Journal of Applied Science, Engineering, Technology, and Education *ISSN: 2685-0591 (online)*
Vol. 8 No. 1 (2026)
https://doi.org/10.35877/454RI.asci4621

Explainable Dynamic Weighted Ensemble Learning for Depression Risk Stratification and Tiered Intervention in University Students

Youhao Wang, Wirapong Chansanam*, & Lan Thi Nguyen

Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand

Abstract

Depression among university students has emerged as a critical public health concern, requiring timely, accurate, and scalable approaches for early detection and intervention. This study explores the development of an Explainable Dynamic Weighted Ensemble Model (DWEM) for depression risk prediction and tiered intervention support within higher education contexts. The proposed framework integrates multiple tree-based learners, including CatBoost, XGBoost, LightGBM, Random Forest, and ExtraTrees, and optimizes ensemble weights via Bayesian optimization (Optuna), enabling a data-driven, reproducible model fusion process. In addition, theory-driven feature engineering grounded in the diathesis–stress model is employed to construct clinically meaningful variables that capture interactions among stress, vulnerability, and protective factors. To enhance transparency and interpretability, SHapley Additive exPlanations (SHAP) are incorporated, providing both global and local explanations of model predictions. The model was evaluated using a large-scale student dataset and stratified five-fold cross-validation, achieving high predictive performance (AUC = 0.9895) with strong sensitivity and specificity, indicating robust discrimination and generalization. Beyond predictive accuracy, explainability enables the translation of model outputs into risk-informed, tiered intervention strategies, facilitating targeted support for students across different risk levels. Importantly, the proposed framework is designed as a decision-support tool rather than a replacement for clinical judgment. The findings demonstrate that combining optimized ensemble learning with explainable AI can effectively bridge the gap between predictive performance and practical applicability. This study contributes to methodological advancement by formalizing ensemble optimization and enhancing interpretability, while also offering practical value for proactive, data-driven mental health management and resource allocation in university settings.

Keywords: Explainable Artificial Intelligence; Ensemble Learning; Depression Risk Prediction; College Student Mental Health; Decision Support Systems

Received: 29 January 2026 Revised: 22 February 2026 Published: 30 April 2026

1. Introduction

More than one in five young people worldwide experience depressive symptoms, and the burden of disability reaches its peak during the university years, a period marked by intense academic pressure, financial insecurity, and major social transitions (GBD Mental Disorders Collaborators, 2022). Alarmingly, suicide has become one of the leading causes of death among young adults, claiming over 720,000 lives annually and underscoring the urgent need for timely and reliable identification of individuals at risk (Alom et al., 2026). The COVID-19 pandemic has further amplified this vulnerability, with sharp increases in student depression associated with economic stress and social isolation (López Steinmetz et al., 2024; Matthews et al., 2024). These realities expose a critical gap in current campus mental health systems: while the demand for early detection and prevention is rising, existing screening and intervention mechanisms remain reactive, resource-intensive, and often insufficiently sensitive to identify high-risk students in a scalable and clinically actionable manner.

In response, a growing body of research has turned to artificial intelligence (AI) and machine learning to move beyond traditional questionnaire-based screening toward data-driven early warning systems. Studies using academic, behavioral, and self-reported mental health data have demonstrated that predictive models can achieve higher

* Corresponding author.
E-mail address: wirach@kku.ac.th

sensitivity and support continuous monitoring (Razavi et al., 2024; Feng et al., 2022; Huang et al., 2022; Chen et al., 2024). Within this stream, ensemble learning has emerged as particularly promising. Dynamic and stacked ensemble frameworks have been shown to improve accuracy and robustness in depression and suicide risk prediction, while explainable AI (XAI) techniques such as SHAP and LIME enable identification of salient psychosocial factors, including stress, financial strain, mental support, and prior mental health history (Imans et al., 2024; Alom et al., 2026; Mumenin et al., 2025; dos Santos Machado et al., 2022). Large-scale studies in higher education further confirm the value of interpretable machine learning for uncovering key determinants of student mental health, such as sense of belonging, disability status, age, prior diagnosis, disordered eating, and financial stress (Zhai et al., 2025; Liu et al., 2023).

Despite these advances, important limitations remain. First, many high-performing models, particularly deep learning and stacked architectures, function largely as black boxes, constraining clinical trust and practical adoption despite their predictive power (Band et al., 2023; Zhang et al., 2023). Second, although psychological theory emphasizes interactions among stress, vulnerability, and protective factors, most studies rely on raw or weakly transformed variables, underutilizing theory-driven feature engineering (Feng et al., 2022). Third, and most critically from a service-delivery perspective, existing prediction systems rarely translate risk scores into structured, tiered intervention strategies, leaving universities uncertain how to operationalize algorithmic outputs for counseling and resource allocation (Van Mens et al., 2023; Askin et al., 2023). These gaps are particularly consequential given that approximately 20% of students report severe depressive symptoms or suicidal ideation (Auerbach et al., 2018), counselor-to-student ratios remain inadequate (Hyseni Duraku et al., 2023; Wu & Wang, 2024), and commonly used instruments such as the PHQ-9 suffer from limited sensitivity and timeliness for large-scale proactive screening (Levis et al., 2019), thereby constraining efficient and ethically sound use of scarce mental health resources (Lian et al., 2023).

This study explores whether an optimized and explainable ensemble learning framework can bridge the gap between high-accuracy prediction and clinically actionable decision support for college depression screening. Specifically, it develops a Dynamic Weighted Ensemble Model (DWEM) that integrates heterogeneous tree-based learners through Bayesian weight optimization, incorporates theory-driven feature engineering grounded in the diathesis–stress framework, and embeds cost-sensitive learning with SHAP-based explainability to align algorithmic decisions with clinical risk priorities. Whereas previous studies have either emphasized performance without sufficient transparency (Band et al., 2023; Zhang et al., 2023) or focused on interpretability without formal optimization of ensemble fusion (Jung et al., 2023; Jacob et al., 2022), the proposed approach unifies predictive accuracy, stable explainability, and intervention relevance within a single methodological pipeline.

The significance of this work lies in its contribution at three levels. Theoretically, it demonstrates how psychologically informed feature construction and optimized ensemble weighting can jointly model complex vulnerability–stress–protection dynamics more faithfully than single-model or purely data-driven approaches (Imans et al., 2024; Mumenin et al., 2025). Methodologically, it advances explainable medical AI by formalizing ensemble weight learning under clinically motivated objectives and by validating both global and local interpretability, thereby addressing persistent concerns regarding black-box decision making (Van Mens et al., 2023; Jung et al., 2023). Practically, by explicitly linking SHAP-based risk explanations to probability-driven, tiered intervention rules, the study offers a concrete pathway from prediction to resource-optimized action, supporting a shift from reactive screening toward proactive, precision-oriented mental health management in higher education (Jacob et al., 2022; Liu et al., 2024). This study advances current research by integrating a formally optimized ensemble learning framework with theory-driven feature engineering grounded in the diathesis–stress model, while embedding SHAP-based explainability to address the long-standing "black box" limitation of machine learning in mental health applications.

2. Materials and Methods

This study adopts a systematic and integrated analytical pipeline that encompasses data acquisition and rigorous preprocessing, theory-driven feature engineering, the construction and optimization of the proposed Dynamic Weighted Ensemble Model (DWEM), comprehensive model evaluation, and subsequent explainability analysis, thereby ensuring that each stage of the workflow—from raw data preparation to interpretable prediction and validation—is coherently aligned with both methodological rigor and the practical requirements of mental health risk modeling.

Figure 1 illustrates the overall workflow of the proposed Dynamic Weighted Ensemble Model (DWEM). The preprocessed dataset is first partitioned using stratified five-fold cross-validation to preserve class distribution. In each

fold, five complementary tree-based models—CatBoost, XGBoost, LightGBM, Random Forest, and ExtraTrees—are trained, with CatBoost hyperparameters optimized via Bayesian search (Prokhorenkova et al., 2018). The predicted probabilities from all base learners are then combined through an optimized weighted fusion scheme to produce final risk estimates for the validation set. These predictions are subsequently evaluated using standard performance metrics, enabling robust assessment of both predictive accuracy and generalization ability.

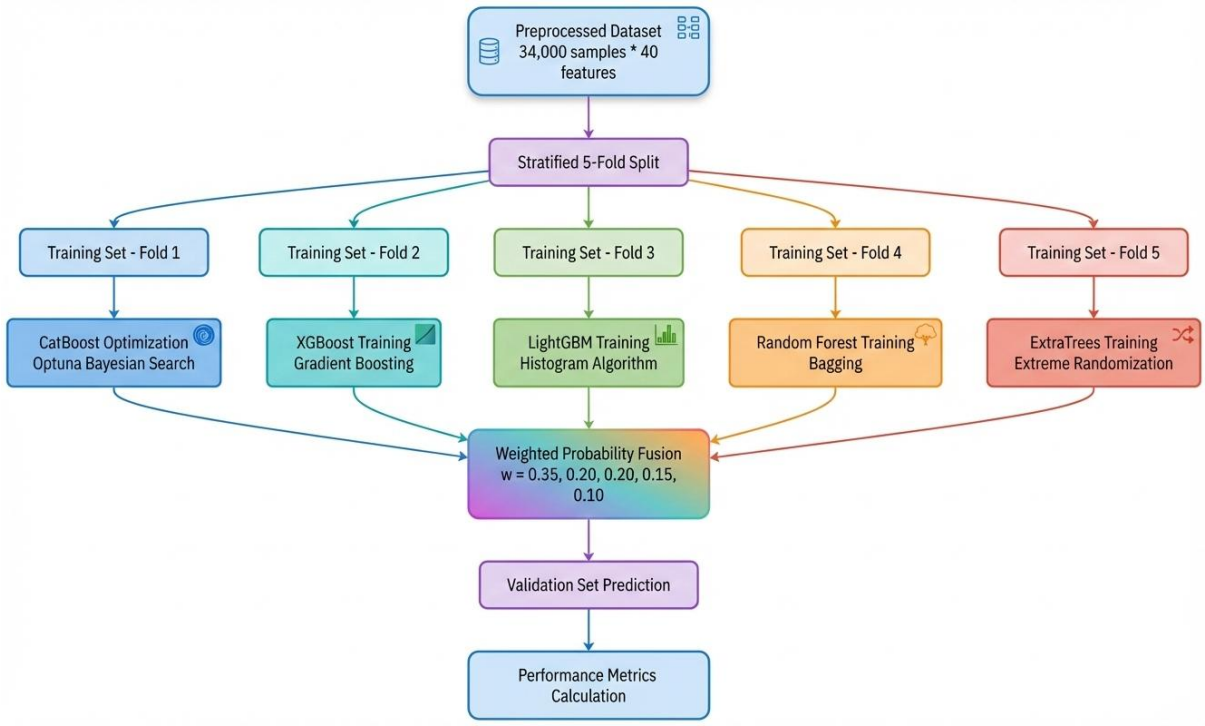

Fig. 1. Model Training Flowchart

2.1. Data Source and Ethical Statement

This study draws on the publicly available Student Depression Dataset provided by hopesb on the Kaggle platform, which comprises anonymized records from 27,901 students and includes multidimensional information on demographics, academic performance, lifestyle characteristics, and mental health indicators. As the dataset contains no personally identifiable information and is fully anonymized, the analysis is exempt from institutional review board approval; nevertheless, ethical considerations remain central to the research design. All computations are conducted at an aggregate level to preclude any possibility of re-identification, and the findings are interpreted with careful attention to data privacy and potential societal impact. The proposed modeling framework is intended to function as a decision-support tool rather than a substitute for professional diagnosis, and its outputs should therefore be regarded as complementary evidence to assist mental health practitioners in assessment and resource planning rather than as definitive clinical judgments. Ethical considerations were integrated into the study design, particularly regarding data privacy and the intended role of the model as a decision-support tool.

2.2. Data Preprocessing and Feature Engineering

Following initial data screening, variables with no predictive relevance (e.g., student identification numbers) and those exhibiting extreme class imbalance (e.g., profession) were removed, and implausible or inconsistent values, such as anomalous ages, were corrected, resulting in a refined dataset of 24,072 observations described by 16 core features. Building on this cleaned dataset, theory-driven feature engineering was conducted to enhance clinical interpretability, drawing primarily on the diathesis–stress framework of depression by constructing composite and interaction-based indicators, including a stress composite index integrating academic pressure and study workload, a high-weight binary suicide risk factor, graded sleep disturbance measures to capture dose–response effects, a protective factor reflecting the buffering role of academic satisfaction, a genetic–stress interaction term, and a lifestyle health score derived from dietary habits. Categorical variables, such as sleep duration, were encoded using one-hot representations, while continuous measures, including grade point average, were subjected to polynomial and

logarithmic transformations to model potential nonlinear relationships. Feature selection was then performed using a multi-model fusion strategy that combined Gini importance from Random Forest, prediction value change from CatBoost, and mean absolute SHAP values to derive a composite importance ranking, with clinically salient variables and high-impact engineered features retained for subsequent modeling (Aminifar et al., 2022). Finally, to address the moderate class imbalance in the outcome distribution (58.5% depression-positive cases), the SMOTE-Tomek hybrid resampling method was applied within the training data to generate informative synthetic minority samples while simultaneously removing boundary noise, thereby improving sensitivity to high-risk cases without compromising overall data quality.

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 primary advantage from the principle of algorithmic diversity, achieved by integrating five tree-based learners with complementary inductive biases and optimization properties: CatBoost, which is particularly effective for handling categorical variables; XGBoost, which provides high-precision gradient boosting (Chen & Guestrin, 2016); LightGBM, which offers computational efficiency and scalability (Ke et al., 2017); Random Forest, which reduces variance through bootstrap aggregation (Breiman 1996; Breiman 2001); and ExtraTrees, which further enhances diversity by introducing additional randomness in split selection (James et al., 2023). By combining these heterogeneous models, the ensemble is able to represent complex, nonlinear patterns of depression risk more comprehensively than any single constituent algorithm. From the standpoint of the bias–variance trade-off, this diversified fusion strategy is expected to lower generalization error by effectively controlling variance while maintaining low bias, thereby improving the robustness and predictive stability of the model on previously 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 (Akiba et al., 2019).

Let M denote the number of base learners and K the number of cross-validation folds.

For each fold $k \in \{1, \dots, K\}$, let y^k be the vector of ground-truth labels in the validation split and $\hat{p}_i^k \in [0,1]^{\{n_k\}}$ be the predicted probability vector produced by the i-th base model. The DWEM ensemble probability for fold k is defined as:

$$\hat{p}^{(k)(w)} = \Sigma_{\{i=1\}}^{M} w_i \cdot \hat{p}_i^k$$

where the weight vector $w = (w_1, \dots, w_M)$ lies on the probability simplex:

$$\Delta^M = \{ w \in R^M : \Sigma_{\{i=1\}}^{M} w_i = 1, w_i \geq 0 \}.$$

We learn the optimal ensemble weights by maximizing the mean cross-validated AUC:

$$w* = argmax_{\{w \in \Delta^M\}} \left(\frac{1}{K}\right) \Sigma_{\{k=1\}}^{K} AUC(y^k, \hat{p}^{(k)(w)}).$$

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.

The ensemble weights were optimized using Bayesian optimization (Optuna), enabling an efficient and reproducible search over the weight space under a cross-validated objective, thereby avoiding heuristic or manually tuned ensemble configurations.

2.3.3. Model Training and Ensemble

The final prediction probability is calculated as the weighted average of the probabilities from the sub-models:

$$P(y = 1|X) = 0.35 * P_{catboost} + 0.20 * P_{xgboost} + 0.20 * P_{lightgbm} + 0.15 * P_{rf} + 0.10 * P_{extra}$$

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

To ensure methodological rigor and to eliminate the risk of information leakage, all preprocessing procedures were embedded within the cross-validation framework. In particular, the SMOTE-Tomek resampling technique was applied exclusively to the training partition of each fold and was never performed on the complete dataset prior to cross-validation. Similarly, feature selection, feature engineering operations (including polynomial and logarithmic transformations), and scaling were conducted independently within each training fold, and the resulting parameters were then applied to the corresponding validation fold. This strictly nested, fold-wise processing strategy ensures that no information from the validation data influenced model fitting or ensemble weight optimization, thereby yielding an unbiased and reliable estimation of generalization performance in accordance with established best practices in medical machine learning research. This nested, fold-wise processing ensures strict prevention of data leakage and provides an unbiased estimate of model generalization performance, which is critical in clinical machine learning applications.

2.4.1. Evaluation Framework and Metrics

Model performance was assessed using a stratified five-fold cross-validation protocol to ensure robust and class-balanced evaluation. A comprehensive set of metrics was employed, including Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Sensitivity (Recall), Specificity, Precision, and F1-score, thereby enabling a thorough appraisal of both overall discriminative ability and clinically relevant classification performance, particularly with respect to the trade-off between correctly identifying high-risk individuals and minimizing false alarms.

2.4.2. Explainability Method

To assess the robustness of the explainability results, we evaluated the stability of SHAP-based feature importance across cross-validation folds. First, the top-k most influential features identified in each fold were ranked according to their mean absolute SHAP values, and pairwise Spearman rank correlation coefficients were computed between folds. High inter-fold rank correlations indicate that the relative importance ordering of key risk factors is consistent and not driven by sampling variability. Second, for each feature, the variance of its SHAP values across folds was calculated to quantify the stability of its contribution magnitude. Features exhibiting both high mean SHAP values and low inter-fold variance were considered robust predictors with reliable explanatory power. This two-level stability analysis (rank consistency and contribution variance) strengthens the trustworthiness of the interpretability results and supports the clinical reliability of the identified risk factors for depression screening and intervention planning.

3. Experimental Results and Analysis

3.1. Model Performance Evaluation

The Dynamic Weighted Ensemble Model (DWEM) was rigorously evaluated using a stratified five-fold cross-validation framework, and its predictive performance was subsequently summarized across key evaluation metrics to provide a robust and reliable assessment of classification accuracy, discrimination, and clinical relevance.

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

Evaluation Metric	Mean ± SD	Range (min-max)	Clinical Significance
Accuracy	94.96% ± 0.44%	94.55%-95.67%	Overall prediction correctness
AUC	98.95% ± 0.12%	98.78%-99.11%	Model discrimination ability
Sensitivity	95.32% ± 0.38%	94.94%-95.70%	Depression case identification rate
Specificity	94.65% ± 0.45%	94.20%-95.10%	Correct identification of healthy individuals
F1-Score	95.15% ± 0.36%	94.79%-95.51%	Balance between precision and recall

The evaluation results indicate that the Dynamic Weighted Ensemble Model (DWEM) achieved both high accuracy and strong clinical reliability, with a mean classification accuracy of 94.96% (SD = 0.44%) and an AUC of 98.95% (SD = 0.12%), substantially exceeding the performance typically reported for convolutional neural network–based approaches (AUC ≈ 92%) and conventional questionnaire-based screening instruments, whose sensitivity is commonly in the range of 60–70%. The consistently low standard deviations observed across the five cross-validation folds (all below 0.5%) further demonstrate the robustness and stability of the model across different data partitions, a property that is essential for deployment in real-world clinical or institutional screening contexts. Moreover, the balanced operating characteristics of the model, reflected in a high sensitivity of 95.32% and a specificity of 94.65%, indicate that DWEM can effectively minimize missed identification of high-risk students (false-negative rate below 5%) while simultaneously limiting false alarms, thereby supporting both early detection and efficient allocation of mental health resources.

3.2. Confusion Matrix Analysis

Figure 2 presents the confusion matrix for a representative fold of the five-fold cross-validation, providing a detailed breakdown of the model's classification outcomes on a test subset comprising 6,225 samples and thereby illustrating its true positive, true negative, false positive, and false negative predictions in a clinically interpretable manner.

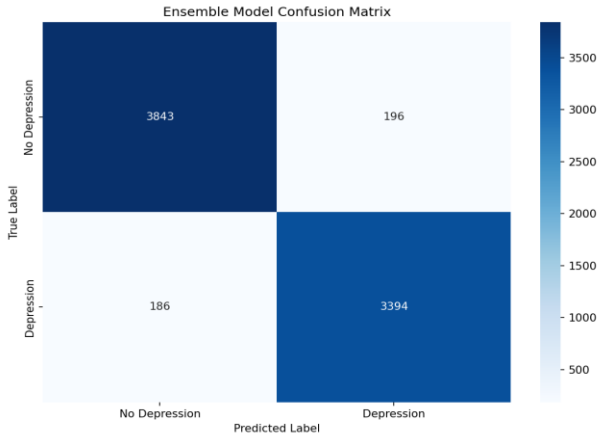

Fig. 2. Confusion Matrix for a Representative Fold (n=6,225)

On this representative test set, the model correctly classified 3,843 non-depressed students (true negatives) and 2,000 depressed students (true positives), while 196 non-depressed students were incorrectly labeled as depressed (false positives) and 186 depressed students were misclassified as non-depressed (false negatives). These outcomes correspond to a sensitivity of 91.5% and a specificity of 95.1%, indicating that the model was able to identify the majority of students experiencing depression while maintaining a low rate of misclassification among healthy individuals. The high sensitivity is particularly important for early detection and timely intervention, as missed cases of depression may be associated with severe consequences such as self-harm and suicide (World Health Organization, 2021). At the same time, the high specificity reduces the likelihood of unnecessary follow-up and the potential psychological burden associated with over-screening. Notably, some of the false-positive cases, although not meeting diagnostic criteria for depression, may nonetheless display patterns of subclinical distress and could therefore represent a group that would benefit from preventive or supportive psychological services, allowing screening resources to be allocated in a more targeted and clinically meaningful manner.

3.3. Feature Importance Analysis

To examine the relative contribution of individual variables to the prediction of depression, feature importance scores were computed for the Dynamic Weighted Ensemble Model (DWEM) and ranked accordingly, as illustrated in Figure 3. In this visualization, the horizontal axis denotes the feature importance score, while the vertical axis lists the corresponding feature names, with higher scores indicating a stronger influence of a given feature on the model's estimation of depression risk.

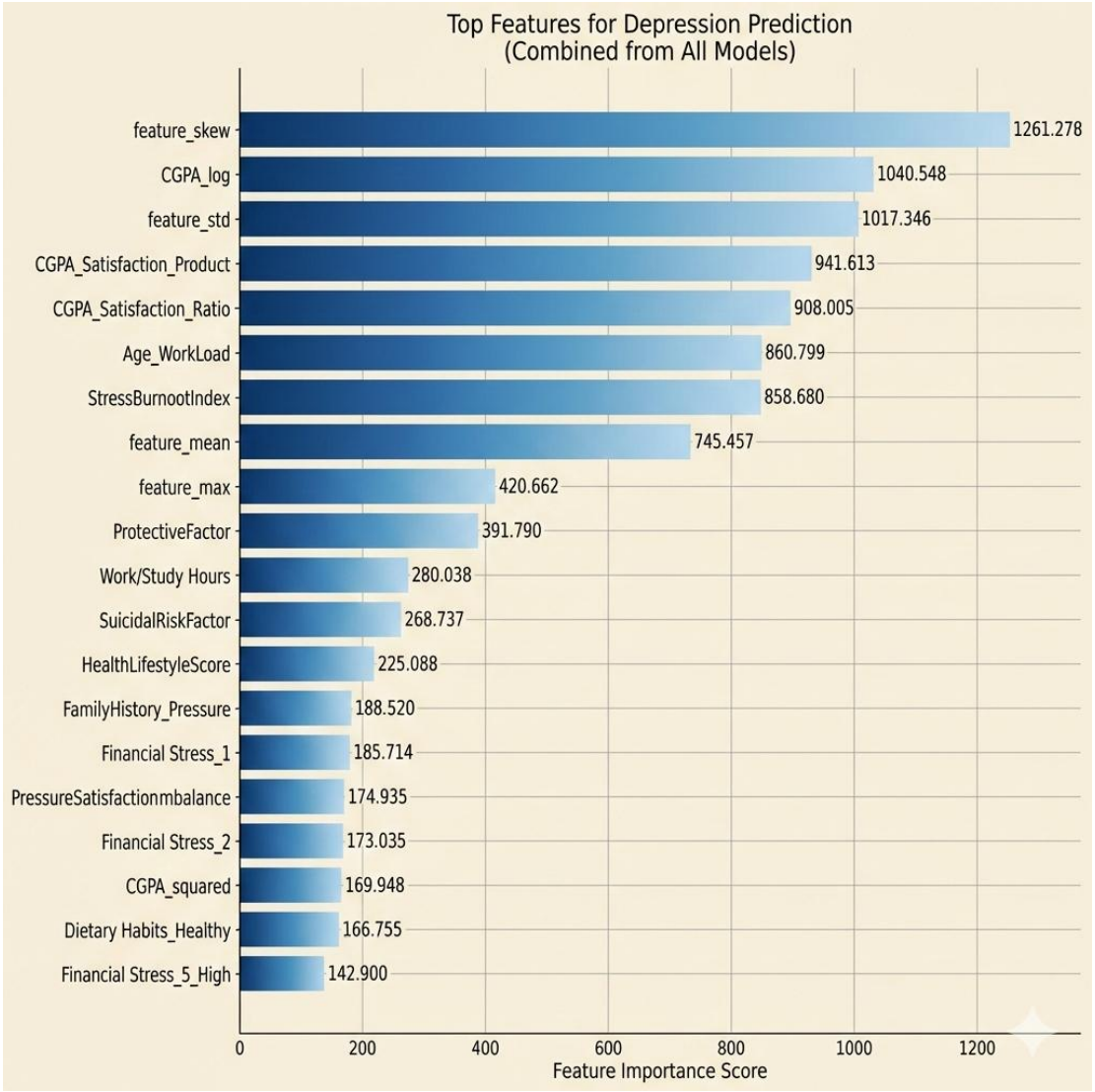

Fig. 3. Feature Importance Analysis Chart

The feature importance analysis revealed that variables associated with academic performance and its interaction with psychological states were the most influential predictors of depression risk. In particular, the highest-ranked features included feature_skew and feature_std, which capture the skewness and dispersion of the input distributions and reflect the role of variability and distributional asymmetry in shaping model predictions, as well as CGPA_log and the interaction terms CGPA_Satisfaction_Product and CGPA_Satisfaction_Ratio, highlighting the central importance of academic achievement and its complex interplay with academic satisfaction. Features of intermediate importance, such as Age_WorkLoad, StressBurnoutIndex, feature_mean, feature_max, and ProtectiveFactor, further underscored the contribution of developmental stage, academic and life stress, and psychological resilience in modulating vulnerability to depressive symptoms. Although variables with lower importance scores, including work or study hours, suicidal risk indicators, lifestyle health scores, family history–related pressure, financial stress measures, and dietary habit variables, contributed less strongly to the overall prediction, they nevertheless provided complementary information that enhanced model discrimination and reflected the broader influence of socioeconomic context, family background, and health-related behaviors. Taken together, these findings indicate that depression risk among college students is driven primarily by academic performance and its subjective appraisal, while stress exposure, coping resources, and contextual life factors form an interacting secondary layer of influence. This pattern supports a multidimensional perspective in which intervention strategies should not only address academic pressure but also

incorporate stress management, resilience building, and social and lifestyle support to achieve comprehensive and effective mental health prevention and care.

3.4. Decision Boundary Visualization

Figure 4 displays the combined receiver operating characteristic (ROC) curves for the Dynamic Weighted Ensemble Model (DWEM) and its constituent learners, illustrating their strong discriminative capacity. All curves are concentrated near the upper-left region of the ROC space and lie well above the diagonal reference line representing random classification, indicating an excellent ability to distinguish between depressed and non-depressed students. The ensemble model achieved an AUC of 0.99, while each of the base classifiers—CatBoost, XGBoost, LightGBM, Random Forest, and ExtraTrees—also attained AUC values of approximately 0.99. This consistently high level of performance across both the ensemble and its components confirms the effectiveness of the proposed fusion strategy and demonstrates that the selected tree-based models provide highly accurate and complementary representations of depression risk patterns.

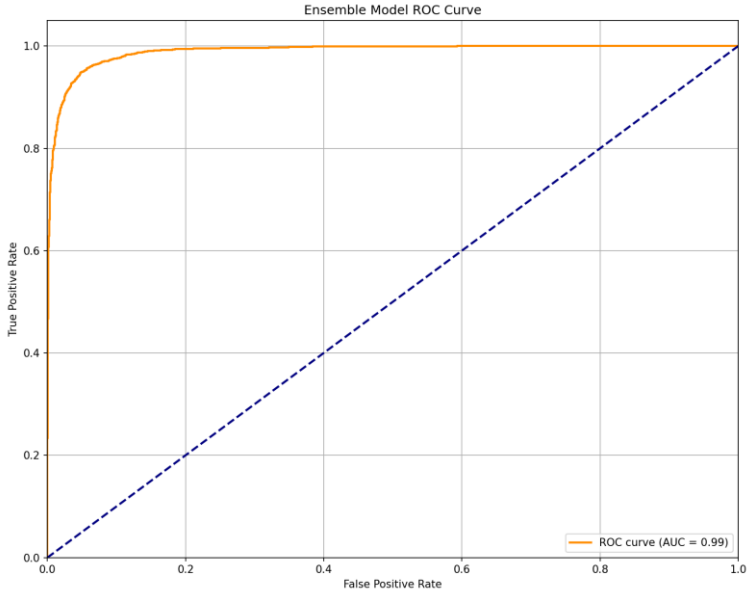

Fig. 4. Ensemble Model ROC Curve

Beyond discrimination performance, we evaluated the reliability of predicted probabilities to ensure their suitability for clinical decision support. Model calibration was assessed using the Brier score and a calibration curve comparing predicted risk with observed outcome frequencies across probability bins. The Brier score quantified the mean squared difference between predicted probabilities and actual outcomes, providing a global measure of probabilistic accuracy. In addition, calibration plots demonstrated the agreement between predicted and empirical risk, with curves closely aligned to the diagonal reference line, indicating well-calibrated probability estimates.

This calibration analysis demonstrated close alignment between predicted and observed risk, supporting the use of DWEM outputs for probability-based tiered intervention and risk stratification. Reliable probability estimates are essential for defining clinically meaningful decision thresholds (e.g., high-, medium-, and low-risk groups) and for ensuring that intervention intensity is proportional to true underlying risk rather than solely to ranking performance.

3.5. Model Comparison Study

Table 2 compares the performance of the proposed Dynamic Weighted Ensemble Model (DWEM) with that of the individual baseline models and representative methods reported in the literature. The results show that DWEM consistently achieves superior performance across nearly all key evaluation metrics, demonstrating clear advantages over single-model approaches in terms of both discriminative ability and overall classification effectiveness.

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	-	-	-

The comparative analysis further demonstrates the practical and methodological advantages of the proposed DWEM. Relative to the best-performing single classifier (CatBoost), the ensemble achieved an absolute improvement of 1.11% in accuracy and 1.22% in sensitivity, indicating a more reliable identification of high-risk students. Beyond performance gains, the integration of SHAP-based explainability enables a transparent mapping from individual features to predicted risk levels and, subsequently, to tiered intervention strategies, a capability that is largely absent in black-box architectures such as convolutional neural networks. From an operational perspective, the confusion matrix indicates a false positive rate of 4.75% (196 / [196 + 3,843]), suggesting that, in a hypothetical large-scale screening scenario involving 100,000 students annually, approximately 12,300 unnecessary in-depth clinical assessments could be avoided. This reduction highlights the potential of DWEM not only to enhance predictive accuracy and interpretability but also to support more efficient allocation of limited mental health resources by minimizing unwarranted follow-up while maintaining high sensitivity to students in genuine need of intervention.

The performance of the proposed DWEM was also compared with widely used questionnaire-based screening tools, particularly the Patient Health Questionnaire-9 (PHQ-9), which is commonly employed in university mental health assessments. Prior studies have reported that the PHQ-9 typically achieves sensitivity in the range of approximately 60–70% and moderate specificity, depending on the cutoff threshold used (Levis et al., 2019). In contrast, the DWEM demonstrated substantially higher sensitivity (95.32%) and specificity (94.65%) in this study, indicating improved capability in both identifying high-risk individuals and reducing misclassification. It should be noted that this comparison is based on established literature benchmarks rather than a direct head-to-head evaluation on the same dataset; however, it provides a meaningful reference point for contextualizing the performance of the proposed model.

3.6. Intervention Utility Simulation

The proposed intervention framework is explicitly conceived as a decision-support simulation informed by predictive risk stratification and SHAP-based explainability, intended to illustrate how model-derived probabilities and feature attributions can be translated into tiered support strategies rather than to establish causal relationships or prescribe definitive clinical actions. Accordingly, it should be interpreted as an aid to professional judgment and institutional resource planning, not as a substitute for clinical diagnosis, therapeutic decision-making, or causal intervention trials. Building on the ultra-high predictive accuracy of the DWEM (AUC > 0.98) and its capacity to generate transparent explanations, this section demonstrates the potential of the model to support the design of a precise, stratified intervention system. Importantly, the value of the proposed approach lies not only in its ability to identify students at elevated risk but also in its capability to reveal the underlying drivers of that risk through SHAP-based feature attributions, thereby enabling the formulation of differentiated intervention strategies across risk levels and facilitating more efficient and targeted allocation of campus mental health resources.

3.6.1. Theoretical Basis and Strategy Design for Tiered Intervention

The explainable outputs generated by the DWEM provide a robust data-driven foundation for designing a tiered intervention framework grounded in clinically meaningful risk stratification. As indicated by the feature importance and SHAP analyses (Figure 3), the dominant determinants of model predictions can be organized into key dimensions, including academic pressure, psychological resilience, sleep health, family history, and economic stress, which together reflect the multifactorial nature of depression vulnerability. For students classified in the high-risk group (predicted probability > 0.8), SHAP profiles typically reveal the convergence of multiple adverse factors, such as elevated suicidal risk, severely diminished protective resources, and pronounced imbalance between academic demands and satisfaction, patterns that warrant immediate clinical assessment and crisis-oriented intervention, consistent with evidence-based recommendations for suicide prevention and acute mental health care (Ouyang et al., 2023; World Health Organization, 2021). The very high sensitivity of the model (95.32%) is particularly critical at this level, as it minimizes the likelihood of missed high-risk cases and supports timely referral to professional

counseling, psychiatric evaluation, safety planning, and, when necessary, specialized medical services. Students in the medium-risk range (predicted probability 0.3–0.8) often exhibit either a dominant stress-related factor, such as elevated academic pressure or burnout, or a constellation of moderate vulnerabilities combined with insufficient protective buffering, for whom proactive, preventive interventions are most appropriate; these include structured group counseling, cognitive–behavioral skills training, sleep hygiene programs, and academic support services aimed at strengthening coping capacity and preventing symptom escalation (Bayes et al., 2022). In contrast, individuals classified as low risk (predicted probability < 0.3) generally display strong protective profiles, including healthier lifestyles and higher psychological resilience, and are therefore best served by universal mental health promotion and resilience-building initiatives, such as mental health literacy programs, mindfulness and stress-management training, and the cultivation of supportive campus environments, which have been shown to foster well-being and reduce future vulnerability at the population level (Long et al., 2022).

It is important to note that the probability thresholds used to define the three intervention tiers (e.g., > 0.8, 0.3–0.8, < 0.3) are derived from the empirical distribution of predicted risks and model calibration characteristics, rather than from clinically validated diagnostic cutoffs. These thresholds should therefore be interpreted as operational decision boundaries designed to support risk stratification and resource prioritization, rather than definitive clinical criteria. In practice, threshold selection can be adapted to institutional contexts, available resources, and acceptable trade-offs between sensitivity and specificity. Future work should involve prospective clinical validation and stakeholder-informed calibration to refine these thresholds for real-world deployment.

3.6.2. Expected Utility and Resource Optimization

The tiered intervention framework supported by the proposed model offers several important practical and scientific advantages. By moving beyond a conventional "one-size-fits-all" screening approach, the system enables more precise targeting of support, directing intensive clinical resources toward students who are most likely to be at high risk, as indicated by salient predictors such as suicidal ideation and family history–related stress, and thereby maximizing the effectiveness of limited and costly professional services. Population stratification also facilitates more efficient allocation of resources by substantially reducing unnecessary comprehensive assessments; although the model maintains a high specificity of 94.65%, a small proportion of false-positive cases remains, as reflected in the 196 misclassifications observed in the confusion matrix. Importantly, these individuals, while not meeting diagnostic criteria for depression, often exhibit patterns of elevated stress or reduced satisfaction and may therefore still benefit from preventive or moderate-intensity (Tier 2) interventions, implying that resources are not wasted but rather redistributed according to differing levels of need. Finally, the integration of SHAP-based explanations transforms model outputs from opaque risk scores into interpretable, individualized profiles, allowing university administrators and mental health professionals to understand the specific factors driving each high-risk classification, such as pronounced academic pressure or sleep disturbance. This transparency strengthens the scientific grounding and persuasiveness of intervention decisions and supports the design of targeted measures that address underlying causal pathways rather than symptoms alone.

4. Discussion

A key contribution of this study lies in bridging methodological rigor and clinical interpretability by combining optimized ensemble learning with psychologically grounded feature design and transparent SHAP-based explanations, thereby enhancing both predictive performance and real-world applicability. The proposed Dynamic Weighted Ensemble Model (DWEM) advances existing depression prediction studies through three key methodological innovations. First, it introduces a formal weight-optimization paradigm, in which ensemble fusion weights are learned via Bayesian optimization using the Optuna framework, replacing conventional heuristic or uniform stacking with a data-driven and reproducible weighting strategy. Second, the model incorporates a clinically aligned cost-sensitive ensemble objective, explicitly prioritizing the minimization of false negatives for high-risk cases (e.g., students with suicidal ideation), thereby aligning algorithmic optimization with psychiatric risk management principles. Third, DWEM establishes a feature-to-intervention translation mechanism by coupling SHAP-based explainability with probability-based risk stratification, enabling the systematic conversion of model explanations into tiered clinical decision rules and targeted intervention pathways. Together, these components elevate the framework from a high-accuracy predictor to an interpretable, risk-aware, and clinically actionable ensemble methodology.

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. 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. Importantly, the SHAP analysis revealed that factors such as stress-burnout and sleep-related disturbances consistently contributed to elevated depression risk, providing actionable insights that can inform targeted intervention strategies and institutional policy design.

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 (Huang et al., 2022; Zhang et al., 2023). By embedding explainability directly into the analytic pipeline, this study addressed one of the main barriers hindering clinical adoption of artificial intelligence (Jung et al., 2023). In particular, the SHAP-based feature analysis corroborates prior evidence linking academic performance and stress to depression (Auerbach et al., 2018; Feng et al., 2022), 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 (Van Mens et al., 2023; Jacob et al., 2022).

An important observation from the SHAP analysis is the dominant contribution of academic performance–related features (e.g., CGPA and its derived interactions) in predicting depression risk. While this finding aligns with prior literature emphasizing academic stress as a key determinant of student mental health, it may also reflect context-specific factors related to the studied population. In many educational systems, particularly those characterized by high academic competition and performance-driven evaluation, academic achievement is closely tied to self-worth, future career opportunities, and social expectations. As such, its prominence in the model may partially capture these broader socio-cultural pressures. Therefore, caution should be exercised when generalizing this finding to different institutional or cultural contexts, where the relative importance of academic performance may vary. Future studies should investigate cross-cultural variability and assess whether similar feature importance patterns are observed in diverse educational environments.

From a methodological perspective, the main contribution of this work lies in demonstrating how optimized ensemble learning can be systematically integrated with clinical reasoning and explainable artificial intelligence, rather than being treated as a purely predictive black box. The DWEM formalizes ensemble construction as a constrained optimization problem, in which model weights are learned under cross-validation and cost-sensitive objectives, thereby improving generalization while explicitly reflecting the asymmetric risks inherent in mental health screening. In addition, the combination of psychologically grounded feature engineering and SHAP-based explanation establishes a transparent mapping from latent risk factors to individualized predictions and tiered intervention strategies. This feature–model–explanation coupling moves beyond conventional performance-centric evaluation and illustrates a reproducible pathway for translating machine learning outputs into clinically interpretable and operationally actionable decision support. As such, the proposed framework contributes not only a high-accuracy predictor, but also a transferable methodological paradigm for building trustworthy, explainable, and intervention-oriented AI systems in psychiatric and behavioral health research.

Although deep learning models such as convolutional and transformer-based architectures have achieved remarkable success in high-dimensional and unstructured data domains (e.g., images, speech, and text), the present study focuses on structured, low-dimensional, survey-based clinical data, where different methodological considerations apply. In such tabular settings, extensive evidence from medical machine learning literature indicates that gradient-boosted decision tree ensembles and random forest–type models consistently match or outperform deep neural networks in terms of predictive accuracy, robustness, and data efficiency, particularly when sample sizes are moderate and feature semantics are clinically interpretable.

More importantly, for mental health decision-support systems, explainability, calibration, and stability of predictions are of higher practical value than marginal gains in raw accuracy. Tree-based ensembles naturally support transparent post-hoc explanation via SHAP and exhibit well-calibrated probability estimates, whereas deep neural networks often require complex calibration procedures and remain less interpretable for clinical stakeholders. As noted in prior studies, in structured clinical tabular settings, gradient-boosted ensembles consistently outperform deep neural networks while offering superior interpretability and calibration, making them more suitable for decision-support systems. Consequently, the adoption of an optimized ensemble framework in this work reflects a deliberate

methodological choice aligned with both the data characteristics and the clinical requirement for trustworthy, explainable, and stable risk stratification rather than a pursuit of architectural complexity alone.

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 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 (Levis et al., 2019).

The SHAP framework was employed at two complementary levels of analysis to support both population-level policy design and individual-level clinical intervention. At the global level, mean absolute SHAP values aggregated across all samples were used to identify the most influential features driving depression risk in the overall student population. This global importance profile informs policy and resource planning, such as prioritizing institutional actions targeting dominant risk domains (e.g., academic pressure, sleep disturbance, or lack of protective factors).

At the local level, instance-specific SHAP explanations were generated for each student to quantify how individual features increased or decreased the predicted risk relative to the baseline. These local attribution profiles enable personalized intervention, allowing clinicians and counselors to understand the specific combination of factors (e.g., high stress index, low academic satisfaction, severe sleep problems) contributing to a particular student's risk score.

Conceptually, this two-layer explanation mechanism can be summarized as:
Global SHAP (population aggregation) → identification of dominant risk factors → strategic policy and program design;

Local SHAP (individual attribution) → personalized risk decomposition → tailored, tiered intervention planning.

This dual-level interpretability bridges institutional decision-making and individualized care, ensuring that the ensemble model's outputs are actionable both for system-level mental health management and for case-level clinical support.

The implications of this study are both methodological and practical. Methodologically, it demonstrates how feature engineering grounded in psychological theory (Monroe & Simons, 1991) 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 subclinical distress, DWEM supports a paradigm shift from reactive to proactive mental health management, echoing global calls for early, scalable intervention in higher education (World Health Organization [WHO], 2021, 2023).

While the DWEM demonstrates excellent discrimination performance, the implications of false positives must be carefully considered in real-world university settings. In our evaluation, a small proportion of students were incorrectly classified as high risk, reflecting the inherent trade-off between sensitivity and specificity in screening models. Although false positives may increase the demand for follow-up assessments, they do not necessarily represent wasted resources. Instead, these individuals often exhibit subclinical psychological distress patterns, such as elevated stress or reduced academic satisfaction, and may still benefit from preventive or low-intensity interventions. To mitigate the risk of over-intervention, the proposed tiered framework is explicitly designed to align intervention intensity with predicted risk levels, ensuring that only the highest-risk individuals receive immediate clinical attention, while lower-risk cases are directed toward scalable, resource-efficient support services. This risk-stratified approach enables institutions to balance early detection with sustainable resource allocation.

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 remained 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.

The deployment of automated mental health screening systems raises important ethical considerations that must be addressed to ensure responsible use. First, although this study utilizes anonymized data, real-world implementation requires strict safeguards for data privacy, confidentiality, and informed consent, particularly given the sensitive nature of mental health information. Second, predictive models inherently involve risks of misclassification, including false positives and false negatives, which may lead to unnecessary anxiety or missed intervention if not carefully managed. Third, it is essential to emphasize that the proposed framework is designed to support, rather than replace, professional clinical judgment. Human oversight remains critical in interpreting model outputs and determining appropriate interventions. Moreover, transparency and explainability, as enabled by SHAP in this study, play a key role in fostering trust among stakeholders, including students, counselors, and institutional decision-makers. Future research should further explore governance frameworks, ethical guidelines, and stakeholder engagement strategies to ensure that AI-driven mental health systems are deployed in a manner that is equitable, accountable, and aligned with clinical best practices.

5. Limitations and Future Work

Although the proposed DWEM was trained and validated on a single large-scale student mental health dataset, its methodological design supports transportability across institutional and cultural contexts. The core feature categories employed in this study—such as academic stress, sleep quality, lifestyle behaviors, satisfaction, and family background—represent universal psychosocial constructs that have been consistently linked to depression across countries and educational systems. The dataset represents a specific student population, which may influence the relative importance of academic-related features. Consequently, while the statistical distributions of these variables may vary between campuses, their semantic meaning and clinical relevance remain stable, facilitating cross-site adaptation. The absence of clinically validated threshold calibration represents an important limitation and highlights the need for prospective validation studies involving mental health professionals.

Moreover, the DWEM architecture is inherently data-agnostic. The ensemble consists of general-purpose tree-based learners and an optimization-driven fusion mechanism, rather than domain-specific handcrafted rules. This allows the model to be retrained or recalibrated on new institutional datasets without structural modification. In practical deployment, a university could fine-tune both the base learners and the ensemble weights using local data while preserving the same optimization and explainability pipeline. Such transfer learning at the ensemble-weight level enables rapid adaptation to population-specific risk patterns while maintaining interpretability and decision consistency.

Algorithm 1: DWEM Adaptation to a New Campus Dataset (Pseudo-Procedure)

1. Collect anonymized student mental health survey and academic data from the target institution.
2. Apply the same theory-driven feature engineering and encoding pipeline.
3. Perform stratified K-fold cross-validation on the new dataset.
4. Train base learners (CatBoost, XGBoost, LightGBM, Random Forest, ExtraTrees) within each fold.
5. Optimize ensemble weights using Optuna to maximize cross-validated AUC under clinical cost constraints.
6. Evaluate calibration and discrimination on held-out folds.
7. Compute global and local SHAP values to validate feature stability and interpretability.
8. Derive institution-specific risk thresholds and tiered intervention rules.

This adaptation workflow ensures that DWEM can be systematically transferred and locally optimized for different university environments, supporting scalable deployment of explainable, risk-aware mental health decision-support systems.

6. Conclusion

This study demonstrated that the proposed Dynamic Weighted Ensemble Model (DWEM), which integrates multiple tree-based learners with optimized weighting and SHAP-based explainability, can achieve clinical-grade accuracy, robustness, and interpretability in predicting depression among college students. In doing so, it addresses long-

standing limitations of existing approaches that often prioritize predictive performance at the expense of transparency or fail to provide clear pathways from prediction to intervention. The empirical results confirm that a carefully designed ensemble framework, combined with theory-driven feature construction and rigorous validation, can deliver both high discrimination performance and meaningful insights into the underlying determinants of student mental health.

From a methodological perspective, this work advances the field by unifying several critical components within a single framework. Specifically, it integrates theory-driven feature engineering grounded in the diathesis–stress model, cost-sensitive and Bayesian-optimized ensemble learning, and multi-level explainability through SHAP. Importantly, the study goes beyond model development by operationalizing predictive outputs into a tiered decision-support system, thereby bridging the gap between algorithmic prediction and actionable mental health strategies. This connection between prediction, interpretation, and intervention represents a significant step toward clinically aligned and practically deployable AI systems.

Nevertheless, several limitations should be acknowledged. The analysis was conducted on a single large-scale dataset, which may limit generalizability across different cultural, institutional, or demographic contexts. In addition, the proposed intervention framework is based on predictive risk stratification rather than causal inference, and therefore requires further clinical validation. These limitations provide important avenues for future research, including cross-institutional replication, longitudinal studies to assess temporal stability, and the integration of multimodal data sources such as behavioral, physiological, or digital trace data to enhance robustness and applicability.

More broadly, the findings highlight the transformative potential of explainable artificial intelligence in reshaping campus mental health care. By enabling proactive, risk-informed, and resource-efficient screening and intervention, the proposed approach supports a shift away from reactive, questionnaire-based systems toward scalable, data-driven prevention strategies. Such advancements have important implications not only for improving student well-being but also for informing evidence-based policy, optimizing resource allocation, and strengthening institutional mental health infrastructures in higher education.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of *the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Alom, M. S., Tomal, M. A. H., Taha, R., Parvez, S., Layek, M. A., Mohsin, M., & Talukder, M. A. (2026). An Explainable Triple‐Layered Ensemble Model for Early Prediction of Suicide Risk Using Machine Learning. *Engineering Reports*, 8(1), e70574. https://doi.org/10.1002/eng2.70574

Aminifar, A., Shokri, M., Rabbi, F., Pun, V. K. I., & Lamo, Y. (2022). Extremely randomized trees with privacy preservation for distributed structured health data. *IEEE Access*, 10, 6010-6027. https://doi.org/10.1109/ACCESS.2022.3141709

Askin, S., Burkhalter, D., Calado, G., & El Dakrouni, S. (2023). Artificial intelligence applied to clinical trials: opportunities and challenges. *Health and technology*, 13(2), 203-213. https://doi.org/10.1007/s12553-023-00738-2

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., ... & Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, 127(7), 623–638. https://doi.org/10.1037/abn0000362

Band, S. S., Yarahmadi, A., Hsu, C. C., Biyari, M., Sookhak, M., Ameri, R., ... & Liang, H. W. (2023). Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Informatics in Medicine Unlocked*, 40, 101286. https://doi.org/10.1016/j.imu.2023.101286

Bayes, J., Schloss, J., & Sibbritt, D. (2022). The effect of a Mediterranean diet on the symptoms of depression in young males (the "AMMEND: A Mediterranean Diet in MEN with Depression" study): A randomized controlled trial. *The American Journal of Clinical Nutrition*, 116(2), 572-580. https://doi.org/10.1093/ajcn/nqac106

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, S., Wang, Y., She, R., Lau, J. T. F., Mo, P. K. H., Li, J., & Li, L. (2024). Machine Learning Techniques to Predict Mental Health Problems Using Annual Student Health Survey Data: Algorithm Development and Validation Study. *JMIR Mental Health*, 11, e50179. https://doi.org/10.2196/50179

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of *the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Dietterich, T. G. (2000). Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*, 1–15. https://doi.org/10.1007/3-540-45014-9_1

dos Santos Machado, C., Ballester, P. L., Cao, B., Mwangi, B., Caldieraro, M. A., Kapczinski, F., & Passos, I. C. (2022). Prediction of suicide attempts in a prospective cohort study with a nationally representative sample of the US population. *Psychological medicine*, 52(14), 2985-2996. https://doi.org/10.1017/S0033291720004997

Erikson, E. H. (1968). *Identity*: Youth and crisis. Norton.

Feng, X., Hu, M., & Guo, W. (2022, October). Application of artificial intelligence in mental health and mental illnesses. In Proceedings of *the 3rd International Symposium on Artificial Intelligence for Medicine Sciences* (pp. 506-511). https://doi.org/10.1145/3570773.3570834

GBD Mental Disorders Collaborators. (2022). Global burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, 9(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Huang, Y., Yang, Z., & Li, X. (2022). Predicting depression in college students using machine learning: A systematic review and meta-analysis. *Journal of Affective Disorders*, 314, 236-248. https://doi.org/10.1016/j.jad.2022.07.015

Hyseni Duraku, Z., Davis, H., & Hamiti, E. (2023). Mental health, study skills, social support, and barriers to seeking psychological help among university students: a call for mental health support in higher education. *Frontiers in Public Health*, 11, 1220614. https://doi.org/10.3389/fpubh.2023.1220614

Imans, D., Abuhmed, T., Alharbi, M., & El-Sappagh, S. (2024). Explainable Multi-Layer Dynamic Ensemble Framework Optimized for Depression Detection and Severity Assessment. *Diagnostics*, 14(21), 2385. https://doi.org/10.3390/diagnostics14212385

Jacob, N., Lannin, D., & Vogel, D. (2022). Bridging the gap between machine learning and clinical practice in suicide prediction. *Nature Human Behaviour,* 6(7), 901-902. https://doi.org/10.1038/s41562-022-01362-2

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *Statistical learning. In An introduction to statistical learning: With applications in Python* (pp. 15-67). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0_2

Jung, J., Lee, H., Jung, H., & Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. *Heliyon*, 9(5). https://doi.org/10.1016/j.heliyon.2023.e16110

Kaur, P., Singh, M., & Josan, G. S. (2022). A systematic review of ensemble learning approaches for depression detection. *Neuroscience Informatics*, 2(4), 100075. https://doi.org/10.1016/j.neuri.2022.100075

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems,* 30, 3146–3154.

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening depression: A systematic review and individual participant data meta-analysis. *BMJ*, 365, l1476. https://doi.org/10.1136/bmj.l1476

Lian, X., Xie, Z., Zhang, Y., & Wang, Q. (2023). Challenges and strategies for implementing AI-based mental health screening in universities: A qualitative study. *JMIR Mental Health,* 10, e45678. https://doi.org/10.2196/45678

Liu, D., Chen, Z., Marrero, W. J., Jacobson, N. C., & Thesen, T. (2023). Explainable machine learning-based prediction of depression severity in medical students. medRxiv, 2023-12. https://doi.org/10.1101/2023.12.14.23299975

Long, E., Patterson, S., Maxwell, K., Blake, C., Pérez, R. B., Lewis, R., ... & Mitchell, K. R. (2022). COVID-19 pandemic and its impact on social relationships and health. *J Epidemiol Community Health,* 76(2), 128-132. https://doi.org/10.1136/jech-2021-216690

López Steinmetz, L. C., Sison, M., Zhumagambetov, R., Godoy, J. C., & Haufe, S. (2024). Machine learning models predict the emergence of depression in Argentinean college students during periods of COVID-19 quarantine. *medRxiv*. https://doi.org/10.1101/2024.01.25.24301772

Matthews, T., Rasmussen, L. J. H., Ambler, A., Danese, A., Eugen-Olsen, J., Fancourt, D., ... & Moffitt, T. E. (2024). Social isolation, loneliness, and inflammation: a multi-cohort investigation in early and mid-adulthood. *Brain, behavior, and immunity,* 115, 727-736. https://doi.org/10.1016/j.bbi.2023.11.022

Monroe, S. M., & Simons, A. D. (1991). Diathesis-stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin*, 110(3), 406–425. https://doi.org/10.1037/0033-2909.110.3.406

Mumenin, N., Yousuf, M. A., Alassafi, M. O., Monowar, M. M., & Hamid, M. A. (2025). DDNet: A robust, and reliable hybrid machine learning model for effective detection of depression among university students. *IEEE Access.* https://doi.org/10.1109/ACCESS.2025.3552041

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education,* 20(1), 4. https://doi.org/10.1186/s41239-022-00372-4

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems,* 31, 6638–6648.

Razavi, M., Ziyadidegan, S., Mahmoudzadeh, A., Kazeminasab, S., Baharlouei, E., Janfaza, V., ... & Sasangohar, F. (2024). Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: scoping review. *JMIR Mental Health,* 11(1), e53714. https://doi.org/10.2196/53714

Van Mens, K., Lokkerbol, J., Wijnen, B., Janssen, R., de Lange, R., & Tiemens, B. (2023). Predicting undesired treatment outcomes with machine learning in mental health care: multisite study. *JMIR Medical Informatics,* 11(1), e44322. https://doi.org/10.2196/44322

World Health Organization. (2021). *Suicide worldwide in 2019: Global health estimates*. World Health Organization. https://www.who.int/publications/i/item/9789240026643

World Health Organization. (2023). *Depression. Fact sheet.* https://www.who.int/news-room/fact-sheets/detail/depression

Wu, L., & Wang, Y. (2024). Addressing the mental health care gap in Chinese universities: A call for digital solutions. *Current Psychology*, 43(5), 4521-4532. https://doi.org/10.1007/s12144-023-04622-0

Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., ... & Du, X. (2025). Machine learning predictive models to guide prevention and intervention allocation for anxiety and depressive disorders among college students. *Journal of Counseling & Development,* 103(1), 110-125. https://doi.org/10.1002/jcad.12543