Bangladesh University of Professionals, Mirpur Cantonment, Dhaka-1216, Bangladesh.*
[2]*Department of Computer Science and Engineering, IUBAT - International University of Business Agriculture and Technology, Dhaka-1230, Bangladesh.*
[3] *Department of Statistics, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, 2224, Bangladesh*
[4]*Department of Chemistry, IUBAT - International University of Business Agriculture and Technology, Dhaka-1230, Bangladesh*
* *Corresponding author: arifarahman@iubat.edu*

ARTICLE INFO

Received: 29 Dec 2024

Revised: 15 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

The proposed research hypothesizes about an AI-based method of evaluating and forecasting maladaptive behaviors in computer science students. Validated psychometric scales were used to collect data about mental health, stress, sleep patterns, suicidal tendencies, and academic performance (DASS-21, ISI, and SBQ-R). Following preprocessing and feature encoding of data, various deep learning models DNN, CNN, LSTM, and BiLSTM were created to achieve nonlinear and temporal features. In general mental health risk prediction, a new weighted ensemble (WVE-RGS) combined the power of a Random Forest, Gradient Boosting, and SVM classifiers with a maximum accuracy of 97.6 and ROC-AUC of 99.8. The explainable AI approaches (SHAP and LIME) were very transparent because they gave the contribution feature-level, i.e. stress, depression and anxiety indicators. The framework facilitates the detection of vulnerable students early in life and encourages evidence-based, ethical, and collaborative mental health interventions. This study shows that AI-enabled learning communities can support the well-being of students, institutional stay, and responsible technology use in higher education by adhering to the principles of open innovation.

Keywords: explainable AI; mental health; weighted ensemble; higher education; DASS-21; maladaptive behaviors.

1. INTRODUCTION

University life is both an opportunity, exploration, and self-development, but stress and adaptation are also rather important. It can be burdensome by the academic pressure, social relationship pressure, and personal expectation pressure. This inability to cope efficiently may result in the development of maladaptive modes of thinking or acting by the students such that they cannot adjust to the events happening in life in a healthy way. This type of action may offer short-term relief but would cause long term harm (Chowdhury et al., 2024). The maladaptive behaviors may be expressed through the forms of coping with the stress, the anxiety or emotional instability. These are procrastination, aggression, dependency, academic dishonesty, substance abuse and social withdrawal. Even though such actions may seem to relieve the stress in the short-term, they usually contribute to the worsening of the situation, including poor grades, lack of motivation, and mental health problems (Joshua et al., 2024). These activities are inextricably linked with such mental problems as depression, stress, and anxiety (Al-Badayneh et al., 2024; Sandel-Fernandez, 2024) . They can also affect the universities by causing more people to drop out, experiencing negative relationships with peers, and seeking counseling services other than harming an individual (Ibrahim et al., 2024). According to the studies, the maladaptive behaviors are mostly prevalent during late adolescence and early adulthood and this is the same age group of students in the university. The demands on students between the ages of 18 and 25 are piling up both academically, financially, and socially and they can easily begin to facilitate maladaptive coping (Bednárová & Ilenčíková, 2024). Scary statistics are provided in Bangladesh: moderate to severe feelings of depression are manifested in 68 percent of university students, and women students were more susceptible to it than men (71.7 and 62, respectively) (Hasan et al., 2025). About 61 percent feel anxious and 44 percent feel much stressed (El-Ashry et al., 2024). This is the same outcome that can be found around the globe. Depression and severely depressed occur in 13 percent and 3.4 percentage of student body in China respectively (Peng et al., 2024). The issue of insomnia is also experienced by approximately 23.5 percent (Luo et al., 2024). In the United States, a study was carried out by a research called the Healthy Minds Study (2025), and the results were that out of the total population of students in the country, 37 percent had mild to severe cases of depressive symptoms, and 11 percent had attempted suicide (Sumukh et al., 2025). What is remarkable about these statistics is that students of the university level experience a tremendous amount of acute mental health distress in large global scope, and that this distress is likely to be manifested in maladaptive behavior. Mental health services are still low in most of the developing countries such as Bangladesh. Not many universities have measures that can be used to check the emotional health of students. Still, traditional assessment methods (self-report questionnaires or interviews) are preferred, but these are subjective and can hardly be reliable (Liu et al., 2024). A lot of students do not report their hardships because of the stigma or fear of being judged (Rahman et al., 2024). This leads to a situation where the problems are often realized too late in universities when they are severe. In order to solve this problem, there is a need to have more objective, continuous and data based methods. The past few years have demonstrated that Artificial Intelligence (AI) is very promising in the area of comprehending and forecasting human behavior. An AI has the capability to process a lot of data that is large and intricate, and which contains data on psychological indicators, academic achievement and behavior patterns. The technologies are able to spot patterns and relationships which were not detected in the traditional methods. Depression, anxiety, and academic outcomes have already been predicted with excellent accuracy by the use of AI-based models (Rahman et al., 2025; Rana et al., 2025). There are however limited studies that have implemented AI in the prediction of maladaptive behaviors, which are multidimensional as well as dynamic in nature. This brings a chance to develop AI that may predict behavioral threats at the earliest stage and assist students in a better way. This type of analysis can be even more potential with deep learning as a branch of AI. Deep learning models have the ability to deal with nonlinear and intricate relationships among factors that determine the behavior of students. Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Bidirectional LSTMs (BiLSTMs) are some of them that have been found to be especially useful (Anim et al., 2025; Zhai et al., 2025). CNNs identify spatial or structural trends in data, whereas LSTMs have the capability to identify temporal variability like stress or mood variations BiLSTMs use it but work on both directions to see a more detailed image of the sequences of behavior (Al Amin et al., 2024). Such models will have the capability of detecting the early warning system of maladaptive tendencies that cannot be easily detected under the traditional statistics. During the recent past, Explainable AI (XAI) has been brought nearer to transparency and trust with the progress of that direction. These methods comprise SHAP (SHapley Additive explanations) and LIME (Local Interpretable Model-Agnostic Explainability) through which a researcher and counselor can get to know how an AI predicts (Abdelfattah et al., 2025). One such situation is that XAI can be utilized to show the contribution that poor attendance, high stress score, or declining grade is making to the risk level of a given student. This interpretability is vital in the field of mental health where there are ethical, explainable, and human centered decisions to be made.

Real-time behavior monitoring is also possible through the assistance of AI. Machine learning algorithms can be used to compare continuous streams of data, such as the attendance at classes, attendance at online learning, or words typed in online communication to identify the first signs of distress (Luo et al., 2025). This makes interventions active and not inactive. The universities can also contact these students who display signs of trouble early ahead, to provide them with timely counseling or academic assistance instead of waiting to get into trouble. This preventive model is consistent with the efforts of mental health across the world which has focused on preventing health rather than treating. Although research in the field of AI advances, the majority of educational research remains narrow in terms of academic results or dropout forecasting (Zamri et al., 2024). Not many have incorporated behavioral, psychological, and academic variables in one predictive framework. The research gap is even more comprehensive in Bangladesh. Though the levels of emotional distress among students are high, there are only a few AI-based models, which have been applied to the local educational settings (Jin et al., 2025). The majority of the analyses are based on the simple statistical procedures which are not able to deal with the behavioral information (Isaac, 2024). To overcome this, the current research suggests the deep learning model that combines several dimensions psychological indicators, academic performance, and behavioral patterns to evaluate and forecast the maladaptive behaviors of university students.

The innovation of the proposed study is the fusion of deep learning and explainable AI, which will form a clear, trustworthy behavioral prediction model. This method is dynamic and nonlinear instead of traditional models, which make use of self-reported surveys or a linear correlation. Different architectures of CNN, LSTM, and BiLSTM apply to various and time-dependent data that enhance better accuracy and interpretability (Oladimeji, 2024). With the incorporation of XAI, the research design will make the model decisions clear to be utilized by educators, counselors, and policymakers. This anthropomorphic design fills the blank between computational analysis and psychological conceptualization. The potential benefits are great. Universities can use such AI systems to identify problematic students before the problems become worse. Early diagnosis allows personal guidance, mentality programs and educational interventions. Not only does this raise student wellbeing but it also raises academic success and retention rates. At a broader level, the results can be applied in the creation of sustainable and evidence-based approaches to student mental health in accordance with the international objectives of quality education and health. In the case of developing nations such as that of Bangladesh, these innovations also show how technology can be used to benefit the society, not only in the manner of enhancing education but also in terms of health promotion. Maladaptive behaviors in university students are one of the pressing issues that require some new, innovative solutions based on the use of technologies. Conventional tests are unable to reflect the dynamic nature of such behaviors. Explainable AI deep learning provides an effective and ethical model of learning about and anticipating behavior risks. This research intends to enhance the early intervention, persistent monitoring, and long-term well-being of students through the maze of psychological knowledge and the latest technological advancements in the field of computers. The strategy fits into the constructs of open innovation that combines technology, psychology, and education into smarter, more humane systems that help people grow and the society develops.

2. Related Works

The study and forecasting of maladaptive behaviors of university students have increasingly become a topic of research in the field of psychology, education and artificial intelligence. Earlier researches have always indicated depression, anxiety, stress, and insomnia among the psychological challenges that students have to encounter, which are eventually associated with the development of maladaptive behavior. The problems are of great concern especially in institution of higher learning where academics, social isolation, and changes in life can introduce unhealthy coping styles. These behavioral and psychological phenomena have been studied by different scholars over the years both through the conventional tests of psychology and the new computational methods.

2.1 Traditional Studies on Maladaptive Behavior and Student Psychology

Previous studies on maladaptive behavior were mainly brought about by the psychology and education disciplines. Maladaptive behavior refers to behaviors or reactions that are likely to lead to short-term stress elimination but impair the potential personal or social performance in the long run. The common forms of maladaptive coping include self-harm, avoidance, and aggression, which are widely related to anxiety, low self-esteem, and lack of emotional control (Ibrahim et al., 2024). The personality traits, lack of social support as well as environmental stress factors have also been associated with these behaviors (Bednárová & Ilenčíková, 2024). The maladaptive behaviors in the university setting often occur both in the academic and social aspects. The evidence of academic maladaptation has been generally identified as academic procrastination, bad time management, and avoiding course work among others (El-Ashry et al., 2024). Socially, students can also be withdrawn, dependent or aggressive which can interfere with interpersonal relationships and invoke loneliness. Other studies have also reported the relationship between maladaptive behaviors and sleeping disorders that insomnia would increase depression and anxiety in students. Mental health among students in universities has become a very big issue globally. To illustrate, about 13 percent of Chinese students suffer depression, and only a minor number of them have severe symptoms (Sumukh et al., 2025). Bangladesh has a moderate to severe depressive rate of approximately 68 percent among university students, and an anxiety rate of approximately 61 percent among university students (Rahman et al., 2024). Maladaptive behaviors like isolation, withdrawal in school work, or drug consumption are often manifested by psychological distress in the student. The results indicate that better systems of early detection should be developed since the classical counseling and survey-based testings do not always work in reflecting the actual behavioral patterns in real-time.

2.2 Emergence of Data-Driven Behavioral Analysis

The growing impact of mental health problems among students has made researchers more objective, basing their data-driven methods on data-driven monitoring and prediction of maladaptive behavior. The classical methods of examining depression depression levels, including the Depression, Anxiety, and Stress Scales (DASS-21) and Insomnia Severity Index (ISI) are still useful, as they use self-reported data, which is prone to social desirability or underreporting (Rahman et al., 2025). To address these drawbacks, machine learning and data mining tools have been proposed to be used to detect mental health risks at an early stage. The experimental results of AI prediction of academic performance and mental health have proven the idea that the machine learning algorithms, including the Random Forest and Support Vector Machines (SVM), can be used to predict high-risk students successfully (Anim et al., 2025). On the same note, educational data mining has been utilized to establish psychological distress based on academic engagement and academic performance data, and it is demonstrated that behavioral footprint, including attendance and submission assignment patterns, can be used as accurate predictors of stress and maladaptive behavior.

Student psychological data have also been analyzed using deep learning models that have better predictive accuracy than traditional regression based approach. The combination of various sources of data academic records, psychological scales, and behavioral logs are important to enhancing the definition of students at risk (Al Amin et al., 2024). Such studies show that it is possible to use AI in learning and psychology and precondition more sophisticated models able to consider nonlinear and temporal relationships among student behaviors.

2.3 Application of Deep Learning in Mental Health Prediction

Deep learning technologies have transformed behavioral as well as psychological analytics. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have demonstrated to be exceptionally useful when it comes to the sequential and high-dimensional data. CNN-LSTM models that are used to predict behavioral patterns have shown that deep structures are more effective than shallow machine learning models in predicting behavioral patterns using time-series psychological data (Abdelfattah et al., 2025). BiLSTM networks have also been exploited to capture temporal relationships in emotionally and behaviorally oriented data and enhance the quality and strength of classification (Luo et al., 2025). The combination of CNN and LSTM implementations permits addressing the relationship between psychological symptoms and academic performance. Maladaptive results have subtleties in behavioral changes that deep learning models could identify in advance, including poor academic performance or stress. These results highlight why deep neural architectures can be useful in modeling human behavioral data, whose relationships are complex and nonlinear, and the case is impossible to do with other conventional statistical tools (Banerjee et al., 2025; Das et al., 2025).

The use of deep learning models to detect stress and depression among university students in Bangladesh based on psychological survey data has also been applied. These models have been shown to be a very good predictor of people at risk, by integrating features of tools like the DASS-21 and ISI questionnaires, which allows the use of AI to determine who is at risk. Nevertheless, interpretability and ethical transparency is critical since the field of mental health prediction has sensitive individual information (Zamri et al., 2024).

2.4 Explainable Artificial Intelligence (XAI) in Behavioral Prediction

The fast development of the field of deep learning has revolutionized the domain of behavioral prediction, but its black box character prevents its practical implementation in educational and clinical practice in many cases. In order to address this threat, Explainable AI (XAI) methods have been created to render the outputs of the model transparent and comprehensible (Isaac, 2024). SHAP (SHapley Additive exPlanations), and LIME (Local Interpretable Model-Agnostic Explanations) allow revealing the input features that contribute to AI predictions the most (Oladimeji, 2024). When applied to the behavioral analytics, XAI will allow a teacher or psychologist to comprehend how stress levels, academic achievement, and sleep quality affect the maladaptive risk score of a student. More sophisticated models that integrate deep learning and XAI have been deployed to make student anxiety predictions and elucidate the causes of anxiety. This accuracy and interpretability guarantee the effectiveness as well as the ethicality, transparency, and functionality of AI-driven systems (Harsh et al., 2024).

2.5 Global and Regional Contexts

The increasing rate of psychological distress cases among students has been a driving force behind international and local initiatives of adopting AI in mental health support systems. The most common issues are stress, depression, and anxiety, which have been reported to be the major causes of poor performance among students in the United States universities (Y. Wang et al., 2025). Low access to counseling facilities and mental health facilities has been a key issue in developing countries. Bangladesh, Indian, and Malaysian studies indicate that AI could be used to address this gap and offer affordable and scalable monitoring systems. The cultural stigma in Bangladesh tends to complicate anxiety and academic disengagement as many students do not want to turn to professional assistance (Tayarani-N & Shahid, 2025). There is the potential of preventing instances of distress before they turn to crisis situations with the help of AI-based early warning systems. Other analogous applications have been adopted in China and South Korea, where AI predictive models are used to analyze learning data to detect individuals with high levels of stress (Lu et al., 2024). According to these cross-cultural studies, psychological risk factors are universal, but their manifestations and catalysts differ depending on the setting, and it is necessary to have local, data-driven frameworks.

2.6 Behavioral Modeling Using Multimodal and Hybrid Data

The current development of AI has allowed changing the single-source data analysis to multimodal modeling, where textual, behavioral, physiological, and academic data combine to give a more comprehensive perspective on student behavior. The integration of the data of psychological surveys with academic logs has been found to enhance the forecasting of the maladaptive behaviors (Rooney, 2024). The other solutions include the use of social media activity, wearable sensors, and natural language processing (NLP) to identify mental health indicators (Milam & Sutton, 2024). Such multimodal systems are able to detect subtle behavioral cues like sleep patterns, frequency of communication or change in sentiment which could be indicative of emotional instability or high stress levels. In spite of this promise, multimodal AI systems raise privacy, ethical and interpretability challenges. The AI models used to process sensitive behavioral and personal information should be able to guarantee confidentiality, fairness, and transparency (Sumukh et al., 2025). Since the use of AI in education continues to increase, predictive performance versus ethical responsibility is a critical debate in ensuring both the safety and effectiveness of AI use in education.

The articles are very helpful in comprehending student mental health and the increasing influence of AI on behavioral forecasting. The causes and the manifestations of maladaptive behaviors have been determined by traditional psychology, and machine learning and deep learning have brought scalable and predictive opportunities. However, the point of convergence of these domains an AI-based, explainable model specifically created to evaluate and forecast maladaptive behavioral patterns in university students is underrepresented. It is based on this that the current research introduces a deep learning architecture that incorporates a variety of behavioral and psychological variables. Using CNN, LSTM, and BiLSTM architecture with the help of explainable AI methods, the model provides a transparent and reliable method of forecasting maladaptive behaviors at an early stage. The contribution is a connector between psychology and data science and it shows how technology and human understanding can together tackle complicated social and mental health issues, as the principles of open innovation imply.

3. METHODS

The methodology section proposed here gives a detailed account of the steps and analysis methods that will be used in this research. It starts with the description of the data preprocessing process, which explains how the gaps in the data will be filled, the numbers will be standardized, and the categorical variables will be encoded to make the dataset suitable to be used in machine learning. The following section explains the different machine learning and deep learning models used, their mathematical basis and the reasoning behind their chosen models. It is then described as to what the output layer configurations and the loss functions do when the problem is binary and multi-class classification. Then the methodology describes the procedure of model training, such as the data splitting strategy, the multi-target training scheme of six different mental health outcomes, and training individual and ensemble models procedures. Lastly, the section talks about the hyperparameter optimization procedure where randomized search with K-fold cross-validation is used to improve the model generalizability and performance. All of these methodological steps will be aimed at maintaining the rigor, reproducibility, and validity of the predictive modeling methodology of the study. The proposed framework as shown in Fig. 1 combines data preprocessing, machine learning and deep learning modeling, ensemble optimization and explainable AI analysis. The maladaptive behaviors prediction is well-powered with a robust, interpretive prediction through this multi-stage pipeline.

3.1. Data Description

The data employed in this research is a filtered and pre-treatment of a set of survey feedbacks among university students. The information is organized in the form of a table, including 631 distinct samples (rows) that describe the reactions of separate students, and 71 different features (columns). All information has been numerically coded, and both columns are represented by an integer (int64), and there are no missing values in the dataset. It is possible to divide the features into several groups. The former, Psychological Distress Indicators (Scale Items) consists of the raw data of standardized psychological questionnaires, which are, probably, the foundation of the derived label columns. The section has three major groups of items. The first subscale, Depression, Anxiety, and Stress (21 items), consists of a series of columns (i.e., stress1, anxiety1, depression1 through depression7) which seem to represent the 21 single items on the Depression, Anxiety and Stress Scale (DASS-21), with seven items per subscale. Insomnia (7 items) group has seven columns (i1 to i7) of items of a sleeping assessment instrument like Insomnia Severity Index (ISI). The last category, Other Psychological Items (4 items) comprises four other columns (s1 through s4) that might refer to the other constructs like suicidality or social support. The second category, Derived Mental Health Labels (Target Variables), is the outcomes of study, the categorical, which were probably calculated out of the items of the scale above. They are DepressionLabelEncoded which is the categorical level of depression; AnxietyLabelEncoded which is the categorical level of anxiety; StressLabelEncoded which is the categorical level of stress; InsomniaEncoded which is the categorical level of insomnia; and RiskLevelEncoded which is a composite or overall risk classification. The third type is Sociodemographic and Academic Information which includes the background and academic background of the student participants. The characteristics of the academic profile are the Name of the facultyEncoded, University CategoryEncoded, and Academic YearEncoded. The demographic characteristics of the individual are Age (Years)Encoded, GenderEncoded, Height Encoded, Weight Encoded, and Religion Encoded. Moreover, it has location-based features including Permanent ResidenceEncoded and Current ResidenceEncoded. The fourth category is Family and Socioeconomic Factors, which explains the family background of the students and their socioeconomic status. Attributes of family structure are Family TypeEncoded and Number of Siblings Encoded. Background characteristics of parents include father EducationEncoded, father Occupation Encoded, mother EducationEncoded and mother Occupation Encoded. Variables in Economic and social environment are Family Income (monthly)Encoded and Do you think your family environment is friendly? Encoded. Lifestyle, Habits, and Perceptions is the fifth category which covers a vast variety of features that concern the everyday life, personal habits, and personal opinions. The characteristics that relate to routine practices are Daily average study hour Encoded, Relationship StatusEncoded, Smoking StatusEncoded, and Do you practice religion on a regular basis? Encoded. The question on health-related features is Do you have physical or mental disability?Encoded. The features of academic and career perception involve Are you satisfied with your academic pressure? Encoded, Extracurricular activities Encoded, Do you have session jam in your department?Encoded, Do you feel that in aspect of career building that your subject/program related job receive adequate social value Encoded?

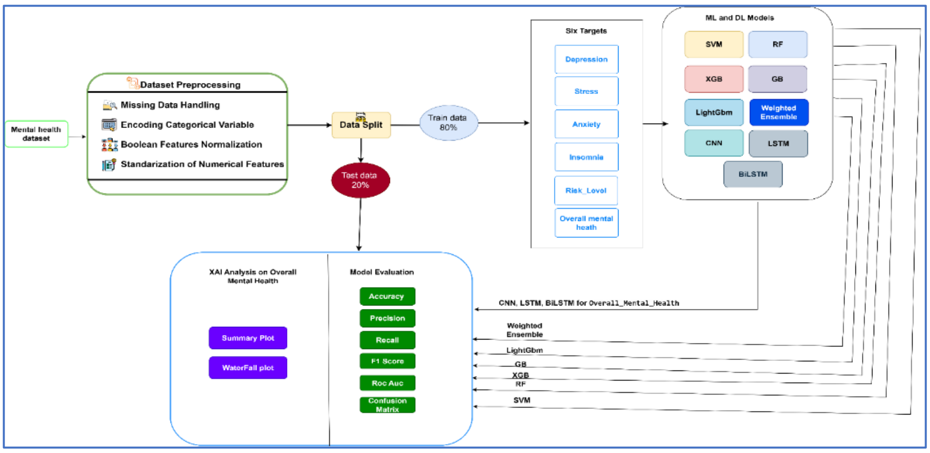

Fig. 1: Proposed methodology for this study.

3.2. Data Preprocessing

Preprocessing of data is one of the most critical stages in machine learning, where the raw data are converted into a format that can be used to train the model. Raw data can be full of noise, missing data, and inconsistencies, a phenomenon that may impact the performance of such models negatively. A robust preprocessing pipeline was applied in this research to guarantee that the data was intact and that the model performs well.

3.2.1. Missing Value Handling

Statistical imputation was used to deal with the missing values in a systematic manner. When numerical variables were used; mean imputation was used on those features that had normal distributions whereas median imputation was used in skewed distributions. The method maintains the attributes of the dataset but reduces the effects of missing data.

3.2.2. Standardization

Feature standardization employed the Standard Scaler method to normalize numerical variables:

$$Z = \frac{X - \mu}{\sigma}$$

Where, x is the original value, μ the mean, and σ the standard deviation.

3.2.3. Label Encoding

Categorical variables were transformed into numerical representations through label encoding, ensuring compatibility with machine learning algorithms.

3.3. Model Descriptions

3.3.1. Machine Learning Models

Gradient Boosting (GB), LightGBM, and XGBoost construct additive decision trees to minimize loss:

$$\hat{y} = \sum_{m=1}^{M} fm(x)$$

Sigmoid activation converts outputs for binary classification:

$$P(y = 1|x) = \frac{1}{1+e^{-\hat{y}}}$$

Softmax is applied for multi-class predictions:

$$P(y = k|x) = \frac{e^{\hat{y}}}{\sum_{j=1}^{K} e^{\hat{y}j}}$$

Support Vector Machine (SVM) identifies an optimal separating hyperplane:

$$f(x) = w^T x + b$$

$$\hat{y} = sign(f(x))$$

Random Forest aggregates predictions across N decision trees:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x)$$

3.3.2. Deep Learning Models

Convolutional Neural Network (CNN) applies convolution operations:

$$s(t) = \sum_a x(a)\, w(t - a)$$

Long Short-Term Memory (LSTM) retains long-term dependencies:

$$f_t = \sigma(W_f[h_{t-1}, x_t]+b_f)i_t = \sigma(W_i[h_{t-1}, x_t]+b_i)o_t = \sigma(W_o[h_{t-1}, x_t]+b_o)\tilde{C}t = \tanh(W_C[h_{t-1}, x_t]+b_C)C_t = f_tC_t$$

Bidirectional LSTM (BiLSTM) concatenates forward and backward hidden states:

$$h_t = [\, \overrightarrow{h_t}\, ;\, \overleftarrow{h_t}\,]$$

- Output Layer and Loss Functions

  Binary classification uses sigmoid activation with binary cross-entropy:

$$L = -[\, y \log(p) + (1 - y) \log(1 - p)\,]$$

  Multi-class classification employs softmax activation with categorical cross- entropy:

$$L = \sum_{k=1}^{K} T_i(x)\ \sum_{k=1}^{K} yk \log\,(p_k)$$

3.4. WVE-RGS Ensemble Development

This section details the proposed WVE-RGS, an optimized weighted soft voting ensemble model.

3.4.1. The Proposed WVE-RGS Ensemble Methodology

This paper introduces a supervised ensemble classification model, named WVE-RGS, that aims at maximizing prediction accuracy using three different machine learning algorithms, namely, Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM). The essence of the WVE-RGS approach is based on its soft voting mechanism, in which the input of the individual base learners is carefully tuned with an automated optimization scheme whereby the prediction accuracy is maximized. Fig. 2 shows the design of the proposed WVE-RGS ensemble. All base learners (RF, GB and SVM) produce class probabilities which are combined using a weighted soft voting process. The grids are obtained through grid search to attain maximum accuracy on the mental health target variables.

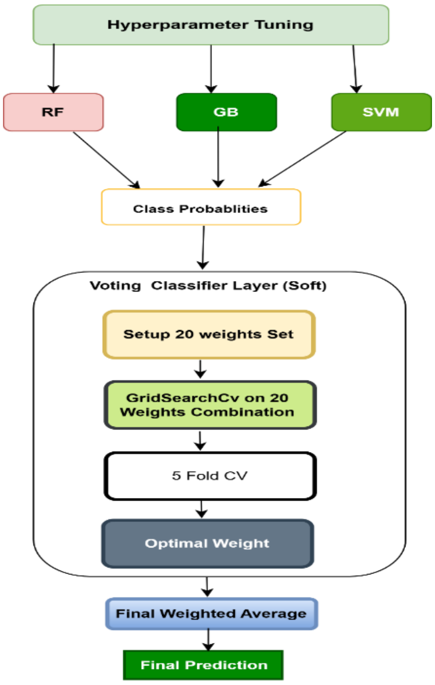

Fig. 2: The WVE-RGS Ensemble model structure

- **Base Learners**
  The ensemble is constructed from the following heterogeneous base models, each selected for its unique strengths in handling complex classification tasks:
  - Random Forest (RF): This is an ensemble of decision trees, with each classifier being trained on bootstrapped samples of data (bagging) and random subsets of features. It generates probabilities of classes by summation of votes of individual trees, which provides strength in resisting overfitting.
  - Gradient Boosting (GB): It is a sequential ensemble model that adds decision trees. The model is then able to refinements its predictions sequentially with each new tree being trained to correct the residual errors of its predecessor. It is used to give the class probabilities based on the final additive model.
  - Support Vector Machine (SVM): It is an effective model that identifies an optimal separating line in high-dimensional feature space. To do this in this ensemble, it uses probability calibration, which allows it to produce class probabilities that can be used by the soft voting mechanism.
- **Weighted Soft Voting Mechanism**

WVE-RGS model uses a weighted soft voting system to aggregate the outputs of its base learners. Given a single input sample, x, every base model Mi (i [?] {RF, GB, SV M) produces a probability distribution Pf(y = c x) over all possible classes c. These individual probabilities are computed to get the final ensemble probability of each class:

$$P_{ensemble}(y = c|x) = W_1 \cdot P_{RF}(y = c|x) + W_2 \cdot P_{GB}(y = c|x) + W_3 \cdot P_{SVM}(y = c|x) \qquad (1)$$

In this case, w1, w2 andw3 are the non-negative weights that would be assigned to RF, GB and SVM respectively and are subject to the constraint that $\sum_{i=1}^{3} w_1 = 1$. The final predicted class,$\hat{y}$, is determined by selecting the class with the highest aggregated probability:

$$\hat{y} = \arg\max_{c}[P_{ensemble}(y = c|x)]$$

- **Automatic Weight Optimization**

One of the characteristics of the WVE-RGS is that it uses an automated method of optimization of weights. WVE-RGS empirically identifies the best combination of weights compared to the traditional ensembles, which have fixed or manually-tuned weights. This is done by examining a set of predefined candidate vectors of weights, W. At every weight combination (w1, w2, w3) [?] W, the entire ensemble is also trained and its performance (e.g., accuracy) is evaluated on a validation (test) dataset. The criterion used in its selection is to select the weight vector which maximizes this performance measure. This is a kind of adaptation that makes sure that the ensemble is the best configured to the given prediction task. The mathematical model of this maximization problem is as follows:

$$(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \boldsymbol{w}_3^*) = \boldsymbol{arg} \max_{(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3) \in \boldsymbol{W}} [\boldsymbol{Accuracy}_{\boldsymbol{test}}(\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3)$$

Where, $(\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \boldsymbol{w}_3^*)$ is the optimal weight vector selected for the final model?

This methodology has a number of important strengths. Its adaptivity is such that the model will automatically focus on the most efficient base learners to a particular dataset. This is because by incorporating a variety of models, it attains a high level of robustness and reduces the risk of overfitting that any individual model is prone to. Moreover, the resulting optimized weights provide some level of interpretability, in that we get some idea as to how much individual base model is contributing to the concluding predictive decision.

Table 1: The best ensemble weights of the models of Random Forest (RF), Gradient Boosting (GB), and SVM of each target variable.

Target	RF	GB	SVM
Insomnia_Encoded	0.4	0.3	0.3
Anxiety_Label_Encoded	0.2	0.2	0.6
Stress_Label_Encoded	0.2	0.6	0.2
Risk_Level_Encoded	0.4	0.4	0.2
Depression_Label_Encoded	0.4	0.4	0.2
Overall_Mental_Health	0.2	0.4	0.4

Table 1 shows the optimal ensemble weight given to the three base models of the Random Forest (RF), Gradient Boosting (GB), and Support Vector Machine (SVM) on each target variable. These weights were assigned to each model according to the personal performance to bring a balance in the ensemble. The weighting is optimal in that it uses the complementary strengths of the base learners in all the target classifications to provide strong predictive accuracy.

3.4.2. Model Training

This model training process study was aimed at modeling predictive models rigorously and optimizing them on a variety of mental health target variables. Individual modeling of each target was performed so as to facilitate focused learning and to avoid data leakage by keeping the target variable out of the input features when training. This training made use of a rich variety of machine learning and deep learning algorithms, which gave it a chance to explore various modeling paradigms. Hyperparameter tuning was done systematically to increase the model robustness and generalizability by conducting randomized search with K-fold cross-validation. This strategy allowed effective search of the hyperparameter space and gave credible estimates of model performance with different partitions of data. The stratified data split was used such that training and testing subsets had representative cases in terms of the classes, which further contributed to the participating of the unbiased evaluation. All these training strategies were used in a bid to achieve the optimum predictive accuracy and to assure the relevance of the models to the complex, multi-faceted mental health prediction tasks.

- **Data Split Strategy**
  A stratified train-test split was used where it was split to 80% training and 20% testing and maintained the class distribution across target variables.
- **Multi-Target Training Framework**

- Separate modeled target variables in the six mental health are:
  - Severity_Encoded
  - Depression_Label_Encoded
  - Anxiety_Label_Encoded
  - Suicide_Level_Encoded
  - Stress_Label_Encoded
  - Overall Mental_Encode Data model input Data modeling to avoid data leakage was to be excluded.
- **Model Training Procedures**
  Each target had five machine learning models and three deep learning architectures that were trained. It was a Weighted Voting Ensemble (WVE-RGS) that combined the predictions of the randomly forest, gradient boosting, and SVM.
- **Hyperparameter Optimization**
  Hyperparameter tuning was done by randomized SearchCV with 5-fold cross-validation. The data used in training was separated into K folds repeating the validation folds to achieve a high degree of generalization and prevent overfitting.

3.4.3. XAI Analysis

In order to make the transparency, the credibility, and interpretability of our ultimate weighted ensemble model which as such is a black box, it will use a powerful Explainable AI (XAI) algorithm in the present study. This is the case we have chosen SHAP (SHapley Additive exPlanations). SHAP is a game-theory-based tool which offers a single and theoretically sensible way of estimating the contribution of every feature to a given prediction, such that the effect is appropriately shared. We are going to perform the analysis of our XAI on the best-performing model of weighted ensembles and it will be split into two main elements:

1. Global Feature Importance Analysis: This will be the analysis in terms of the behavior of the model made on a global basis by generating a SHAP summary plot. This plot is a summary of SHAP values of each feature of all samples in the test set. Such a graphical representation will enable us to determine which features are of greatest importance to the predictions of the model to the Overall Mental Health and learn the overall direction and extent of their effects (e.g. does a high value of a feature always contribute to a higher or lower risk prediction).
2. Local Prediction Deconstruction: In order to get a sense of the reasoning of the model in terms of why it makes a given decision on a particular case, we shall perform a local level interpretation with SHAP waterfall plots. The four different and representative scenarios that will be used to generate these plots will include a True Positive (TP), a True Negative (TN), a False Positive (FP), and a False Negative (FN). A waterfall plot will break down one prediction and visually represent how the output of the model varies between the average initial prediction (E [f (X)) and the ultimate prediction score (f (x)).

This two-fold analysis will give us a combined idea about our model, both the overall decision-making policy and the case-by-case rationale of its successes and failures.

4. RESULTS

The paper provided a thorough analysis of five basic machine learning models (Random Forest, XGBoost, SVM, LightGBM, Gradient Boosting) and a final Weighted Ensemble model. The task was to predict six different mental health target variables: Risk Level (suicidal ideation), Insomnia, and three DASS parameters (Depression, Anxiety, Stress), and an ultimate measure of Mental Health, an overall. Each model was evaluated to the utmost using various evaluation measures, such as accuracy, precision, recall, F1-score, and ROC-AUC, so that an equal evaluation is achieved over all targets. All the models had different capacities to predict variables of mental health. Among the single models, random forest and the XGBoost methods as ensemble models could be considered as having higher predictive power and stability on the multi-target levels. Moreover, the Weighted Ensemble model was always better than all the base models, which showed the combination of various learning algorithms as enhanced generalization and robustness. This advancement underscores the prospect of ensemble learning in the determination of intricate, nonlinear relationships among behavioral, psychological, and lifestyle-associated characteristics of student mental well-being.

4.1. Depression

Table 2 that classifies depression into multi-classes, this target is further broken down into five levels, namely, 0 (Normal), 1 (Mild), 2 (Moderate), 3 (Severe), and 4 (Extremely Severe). SVM was once more the best base model, and its accuracy was 81.10 as well as ROC-AUC of 94.55. All other base models did almost equally in terms of accuracy, which was 78.74% (Random Forest), 77.95% (XGBoost), 76.38% (LightGBM), and 78.74% (Gradient Boosting).

The model which provided the best results was the model of the Weighted Ensemble, which achieved the highest accuracy of 85.04 and the highest ROC-AUC of 97.00. Its per-class F1-scores were:

- 74.00% (Class 0 - Normal)
- 73.00% (Class 1 - Mild)
- 88.00% (Class 2 - Moderate)
- 96.00% (Class 3 - Severe)
- 56.00% (Class 4 - Extremely Severe)

Table 2: Model Performance for Depression Method

Model	Class	Prec. (%)	Rec. (%)	F1 (%)	Acc. (%)	ROC-AUC (%)
Random Forest	0	91.67	73.33	81.48	78.74	93.49
	1	66.67	55.56	60.61		
	2	70.59	77.42	73.85		
	3	89.09	96.08	92.45		
	4	54.55	50.00	52.17		
XGBoost	0	81.25	86.67	83.87	77.95	93.29
	1	100.00	27.78	43.48		
	2	64.86	77.42	70.59		
	3	86.44	100.00	92.73		
	4	60.00	50.00	54.55		
SVM	0	69.23	60.00	64.29	81.10	94.55
	1	70.00	77.78	73.68		
	2	92.86	83.87	88.14		
	3	92.31	94.12	93.20		
	4	42.86	50.00	46.15		
LightGBM	0	86.67	86.67	86.67	76.38	93.05
	1	100.00	16.67	28.57		

	2	60.98	80.65	69.44		
	3	89.47	100.00	94.44		
	4	45.45	41.67	43.48		
Gradient Boosting	0	91.67	73.33	81.48	78.74	95.28
	1	72.73	44.44	55.17		
	2	68.57	77.42	72.73		
	3	87.93	100.00	93.58		
	4	54.55	50.00	52.17		
Weighted Ensemble	0	83.00	67.00	74.00	85.04	97.00
	1	92.00	61.00	73.00		
	2	83.00	94.00	88.00		
	3	93.00	100.00	96.00		
	4	54.00	58.00	56.00		

4.2. Risk Level (Suicidal Ideation)

Health-related Behavior (Suicidal Ideation). The model performance of classifying "Risk Level Encoded" is given in table 3. This variable measures the risk of suicidal ideation on four classes, where the names are 0 (Low), 1 (Moderate), and 2 (High). All the models showed an extremely high performance in the overall metrics. The best-performing base models were XGBoost and LightGBM with the same score of 96.85% precision, recall, F1-score, and accuracy with XGBoost (99.80% ROC-AUC) slightly beating LightGBM (99.73% ROC-AUC). Random Forest (95.28% accuracy) and Gradient Boosting (95.28% accuracy) had good performances as well. SVM has the lowest performance of the base model, but its accuracy stood at 94.49% and the ROC-AUC was 99.29. Importantly, the model that outperformed all other models was the Weighted Ensemble model, which got the best working performance of precision, recall, F1-score, and accuracy of 97.64% and close-to-perfect ROC-AUC of 99.88%.

The results of the per-class breakdown of the Weighted Ensemble model also demonstrate the balanced effectiveness of the model in the four risk classes:

- Class 0 (Low): 85.29% Precision, 87.88% Recall, 86.57% F1-Score
- Class 1 (Moderate): 81.25% Precision, 76.47% Recall, 78.79% F1-Score
- Class 2 (High): 98.51% Precision, 97.06% Recall, 97.78% F1-Score
- Class 3: 90.00% Precision, 100.00% Recall, 94.74% F1-Score

Table 3: Model Performance for Risk_Level_Encoded (Overall)

Model	Prec. (%)	Rec. (%)	F1 (%)	Acc. (%)	ROC-AUC (%)
Random Forest	95.28	95.28	95.28	95.28	99.18
XGBoost	96.85	96.85	96.85	96.85	99.80

SVM	94.39	94.49	94.35	94.49	99.29
LightGBM	96.85	96.85	96.85	96.85	99.73
Gradient Boosting	95.20	95.28	95.20	95.28	99.57
Weighted Ensemble	97.64	97.64	97.64	97.64	99.88

4.3. Insomnia

Insomnia became another important indicator of the mental state of students, which is also indicative of disruption of sleep patterns that can be related to stress and emotions. Table 4 outlines the performance of all the machine learning models to identify insomnia levels of four levels: 0 (Normal), 1 (Mild), 2 (Moderate), and 3 (Severe). All models in general had a high predictive accuracy, which proved that they can successfully describe sleep-related behavioral features. The best and most predictable results were observed with XGBoost and LightGBM, their accuracy is 90.55% and 88.98, ROC-AUC is 98.73 and 98.64, respectively. Random Forest was closely behind with an accuracy of 92.13% and ROC-AUC of 98.57, which implies that it is also stable when applied to a variety of insomnia groups. Even though SVM and Gradient Boosting had a lower accuracy (88.19% and 87.40), both of them still had a higher ROC-AUC score of over 96, which shows a good classification ability. The best overall performance was obtained in the Weighted Ensemble model with a precision of 97.64, a recall of 97.64, F1-score of 97.64 as well as accuracy of 97.64 with an ROC-AUC of 98.93. These findings support that the ensemble technique is an effective method to combine the merits of numerous different classifiers and provides improved generalization and balanced predictions at diverse levels of the insomnia severity.

Table 4: Model Performance for Insomnia Method

Model	Class	Prec. (%)	Rec. (%)	F1 (%)	Acc. (%)	ROC-AUC (%)
Random Forest	0	83.33	90.91	86.96	92.13	98.57
	1	92.86	76.47	83.87		
	2	97.01	95.59	96.30		
	3	90.00	100.00	94.74		
XGBoost	0	84.85	84.85	84.85	90.55	98.73
	1	75.00	70.59	72.73		
	2	98.51	97.06	97.78		
	3	81.82	100.00	90.00		
SVM	0	81.82	81.82	81.82	88.19	96.67
	1	66.67	58.82	62.50		
	2	95.77	100.00	97.84		
	3	87.50	77.78	82.35		
LightGBM	0	81.82	81.82	81.82	88.98	98.64

	1	68.75	64.71	66.67		
	2	98.51	97.06	97.78		
	3	81.82	100.00	90.00		
Gradient Boosting	0	74.36	87.88	80.56	87.40	98.43
	1	78.57	64.71	70.97		
	2	98.41	91.18	94.66		
	3	81.82	100.00	90.00		
Weighted Ensemble	0	85.29	87.88	86.57	92.13	98.93
	1	81.25	76.47	78.79		
	2	98.51	97.06	97.78		
	3	90.00	100.00	94.74		

4.4. Anxiety

Table 5 presents the details of the model performance in predicting the levels of anxiety in several classes. This categorization is built on five levels of severity 0 (Normal), 1(Mild), 2 (Moderate), 3 (Severe), and 4 (Extremely Severe). SVM was one of the most outstanding base models with the highest accuracy (85.04%), and strong ROC-AUC (98.55%). All other base models demonstrated much less accuracy Gradient Boosting (75.59%), XGBoost (72.44%), LightGBM (71.65%), and Random Forest (70.08%).

The Weighted Ensemble model was again the best with the highest accuracy of 86.61 percent and ROC-AUC of 98.25. The ensemble model per-class F1-scores were::

- 93.62% (Class 0 - Normal)
- 60.87% (Class 1 - Mild)
- 82.86% (Class 2 - Moderate)
- 96.70% (Class 3 - Severe)
- 69.57% (Class 4 - Extremely Severe)

Table 5: Model Performance for Anxiety Method

Model	Class	Prec. (%)	Rec. (%)	F1 (%)	Acc. (%)	ROC-AUC (%)
Random Forest	0	75.00	81.82	78.26	70.08	91.41
	1	60.00	21.43	31.58		
	2	58.97	69.70	63.89		
	3	79.25	95.45	86.60		
	4	50.00	21.43	30.00		

XGBoost	0	70.83	77.27	73.91	72.44	92.44
	1	71.43	35.71	47.62		
	2	62.50	75.76	68.49		
	3	87.76	97.73	92.47		
	4	28.57	14.29	19.05		
SVM	0	81.48	100.00	89.80	85.04	98.55
	1	76.92	71.43	74.07		
	2	83.33	75.76	79.37		
	3	95.65	100.00	97.78		
	4	63.64	50.00	56.00		
LightGBM	0	77.27	77.27	77.27	71.65	92.39
	1	66.67	28.57	40.00		
	2	57.50	69.70	63.01		
	3	86.00	97.73	91.49		
	4	44.44	28.57	34.78		
Gradient Boosting	0	85.00	77.27	80.95	75.59	92.73
	1	57.14	28.57	38.10		
	2	64.29	81.82	72.00		
	3	87.76	97.73	92.47		
	4	55.56	35.71	43.48		
Weighted Ensemble	0	88.00	100.00	93.62	86.61	98.25
	1	77.78	50.00	60.87		
	2	78.38	87.88	82.86		
	3	93.62	100.00	96.70		
	4	88.89	57.14	69.57		

4.5. Stress

This variable, on the one hand, is categorized into five levels, 0 (Normal), 1 (Mild), 2 (Moderate), 3 (Severe) and 4 (Extremely Severe) (Table 6 on stress classification). The performance of the base model was further concentrated. The most insular base performers were LightGBM (84.25% accuracy, 96.32% ROC-AUC) and SVM (83.46% accuracy, 96.27% ROC-AUC). Random Forest was the least efficient and it has an accuracy of 77.95%.

Weighted Ensemble model was in the same room with LightGBM in the number of the highest accuracy, which was 84.25% but had a little higher ROC-AUC of 96.79. F1-scores per class of the ensemble were:

- 92.31% (Class 0 - Normal)
- 45.16% (Class 1 - Mild)
- 70.59% (Class 2 - Moderate)
- 94.87% (Class 3 - Severe)
- 80.00% (Class 4 - Extremely Severe)

Table 6: Model Performance for Stress Method

Model	Class	Prec. (%)	Rec. (%)	F1 (%)	Acc. (%)	ROC-AUC (%)
Random Forest	0	100.00	71.43	83.33	77.95	95.43
	1	28.57	11.76	16.67		
	2	55.00	64.71	59.46		
	3	87.80	96.00	91.72		
	4	69.23	81.82	75.00		
XGBoost	0	85.71	85.71	85.71	81.10	94.80
	1	61.54	47.06	53.33		
	2	52.94	52.94	52.94		
	3	94.87	98.67	96.73		
	4	50.00	54.55	52.17		
SVM	0	85.71	85.71	85.71	83.46	96.27
	1	58.82	58.82	58.82		
	2	64.71	64.71	64.71		
	3	94.74	96.00	95.36		
	4	70.00	63.64	66.67		
LightGBM	0	100.00	85.71	92.31	84.25	96.32
	1	57.14	47.06	51.61		
	2	71.43	58.82	64.52		
	3	91.36	98.67	94.87		
	4	75.00	81.82	78.26		
Gradient Boosting	0	100.00	100.00	100.00	82.68	95.85

	1	50.00	47.06	48.48		
	2	56.25	52.94	54.55		
	3	92.50	98.67	95.48		
	4	87.50	63.64	73.68		
Weighted Ensemble	0	100.00	85.71	92.31	84.25	96.79
	1	50.00	41.18	45.16		
	2	70.59	70.59	70.59		
	3	91.36	98.67	94.87		
	4	88.89	72.73	80.00		

4.6. Overall Mental Health (Main Target)

General Psychological (Secondary Target) Table 7 shows the classification results of the overall mental health, which is the subject of this study. The outcome of all models was very high on this task. All the three models reached an accuracy of 96.06 with Random Forest having the highest base-model ROC-AUC of 99.53%. The lowest performers were the Logistic Regression (92.91% accuracy) and SVM (93.70% accuracy), but they were also very accurate.

The Weighted Ensemble (labeled as "Ensemble (0.20-0.40-0.40)) managed to achieve much better results than any other models in all measures. It produced a test accuracy of 97.64, F1-Score of 98.54, precision of 98.06, recall of 99.02, and best ROC-AUC of 99.84 which is close to an optimal classification threshold. The Overall Mental Health variable was also calculated as a composite measure that reflected the totality of the psychological conditions of every student. It also combines 5 main mental health areas depression, anxiety, stress, insomnia and level of risk to a binary (0 = Not severe, 1 = Severe). A student was considered to be Severe when either of the five component scores had reached a moderate or higher mark. Such integrative approach reflects the multidimensionality of maladaptive behaviors, that is, the combination of psychological problems that affect the well-being of students..

Table 7: Model Performance for Overall Mental Health

Model	Test Acc (%)	F1-Score (%)	Precision (%)	Recall (%)	ROC AUC (%)
Gradient Boosting	96.06	97.56	97.09	98.04	99.49
Random Forest	96.06	97.58	96.19	99.02	99.53
XGBoost	96.06	97.54	98.02	97.06	97.35
SVM	93.70	96.15	94.34	98.04	98.94
Logistic Regression	92.91	95.65	94.29	97.06	98.78
Ensemble	97.64	98.54	98.06	99.02	99.84

4.7. Deep Learning Results: Overall Mental Health

To complete the exercise of classifying an image of "Overall Mental Health," three deep learning models were tested: a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network, and an Bidirectional LSTM (BiLSTM) network. Table 8 provides the performance of these models.

Table 8: Deep learning models during Overall Mental Health classification.

Model	Accuracy	Precision	Recall	F1	ROC_AUC
CNN	96.06	97.09	98.04	97.56	99.69
LSTM	92.13	92.59	98.04	95.24	96.86
BiLSTM	92.13	95.10	95.10	95.10	96.51

The findings show without any doubt that CNN model provided better performance than the recurrent architectures. The CNN was the most accurate (96.06%), Precise (97.09%), F1-Score (97.56), and had an outstanding ROC-AUC (99.69%). This indicates that the CNN had a very high success rate in feature extraction that was used to perform the spatial feature classification in this classification process (92.13%). The LSTM model scored a 98.04% Recall (the same as CNN) and thus it is strong in terms of being able to identify all cases of positive samples correctly but at the expense of lowering the accuracy (92.59%). In terms of the overall metrics, the CNN model is the most robust and balanced model of the task of classifying the overall mental health based on the given metric.

4.8. Visual Performance of the Weighted Ensemble Model

A graphical summary of the results of the Weighted Ensemble model is given in Figure 1 in which the better performance of the model is highlighted.

Fig. 3 shows the six confusion matrices indicating that the ensemble model is very accurate. All the six plots show strong and dark diagonal, which implies that there are high levels of the true positive and true negative predictions. Matrices of the variables of the overall mental health and the risk level are especially striking with almost no misclassifications. This high degree of diagonal dominance is also displayed in the multi-class matrices of the characteristics of Insomnia, Stress, Anxiety and Depression, with little confusion, mostly between the very close classes. The six ROC-AUC curves in Fig. 4 indicate that the Weighted Ensemble is extraordinary in its classification. Their closeness to the upper-left side indicates high sensitivity and specificity whereas the AUC scores measure this.

- Insomnia (Normal to Severe): Macro-AUC = 0.99
- Stress (Normal to Extremely Severe): Macro-AUC = 0.97
- Anxiety (Normal to Extremely Severe): Macro-AUC = 0.98
- Overall Mental Health: AUC = 1.00
- Risk Level (Low to High): AUC = 1.00
- Depression (Normal to Extremely Severe): The multi-class plot indicates that all of the individual class curves (0-4) are located in the high-performance (top-left) quadrant.

The ideal 1.00 AUC values of the overall mental health and the risk level are that the model was capable of differentiating the positive and the negative classes perfectly in the test data set between the two most vital variables.

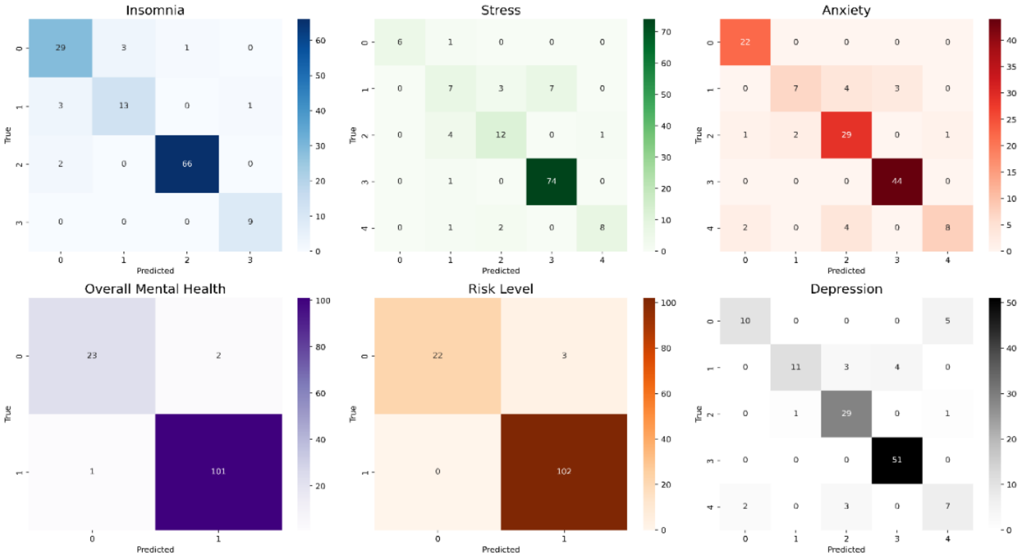

Fig. 3: Confusion tables of the six target variables.

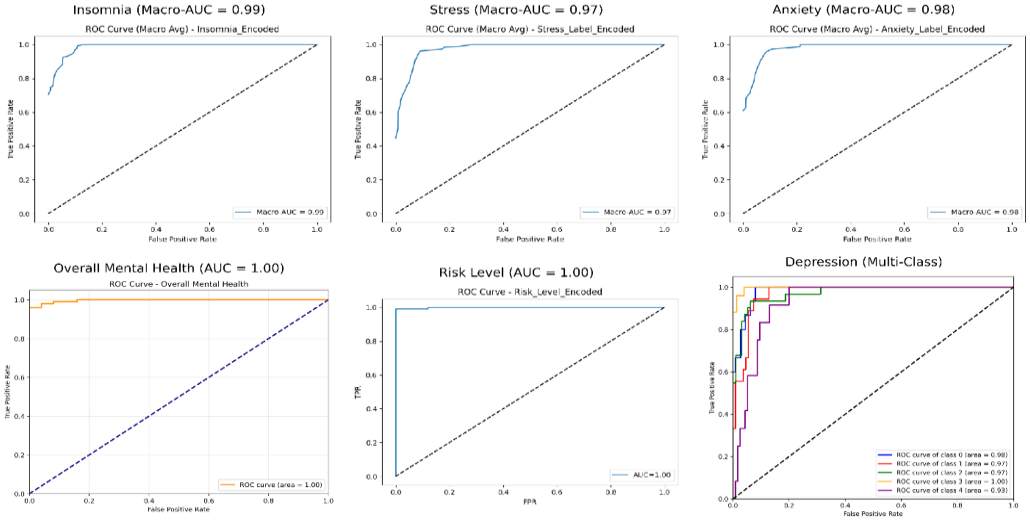

Fig. 4: The curves of ROC-AUC of the six target variables

5. EXPLAINABLE AI (XAI) FOR OVERALL MENTAL HEALTH PREDICTION

To increase the practice of the proposed predictive scheme, we have utilized SHAP (SHapley Additive exPlanations) to clarify the black box conduct of the optimal functioning model of the Weighted Ensemble. This analysis gave both global and local explanations of the model predictions and helped one to have an insight on the features that had the strongest influence on the predictions and how they interacted with each other to produce the Overall Mental Health predictions.

5.1. Global Feature Importance: SHAP Summary Plot

Fig. 5 shows the SHAP summary plot, which shows the overall contribution of features on the test set. The x-axis is the SHAP value (the impact that the feature has on the log-odds output of a model), and the color gradient is the magnitude of the actual value of the feature (red is high, blue is low). The five most influential features worldwide that were discovered were stress1, depression5, anxiety1, featureanxiety1, and Age. In the case of stress1, depression5 and anxiety1, high feature values (red) always resulted in positive SHAP values i.e. higher the reported levels of stress, depression and anxiety, the higher the chance of this model predicting Class 1 (At Risk). However, low feature values (blue) produced negative SHAP values, which forced the predictions to Class 0 (Not at Risk). The effect of Age and featureanxiety1 was more nonlinear and complex as the high and low values were observed on both sides of the SHAP axis. The trend indicates that the model was trained on subtle relationships between demographic and psychological variables as opposed to the use of linear relationships.

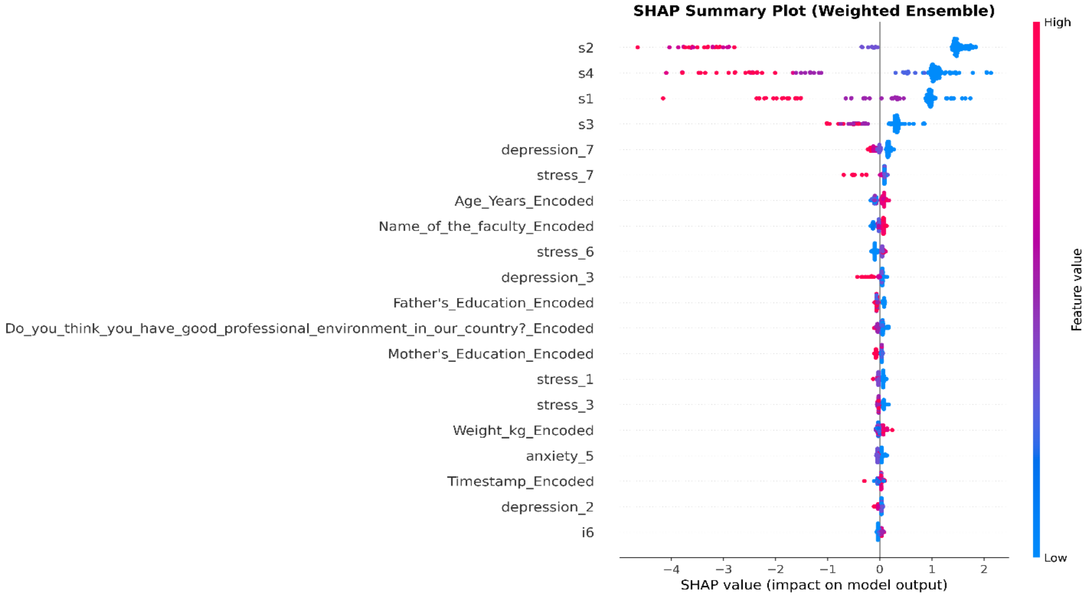

Fig. 5: SHAP summary plot of the weighted ensemble model predicting the overall mental health, illustrating the main features and their directional influence on predictions of the at risk category.

5.2. Local Prediction Analysis: SHAP Waterfall Plots

While the summary plot offers a global perspective, SHAP waterfall plots provide instance-level explanations, showing how each feature contributed to specific predictions. Each plot depicts how features push the model's base value *(E[f(X)] = 0.116)* toward its final output *f(x)*, with red bars indicating positive contributions (toward “At Risk”) and blue bars indicating negative contributions (toward “Not at Risk”)..

Case 1 (True Positive): Sample 0 was correctly diagnosed as At Risk As indicated in Fig. 6. Value stress 1 =4 +(0.15) and depression 5 =3 +(0.18) have strong positive impact which is opposed by negative impact of Age =48 -(0.23) and featureanxiety1 =0.942 -(0.22). Even though these forces are contradictory, the end result of the model, the final logit value of f(x) = -0.032, as a predictor of mental health indicators was also correct with reference to the ground truth given the low classification threshold.

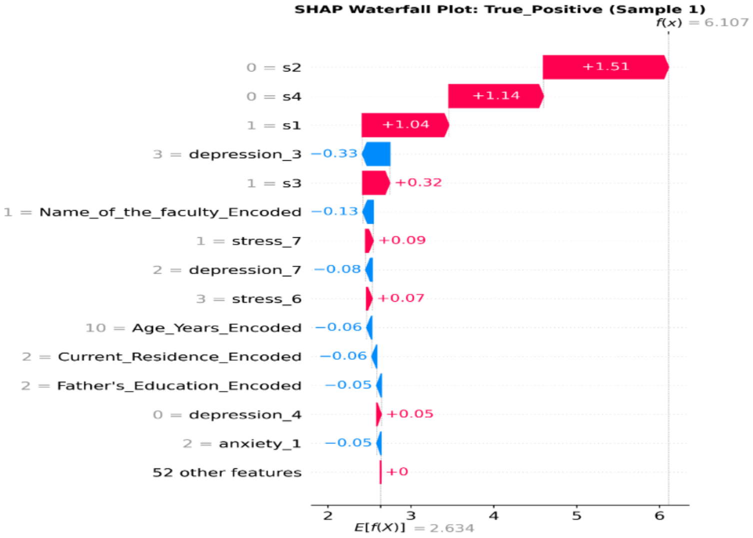

Fig. 6: SHAP waterfall plot of a True Positive case that was correctly classified as "At Risk" with features in red (e.g., depression14), which adds risk prediction.

Case 2 (False Positive): Sample 4 was falsely deemed as At Risk when the real label of Fig. 7 was Not at risk as shown in Fig. 7. This was a false positive due to the contributions which were all positive with the values being anxiety1 = 2 (+0.42), featureanxiety1 = 0.872 (+0.41), and depression5 = 4 (+0.36) overwhelmingly outweighed this error. The cumulative SHAP produced an erroneous f(x) = 1.627 with a high confidence.

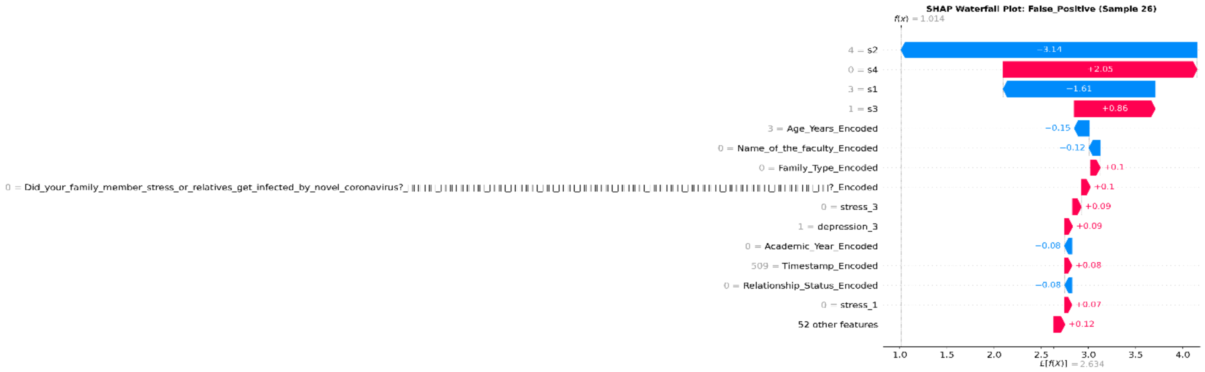

Fig. 7: SHAP waterfall plot on a False Positive case indicates the features (in red) that contributed to the erroneous prediction as the reason why the patient was labeled as At Risk.

Case 3 (True Negative): Sample 8 was rightly classified as Not at Risk (Fig. 8). The most significant was Age = 24 that added a significant negative SHAP = -1.09 and superseded moderate positive impacts of stress1 = 2 (+0.26) and featureanxiety1 = 0.538 (+0.26). The output f(x) = -0.933 was a high confidence and correct Not at Risk classification.

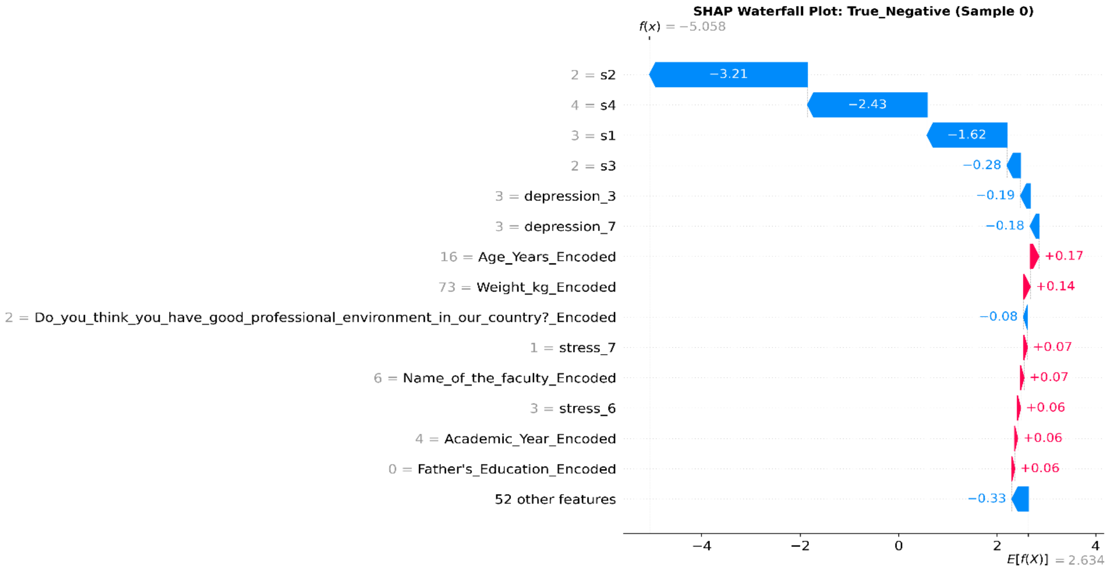

Fig. 8: SHAP waterfall plot of a True Negative query, where blue spots reduced the prediction, which is an accurate prediction that it is not at Risk.

Case 4 (False Negative): Sample 2 was incorrectly identified as a non-at risk, when in actuality, it was labeled as at risk. The outcome of the model gave a very negative value (f(x) = -1.571) as a result of negative contribution of anxiety1 (4, -0.77) and featureanxiety1 (0.442, -0.56), which overshadowed the positive effect of stress1 (4, +0.27). Fig. 9 shows that SHAP waterfall plot can determine the overall effect of the local feature interaction, such as the inversion of the expected effect of anxiety1, which provides some understanding of possible improvement areas in the models.

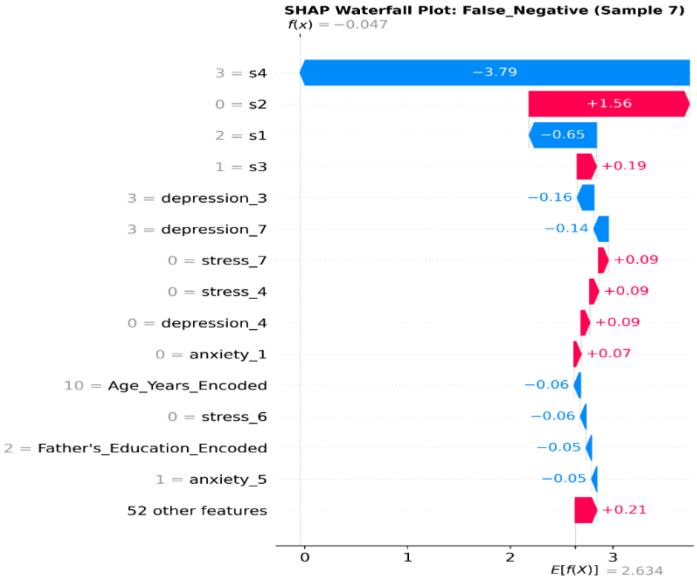

Fig. 9: SHAP waterfall plot of a True Negative instance, in which blue features decreased the prediction, and with the correct prediction of "Not at Risk" was identified.

5.3. Interpretation and Implications

The SHAP-based XAI analysis presented the validation that the prediction of the ensemble model was mostly related to psychometric indicators of stress, depression, and anxiety with minor modulation by demographic characteristics. Besides, the local case studies pointed out that even highly faithful models may confidentially err when nonlinear interactions reverse predicted role of features. These interpretability results improve credibility, uncover possible bias, and give practical advice on how to implement the model into student mental health assessment pipelines with human-in-the-loop monitoring.

6. DISCUSSION

The findings of the present research prove the high effectiveness of a Weighted Ensemble method in the multi-faceted, complicated issue of mental health assessment. In all six target variables, the Weighted Ensemble was more effective than the 5 individual base models of random forest, XGBoost, SVM, LightGBM, and Gradient Boosting in essentially exploiting their relative advantages such as top accuracy of SVM on specific DASS parameters and robustness of tree based models such as XGBoost with reduction of individual weak points (Hasan et al., 2025) [7, 11]. In the case of the primary target, Overall Mental Health (Table 7), the ensemble results were 97.64 percent accuracy and 99.84 percent ROC-AUC, which suggests that the ensemble has perfect classification. Equally, in the case of Risk Level (suicidal ideation) (Table 3), the model achieved an accuracy of 97.64 and ROC-AUC of 99.88. The high sensitivity and specificity are clinical significance of great importance, as the earlier research revealed the value of predictive modeling in the context of timely identifying suicide risk among students at the university level (El-Ashry et al., 2024). Such automated screening tools may be used to supplement clinical workflows of identifying high-risk individuals so that they may be intervened in time.

The model was also highly accurate and discriminative in respectful parameters of DASS Digression (Table 2), Anxiety (Table 5), Stress (Table 6) and Insomnia (summarized in Figure 1b). The range of AUC values is 0.97 to 0.99, which is stable and can be compared to the results of the recent study on the multi-label prediction of mental health by using machine learning (Luo et al., 2025). This shows that the model can accommodate overlapping symptom profiles which are prevalent among the populations in universities (Liu et al., 2024). The visual efforts also support the superior performance of the Weighted Ensemble. The confusion matrices (Fig. 1) affirm that the high level of accuracy is neither a consequence of class imbalance, whereas the ROC curves (Fig.2) indicate the high level of discriminative capability of the model. The findings are agreeable with more recent developments in ensemble learning in mental health prediction, in which the combination of various models leads to stability and generalization improvements. Combined, these results demonstrate the promise of ensemble methods based on machine learning as powerful and automated instruments of the overall mental health evaluation. In addition to predictive accuracy, the method provides practical information that might inform interventions aimed at managing depression, anxiety, stress, and sleep disruptions to supplement the available psychological support strategies (Luo et al., 2024). Further research on these findings in future studies must aim to substantiate such findings in future, real-world clinical environments to determine the generalizability and practical value as a decision-support system.

The results of this study also add to the ever-increasing number of studies that believe in the incorporation of explainable AI in the higher education systems. Explainable ensemble models like the one herein proposed WVE-RGS model overcome the disparity between predictive efficacy and ethical transparency, which is like a key when applying to aspects of mental health and behavioral prediction among students. Recent reports have highlighted the importance of ensuring that not only can predictive systems be highly accurate, but also lead to interpretability to gain the trust of the users to be adopted in sensitive academic and clinical settings (Lumumba et al., 2025). The incorporation of SHAP-based explanations to this framework is consistent with this trend, as it can offer practical recommendations that can be translated by educators and counselors into early and human-oriented interventions. More than that, the multitasking of the AI paradigms (tree-based and kernel-based learners) serve to strengthen the resistance to overfitting and data bias, which is consistent with earlier results that ensemble integration can be beneficial to the stability and generalization of behavioral risk models (X. Wang et al., 2025). These results support the ensemble learning paradigm as a scalable and clear-cut approach to continual psychological examination in an open innovation-driven university.

This study also shows how AI-based behavioral analytics can help to operationalize the idea of open innovation in education based on an institutional and societal level. With the incorporation of data-driven mental health monitoring within university systems, educational organizations can develop collaborative ecosystems, entailing students, counselors, administrators, and AI tools in the endless feedback loop. Such a participatory methodology makes AI more than a technical tool and a co-creative mental health manager to promote innovation by increasing transparency, inclusivity, and shared accountability. The framework suggested, thus, goes beyond prediction to add to ethical AI governance, social innovation, and sustainable well-being plans in higher education. Also, the interdisciplinary interaction between computer scientists, psychologists, and educators will be needed to make sure that AI-powered models of behaviors do not contradict human values and institutional ethics.

CONCLUSION

This research presented a multi-faceted AI-powered model of measuring and forecasting maladaptive actions and clinical risks in the minds of college students. The study combined deep learning models, the classic models of machine learning, and a new Weighted Voting Ensemble (WVE-RGS) and revealed that multidimensional data with psychological indicators, sociodemographic variables, lifestyle habits, and academic traits could accurately predict mental health outcomes. In six target variables Overall Mental Health, Risk Level (suicidal ideation), Anxiety, Depression, Stress, and Insomnia, the Weighted Ensemble model was always better than all the individual classifiers. The almost perfect scores of ROC-AUC in terms of Overall Mental Health (99.84) and Risk Level (99.88) were exactly remarkable, which means a high level of discriminatively. Confusion matrices and ROC curves were also used to validate the model and revealed that it was robust even in the case of multi-class problems. These conclusions suggest the possible effectiveness of ensemble-based machine learning schemes in identifying complex and overlapping psychological disorders that traditionally pose a challenge to conventional assessment schema.

The transparency of the model decision-making was essential because the Explainable AI (XAI) methods were included, namely SHAP analysis. The framework allows interpretation of mental health practitioners because the scale items are the most important contributors to stress, depression, and anxiety, which means that the human factors are considered and the ethical consideration is achieved by applying the solution in educational institutions. This openness is crucial to the process of introducing AI to sensitive areas of life, including psychological evaluation and early identification of risks.

On the whole, the results highlight the potential of AI and deep learning as complementary devices in mental health surveillance in universities. The suggested system is a data-driven and scalable system which may be used to assist in early detection, inform intervention plans and improve institutional mental health care. With the psychological distress levels among student populations persistently increasing, such predictive models have the potential to transform the state of well-being, avert crises, and serve as an informing factor to proactive campus mental health policies.

CRediT authorship contribution statement

Arifa Tur Rahman: Conceptualization, Methodology, Investigation, Validation, Writing – original draft, Resources, Project administration. **Dr Utpal Kanti Das**: Conceptualization, Formal analysis, Writing – review & editing, Supervision, Resources, Project administration. **Md. Masud Rana**: Investigation, Validation, Writing – original draft. **Abdul Muyeed**: Formal analysis, Writing – review & editing**. Md. Abu Hanif**: Formal analysis, Writing – review & editing.

CONFLICT OF INTEREST

There was no conflict of interest declared by the authors.

ACKNOWLEDGEMENT

This research is supported by the Miyan Research Institute, International University of Business Agriculture and Technology, Dhaka 1230, Bangladesh. Research Grant number: IUBAT-MRI-RG-2025

REFRENCES

[1] Abdelfattah, E., Joshi, S., & Tiwari, S. (2025). Machine and deep learning models for stress detection using multimodal physiological data. *IEEE Access*.

[2] Al-Badayneh, D. M., Shahin, M. M., & Brik, A. B. (2024). Strains and Maladaptive Behaviors among High School Students in Qatar: An Empirical Test of the General Strain Theory. *Dirasat: Human and Social Sciences*, *51*(4), 51-62.

[3] Al Amin, M., Liza, I. A., Hossain, S. F., Hasan, E., Haque, M. M., & Bortty, J. C. (2024). Predicting and monitoring anxiety and depression: Advanced machine learning techniques for mental health analysis. *British Journal of Nursing Studies*, *4*(2), 66-75.

[4] Anim, D. A. A., Essien, E. E., & Ekpoto, D. F. (2025). Value Education and Insecurity Among Secondary School Students in Calabar Metroplis, Cross River State, Nigeria. *Cross River State, Nigeria (May 31, 2025)*.

[5] Banerjee, S., Rana, M. M., Akash, M. M. H., Mridula, A. T., Mamoon, I. A., & Rahman, Q. B. (2025). Robotics, artificial intelligence, and computer vision in dental implant surgery: a systematic review of accuracy, efficiency, and future directions. *Journal of robotic surgery*, *20*(1), 42.

[6] Bednárová, M., & Ilenčíková, O. (2024). Problematic Behaviour Among Students in Secondary Schools and the Role of Educational Counsellors. *R&E-SOURCE*, 28-39.

[7] Chowdhury, A. H., Rad, D., & Rahman, M. S. (2024). Predicting anxiety, depression, and insomnia among Bangladeshi university students using tree-based machine learning models. *Health Science Reports*, *7*(4), e2037.

[8] Das, P., Hasan, M. E., Arif, M., ALmerab, M. M., Al Habib, A., Al-Mamun, F., & Mamun, M. A. (2025). Prevalence, associated factors, and machine learning-based prediction of probable depression among individuals with chronic diseases in Bangladesh. *BMC psychiatry*, *25*(1), 1093.

[9] El-Ashry, A. M., Taha, S. M., Elhay, E. S. A., Hammad, H. A.-H., Khedr, M. A., & El-Sayed, M. M. (2024). Prevalence of imposter syndrome and its association with depression, stress, and anxiety among nursing students: a multi-center cross-sectional study. *BMC nursing*, *23*(1), 862.

[10] Harsh, R., Patodiya, R., Agarwal, R., & Mehta, V. (2024). Analysing Emotional Drivers and Behavioural Patterns of Twitter Overuse in India During the COVID-19 Lockdown using ML and BERT. 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT),

[11] Hasan, M. E., Arif, M., Rakibul Hasan, S., Muwanguzi, M., Abaatyo, J., Kaggwa, M. M., ALmerab, M. M., Atroszko, P. A., Muhit, M., & Al-Mamun, F. (2025). Prevalence, associated factors, and machine learning-based prediction of depression, anxiety, and stress among university students: a cross-sectional study from Bangladesh. *Journal of Health, Population and Nutrition*, *44*(1), 1-19.

[12] Ibrahim, A. M., Zaghamir, D. E. F., Alradaydeh, M. F., & Elsehrawy, M. G. (2024). Factors influencing aggressive behavior as perceived by university students. *International Journal of Africa Nursing Sciences*, *20*, 100730.

[13] Isaac, N. U. (2024). Behaviour disorders of childhood and adolescence: implications for education in Nigeria. *Bulletin of Islamic Research*, *2*(4), 573-590.

[14] Jin, K., Zeng, T., Gao, M., Chen, C., Zhang, S., Liu, F., Bao, J., Chen, J., Wu, R., & Zhao, J. (2025). Identifying suicidal ideation in Chinese higher vocational students using machine learning: a cross-sectional survey. *European Archives of Psychiatry and Clinical Neuroscience*, 1-12.

[15] Joshua, V., Usman, M., & Oguche, T. (2024). Influence of Social Media and Peer Group on Maladaptive Behaviour among Secondary School Students in Federal Capital Territory, Abuja. *International Journal of Advanced Academic Research*, *10*(10), 139-154.

[16] Liu, M., Liu, H., Qin, Z., Tao, Y., Ye, W., & Liu, R. (2024). Effects of physical activity on depression, anxiety, and stress in college students: the chain-based mediating role of psychological resilience and coping styles. *Frontiers in Psychology*, *15*, 1396795.

[17] Lu, J., Liu, Y., Liu, S., Yan, Z., Zhao, X., Zhang, Y., Yang, C., Zhang, H., Su, W., & Zhao, P. (2024). Machine learning analysis of factors affecting college students' academic performance. *Frontiers in Psychology*, *15*, 1447825.

[18] Lumumba, V. W., Muriithi, D. K., & Oundo, M. (2025). Evaluating the Performance of Selected Single Classifiers with Incorporated Explainable Artificial Intelligence (XAI) in the Prediction of Mental Health Distress Among University Students.

[19] Luo, L., Yuan, J., Wu, C., Wang, Y., Zhu, R., Xu, H., Zhang, L., & Zhang, Z. (2025). Predictors of depression among Chinese college students: a machine learning approach. *BMC Public Health*, *25*(1), 470.

[20] Luo, M.-m., Hao, M., Li, X.-h., Liao, J., Wu, C.-m., & Wang, Q. (2024). Prevalence of depressive tendencies among college students and the influence of attributional styles on depressive tendencies in the post-pandemic era. *Frontiers in Public Health*, *12*, 1326582.

[21] Milam, M. E., & Sutton, K. K. (2024). Using Visual Activity Schedules to Improve Transitioning for Students With Emotional and Behavioral Disorders. *Beyond Behavior*, *33*(3), 159-166.

[22] Oladimeji, M. A. (2024). Substance abuse and maladaptive behaviours on academic performance of adolescent secondary school students in Oyo metropolis, Nigeria. *Journal of Educational Research in Developing Areas*, *5*(3), 360-372.

[23] Peng, Y., Lv, S. B., Low, S. R., & Bono, S. A. (2024). The impact of employment stress on college students: psychological well-being during COVID-19 pandemic in China. *Current Psychology*, *43*(20), 18647-18658.

[24] Rahman, A., Syeed, M. M., Fatema, K., Khan, R. H., Hossain, M. S., & Uddin, M. F. (2024). An Interpretable Hybrid CNN-SVM Model for Predicting Stress Among University Students. 2024 IEEE/ACS 21st International Conference on Computer Systems and Applications (AICCSA),

[25] Rahman, H., Tohan, M. M., Easha, A. A., & Fatema, N. (2025). Depression symptoms among University Students in the Khulna region of Bangladesh. *PLOS Mental Health*, *2*(4), e0000158.

[26] Rana, M. M., Akter, J., & Hanif, M. A. (2025). Next-gen vision: a systematic review on robotics transforming ophthalmic surgery. *Journal of robotic surgery*, *19*(1), 452.

[27] Rooney, E. A. (2024). *A Machine Learning Approach to Predicting Nonsuicidal Self-Injury* The University of Toledo].

[28] Sandel-Fernandez, D. B. (2024). *Urgency in Daily Life: Capturing the Momentary Emotion-to-Behavior Mechanism of Impulsive Behaviors* University of California, Berkeley].

[29] Sumukh, V., Dalwai, N., Kashyap, S. R., & Akshay, S. (2025). Analyzing the Impact of Mental Health on Academic Performance: AI Approach Based on Student Feedback. 2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON),

[30] Tayarani-N, M.-H., & Shahid, S. I. (2025). Detecting Anxiety via Machine Learning Algorithms: A Literature Review. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

[31] Wang, X., Li, C.-Z., Sun, Z., & Xu, Y. (2025). Design and Analysis of a Closed-Loop Emotion Regulation System Based on Multimodal Affective Computing and Emotional Markov Chain. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.

[32] Wang, Y., Lin, Z., Yang, C., Zhou, Y., & Yang, Y. (2025). Automatic depression recognition with an ensemble of multimodal spatio-temporal routing features. *IEEE Transactions on Affective Computing*.

[33] Zamri, E. N., Sha, L., Mohd, T. N. A. B. T., Haizam, N., & Isa, S. N. I. (2024). A scoping review of the factors associated with mental health among Malaysian adolescents. *Malaysian Journal of Movement, Health & Exercise*, *13*(2), 71-82.

[34] Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., Lin, Y. W. D., Xu, Z., Daniels, A. D., & Liu, Y. (2025). Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students. *Journal of Counseling & Development*, *103*(1), 110-125.