Grenze International Journal of Engineering and Technology, June Issue

Machine Learning-based Multi-Module System for Student Depression Prediction and Behavioural Analysis

Abhishek S[1], Manoj Kumar V[2], Sneha George[3] and T. Jemima Jebaseeli[4]
[1-3]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India
Email: abhisheks@karunya.edu.in, manojkumarv@karunya.edu.in, snehageorge@karunya.edu
[4]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India
Email: jemi.jeba@gmail.com

***Abstract*— More students are being affected, and they are finding themselves depressed, stressed, anxious, having sleeping issues as well as unhealthy digital behaviour patterns that directly impact their study life and overall wellbeing. Manual intervention Traditional screening processes are known to be time consuming, inconsistent and not always linear. This research proposes an intelligent mental health decision support system, which is inclusive of Machine Learning (ML), behavioural analytics, and AI-assisted recommendation strategies in the early detection of depression risk and digital wellbeing conditions among students. The proposed system is made up of two large prediction modules. The first module approximates the risk of depression with the help of personal, academic, lifestyle, and family related variables age, gender, city, degree, CGPA, academic stress, financial stress, sleep, diet, work-study hours, suicidal thoughts, and family history of mental illness. The second module measures digital wellbeing by behavioural attributes such as average time spent on the screen, frequency of switching between apps, amount of time spent asleep, the number of notifications, time spent on social media, focus score, mood score and the level of anxiety. The system also has a rule-based overriding process, which is used to detect emergency cases particularly in cases where there are suicidal thoughts or the presence of powerful stressful conditions. A recommendation engine based on AI will provide personalized mental health recommendations, lifestyle recommendations, information on support, and motivational content depending on the level of predicted risk. There is also storage of historical prediction, behavioural timelines and risk trend visualization, which provides long term monitoring and analysis. The suggested system is a scalable model offering an efficient and easy to use model of early mental health evaluation and constant behavioural monitoring.**

***Index Terms*— Mental health, depression, digital wellbeing, behavioural analysis, machine learning, stress analysis, AI recommendation, predictive analytics.**

I. Introduction

The psychological well-being of college students has been a significant issue in the last few years due to increasing academic stress, financial struggles, social loneliness, social and digital addiction, insomnia,

Grenze ID: 01.GIJET.12.2.541_26

and emotional burnout. Depression, anxiety and behavioural disorders considerably diminish concentration, academic performance, motivation as well as social wellbeing. Conventional approaches to mental health screening tend to rely on the use of questionnaires, clinical interview, and counsellor observations that might require some time and might be hard to generalize to a large student body. The recent developments in the field of machine learning have opened up the prospects of automated mental health prediction tools which can quickly detect at-risk students even at a tender age. It has been shown that demographic, behavioural, and academic features of depression, stress and anxiety levels can be effectively classified using interpretable machine learning methods [1]. Other predictive models have been used to assign intervention measures to anxiety and depressive disorders to college students [2]. The discovery of concealed data related to mental health risks among students has been done using apriori based techniques of data mining [3]. In the same manner, ML has been used to determine the most significant predictors of depression in college students that include the academic stress, sleep quality, financial burden and the family history [4]. These advances suggest that systems based on data can offer quicker and dependable mental health appraisal than conventional systems.

Besides predicting depression, the use of digital behaviour and gadgets have emerged as majorpredictors of student wellbeing. Mental health and productivity may be adversely influenced by excessive screen time, excessive use of social media, high frequency of notifications, sleeping poor habits, and switching apps constantly. According to the recent research, students have a close association between gadget addiction and unhealthy digital practices and stress, anxiety, and depressive symptoms [9]. The relevance of behavioural cues in detecting emotional instability and stress states has also been established with real-time stress monitoring methods based on wearable devices and machine learning [10]. The researchers have also highlighted the importance of interpretable models to enable the user to understand how a prediction has been made and the behavioural characteristics that possess most influence in increasing the risk of mental health [5]. Such results demonstrate the increased concern of coherent systems integrating the analysis of depression, behaviour monitoring, and intelligent recommendation systems.

The purpose of the proposed research for depression prediction, digital wellbeing analysis, historical monitoring, and AI-assisted recommendations suggesting that a ML-based mental health decision support system needs to be developed targeting the students. The system employs a number of the input characteristics that include, age, gender, academic pressure, financial stress, CGPA, dietary habits, time spent in sleeping, social media use, screen time, mood score, focus score, and level of anxiety to come up with predictive results. Compared to most of the literature under existence in which several researchers concentrate on only one dimension of mental health, the proposed method incorporates both the mental and behavioural dimensions within a single framework. A rule-based override mechanism in case of extreme situations related to suicidal thoughts and severe stress conditions is also part of the system, as in that way, high-risk cases will not be overlooked. Moreover, the system offers the visual data analytics like prediction history, wellbeing timelines and risk distribution charts that will be useful to promote long-term monitoring. Just like most recent studies on machine learning-based depression identification and mental health diagnosis [6][8], the model suggested here is intended to offer meaningful mental health assistance in learning settings correctly, scalable, and easily. When the system is combined with predictive analytics and recommendation generation and behavioural tracking, it will be able to assist institutions to identify vulnerable students earlier and offer interventions in time before the situation becomes serious. The key contributions are as follows.

- Proposes a unified system combining depression prediction and digital wellbeing analysis
- Introduces a hybrid approach using machine learning with rule-based safety overrides
- Incorporates behavioural and digital features for holistic assessment
- Provides AI-based personalized recommendations for students
- Supports long-term monitoring through historical tracking and visual analytics

II. Literature Survey

A. Student Mental Health Interpretable Machine Learning

Recent research has also paid attention to interpretable ML methods to enhance prediction system channels in student mental health. The models identified by researchers to determine the potential risk factors that are significant to students include decision trees, random forests, gradient boosting and explainable AI models. The main factors that have been widely referred to as the causes of depression and anxiety include academic pressure, economic issues, sleep level, family background, and isolation. The importance of the interpretable models is that it enables the users and the healthcare experts to know the rationale behind a given prediction.

These systems enhance confidence, openness and functionality within institution of learning. These measures have demonstrated good results in determining the high risk of mental health students [1][4][8].

B. Deep Learning based depression and anxiety detection models

A number of ML methods are also suggested to detect the condition of depression and anxiety disorders among college students. The most popular methods that are used in recent research include logistic regression, support vectors machines, decision trees, random forests, XGBoost and neural networks. These algorithms will process demographics, behaviours, academic and emotional features in order to forecast mental illness. Research has revealed that machine learning models can be accurate and screened faster compared to the manual one. Predictive systems have also been studied by researchers on how to allocate resources to counselling and earlier interventions on those students with severe mental cases. This is done by using such approaches so as to minimize the delay in recognizing vulnerable students [2][4][6].

C. Digital Wellbeing Analysis and Behavioural Analysis

Digital behaviour has been revealed as a critical aspect of mental health assessment of students. Scholars have discovered a close connection between the screen and social media use, the frequency of switching apps, the duration of sleep and emotional wellbeing. Uncontrolled use of smartphones and other devices may result in stress, anxiety, lack of concentration and low academic results. Research based on gadget addiction prediction and digital behaviour analysis has demonstrated that machine learning will be able to detect unhealthy behavioural patterns. It has also been found out that real time stress monitoring through wearable device has also become an effective device of identifying emotional instability and behavioural risk among students [9][10].

D. Artificial Intelligence Proposal and Assistance

Most of the new systems do not only forecast mental illnesses but also offer individual recommendation to students. Artificial intelligence-oriented recommendation systems produce the recommendations connected to lifestyle change, stress management, sleep, physical activity, counsellorship, and crisis management. The systems enhance the usability of prediction models by assisting the user in comprehending what to do next until they get the result of a mental health risk. With the help of recommendation engines, both low and high-risk students can be promoted by giving them specific advice depending on the personal circumstances. They are becoming more significant within the learning institutions where resources of mental health support are scarce [1][2][3].

E. Improving Trends and Future Directions

The current studies reveal a change in using more centralized machine learning system to safer and more scalable federated and deep learning systems. Federal learning gives mental health models a chance to be trained on numerous devices or institutions without sharing personal information. Graph neural networks, wearable sensors, real-time monitoring, and contextual learning models are other areas of research by the researchers to enhance the performance of predictions. It is likely that in the future, behavioural monitoring, performance, emotional analysis, and recommendation system would all be put into one system. The developments can cause student mental health systems to be more precise, confidential, and efficient [5][7][10].

F. Novel Contribution of the Proposed System

The proposed system introduces a unified mental health decision support framework that integrates depression prediction and digital wellbeing analysis within a single platform. It combines ML models with a rule-based override mechanism to ensure accurate and safety-critical predictions, especially in high-risk conditions such as suicidal thoughts and severe stress. The system uniquely incorporates behavioural features like screen time, app usage, and sleep patterns alongside psychological factors, enabling assessment. Additionally, it provides AI-driven personalized recommendations and maintains historical records with visual analytics for continuous monitoring. This end-to-end integration of prediction, behavioural analysis, recommendation, and long-term tracking distinguishes the proposed system from existing approaches.

III. Proposed Methodology

A. Data Collection and Processing

The proposed system has the ability to gather both the psychological and behavioural data of the students by using structured input forms. Some of the resultant depression-related inputs are age, gender, city, degree, academic pressure, financial stress, CGPA, work-study hours, dietary and sleep duration, suicidal thoughts as well as family history of mental illness. Some of the behavioural inputs include screen time, frequency of switching apps, hours of sleep, number of notifications, amount of time using social media, mood level, level of focus and level of anxiety. This is because before the prediction, these values are validated so as not to miss or invalidate data. Then, the gathered information is transformed into a computer-readable one to continue the preprocessing and scaling, features selection, and prognostication.

TABLE I: DATASET SUMMARY

Property	Value
Dataset Size	27,870 samples
Features Used	11 features
Target	Depression (0 / 1)
Train/Test Split	80% / 20%

As shown in Table I, the depression prediction model was trained on a dataset containing 27,870 student records. The dataset included 11 selected predictive features related to academic stress, lifestyle habits, and psychological conditions. The target variable represents whether the student is experiencing depression or not. The dataset was divided into 80% training data and 20% testing data using stratified sampling.

B. Preprocessing Data and Features Engineering

Preprocessing is to achieve consistency in the data collected and obtain a higher accuracy in prediction. Normative scales of numerical values like age, CGPA, sleep hours, and time at the screen are obtained using the scaling methods. Label encoding is used to encode categorical traits like city, degree, eating habits and sleep duration. The selected features are used to reduce the quantity of important attributes being used by the models that are being trained. The results of this process are the reduction of noise, increased efficiency of models, and less unwarranted computation. The department of depression prediction and behavioural wellbeing prediction have separate preprocessing pipelines due to the variety of inputs needed by the different modules and the methodological differences in the process of scaling they use.

The ML models were trained using a structured pipeline consisting of preprocessing, feature selection, model training, and evaluation stages. For the depression prediction model, missing values were handled using median imputation, and categorical variables such as sleep duration and dietary habits were converted using ordinal encoding. Numerical features were standardized before training the classifiers. For the behavioural wellbeing prediction module, all features were numerical and no missing values were observed, therefore only feature scaling was applied. Multiple machine learning models were trained and compared, and the final models were selected based on cross-validated performance metrics.

TABLE II: CROSS-VALIDATION PERFORMANCE

Metric	Value
ROC-AUC	92.10% ± 0.53%
Accuracy	84.42% ± 0.62%
Precision	87.97% ± 0.64%
Recall	85.00% ± 0.67%
F1-Score	86.46% ± 0.54%

During the model training phase, the dataset was divided using an 80:20 train–test split with stratification to preserve the class distribution of depression labels. Multiple classification algorithms were trained and evaluated using 10-fold stratified cross validation to ensure reliable performance estimation. The evaluation metrics included accuracy, precision, recall, F1-score, and ROC-AUC. As shown in Table II, the cross-validation results showed strong and consistent performance with an average ROC-AUC of 92.10%, accuracy of 84.42%, precision of 87.97%, recall of 85.00%, and F1-score of 86.46%.

C. Mathematical Formulation of Models

The depression prediction module follows a supervised classification approach. Let the input feature vector be:

$$X = (x_1, x_2, x_3, \ldots, x_n) \tag{1}$$

where each x_i represents features such as academic pressure, CGPA, sleep duration, and stress levels. The classification model predicts probability is given in equation (2).

$$P(Y = 1 \mid X) = f(X) \tag{2}$$

where:

- $Y = 1 \rightarrow$ High Risk
- $Y = 0 \rightarrow$ Low Risk

The final prediction is shown in equation (3).

$$\hat{Y} = \begin{cases} 1, & \text{if } P(Y = 1 \mid X) > \theta \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

For behavioural wellbeing prediction (regression model):

$$\text{Wellbeing Score} = f(x_1, x_2 \ldots, x_n) \tag{4}$$

where inputs include: screen time, sleep duration, notifications, mood score, and anxiety level. The predicted score is mapped into categories:

$$Category = \begin{cases} Excellent & 80 - 100 \\ Good & 60 - 79 \\ Moderate & 40 - 59 \\ Poor & 20 - 39 \\ Critical & 0 - 19 \end{cases} \tag{5}$$

D. Depression at risk Forecasting Module

As shown in Figure 1, the depression prediction module takes academic, emotional, and lifestyle related variables into consideration to establish whether a student is highly at risk or not of depression.

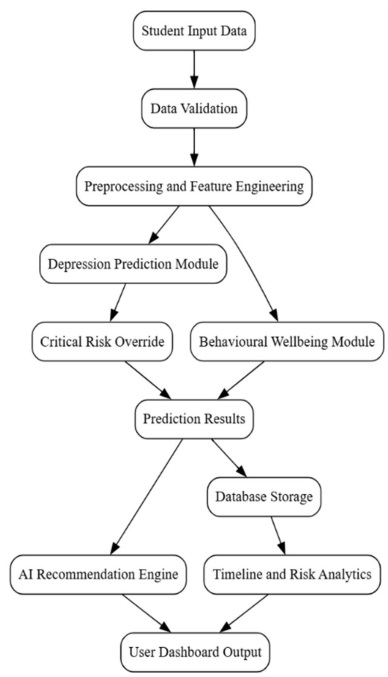

Figure 1. System Architecture

The processed features are generated into a trained ML classification model which predicts an actual label and a score. Probability value of both the depressed and the non-depressed classes is also calculated by the module. Also, there is the rule-based override mechanism which is incorporated in the system. The system automatically recurs a high-risk prediction in case the suicidal thoughts were identified or the combination of severe academic stress, financial stress, and poor sleep was identified to be at a risk of safety.

E. Prediction of behavioural wellbeing Module

The behavioural wellbeing module measures digital habits and emotional wellbeing to have an approximation of a digital wellbeing score of 0-100. A trained regression model is used to analyse inputs, that is, the number of apps switches, duration of sleep, number of app notifications, mood scale, anxiety level, and focus scale.

Depending on the predicted score, the student is categorized under one of five wellbeing categories: Excellent, Good, Moderate, Poor, or Critical. This module makes the users realize the consequences of digital behaviour on mental health and create an extra dimension of wellbeing testing and not just a depression prediction.

F. AI-driven Recommendation and Decision support

The system will produce AI-based recommendations based on the risk profile of a particular student after prediction. In order to make personalized recommendations, the recommendation engine provides it with the factors including the level of depression risk, the quality of sleep, and the level of stress, possibilities of suicide, and the history of family. These proposals consist of lifestyle changes, stress coping in academics, counselling information, self-management practices and crisis support information. Recommendation system also forms inspirational messages and actions that may be taken by the students to enhance their mental status. This feature will make the prediction system an entire decision support platform as opposed to a risk detection model.

G. Visual Analytics and Historical Tracking

All the results of the prediction are stored in a database by the proposed system to be analyzed in the future. The history of depression prediction and behavioural wellbeing history is kept as separate ones. The information on prediction labels, confidence scores, timestamps, sleep, mood score, anxiety level, and wellbeing are all documented in the past. Interactive visualizations that are created using these stored values include trend charts, risk distribution graphs, graphs on confidence as well as wellbeing personal timelines. These visual analytics can be used to track variations in mental health conditions in the students over time as well as in administrators. This is because the system is applicable with continuous monitoring and early intercession as it is a long-term tracking feature.

IV. RESULTS AND DISCUSSION

The performance of the depression prediction module is further illustrated using graphical analysis similar to the behavioural wellbeing module. The comparison graph presents multiple classification models including Logistic Regression, Random Forest, and XGBoost evaluated using Accuracy, Precision, Recall, F1-Score. The graphical visualization clearly shows that Logistic Regression consistently achieves the best balance across all evaluation metrics.

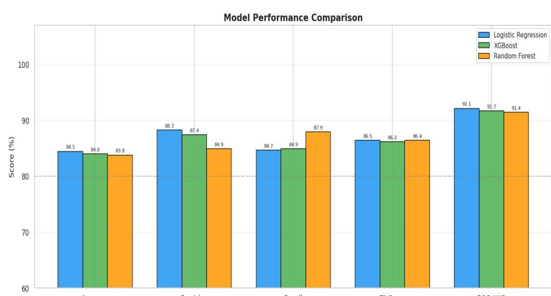

Figure 2. Model Performance Comparison using Classification Models

As shown in Figure 2, the Logistic Regression demonstrates superior overall performance compared to Random Forest and XGBoost by achieving the highest precision and ROC-AUC while maintaining a well-balanced recall and F1-score. Although the competing models exhibit comparable accuracy, Logistic Regression consistently provides a more stable trade-off between correctly identifying depressed cases and minimizing false predictions. This balanced performance is particularly important in mental health applications, where both false negatives and false positives carry significant implications. Hence, the results confirm that Logistic Regression is the most reliable and effective model, making it suitable for final deployment in the proposed depression prediction system.

TABLE III: CLASSIFICATION MODEL PERFORMANCE COMPARISON FOR DEPRESSION PREDICTION

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	ROC-AUC (%)
Logistic Regression	84.46	88.27	84.70	86.45	92.14
Random Forest	83.80	84.93	87.92	86.40	91.42
XGBoost	84.03	87.44	84.92	86.16	91.69

As shown in Table III, the multiple classification algorithms were evaluated for depression prediction including Logistic Regression, Random Forest, and XGBoost. Logistic Regression achieved the highest ROC-AUC score of 92.14%, with an accuracy of 84.46%, precision of 88.27%, recall of 84.70%, and F1-score of 86.45%. Although Random Forest achieved slightly higher recall (87.92%), Logistic Regression provided the best overall balance across all evaluation metrics and was therefore selected as the final model.

TABLE IV: REGRESSION MODEL PERFORMANCE COMPARISON FOR BEHAVIORAL WELLBEING PREDICTION

Model	MAE	RMSE	R^2 Score
Linear Regression	0.1470	0.1804	0.9994
Gradient Boosting	0.6724	1.0416	0.9815
Random Forest	0.9388	1.4411	0.9645

As shown in Table VI, the three regression models were evaluated for predicting digital wellbeing scores: Linear Regression, Gradient Boosting, and Random Forest. Linear Regression achieved the best performance with the lowest error values (MAE = 0.1470, RMSE = 0.1804) and the highest R^2 score of 0.9994, indicating that the model explains nearly 99.9% of the variance in the digital wellbeing score. Therefore, Linear Regression was selected as the final regression model for behavioural wellbeing prediction.

V. CONCLUSION

The proposed system is a very efficient and smart system of assessing student mental health through a combination of identifying and forecasting depression, digital wellbeing, AI-generated recommendations, and tracking the past. It is an effective system; it combines the academic, behavioural, emotional, and lifestyle-based factors to mark students who are at a risk of depression or low digital wellbeing. The depression prediction module of the system exhibited good classification and the behavioural module was also found to predict the score of digital wellbeing. Adding normal cases of suicidal thoughts and high levels of stress enhanced the level of reliability of the system with rule-based safety overrides. Moreover, the system was further complemented by visual analytics and timeline tracking, as well as the storage of the data in databases, which enabled the system to be more beneficial with regard to making decisions in the long term. This recommendation engine also increased the overall usability of the framework because it offered customized recommendations and self-care advice. Thus, the offered system can be an efficient mental health decision support system in educational institutions and among students.

REFERENCES

[1] Chowdhury, Meghna Roy, et al. “Predicting and understanding college student mental health with interpretable machine learning.” Proceedings of the ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies. 2025.

[2] Zhai, Yusen, et al. “Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students.” Journal of Counseling & Development 103.1 (2025): 110-125.

[3] Fu, You, Fang Ren, and Jiantao Lin. “Apriori algorithm based prediction of students’ mental health risks in the context of artificial intelligence.” Frontiers in Public Health 13 (2025): 1533934.

[4] Luo, Lin, et al. “Predictors of depression among Chinese college students: a machine learning approach.” BMC Public Health 25.1 (2025): 470.

[5] Ebrahimi, Maryam, et al. “The transition from centralized machine learning to federated learning for mental health in education: A survey of current methods and future directions.” arXiv preprint arXiv:2501.11714 (2025).

[6] Fadhilla, Mutia, et al. “Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa.” Journal of Informatics Management and Information Technology 5.1 (2025): 40-47.

[7] Zhang, Xiaoyi, et al. “Optimizing multi label student performance prediction with GNN-TINet: A contextual multidimensional deep learning framework.” PLoS one 20.1 (2025): e0314823.

[8] Karimian, Mehran. “A short review on diagnosing and predicting mental disorders with machine learning.” International Journal of Applied Data Science in Engineering and Health 1.1 (2025): 20-27.

[9] Dharnasi, Prasad. “Electronic gadget addiction prediction using machine learning.” International Journal of Research Publications in Engineering, Technology and Management (IJRPETM) 9.2 (2026): 500-505.

[10] Ta, Alan, et al. “Real-time stress monitoring, detection, and management in college students: A wearable technology and machine-learning approach.” arXiv preprint arXiv:2505.15974 (2025).

[11] Wang, Fang. “Application of Intelligent Medical Treatment in Long-Term Mental Health Monitoring and Early Warning of College Students.” Informatica 49.13 (2025).

[12] Liu, Sannyuya, et al. “Academic stress detection based on multisource data: a systematic review from 2012 to 2024.” Interactive Learning Environments 33.3 (2025): 1823-1849.

[13] Alhasani, Mona, and Rita Orji. “Promoting stress management among students in higher education: Evaluating the effectiveness of a persuasive time management mobile app.” International Journal of Human–Computer Interaction 41.1 (2025): 219-241.
[14] Lu, Zhao, and Lingyu Zhou. “Entertainment robots based on internet of things and artificial intelligence in the process of student stress and emotion recognition.” Entertainment Computing 52 (2025): 100786.
[15] Roy, Sharmistha, Ashis Kumar Biswas, and Manoj Sharma. “Stress, anxiety, and depression as psychological distress among college and undergraduate students: a scoping review of reviews guided by the socio-ecological model.” Healthcare. Vol. 13. No. 16. MDPI, 2025.