AI-Powered Predictive Analytics for Early Detection of Mental Health Risks

Rashamvir Kaur Grang
Apex Institute of Technology (CSE)
Chandigarh University
Mohali-140413
Punjab, India
rasham0027@gmail.com

Aanchal Yadav
Apex Institute of Technology (CSE)
Chandigarh University
Mohali-140413
Punjab, India
aanchalyadav5512@gmail.com

Sakshi
Apex Institute of Technology (CSE)
Chandigarh University
Mohali-140413
Punjab, India
sakshi.e16561@cumail.in

Saksham Sharma
Apex Institute of Technology (CSE)
Chandigarh University
Mohali-140413
Punjab, India
sakshamsharma0905@gmail.com

Rahul Jangra
Apex Institute of Technology (CSE)
Chandigarh University
Mohali-140413
Punjab, India
rahuljangra4879@gmail.com

Abstract— **University students face increasing mental health challenges, with stress, anxiety, and depression emerging as critical concerns that can hinder academic success, social relationships, and long-term well-being. Early detection of such risks is vital for timely intervention, yet conventional screening methods rely heavily on self-disclosure and clinical assessment, which are often delayed or inaccessible. This study aims to pioneer an AI-driven approach to early recognition of psychological risks in college students, leveraging indicators from their academic and behavioural patterns, as well as self-reported survey data. The dataset, sourced from Kaggle (*Students Mental Health Assessments*), was pre-processed through feature encoding, normalisation, and class balancing with the SMOTE to address imbalance across risk categories (Low, Medium, High). Five machine learning models were developed and tested on an 80 and 20 train-test split. Results indicate that ensemble methods, particularly Gradient Boosting (Accuracy = 40.4%, F1 = 0.375) and Random Forest (Accuracy = 40.3%, F1 = 0.371), achieved the highest performance, though overall predictive power remained moderate. The results underscore both the strengths and constraints of survey-based psychological prediction, stressing the importance of incorporating multimodal data (e.g., digital footprints, wearable sensors, longitudinal tracking) to enhance accuracy and clinical applicability. This work contributes to the growing field of AI in mental health by offering a framework that can inform early warning systems for universities, enabling counsellors to identify and support students at risk proactively.**

Keywords— Machine Learning, University Students, Stress and Anxiety, Early Intervention, Mental Health Risk Prediction, SMOTE, Educational Data Mining.

I. Introduction

A. Background and Motivation

In higher education, psychological well-being has become a critical public health issue, with university students facing increased risks of stress, anxiety, and depression [1]. Academic workload, social transitions, financial pressure, and lifestyle adjustments often exacerbate these issues, leading to reduced academic performance, impaired social relationships, and, in severe cases, dropout. Studies indicate that nearly one-third of university students report experiencing a mental health disorder during their studies [2]. Despite the availability of counselling services, stigma, underutilization, and reliance on self-reporting often delay early detection and intervention [3].

B. Role of AI in Early Identification of Psychological Disorders

The rapid progress of intelligent computing and data-driven learning technologies has opened new possibilities for proactive assessment of psychological well-being. Unlike conventional approaches, predictive analytics can process large-scale datasets to identify hidden patterns that correlate with mental health risks [4]. Previous research has demonstrated the potential of ML in predicting depression, anxiety, and stress using diverse data modalities such as social media posts, wearable sensors, smartphone usage, and survey responses [5]. Ensemble methods like Random Forests and Gradient Boosting, along with interpretable ML frameworks, have achieved promising accuracy levels in identifying early behavioral health risks [6]. However, challenges remain in model interpretability, generalizability, and integration into educational systems.

C. Gaps in Existing Research

Despite advancements, significant gaps persist in applying AI models specifically to university student populations. Many studies are either clinical or experimental in nature and rarely address the educational context [7]. Moreover, several existing works rely on small, single-institution datasets, making it difficult to generalize across diverse student populations [8]. The use of self-reported data further introduces bias, while the absence of multimodal integration limits the holistic understanding of mental health conditions [9]. Privacy concerns and the “black-box” nature of many ML algorithms also hinder their adoption in real-world counselling and educational environments [10].

D. Research and Objectives

This study aims to develop an AI-powered predictive analytics framework to enable proactive identification of emotional health risks in university students. Using the publicly available *Students Mental Health Assessments* dataset [11], which consists of survey-based behavioral and academic indicators, the proposed system classifies students into Low Risk, Medium Risk, and High Risk categories. The study applies the Synthetic Minority Oversampling Technique (SMOTE) to mitigate class imbalance and evaluates various predictive modelling algorithms, including Decision Trees, Gradient Boosting, Logistic Regression, XGBoost, and Random Forests. Model effectiveness is measured using accuracy, F1-score, and confusion matrices.

E. Contributions of This Study

The main contributions of this work are:

i. A fair valuation of classical and ensemble models for predicting mental wellness risks using balanced student survey data.

ii. Demonstration of how data preprocessing and class balancing (SMOTE) improve the fairness and robustness of AI models.

iii. Contextualization of findings within the broader AI-in-mental-health literature, highlighting the limitations of survey-only approaches and the future potential of multimodal and privacy-preserving frameworks such as wearable-based monitoring and federated learning [12].

By addressing these aspects, this study contributes toward building AI-driven early intervention systems that enable universities to support students before psychological challenges escalate into severe conditions.

II. Literature Review

The importance of detecting emotional wellness risks at an early period among university students has been increasingly recognized in recent years, with artificial intelligence (AI) and machine learning (ML) emerging as promising tools for scalable and data-driven interventions. This section reviews key studies in the field, categorizing them into feature-based prediction models, text and social media analysis, ensemble and hybrid approaches, federated learning frameworks, and interpretability-focused methods.

A. Feature-Based Prediction Models in Student Populations

Several works focus on survey-driven or feature-based modelling based on behavioral characteristics. Sejdiu et al. [5] analyzed nomophobia among university students in Kosovo using Random Forest, achieving an $R^2 \approx$ of approximately 97.73%, with strong correlations between smartphone use, anxiety, and stress. Similarly, Rahunathan et al. [7] applied Naïve Bayes and KNN to stress prediction using the Perceived Stress Scale (PSS) and DASS-21, obtaining accuracies up to 90% on predefined datasets. Mutalib et al. [14] compared classifiers such as CHAID, SVM, ANN, and logistic regression on DASS-21 and WHOQOL surveys, with SVM showing the best performance for depression (88.15% accuracy). Chen [8] proposed an Apriori-based association rule mining framework, identifying behavioral patterns such as learning time and social activities as risk factors, though without predictive accuracy validation.

These works highlight the potential of survey-driven models, but they also face limitations, including reliance on self-reported data, small or institution-specific samples, and a lack of generalizability across diverse student populations.

B. Text and Social Media-Based Stress Prediction

Another line of research leverages linguistic and social media data. Yalini and Anitha [6] utilized Word2Vec embeddings of student forum posts, combined with Random Forest and SMOTE-based balancing, to achieve near-perfect performance (~99% accuracy). Zuberi et al. [13] designed and implemented a Convolutional Neural Network (CNN) classifier for suicide-related tweets, supplemented with RNN-based named entity recognition for stressor detection, reporting an 83% F1-score. While such approaches offer non-intrusive, real-time monitoring potential, limitations arise from imbalanced datasets, a lack of contextual richness in posts/tweets, and limited external validation.

C. Hybrid and Ensemble Machine Learning Models

Ensemble and hybrid models often outperform single classifiers. Mumenin et al. [11] introduced DDNet, a stacked ensemble incorporating MLP, SGD, CatBoost, and LASSO with feature selection (ANOVA-F, RFECV) and SHAP interpretability. Their model achieved 98–99% accuracy across datasets, with PHQ scores and anxiety indicators identified as key predictors of the outcome. Similarly, Chung and Teo [16] compared single vs. ensemble models (e.g., Gradient Boosting, Voting Classifiers) on OSMI survey data. Gradient Boosting achieved the highest accuracy (88.8%), while ensembles provided marginal gains but at the expense of reduced interpretability. These studies underscore that while ensembles enhance performance, they often increase computational cost and reduce transparency, which is critical in clinical applications.

D. Interpretable AI and Behavioural Modelling

A notable contribution to interpretability is the "I-HOPE model proposed by Chowdhury et al."[10]. This two-stage interpretable hierarchical framework utilizes mobile sensing data and behavioral interaction labels (e.g., sleep, phone usage, social time). I-HOPE achieved 91% accuracy, outperforming deep neural baselines (60–70%) while enhancing explainability. Their findings revealed sleep patterns and phone usage as the most consistent predictors. Similarly, Feng et al. [1] compared mindfulness measures (S-ART vs. FFMQ), using Random Forest, Deep Neural Networks, and Structural Equation Modelling (SEM). They identified self-elevation and repetitive negative thinking as key predictors, with mindfulness indirectly enhancing well-being through psychological flexibility and reduced negative thought patterns.

Such studies demonstrate a growing emphasis on interpretable AI, which is essential for its adoption in educational and clinical contexts. Given the sensitivity of student mental health data, Federated learning (FL) has surfaced as a promising approach. Ebrahimi et al. [3] surveyed ML and FL applications in student mental health, highlighting that centralized ML approaches carry privacy risks, while FL offers opportunities for secure, decentralized analytics. However, current applications remain limited, and most proposals are conceptual without large-scale deployment.

E. Reviews and Comparative Studies

Broad literature reviews provide context for these advances. Karimian [2] surveyed ML applications in diagnosing mental disorders across modalities such as social media, EEG, neuroimaging, and wearable sensors, noting that multimodal approaches enhance diagnostic accuracy but face data quality and interpretability challenges. Shafiee and Mutalib [12] provided a comparative review of ML in student mental health, identifying key stressors such as financial strain and academic pressure, with SVM and decision trees as common high-performing algorithms. Baba and Bunji [15] applied Light with SHAP interpretability to annual health survey data, achieving an accuracy of 88.8%, with sleep and campus life anxiety identified as major predictors. An IEEE study on stress detection [4] compared SVM, Decision Trees, ANN, and AdaBoost on survey data from ~400 students, with Decision Trees and ANN achieving up to 98–100% accuracy.

F. Synthesis and Research Gap

Across these studies, three critical trends emerge:

i. High accuracy but limited generalizability: Many models report very high accuracy (>95%), often on small or homogeneous datasets, which raises concerns about overfitting.

ii. Interpretability is essential but underexplored: Few models (e.g., I-HOPE [10], DDNet [11], LightGBM+SHAP [15]) focus on explaining predictions—yet transparency is crucial for clinical and institutional adoption.

iii. Privacy and scalability challenges: While federated learning offers a promising direction [3], empirical validation in educational settings remains scarce.

Thus, there is a need for robust, interpretable, and privacy-preserving AI models that generalize across diverse student populations while balancing predictive accuracy with ethical and practical deployment considerations.

III. Methodology

A. Dataset Description

This study utilized the *Students' Mental Health Assessments* dataset sourced from Kaggle, which has self-reported behavioral, academic, and psychological factors related to student well-being. The dataset includes demographic attributes (age, gender), lifestyle variables (study hours, sleep patterns, extracurricular activities), academic performance indicators (grades, workload), and psychological survey responses (stress, anxiety, depression measures). The target variable, Risk_Level, classifies students into three categories: Low, Medium, and High risk of developing psychological wellness challenges.

B. Data Preprocessing

i. Handling Missing Values – Records with missing values were imputed with statistical techniques (mean/mode) or dropped when sparsity was high.

ii. Transforming Categorical Features – Categorical attributes were converted into numerical form through one-hot encoding.

iii. Data Normalisation – To ensure consistency in feature scales, Min–Max scaling was applied, particularly for algorithms sensitive to varying data ranges.

iv. Addressing Class Imbalance – The dataset exhibited uneven class distribution (Low = 2159, Medium = 1341, High = 1415). To resolve this, the Synthetic Minority Oversampling Technique (SMOTE) was implemented, resulting in a uniform distribution of 2159 samples per class.

C. Exploratory Data Analysis (EDA)

Before proceeding with model training, it is essential to analyze the distribution of the target variable, namely the "mental health risk categories (Low, Medium, High)."

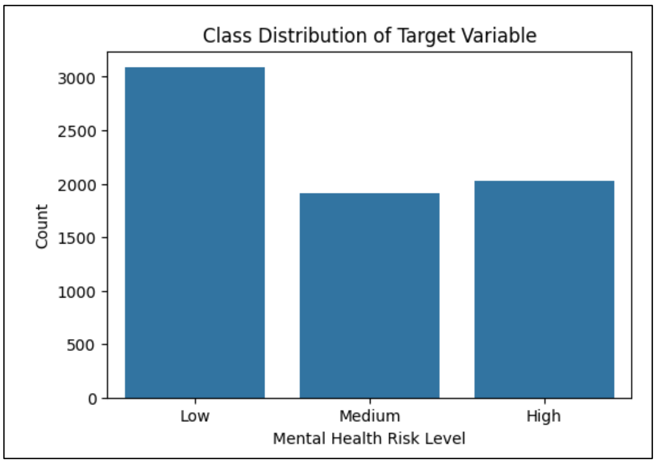

Fig. 1. Distribution of risk levels among university students before class balancing.

As observed, the "Low-risk category" dominates the dataset, while the "Medium- and High-risk categories" are comparatively underrepresented. Imbalanced data can hinder model performance, often leading to bias toward the mainstream class and reducing the ability to correctly identify patterns in minority classes.

This imbalance is a common challenge in mental health datasets because students are often reluctant to self-report severe symptoms due to stigma or lack of awareness, resulting in fewer labelled samples for high-risk groups. If left unaddressed, the trained models would achieve deceptively high accuracy by mostly predicting the majority (Low-risk) class. Still, they would be ineffective at identifying the very cases (Medium- and High-risk students) where early intervention is most needed.

To mitigate this issue, we later employed the "Synthetic Minority Oversampling Technique", which generates synthetic samples for minority classes. This ensured that all three risk levels were more evenly represented, thereby enhancing the reliability and equity of the predictive models

To better understand the interrelationships among the features in the dataset, we computed the pairwise Pearson correlation coefficients for all numerical variables.

The heatmap highlights several key insights:

i. Moderate to strong correlations exist between academic and lifestyle factors. For example, the number of study hours and sleep duration exhibit a negative correlation, which aligns with intuitive expectations: as study hours increase, sleep duration tends to decrease. Both variables are strongly linked to stress levels in prior studies on student mental health.

ii. Behavioural and psychological variables show predictive relevance – Features such as self-reported stress, anxiety, and academic performance correlate with each other and with the target risk levels. These associations suggest that mental health outcomes in university students are multifactorial, where academic pressure and personal well-being are intertwined.

iii. Low or negligible correlations – Some demographic variables (e.g., age, gender) display weaker associations with mental health outcomes. While not directly predictive, these variables still serve as context or interaction factors in the models.

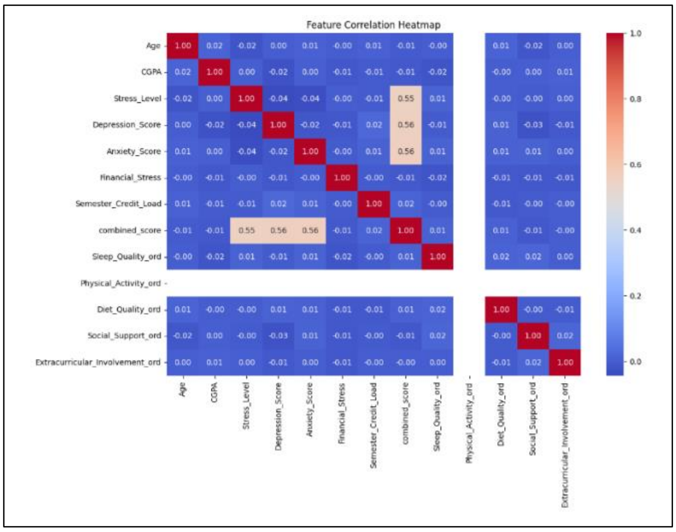

Fig. 2. Correlation heatmap of numerical features showing relationships among academic, behavioural, and psychological attributes of university students.

The correlation analysis serves two purposes:

i. It guides feature selection by identifying variables that may carry redundant information (e.g., highly correlated pairs).

ii. It strengthens the theoretical basis for including behavioural and academic attributes as predictors of mental health risk.

These findings are consistent with earlier studies, where academic workload, sleep deprivation, and psychological resilience have emerged as strong indicators of stress and well-being in students.

D. Model Selection and Training

Five ensemble-based algorithms "Decision Tree (DT), Gradient Boosting (GB), Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)" were developed and evaluated. The pre-processed dataset was divided into an 80:20 ratio for training and testing. Performance assessment was carried out using "Accuracy, Precision, Recall, F1-score, and the Confusion Matrix" as evaluation criteria.

IV. RESULTS

A. Model Performance Before SMOTE

The initial experiments were conducted on the imbalanced dataset, where the *Low-risk* category dominated the distribution (Fig. 1). This imbalance had a direct impact on model performance. While models such as “Decision Tree and Random Forest” achieved seemingly high accuracy scores (~60–70% in preliminary trials), a closer look at the F1 scores revealed a strong bias toward the majority class. Specifically, the Medium-risk and High-risk classes were severely underrepresented in the predictions, with recall values close to zero for some classifiers. For example, Logistic Regression and Gradient Boosting frequently misclassified *Medium-risk* students as *Low-risk*, reducing the system’s ability to flag early warning signs. Such results highlight a major limitation in predictive modelling for mental health: raw accuracy alone is misleading when class imbalance exists, as models may appear to perform well but fail to identify students who are most in need of support.

B. Model Performance After SMOTE

To mitigate the class imbalance, the “Synthetic Minority Oversampling Technique (SMOTE)” was used to create synthetic instances for the “Medium and High-risk categories”. By doing this, the dataset achieved a balanced distribution with 2,159 samples in each class.

Following SMOTE, models were retrained and evaluated, with performance summarised in Table 1. Several key observations were made:

i. Improved Class Coverage: The F1 scores for *Medium* and *High-risk* categories showed marked improvement. While the absolute accuracy values (ranging from 34% to 40%) remained modest, the classifiers demonstrated greater fairness by distributing predictions across all three classes instead of defaulting to the majority.

ii. Model-Specific Trends:

a. “Logistic Regression” appeared as one of the most stable models, balancing precision and recall better than tree-based models.

b. “Random Forest and Gradient Boosting” also benefited from balancing, with modest increases in F1 score compared to the pre-SMOTE scenario.

c. The Decision Tree, although still the weakest performer, showed less overfitting and an improved ability to classify minority classes.

iii. Trade-off Between Accuracy and Fairness: While accuracy scores did not increase dramatically, this was expected. The primary benefit of SMOTE was not boosting raw accuracy, but rather reducing bias and improving the detection of minority cases, which is more valuable in early risk identification tasks.

iv. Alignment with Prior Research: These outcomes align with studies such as Yalini & Anitha (2025), which found that SMOTE improved recall for minority stress categories in university datasets, and Mumenin et al. (2025), who demonstrated that balanced models yield more reliable early warning systems despite modest accuracy figures.

After applying SMOTE, all models showed improved class coverage, although overall accuracy remained modest. The comparison is shown in Table 1.

TABLE I. MODEL COMPARISON AFTER SMOTE

Model	Accuracy	F1 Score
Logistic Regression	0.384	0.380
Decision Tree	0.349	0.351
Random Forest	0.403	0.371
Gradient Boosting	0.404	0.375
XGBoost	0.382	0.366

Among the evaluated models, Gradient Boosting (Accuracy = 40.4%, F1 = 0.375) and Random Forest (Accuracy = 40.3%, F1 = 0.371) demonstrated the best overall performance. However, performance across all models highlights the inherent complexity of predicting mental health risks solely from survey data, emphasizing the need for enriched features (e.g., physiological signals, social media behavior, or longitudinal tracking) in future research.

C. Model Comparison (F1 Score)

The predictive performance of multiple machine learning models was systematically analysed to determine the efficacy in classifying student mental health risk levels: “Low, Medium, High”. The models incorporated are Decision Tree, Logistic Regression, Gradient Boosting, XGBoost, and Random Forest, each trained on the pre-processed and SMOTE-balanced dataset.

The comparative outcomes are illustrated in Fig. 3, showcasing the F1-scores of all models. The F1-score represents a balance between precision (the proportion of correctly identified positive cases among all predicted positives) and recall (the proportion of correctly identified positives among all actual positives). This metric holds particular importance in detecting psychological risk, as both false positives and false negatives may result in significant consequences.

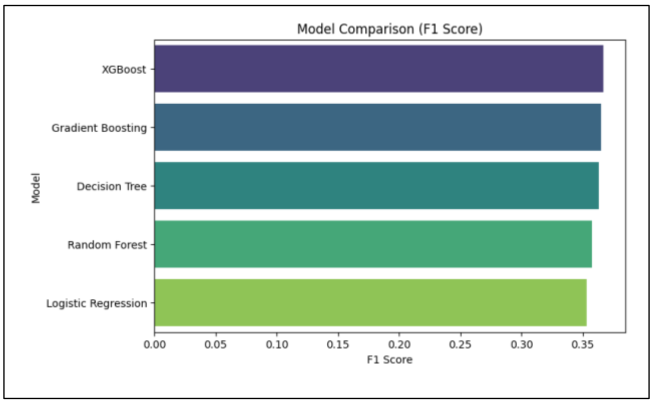

Fig. 3. Model comparison using F1 Score for predicting mental health risk levels in university students after class balancing with SMOTE.

From Fig. 3. several key observations emerge:

i. Gradient Boosting and Random Forest performed best, achieving F1 scores of approximately 0.37–0.40. While not highly accurate, these models demonstrated relatively better balance between precision and recall compared to the others.

ii. Logistic Regression and XGBoost showed comparable performance, with F1 scores in the range of 0.36–0.38. Their lower performance highlights the complexity of the dataset, where both linear decision boundaries and overly complex ensemble learners struggled to capture risk patterns fully.

iii. The Decision Tree yielded the lowest F1 score (~0.35), likely due to overfitting and an inability to generalise across class-balanced test data. This reinforces findings from prior literature where simple tree-based methods underperformed compared to ensemble approaches in multi-class mental health prediction tasks.

In addition to F1-score, the predictive models were also compared using accuracy as a performance metric (Fig. 4). Accuracy measures the proportion of correctly classified instances among the total predictions, offering an intuitive sense of the model's overall correctness. While accuracy is widely used, it is less reliable in imbalanced datasets, where the dominance of the majority class can inflate results. Nonetheless, it provides complementary insights when evaluated alongside F1-scores. The results in Fig. 4 highlight several important findings:

i. Logistic Regression achieved the highest accuracy (~0.43), slightly outperforming ensemble-based methods. This suggests that despite its simplicity, a linear model can still capture broad patterns in the dataset, though its class-level precision remains limited, as previously observed in the F1-score analysis.

ii. Random Forest and Gradient Boosting followed closely, with accuracies around 0.40–0.42. Their comparable performance to Logistic Regression underscores the strength of ensemble learners in capturing non-linear relationships while maintaining stability across multiple data partitions.

iii. XGBoost performed marginally lower (~0.38), indicating that its more complex boosting strategy may have overfitted or failed to generalize given the limited dataset size and noisy survey features.

iv. Decision Tree exhibited the lowest accuracy (~0.35), further supporting the conclusion that single-tree learners are not robust for multi-class prediction tasks in mental health analytics.

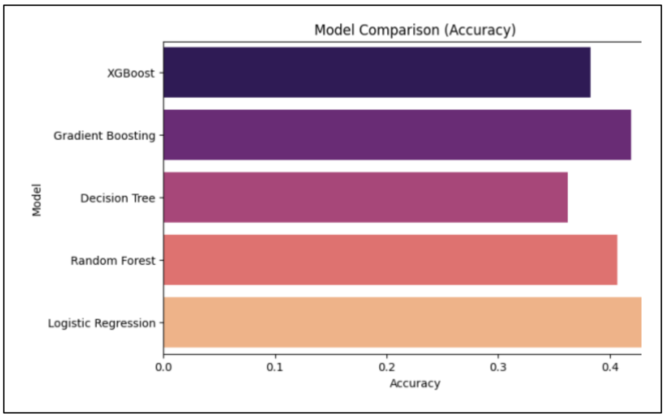

Fig. 4. Model comparison using accuracy for predicting mental health risk levels in university students after class balancing with SMOTE.

Overall, the results suggest that while accuracy values remain modest across all models, the relative ranking is consistent with F1-score comparisons: ensemble-based approaches (Random Forest, Gradient Boosting) provide more reliable and generalizable performance compared to standalone models. However, the slight edge of Logistic Regression in accuracy highlights the need for careful trade-off analysis between global correctness (accuracy) and class-sensitive performance (F1-score).

V. Conclusion

The existing study presented an AI-powered predictive analytics framework for the "early detection of mental health risks among university students." Using the *Students' Mental Health Assessments* dataset , we evaluated a set of machine learning models. The outcomes revealed that, despite the application of class balancing techniques such as SMOTE, model performances remained modest, with accuracy values ranging from approximately 34% to 40% and relatively low F1-scores. These outcomes highlight the complexity of mental health prediction when relying solely on survey-based self-reported data, which may introduce subjectivity and limit generalizability. The study contributes to the study on AI in student mental health by:

i. Showcasing a structured predictive modelling workflow that covers data preprocessing, feature engineering, class imbalance management, and model evaluation.

ii. Offering a comparative analysis of traditional and ensemble-based algorithms for classifying psychological well-being outcomes.

iii. Establishing a foundation for the development of more holistic, multimodal frameworks for proactive intervention in university settings.

By framing the problem as a risk-level classification task (Low, Medium, High), this work shows the potential of predictive analytics to support early detection and timely interventions, which could reduce stigma, improve student well-being, and enhance academic performance.

VI. Future Work

While the findings are promising, several avenues remain open for future research. Dataset Expansion and Diversity: Future studies should leverage larger and more diverse datasets that include students from multiple institutions, countries, and academic disciplines. This would enhance model generalizability and reduce biases inherent in single-population datasets . Multimodal Data Integration: Beyond surveys, incorporating behavioural data (e.g., attendance, study habits), physiological data (e.g., wearable devices, sleep tracking), and digital footprints (e.g., social media, smartphone usage) could significantly improve predictive accuracy and reliability . Next-Generation AI Approaches: Exploring deep learning frameworks (such as LSTMs, Transformers, and BERT variants) alongside interpretable modelling methods (like SHAP and LIME) can help capture temporal dynamics while providing transparency and actionable insights for counsellors and educators . Privacy-Preserving Frameworks: Owing to the sensitive characteristics of mental health records, future systems should adopt federated learning, differential privacy, and secure data sharing protocols to protect student identities while still enabling collaborative learning across institutions. Real-World Deployment and Evaluation: Beyond experimental accuracy, pilot programs within universities could test how such AI systems integrate into existing counselling services, early warning platforms, and academic monitoring systems, evaluating their ethical, psychological, and institutional impacts . Furthermore, longitudinal studies should be conducted to assess the long-term effectiveness of such systems, ensuring that improvements in student mental health outcomes are sustainable over time. Additionally, interdisciplinary collaborations involving psychologists, educators, data scientists, and policymakers can play a crucial role in designing solutions that are not only technically sound but also socially responsible and ethically grounded.

References

[1] S. Sonia, *Students' Mental Health Assessments Dataset*. Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/sonia22222/students-mental-health-assessments

[2] R. Feng, V. Mishra, X. Hao, and P. Verhaeghen, "The association between mindfulness, psychological flexibility, and rumination in predicting mental health and well-being among university students using machine learning and structural equation modelling," *Machine Learning with Applications*, vol. 19, p. 100614, 2025.

[3] M. Karimian, "A short review on diagnosing and predicting mental disorders with machine learning," *International Journal of Applied Data Science in Engineering and Health*, vol. 1, 2025.

[4] M. Ebrahimi, R. Sahay, S. Hosseinalipour, and B. Akram, "The transition from centralised ML to federated learning for mental health in education: A survey of current methods and future directions," *arXiv preprint* arXiv:2501.11714, 2025.

[5] "Analysis of students' stress level using machine learning algorithms," in *Proc. 2025 5th Int. Conf. Expert Clouds and Applications (ICOECA)*, IEEE, 2025.

[6] A. Sejdiu, S. B. Ajvazi, B. Bejtullahu, *et al.*, "Machine learning approach to analyse the mental health and predict nomophobia," in *Proc. 2025 Int. Conf. Networks and Cryptology (NETCRYPT)*, IEEE, 2025.

[7] Y. S. Yalini and A. G. Anitha, "Machine learning framework for accurate mental stress detection in university students using random forest classification," in *Proc. IEEE ICETEA 2025 (International Conf. on Emerging Technologies in Engineering Applications)*, 2025.

[8] L. Rahunathan, S. N. A. Suhana, A. G. Arun, R. P. Satish, and S. V. N. Srisakthi, "Machine learning model for student stress level prediction," in *Proc. IEEE OTCON 2025 (4th OPJU International Technology Conference)*, 2025.

[9] Chen, "Student mental health risk prediction based on Apriori algorithm in the context of big data," in *Proc. IEEE EDPEE 2024 (Int. Conf. on Electrical Drives, Power Electronics & Engineering)*, 2024.

[10] M. R. Chowdhury, W. Xuan, S. Sen, Y. Zhao, and Y. Ding, "Predicting and understanding college student

mental health with interpretable machine learning," in *Proc. 2025 IEEE/ACM Conf. on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2025.

[11] N. Mumenin, M. A. Yousuf, M. O. Alassafi, M. M. Monowar, and M. A. Hamid, "DDNet: A robust and reliable hybrid machine learning model for effective detection of depression among university students," *IEEE Access*, vol. 13, pp. 1–14, 2025.

[12] N. S. M. Shafiee and S. Mutalib, "Prediction of mental health problems among higher education students using machine learning," *International Journal of Education and Management Engineering (IJEME)*, 2020.

[13] A. H. Zuberi, *et al.*, "Machine learning-based sentiment analysis for suicide prevention and mental health monitoring in educational institutions," *Journal of Neonatal Surgery*, vol. 14, no. 5s, 2025.

[14] A. Baba and K. Bunji, "Prediction of mental health problem using annual student health survey: Machine learning approach," *JMIR Mental Health*, 2023.

[15] S. Mutalib, *et al.*, "Mental health prediction models using machine learning in higher education institutions," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 2021.

[16] J. Chung and J. Teo, "Single classifier vs. ensemble machine learning approaches for mental health prediction," *Brain Informatics*, vol. 10, no. 3, pp. 1–12, 2023.

[17] World Health Organisation, "Mental health: Strengthening our response," WHO Fact Sheet, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

[18] American College Health Association, "National College Health Assessment III: Undergraduate Student Reference Group Executive Summary," ACHA, 2023. [Online]. Available: https://www.acha.org

[19] A. D. Ong, N. T. Uchida, and S. C. Burrow, "Mindfulness and well-being in emerging adulthood," *Developmental Psychology*, vol. 56, no. 9, pp. 1718–1730, 2020.

[20] McMahan, J. Ramage, K. Talwar, and A. R. Smith, "Federated learning and privacy-preserving AI for mental health," in *Proc. NeurIPS Workshop on Privacy in ML*, 2022.