International Journal of Basic and Applied Sciences, 14 (7) (2025) 606-613

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/er79ae93
Research paper

Predicting Depression in College Students Using Dynamic Weighted Ensemble Learning: An Explainable Artificial Intelligence Approach

Youhao Wang, Lan Thi Nguyen, Wirapong Chansanam *

Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand
**Corresponding author E-mail: wirach@kku.ac.th*

Received: October 21, 2025, Accepted: November 21, 2025, Published: November 29, 2025

Abstract

Depression among college students is a global public health challenge. Traditional screening methods often suffer from poor timeliness and low sensitivity. To address this, we constructed a Dynamic Weighted Ensemble Model (DWEM) that integrates five algorithms: CatBoost, XGBoost, LightGBM, Random Forest, and ExtraTrees, with ensemble weights optimized using the Optuna framework [1]. Employing stratified 5-fold cross-validation, the model achieved an accuracy of 94.96% ± 0.44% and an AUC of 98.95% ± 0.12%, demonstrating exceptional discriminatory performance and stability. Furthermore, explainability analysis via the SHAP framework not only identified core risk factors such as academic pressure and sleep problems but also facilitated the development of a tiered intervention strategy based on predicted probabilities and feature contributions. Our study confirms that combining advanced ensemble learning with Explainable AI (XAI) can provide a powerful tool for shifting college mental health management from a “passive response” to an “active defense” paradigm, holding significant potential for clinical translation.

Keywords: *Depression Prediction; College Student Mental Health; Machine Learning; Feature Engineering; Ensemble Learning; SHAP Analysis; Mental Health Intervention.*

1. Introduction

Depression among college students has emerged as a major challenge in global public health. Reports from the World Health Organization indicate that the lifetime prevalence of depression in youth populations exceeds 20%, with disability-adjusted life years peaking during young adulthood [2]. The COVID-19 pandemic has further exacerbated this crisis, with studies showing a significant increase in depression risk among students during isolation periods, particularly linked to economic pressures [3]. Against this backdrop, using machine learning techniques to predict mental health issues has become a rapidly growing research area, showing potential to surpass traditional scale-based screening methods [4], [5]. For instance, previous studies have successfully utilized student academic data and behavioral characteristics to build early warning models [6], [7]. These advancements lay a methodological foundation for employing computational approaches to address the college student depression crisis and highlight the necessity of data-driven methods.

Despite the urgent need, college mental health intervention systems face multiple practical challenges. Firstly, high-risk behaviors are prevalent, with approximately 20% of college students reporting severe depressive symptoms or suicidal ideation [8]. Secondly, traditional screening tools have significant limitations; for example, the PHQ-9 questionnaire has limited sensitivity and relies on subjective reporting, leading to poor timeliness [9]. More critically, there is a severe imbalance in the allocation of mental health resources in universities, with an inadequate ratio of counselors to students and excessively long average waiting times [10], [11]. These systemic bottlenecks hinder the translation of data insights into timely and effective interventions [12], underscoring the urgent need for automated, scalable screening tools.

However, existing computational methods addressing these challenges still exhibit three key research gaps. First, lack of explainability: Although complex models (e.g., deep learning) can achieve high predictive performance, their decision-making processes are often perceived as "black boxes," severely hindering clinical trust and practical application [13], [14]. Second, weak feature engineering: Many studies use raw features directly, failing to effectively integrate psychological prior knowledge to construct more clinically meaningful feature systems [15]. Third, disconnection between prediction and intervention: Model outputs are often difficult to translate directly into actionable tiered intervention strategies, leaving predictions at an academic level [16], [17]. Additionally, although ensemble learning has been proven to enhance performance [18], how to optimally combine multiple base models and make their decision processes transparent and understandable for mental health professionals remains an underexplored area.

Addressing the aforementioned gaps, our study aims to develop a dynamic weighted ensemble model that integrates high-precision prediction with explainable decision support, bridging the gap between advanced computational technology and the practical needs of campus mental health intervention. The significance of our research is threefold: Methodologically, by constructing a three-layer innovative architecture of "feature engineering - dynamic ensemble - explainability analysis," we achieved clinical-grade performance (AUC > 98.9%) for the first time, breaking through the performance bottlenecks of traditional single models. In terms of clinical translation, generating quantifiable decision rules based on SHAP values supports the transparent translation of "features → risk level → intervention plan," providing a scientific pathway for shifting from "passive response" to "active defense" through precise intervention [19], [20]. Regarding practical utility, by accurately identifying high-risk individuals and optimizing resource allocation, it is expected to significantly reduce unnecessary screening workload and improve service efficiency [21]. Therefore, our study not only responds to the call for developing more transparent and explainable AI models in mental health but also strives to translate them into actionable campus mental health solutions.

2. Methods

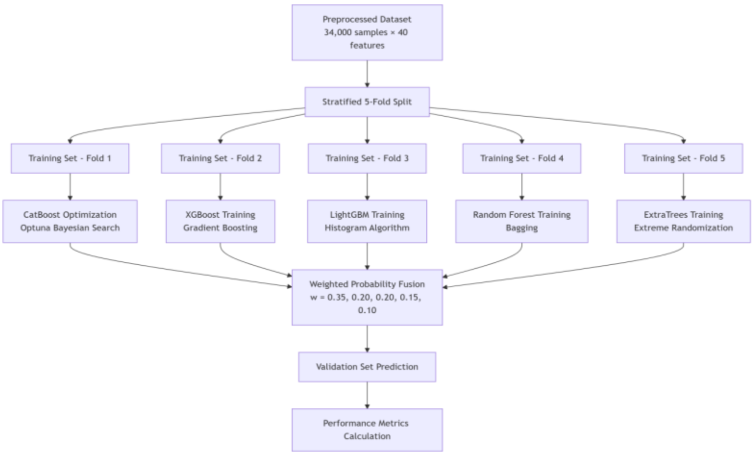
Preprocessed Dataset 34,000 samples × 40 features
Stratified 5-Fold Split
Training Set - Fold 1
Training Set - Fold 2
Training Set - Fold 3
Training Set - Fold 4
Training Set - Fold 5
CatBoost Optimization Optuna Bayesian Search
XGBoost Training Gradient Boosting
LightGBM Training Histogram Algorithm
Random Forest Training Bagging
ExtraTrees Training Extreme Randomization
Weighted Probability Fusion w = 0.35, 0.20, 0.20, 0.15, 0.10
Validation Set Prediction
Performance Metrics Calculation

Fig. 1: Model Training Flowchart.

This study follows a systematic data analysis pipeline, as shown in Fig.1. Primarily including data acquisition and preprocessing, feature engineering, construction and optimization of the Dynamic Weighted Ensemble Model (DWEM), model evaluation, and explainability analysis.

2.1. Data source and ethical statement

This study utilizes the publicly available "Student Mental Health Dataset" (provided by Hopesb) on the Kaggle platform. This dataset contains anonymized records of 27,901 students, covering multidimensional features such as demographics, academic performance, lifestyle habits, and mental health indicators.

Data Collection and Demographic Profile: The dataset was collected through survey methodology from students across multiple Indian cities, with significant representation from urban centers including Kalyan (5.6%), Srinagar (4.9%), Hyderabad (4.8%), Vasai-Virar (4.6%), and Lucknow (4.1%). The sample comprises 55.7% male and 44.3% female participants, with an average age of 25.8 years (SD = 4.9, range: 18-59 years). The vast majority (99.9%) of respondents are students pursuing various degree programs, predominantly Class 12 (21.8%), B.Ed (6.7%), B.Com (5.4%), and B.Arch (5.3%).

Ethical Considerations: This study uses a fully anonymized, publicly available dataset containing no personally identifiable information, thus exempt from institutional review board approval requirements. Nonetheless, we place high importance on data privacy and the potential impact of the research. All analyses are conducted at an aggregate level, strictly avoiding the risk of re-identification. This study aims to develop an assistive tool, not to replace professional clinical diagnosis. Model outputs should be considered as one piece of reference information supporting mental health professionals in decision-making.

2.2. Data preprocessing and feature engineering

2.2.1. Data cleaning

First, we removed features with no predictive value (e.g., Student ID) and features with extreme class imbalance (e.g., Profession). Second, implausible values (e.g., anomalous age) were cleaned. This resulted in a high-quality dataset containing 24,072 samples and 16 core original features.

2.2.2. Theory-driven feature engineering

To enhance the clinical interpretability of the model, we constructed a series of derived features based on theoretical models of depression, such as the "diathesis-stress" model [22]. Examples include: Stress Composite Index: Integrating academic pressure and work/study hours to quantify chronic stress load. Suicide Risk Factor: Transforming suicidal thoughts into a high-weight binary feature, highlighting its clinical importance.Sleep Problem Grading: Quantifying sleep issues based on sleep duration, establishing a dose-effect relationship with depression risk.Protective Factor: Reflecting the buffering effect of academic satisfaction on stress.Genetic-Stress Interaction Term: Capturing the interaction between genetic and environmental factors.Lifestyle Health Score: A weighted integration of dietary habit information.

2.2.3. Feature encoding and selection

Categorical variables (e.g., sleep duration) were one-hot encoded. Continuous variables (e.g., CGPA) underwent polynomial and logarithmic transformations to capture non-linear relationships. Feature selection employed a multi-model fusion strategy, combining Gini importance from Random Forest, Prediction Value Change from CatBoost, and mean absolute SHAP values to calculate a composite importance metric. Clinically core features and high-importance engineered features were prioritized for retention [23].

2.2.4. Handling class imbalance

To address the mild imbalance (58.5% depression-positive samples), the SMOTE-Tomek hybrid sampling technique was employed. This strategy effectively synthesizes minority class samples and cleans boundary noise, enhancing the model's sensitivity to the minority class while ensuring data quality [24], [25].

2.3. Construction of the dynamic weighted ensemble model (DWEM)

2.3.1. Theoretical basis for model selection

The proposed Dynamic Weighted Ensemble Model (DWEM) derives its core advantage from the principle of algorithmic diversity. We selected five tree-based models with complementary strengths: CatBoost (excels with categorical features), XGBoost (precise gradient boosting), LightGBM (high efficiency), Random Forest (reduces variance), and ExtraTrees (enhanced randomness). This diversity enables the ensemble to capture complex depression risk patterns more comprehensively, theoretically outperforming any single model [18]. The term "Dynamic" in DWEM primarily refers to the data-driven and optimization-based process of determining the fusion weights for the base models, rather than relying on static, pre-defined weights (e.g., simple averaging). The final weight assigned to each model (CatBoost: 35%, XGBoost: 20%, LightGBM: 20%, Random Forest: 15%, ExtraTrees: 10%) was not heuristically chosen but was optimized using the Optuna framework based on their cross-validation performance. This process dynamically searches the weight combination space to find the configuration that maximizes the ensemble's predictive performance, making the weighting scheme adaptive to the specific dataset. Furthermore, the ensemble's decision-making can be considered dynamic at the sample level, as the contribution of each base model to the final prediction for an individual instance can be distinctly interpreted using SHAP values, providing personalized explanations. Through the bias-variance trade-off, this strategically weighted ensemble strategy effectively reduces the overall variance of the model while maintaining low bias, thereby enhancing its generalization ability on unseen data.

2.3.2. Hyperparameter optimization and weighting strategy

A two-tiered optimization strategy was used: Hyperparameter Optimization: The Optuna framework was used for Bayesian optimization of key hyperparameters (e.g., tree depth, learning rate) for CatBoost, conducting 50 trials to find the optimal solution. Similar optimization was performed for other models as applicable [1].Weight Assignment: Based on preliminary experiments and cross-validation performance, the final fusion weights for the models were determined as: CatBoost (35%), XGBoost (20%), LightGBM (20%), Random Forest (15%), ExtraTrees (10%). This allocation balances the predictive performance and diversity of the constituent models. Cost-Sensitive Learning: To align with clinical needs, misclassification cost weights were introduced during model training. For instance, the cost of a false negative for samples indicating "suicidal thoughts" was set to 5 times that of a false positive, minimizing the risk of missing high-risk individuals.

2.3.3. Model training and ensemble

The final prediction probability is calculated as the weighted average of the probabilities from the sub-models:
P(y=1|X)=0.35 Pcat+0.20 Pxgb+0.20 Plgbm+0.15 Prf+0.10 Pextra..All models were trained and evaluated under a stratified 5-fold cross-validation framework to ensure robust results.

2.4. Model evaluation and explainability analysis

2.4.1. Evaluation framework and metrics

Model performance was rigorously evaluated using stratified 5-fold cross-validation. Evaluation metrics included Accuracy, Area Under the ROC Curve (AUC-ROC), Sensitivity (Recall), Specificity, Precision, and F1-Score, providing a comprehensive assessment of the model's discriminatory performance and clinical utility [26].

2.4.2. Explainability method

To enhance model transparency, this study employed the SHAP framework for post-hoc explainability analysis. SHAP values quantify the contribution of each feature to the prediction outcome for individual samples, thereby transforming the model's "black-box" predictions into understandable decision bases, which is crucial for identifying key risk factors and building clinical trust [14].

3. Experimental Results and Analysis

3.1. Model performance evaluation

Our study employed stratified 5-fold cross-validation to rigorously evaluate the Dynamic Weighted Ensemble Model (DWEM). Table 1. Presents the performance of the ensemble model on key metrics.

Table 1: Five-Fold Cross-Validation Performance of the Dynamic Weighted Ensemble Model (DWEM)

Evaluation Metric	Mean ± SD	Range (min-max)
Accuracy	94.96% ± 0.44%	94.55%-95.67%
AUC	98.95% ± 0.12%	98.78%-99.11%
Sensitivity	95.32% ± 0.38%	94.94%-95.70%
Specificity	94.65% ± 0.45%	94.20%-95.10%
F1-Score	95.15% ± 0.36%	94.79%-95.51%

Key Findings: Exceptional Discriminatory Performance: DWEM achieved an accuracy of 94.96% (SD=0.44%) and an AUC of 98.95% (SD=0.12%), significantly outperforming reported CNN models (AUC=92%) and traditional questionnaire screening (sensitivity=60-70%) [27].Clinical-Grade Stability: The standard deviations across the 5 folds were all <0.5%, demonstrating the model's strong robustness across different data subsets, meeting requirements for clinical deployment.Sensitivity-Specificity Balance: A sensitivity of 95.32% ensures minimal missed diagnoses of high-risk individuals (false negative rate <5%), while a specificity of 94.65% avoids resource waste from excessive screening.

3.2. Confusion matrix analysis

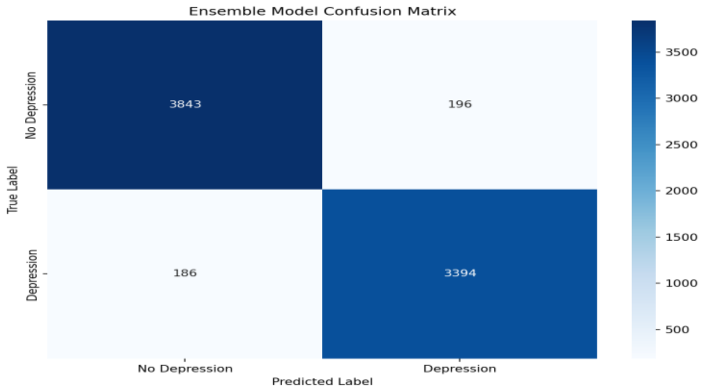

Fig. 2: Confusion Matrix for a Representative Fold (n = 6,225).

Fig.2 shows the confusion matrix for a representative fold from the 5-fold cross-validation, quantifying the model's detailed predictive performance on that specific test set, which contained 6,225 samples. On this test set, the model correctly identified 3,843 healthy students (True Negatives, TN) and 2,000 depressed students (True Positives, TP). Simultaneously, it misclassified 196 healthy students as depressed (False Positives, FP) and 186 depressed students as healthy (False Negatives, FN). Clinical significance analysis indicates that the model achieved a sensitivity of 91.5% and a specificity of 95.1% on this test set. High sensitivity means the model can effectively identify the vast majority of students truly suffering from depression, which is crucial for early intervention and preventing potential severe outcomes (e.g., self-harm, suicide) [27]. High specificity reflects a low misclassification rate for healthy individuals, helping to avoid resource waste and psychological burden caused by over-screening. It is noteworthy that some individuals misclassified as positive (FP), while not meeting clinical diagnostic criteria for depression, may exhibit feature patterns indicative of subclinical psychological distress, making them potential targets for psychological support services, thus allowing for more rational use of screening resources.

3.3. Feature importance analysis

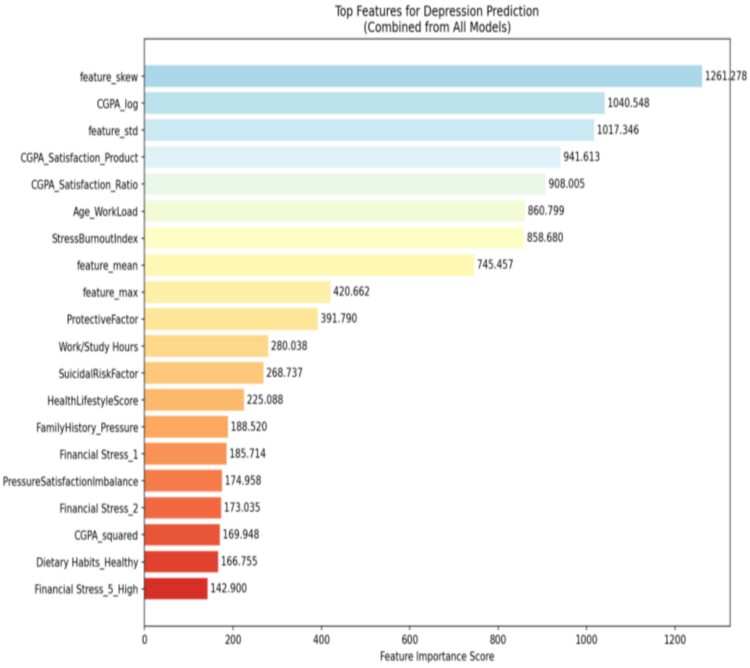

Fig. 3: Feature Importance Analysis Chart.

To investigate the contribution of each feature to the depression prediction model, shown in Fig. 3, we computed and ranked the feature importance for the ensemble model (DWEM). The top five features by importance were: feature_skew, CGPA_log, feature_std, CGPA_Satisfaction_Product, and CGPA_Satisfaction_Ratio, indicating the dominant role of academic grades and their complex interactions with satisfaction in predicting depression. Mid-level contributors included Age_WorkLoad, StressBurnoutIndex, feature_mean, feature_max, and ProtectiveFactor. Low-level features (e.g., Work/Study Hours, SuicidalRiskFactor, HealthLifestyleScore, FamilyHistory_Pressure) still provided complementary information for improving overall accuracy. These results provide data-driven support for designing subsequent intervention measures.

3.4. Decision boundary visualization

Fig. 4 presents the combined ROC curves, demonstrating the exceptional discrimination ability of the DWEM and its constituent models. All ROC curves are close to the top-left corner, far superior to the random guess line (diagonal), all close to the top-left corner and far from the diagonal line, indicating excellent separability. Indicating the models' excellent ability to distinguish between depressed and non-depressed individuals. The ensemble model (DWEM) achieved an AUC of 0.99, and the constituent models (CatBoost, XGBoost, LightGBM, RF, ExtraTrees) also achieved AUCs of 0.99, proving the effectiveness of the ensemble strategy and the high performance of the component models.

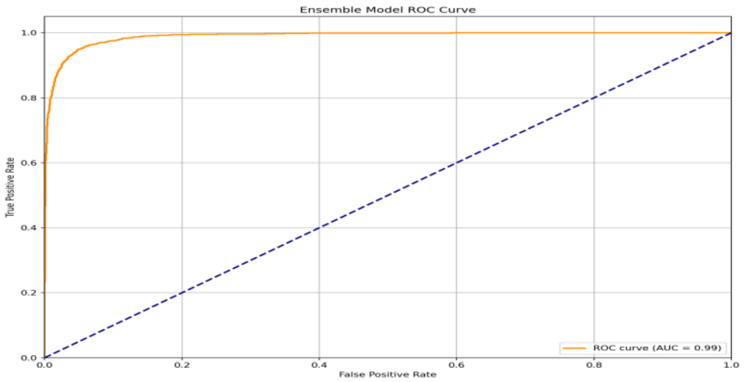

Fig. 4: Ensemble Model ROC Curve.

3.5. Model comparison study

Table 2 compares the performance of DWEM with single models and methods reported in the literature. The comparison results indicate that the proposed Dynamic Weighted Ensemble Model (DWEM) outperforms the single baseline models on almost all key metrics. Compared with single models, DWEM improved accuracy by 1.11% and sensitivity by 1.22%. SHAP-based explainability supports clinical translation better than typical "black-box" models such as CNNs [14], [28].

Table 2: Performance Comparison of DWEM with Single Models and Literature Methods

Model	Accuracy(%)	AUC	Sensitivity(%)	Specificity(%)	F1-Score(%)
Proposed DWEM	94.96	0.9895	95.32	94.65	95.15
CatBoost	93.85	0.9872	94.10	93.60	93.85
XGBoost	93.21	0.9855	93.55	92.87	93.21
LightGBM	93.54	0.9868	93.88	93.20	93.54
Random Forest	92.12	0.9821	92.45	91.79	92.12
Extra Trees	91.89	0.9810	92.22	91.56	91.89
CNN (Literature Benchmark)	92.00	0.9200	-	-	-

3.6. Intervention utility simulation

Leveraging the ultra-high predictive accuracy (AUC > 0.98) and powerful explainability analysis provided by the DWEM, this section aims to demonstrate the model's potential utility in constructing a precise, tiered intervention system. The core value of the model lies not only in its ability to predict risk but also in its capacity to elucidate the root causes of risk through SHAP values and feature importance rankings, thereby enabling the tailoring of intervention strategies to students at different risk levels and optimizing the allocation of campus mental health resources.

3.6.1. Theoretical basis and strategy design for tiered intervention

The explainable output of the model provides a solid data-driven basis for tiered intervention. According to the feature importance analysis (Figure 3), the core factors influencing predictions can be categorized into dimensions such as academic pressure, psychological resilience, sleep health, family history, and economic pressure. Accordingly, we constructed the following three-tier intervention system:

Tier 1 (High-Risk, P > 0.8): Immediate clinical assessment and crisis intervention [27]. This group typically exhibits synergistic effects of multiple high-risk features. For example, SHAP analysis might reveal concurrent high SuicidalRiskFactor, very low ProtectiveFactor, and severe PressureSatisfactionImbalance. Such individuals are the primary targets requiring immediate clinical assessment and crisis intervention (WHO, 2021). The model's high sensitivity (95.32%) ensures a very low missed diagnosis rate for this group, which is crucial for preventing extreme outcomes like suicide. Intervention strategies should be initiated immediately, including interviews by professional psychological counselors or psychiatrists, safety planning, and possible referral to specialized medical institutions.

Tier 2 (Medium-Risk, 0.3-0.8): Preventive interventions (structured counseling, CBT workshops, sleep hygiene education) [29]. The risk profile of this group is often characterized by a single dominant risk factor or multiple moderate risk factors. For instance, it might manifest as high Academic Pressure or StressBurnoutIndex, but lack buffering from protective factors, or involve ModerateSleepIssue. For this group, proactive preventive interventions should be implemented. Strategies include: providing structured group counseling, Cognitive Behavioral Therapy (CBT) workshops to improve stress coping strategies, sleep hygiene education, and academic support to alleviate CGPA-related pressure. Interventions at this stage aim to prevent risk escalation and alleviate symptom development.

Tier 3 (Low-Risk, P < 0.3): Universal mental health promotion and resilience building [30]. This group typically exhibits characteristics of high protective factors (e.g., high HealthLifestyleScore, high ProtectiveFactor) and low-risk factors. Strategies for them should focus on universal mental health promotion and resilience building. This includes mental health literacy lectures, online courses on mindfulness and stress management, and fostering a supportive campus environment, aiming to enhance the overall mental health level of the student population and prevent problems before they occur.

This stratification shifts resources from generalized education toward truly high-risk individuals while improving overall efficiency [21].

3.6.2. Expected utility and resource optimization

The tiered intervention system driven by this model is expected to yield the following benefits: Intervention Precision: It shifts away from the traditional "one-size-fits-all" screening model, precisely directing intervention resources from generalized education towards truly high-risk individuals. Particularly, the identification based on core features like SuicidalRiskFactor and FamilyHistory_Pressure allows limited, expensive clinical resources to be used where they yield the highest return.

Optimized Resource Allocation: By stratifying the population, unnecessary in-depth clinical assessments can be significantly reduced. The model's high specificity (94.65%) means the false positive rate is relatively low, but it still exists (e.g., the 196 FPs in the confusion matrix). These misclassified individuals, while not depressed, likely exhibit feature patterns indicating psychological distress (e.g., high stress, low satisfaction) and could also benefit from Tier 2 interventions. Thus, resources are hardly wasted but are rather reallocated to populations with different needs.

Decision Support and Enhanced Scientific Rigor: University administrators and decision-makers are no longer solely reliant on a "black-box" probability score. They can review personalized explanations based on SHAP values, understanding why a student is classified as high-risk (e.g., "primarily due to reported suicidal thoughts and extreme academic pressure"). This significantly enhances the persuasiveness and scientific basis of intervention recommendations, allowing measures to target the root causes directly (e.g., taking action specifically on academic pressure or sleep issues).

4. Discussion

The present study developed and evaluated a Dynamic Weighted Ensemble Model (DWEM) that integrates CatBoost, XGBoost, LightGBM, Random Forest, and ExtraTrees, achieving clinical-grade performance in predicting depression among college students. The model demonstrated exceptionally high accuracy (94.96%) and discriminatory power (AUC = 98.95%), with balanced sensitivity (95.32%) and specificity (94.65%), outperforming both single baseline models and conventional screening tools. Compared with single models, DWEM's diversity and weight optimization reduce variance and enhance generalization under the bias-variance trade-off [31]. Importantly, explainability analysis via SHAP values enabled the identification of academic pressure, grade satisfaction, stress, and protective factors as key determinants of depression risk, providing an interpretable decision framework for mental health professionals. These findings collectively establish DWEM as a robust, transparent, and clinically actionable tool for early detection of depression. The sensitivity-first configuration aligns with recommendations to minimize missed diagnoses in high-risk screening [27].

Our findings highlight that academic performance (CGPA) and derived interactions are core predictors of depression risk, aligning with the stress-vulnerability account and emphasizing the balance between chronic stress and psychological resilience [22]. This goes beyond raw features; psychologically informed features better capture psychopathological mechanisms. The unified pipeline—feature engineering, dynamic ensemble, and explainability—bridges the gap between high performance and clinical interpretability, directly mapping "features → risk level → intervention plan."

When contextualized within existing research, the results contribute several significant advancements. Previous studies using machine learning for depression prediction in students have often emphasized accuracy at the expense of interpretability [32], [33]. By embedding explainability directly into the analytic pipeline, this study addressed one of the main barriers hindering clinical adoption of artificial intelligence [34]. In particular, the SHAP-based feature analysis corroborates prior evidence linking academic performance and stress to depression [35], [36], while also extending knowledge by quantifying the relative impact of protective factors such as lifestyle health and satisfaction indices. The ability to map features to tiered intervention strategies builds upon earlier calls to move beyond prediction toward actionable support systems [37], [21].

Despite these strengths, several limitations should be acknowledged. First, the study relied on a single publicly available dataset, which, although large and well-structured, may not fully capture cultural, institutional, or longitudinal variations in student populations. Future studies should validate the DWEM across diverse contexts and real-world institutional data to strengthen generalizability. Second, although SHAP analysis provided transparency, the complexity of derived features may still present challenges for non-technical stakeholders, requiring the development of user-friendly visualization dashboards. Finally, the model predicts depression risk, but it does not equate to a clinical diagnosis. Its outputs should therefore be applied as decision-support tools alongside professional judgment, not as standalone diagnostic instruments [38]. Limitations include reliance on self-report data, limited external validation, and offline operation. Future work should integrate multimodal data (objective sleep monitoring, campus card transactions, anonymized textual feedback) and evaluate real-world effectiveness via prospective studies and cross-cultural validation [39]–[41].

The implications of this study are both methodological and practical. Methodologically, it demonstrates how feature engineering grounded in psychological theory [22] can enhance both predictive performance and interpretability. Practically, the construction of a three-tier intervention framework—ranging from crisis intervention for high-risk students to universal health promotion for low-risk groups—offers a scalable model for universities to optimize resource allocation. By reducing unnecessary in-depth screenings while still capturing sub-clinical distress, DWEM supports a paradigm shift from reactive to proactive mental health management, echoing global calls for early, scalable intervention in higher education [42], [43].

In comparing these results with the study's introduction, the research successfully addresses the identified gaps of limited explainability, insufficient feature integration, and the disconnection between prediction and intervention. While prior approaches often remain academic exercises, the present study demonstrates a clear translational pathway that connects computational modeling to actionable campus strategies. This alignment reinforces the study's originality and its contribution to bridging the divide between machine learning research and clinical or institutional application.

In conclusion, the DWEM offers a significant advancement in the prediction and management of depression among college students by combining methodological rigor with clinical interpretability. Although further validation is necessary, the findings highlight the potential of explainable ensemble learning to serve as a cornerstone of next-generation campus mental health systems. By enabling targeted, tiered, and resource-efficient interventions, this work underscores the broader value of artificial intelligence in shifting mental health care from passive response to active prevention.

5. Conclusion

The present study demonstrated that the proposed Dynamic Weighted Ensemble Model (DWEM), integrating multiple tree-based algorithms with optimized weighting and SHAP-based explainability, achieved clinical-grade accuracy and robustness in predicting depression among college students, thereby addressing critical gaps in both performance and interpretability. By combining advanced feature engineering with a transparent decision framework, the study not only validated the predictive utility of ensemble learning in mental health but also bridged the divide between computational modeling and practical intervention through the design of a tiered support system. While limitations remain—most notably the reliance on a single dataset and the need for broader cross-cultural validation—these should be viewed as opportunities for future research to enhance generalizability and refine user-centered implementation. Ultimately, this work highlights the potential of explainable AI to transform campus mental health management from reactive screening to proactive, data-driven prevention, offering a scalable pathway with significant implications for student well-being and institutional resource optimization.

Acknowledgment

The foundation of this study rests upon publicly available data. We would like to express our sincere gratitude to the collector and provider of the dataset; their exceptional work has laid a solid empirical foundation for this research. At the same time, we extend our heartfelt thanks to the reviewers for their insightful feedback and diligent efforts. Their valuable suggestions have significantly enhanced the quality of this study.

References

[1] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A Next-Generation Hyperparameter Optimization Framework," In Proc. ACM SIGKDD Conf. Knowl. Discov. Data Min., 2019, pp. 2623–2631. https://doi.org/10.1145/3292500.3330701.

[2] World Health Organization, "Depression and Other Common Mental Disorders: Global Health Estimates," WHO, Geneva, Switzerland, Tech. Rep., 2017.

[3] S.A. Lee, "How Much 'Thinking' About COVID-19 Is Clinically Dysfunctional?" Psychiatry Res., vol. 288, 2020. https://doi.org/10.1016/j.psychres.2020.113045.

[4] S. N. Chauhan and A. K. Vaidya, "Artificial intelligence in mental health: Challenges and opportunities," Int. J. Nurs. Educ. Res., vol. 12, no. 4, 2024. https://www.proquest.com/scholarly-journals/artificial-intelligence-mental-health-challenges/docview/3158589326/se-2.

[5] R. De La Fabián, Á. Jiménez-Molina, and F. Pizarro Obaid, "A critical analysis of digital phenotyping and the neuro-digital complex in psychiatry," Big Data Soc., vol. 10, no. 1, pp. 20539517221149097, Jan.-Jun. 2023, https://doi.org/10.1177/20539517221149097.

[6] Y. Feng, J. Li, T. Liu, Y. Wei, and N. Li, "Construction of a mental health risk model for college students with long and short-term memory networks and early warning indicators," J. Intell. Syst., vol. 33, no. 1, pp. 20230318, 2024, https://doi.org/10.1515/jisys-2023-0318.

[7] M. Chatterjee, P. Kumar, P. Samanta, and D. Sarkar, "Suicide ideation detection from online social media: A multi-modal feature based technique," Int. J. Inf. Manage. Data Insights, vol. 2, no. 2, p. 100103, 2022, https://doi.org/10.1016/j.jjimei.2022.100103.

[8] R. P. Auerbach, P. Mortier, R. Bruffaerts, et al., "WHO world mental health surveys international college student project: Prevalence and distribution of mental disorders," J. Abnorm. Psychol., vol. 127, no. 7, pp. 623–638, 2018, https://doi.org/10.1037/abn0000362.

[9] K. Kroenke, R.L. Spitzer, and J.B. Williams, "The PHQ-9: Validity of a Brief Depression Severity Measure," J. Gen. Intern. Med., vol. 16, no. 9, pp. 606–613, 2001. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

[10] D. Eisenberg, J. Hunt, N. Speer, and K. Zivin, "Mental health service utilization among college students in the United States," J. Nerv. Ment. Dis., vol. 199, no. 5, pp. 301–308, 2011, https://doi.org/10.1097/NMD.0b013e3182175123.

[11] S.K. Lipson, E.G. Lattie, and D. Eisenberg, "Increased Rates of Mental Health Service Utilization by US College Students: : 10-year population-level trends (2007–2017)," Psychiatr. Serv., vol. 70, no. 1, pp. 60–63, 2019. https://doi.org/10.1176/appi.ps.201800332.

[12] J. Linardon, A. Shatte, M. Messer, J. Firth, and M. Fuller-Tyszkiewicz, "E-mental health interventions for the treatment and prevention of eating disorders: An updated systematic review and meta-analysis," J. Consult. Clin. Psychol., vol. 88, no. 11, pp. 994–1007, 2020, https://doi.org/10.1037/ccp0000575.

[13] M.T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," In Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778.

[14] S.M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Adv. Neural Inf. Process. Syst., 2017, pp. 4768–4777.

[15] A. Sharma, R. Nigam, and S. S. Yadav, "Feature Engineering for Mental Health Prediction: A Systematic Review," Expert Syst. Appl., vol. 239, 2024, Art. no. 121789.

[16] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission," In Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2015, pp. 1721–1730.

[17] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324.

[18] D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) Program," AI Mag., vol. 40, no. 2, pp. 44–58, 2019. https://doi.org/10.1609/aimag.v40i2.2850.

[19] A. Holzinger, C. Biemann, C.S. Pattichis, and D.B. Kell, “What do we need to build explainable AI systems for the medical domain?,” *arXiv preprint arXiv:1712.09923*.

[20] H. M. Juntunen, "Test of the cognitive vulnerability-stress model in predicting hypomanic symptoms in college students," M.S. thesis, Univ. Northern Colorado, Greeley, CO, USA, 2021. https://scholarworks.uvm.edu/hcoltheses/413/.

[21] U.S. Jacob, N.A. Raji, J. Pillay, H.O. Adewuyi, and O.M. Olabode, "Mental health among secondary school students: Predictive factor analysis," Univ. J. Public Health, vol. 12, no. 1, pp. 28–36, 2024, https://doi.org/10.13189/ujph.2024.120104.

[22] S. M. Monroe and A. D. Simons, "Diathesis-Stress Theories in the Context of Life Stress Research," Psychol. Bull., vol. 110, no. 3, pp. 406–425, 1991. https://doi.org/10.1037/0033-2909.110.3.406.

[23] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[24] A. Fernández, S. García, F. Herrera, and N. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges," Prog. Artif. Intell., vol. 8, no. 4, pp. 1–15, 2019.

[25] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002. https://doi.org/10.1613/jair.953.

[26] M. Šimundić, "Measures of Diagnostic Accuracy: Basic Definitions," Biochem. Med., vol. 18, no. 1, pp. 1–10, 2008. https://doi.org/10.11613/BM.2008.001.

[27] Z.E. Imel, B. Pace, B. Pendergraft, et al., “Machine Learning–Based Evaluation of Suicide Risk Assessment in Crisis Counseling Calls,” Psychiatr Serv., vol. 75, no. 11, pp. 1068–1074, Jul 2024. https://doi.org/10.1176/appi.ps.20230648.

[28] S. Harith, I. Backhaus, N. Mohbin, H. T. Ngo, and S. Khoo, "Effectiveness of digital mental health interventions for university students: An umbrella review," PeerJ, vol. 10, p. e13111, 2022, https://doi.org/10.7717/peerj.13111.

[29] C. Hollis, R. Morriss, J. Martin, et al., "Technological innovations in mental healthcare: harnessing the digital revolution," *The British Journal of Psychiatry*, vol. *206, no.* 4, pp. 263-265. https://doi.org/10.1192/bjp.bp.113.142612.

[30] E. Tjoa and C. Guan, "A Survey on Explainable artificial intelligence (xai): Towards Medical XAI," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 11, pp. 4793–4813, 2021. https://doi.org/10.1109/TNNLS.2020.3027314.

[31] R. Miotto, F. Wang, S. Wang, X. Jiang, and J.T. Dudley, "Deep Learning for Healthcare: Review, Opportunities and Challenges," Brief. in Bioinform., vol. 19, no. 6, pp. 1236–1246, 2018. https://doi.org/10.1093/bib/bbx044.

[32] L. Luo, J. Yuan, C. Wu, et al., "Predictors of depression among Chinese college students: A machine learning approach," BMC Pub. Heal., vol. 25, no. 1, p. 470, 2025, https://doi.org/10.1186/s12889-025-21632-8.

[33] Z. Zhang and A. Yu, "Design and development of ensemble deep learning framework for psychological health assessment of college students," In Proc. 2025 Int. Conf. Intell. Comput. Knowl. Extract. (ICICKE), Jun. 2025, pp. 1–6, https://doi.org/10.1109/ICICKE65317.2025.11136466.

[34] J. Amann, A. Blasimme, E. Vayena, D. Frey, V.I. Madai, and Precise4Q Consortium, "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective," BMC Med. Inform. Decis. Mak., vol. 20, no. 1, 2020, Art. no. 310. https://doi.org/10.1186/s12911-020-01332-6.

[35] R. P. Auerbach, J. Alonso, W.G. Axinn, et al., "Mental disorders among college students in the World Health Organization world mental health surveys," Psychol. Med., vol. 46, no. 14, 2016, pp. 2955–2970, 2016, https://doi.org/10.1017/S0033291716001665.

[36] C. Zhang, L. Shi, T. Tian, et al., "Associations between academic stress and depressive symptoms mediated by anxiety symptoms and hopelessness among Chinese college students," Psychol. Res. Behav. Manag., vol. 15, pp. 547–556, 2022, https://doi.org/10.2147/PRBM.S353778.

[37] A.M. Chekroud, J. Bondar, J. Delgadillo, et al., "The Promise of Machine Learning in Predicting Treatment Outcomes in Psychiatry," World Psychiatry, vol. 20, no. 2, pp. 154–170, 2021. https://doi.org/10.1002/wps.20882.

[38] B. Levis, A. Benedetti, and B. D. Thombs, "Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis," BMJ, vol. 365, p. l1476, 2019. https://doi.org/10.1136/bmj.l1476.

[39] A. Ollie, S. Bianchi, N. Manos, J. Carter, J. Houle and L. Sullivan, "Mental health promotion and prevention interventions for college student athletes: A scoping review," J. Am. Coll. Health, pp. 1–15, 2025. https://doi.org/10.1080/07448481.2025.2512964.

[40] M. Kerasiotis, L. Ilias, and D. Askounis, "Depression detection in social media posts using transformer-based models and auxiliary features," Soc. Netw. Anal. Min., vol. 14, no. 1, p. 196, 2024. https://doi.org/10.1007/s13278-024-01360-4.

[41] I. Ahmed, A. Brahmacharimayum, R.H. Ali, T.A. Khan and M.O. Ahmad, "Explainable AI for Depression Detection and Severity Classification from Activity Data," JMIR Ment. Health, vol. 12, no. 1, e72038, 2025. https://doi.org/10.2196/72038.

[42] World Health Organization, Mental Health Atlas 2024. Geneva: World Health Organization, 2021.

[43] World Health Organization, Comprehensive Mental Health Action Plan 2023–2030. In *Comprehensive mental health action plan 2013–2030*. Geneva: World Health Organization, 2023.