

RESEARCH

Open Access



An artificial intelligence tool to assess the risk of severe mental distress among college students in terms of demographics, eating habits, lifestyles, and sport habits: an externally validated study using machine learning

Lirong Zhang^{1*}, Shaocong Zhao¹, Zhongbing Yang², Hua Zheng^{3*} and Mingxing Lei^{4,5*}

Abstract

Background Precisely estimating the probability of mental health challenges among college students is pivotal for facilitating timely intervention and preventative measures. However, to date, no specific artificial intelligence (AI) models have been reported to effectively forecast severe mental distress. This study aimed to develop and validate an advanced AI tool for predicting the likelihood of severe mental distress in college students.

Methods A total of 2088 college students from five universities were enrolled in this study. Participants were randomly divided into a training group (80%) and a validation group (20%). Various machine learning models, including logistic regression (LR), extreme gradient boosting machine (eXGBM), decision tree (DT), k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM), were employed and trained in this study. Model performance was evaluated using 11 metrics, and the highest scoring model was selected. In addition, external validation was conducted on 751 participants from three universities. The AI tool was then deployed as a web-based AI application.

Results Among the models developed, the eXGBM model achieved the highest area under the curve (AUC) value of 0.932 (95% CI: 0.911–0.949), closely followed by RF with an AUC of 0.927 (95% CI: 0.905–0.943). The eXGBM model demonstrated superior performance in accuracy (0.850), precision (0.824), recall (0.890), specificity (0.810), F1 score (0.856), Brier score (0.103), log loss (0.326), and discrimination slope (0.598). The eXGBM model also received the

*Correspondence:

Lirong Zhang
22674481@qq.com
Hua Zheng
20130296@cqnu.edu.cn
Mingxing Lei
leimingxing@301hospital.com.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

highest score of 60 based on the evaluation scoring system, while RF achieved a score of 49. The scores of LR, DT, and SVM were only 19, 32, and 36, respectively. External validation yielded an impressive AUC value of 0.918.

Conclusions The AI tool demonstrates promising predictive performance for identifying college students at risk of severe mental distress. It has the potential to guide intervention strategies and support early identification and preventive measures.

Keywords Artificial intelligence, Mental distress, Machine learning, College students, Application

Background

The mental well-being of college students has become a growing concern due to the increasing prevalence and negative impact of mental distress [1–3]. The college years are a critical period when young adults face various challenges and transitions that can significantly impact their mental health. Studies have shown that college students experience high rates of mental distress, including anxiety, depression, and other psychological disorders [1], and there was a notable rise in self-reported psychological distress. Severe mental distress, including severe anxiety or depression [4], has been linked to several negative outcomes such as poor academic performance, decreased social engagement, and an increased risk of substance abuse [5, 6].

Accurately predicting the likelihood of mental issues among college students is crucial for early intervention and prevention [7–9]. Recent advancements in artificial intelligence (AI) and machine learning techniques have shown great promise in the field of mental health [7–9]. These technologies have the potential to revolutionize the prediction and prevention of mental health among college students. AI algorithms can process large amounts of data [10], including demographic information, lifestyle factors, and psychological parameters, to develop predictive models with high accuracy and reliability. Additionally, AI tools can provide personalized risk assessments and recommendations, facilitating targeted interventions and support [10–13]. Several studies have explored the use of AI in predicting mental health problems among college students [14–16]. These studies have shown favorable results, with AI algorithms achieving relatively high levels of accuracy in identifying individuals at high risk for mental issues, such as negative mental well-being traits, mental health problems, severe depressive symptoms, suicidal ideation, and perceived stress [7–9, 14–16]. However, there have been no specific AI models reported for predicting severe mental distress currently.

Therefore, the main objective of this study was to establish an advanced AI tool specifically for predicting the risk of severe mental distress among university students, and internally and externally assess the performance of the AI tool. The findings of this study would

have important implications for early intervention and preventive measures in college mental health.

Methods

Participants and study design

This study analyzed 2088 college students from five universities between September, 2021 and May, 2023. We recruited college students who volunteered to participate in a survey. The survey included questions about participants' basic demographics, exercise and eating habits, lifestyle, sleep quality, and mental health status [17]. The questionnaire was presented in Chinese, and the same language versions questionnaires applied uniformly on all participants. The questionnaire was distributed online at these various universities. Participants were excluded if they had a previous diagnosis of anxiety or depression, or were unwilling to participate. All participants were randomly divided into a training group and a validation group in an 8:2 ratio. The training group was used to develop models, while the validation group was used to test the models internally. External validation was conducted on 751 participants from three universities between May and June 2023. The survey distributed was identical to the one used for model development, ensuring consistency in the data collection process. Notably, the online survey was anonymous and did not collect any personal information. See Fig. 1 for a visual representation of the study design. The study was approved by the Academic Committee and Ethics Board of the Xiamen University of Technology, and all participants provided informed consent. This study was conducted in accordance with the Declaration of Helsinki and reported following the TRIPOD Checklist [18].

Data collection

In this study, data on participants' age, gender, grade, marital status, drinking and smoking habits, dietary preferences (such as low salt and oil, fatty foods, red meat, barbecued foods, vegetables, and fruits. In addition, a dietary questionnaire is presented as Supplementary File 1), monthly expenses, daily sedentary time, frequency of exercise per week, presence of chronic diseases, and sleep quality were collected, after reviewing literature and counseling experts. An explanation of why the selected variables were chosen in this study is summarized in

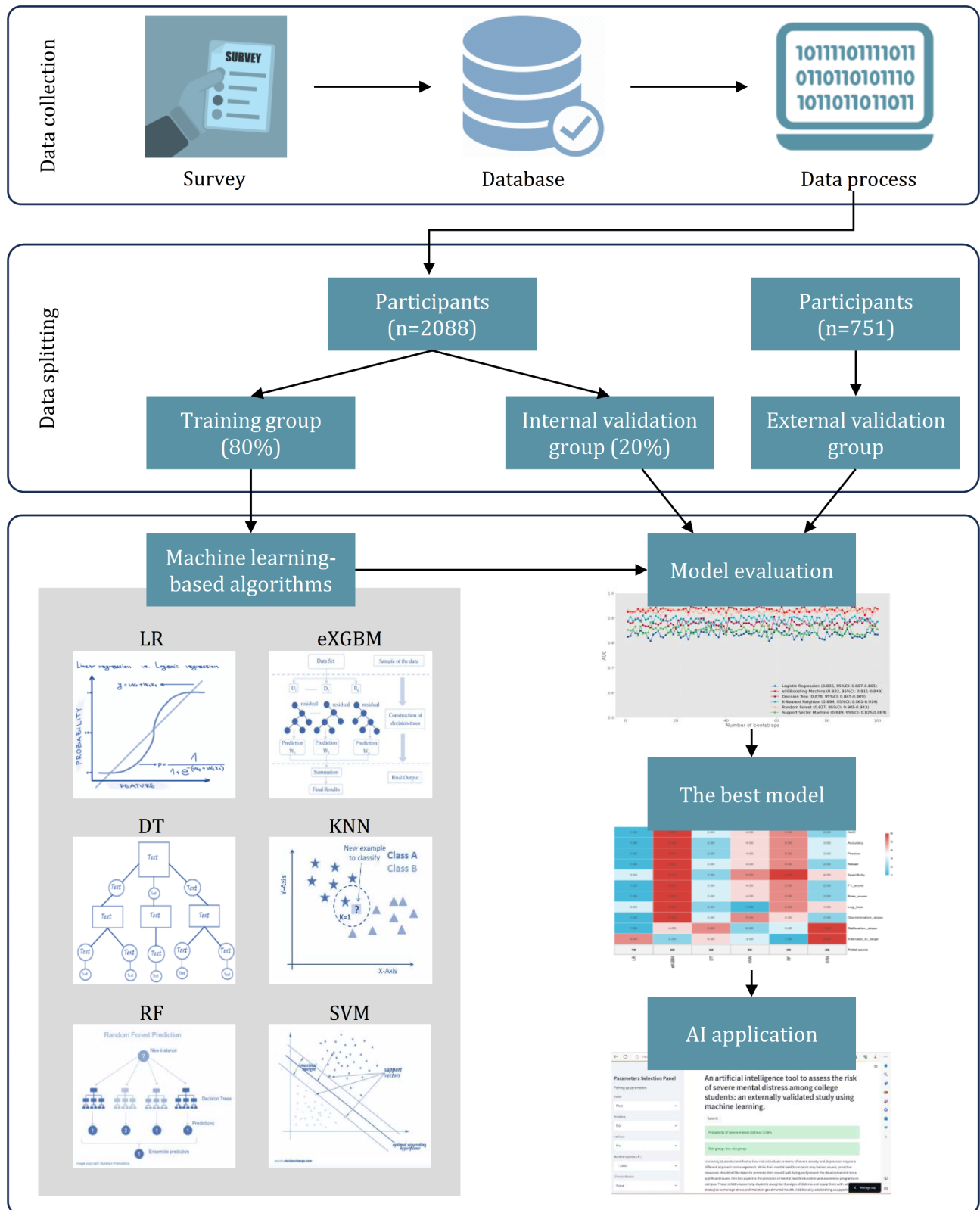


Fig. 1 Design of the study. LR, logistic regression; eXGBM, extreme gradient boosting machine; DT, decision tree; KNN, K-nearest neighbor; RF, random forest; SVM, support vector machine

Supplementary Table 1. Chronic diseases considered in the study included hypertension, diabetes, congenital heart disease, chronic kidney disease, chronic lung disease, chronic liver disease, previous cerebral infarction, rheumatoid arthritis, multiple sclerosis, Parkinson's disease, thyroid disorders, inflammatory bowel disease. The sleep quality of college students was assessed using the Pittsburgh Sleep Quality Index (PSQI), a widely-used questionnaire that measures various aspects of sleep quality [19]. The PSQI consists of 19 items that evaluate factors such as sleep duration, disturbances, latency, efficiency, medication usage, and daytime dysfunction. Each item is scored on a scale from 0 to 3, with a total score ranging from 0 to 21. Higher scores indicate poorer overall sleep quality. The Chinese version of PSQI has been pre-validated in Chinese university students [20].

Definition of the outcome

The severity of anxiety was evaluated with the general anxiety disorder-7 (GAD-7), and the severity of depression was evaluated with the patient health questionnaire-9 (PHQ-9). The GAD-7 and PHQ-9 are widely used self-report questionnaires [21, 22]. Both scales consist of several items that are scored on a scale from 0 to 3, with higher scores indicating greater symptom severity. A score of 15 or above was regarded as severe anxiety or depression in both scales. They were valuable tools for screening, diagnosing, and monitoring anxiety and depression in individuals. The reliability of GAD-7 and PHQ-9 was pre-validated in Chinese populations [23, 24]. In this study, severe mental distress in this study was defined as participants with severe anxiety or depression [4].

Data preparation

In order to ensure the smooth development and validation of machine learning-based models, a comprehensive data preprocessing pipeline was employed in this study. The pipeline utilized the scikit-learn library (version 1.1.3) to achieve data standardization. Additionally, to address the challenge of imbalanced data distribution and improve the robustness of our models, we employed the Synthetic Minority Oversampling Technique in conjunction with Tomek Links Undersampling Techniques [11–13, 25, 26]. This resampling technique, known as SMOTETomek, effectively balanced the proportions of outcome classes within the training and validation groups. The SMOTETomek was selected to address data imbalance because it combines SMOTE, which generates synthetic minority class samples, and Tomek Links, which remove borderline or noisy instances, resulting in a balanced and cleaner dataset. This approach reduces overfitting by eliminating overlapping instances, enhances class separability, and is particularly effective in

complex datasets where the minority class is dispersed. In addition, by implementing a stratified strategy, we ensured consistency in these proportions.

Modeling

In this study, a wide range of machine learning techniques were employed for modeling purposes. These techniques included logistic regression (LR), extreme gradient boosting machine (eXGBM), decision tree (DT), k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM). All models were trained and optimized using the same input features identified through subgroup analysis of university students with and without severe mental distress. The process of hyperparameter tuning for our machine learning models was meticulously designed to ensure optimal performance while maintaining a balance between complexity and generalization. Initially, we established wide ranges for each hyperparameter, informed by extensive literature reviews and empirical evidence [27]. This approach enabled a thorough exploration of potential values. For instance, we set the depth of decision trees to range from 2 to 100. To navigate these ranges, we employed a combination of grid search and random search techniques; grid search was used for smaller, discrete hyperparameter sets, while random search covered larger, continuous ranges. The performance of each model configuration was rigorously evaluated using k-fold cross-validation, typically with k set to 5 or 10, depending on the dataset's size. By employing this approach, we ensured the selection of well-performing models while avoiding both underfitting and overfitting. The machine learning algorithms were implemented using Python (version 3.9.7), and hyperparameter tuning was conducted using scikit-learn (version 1.2.2).

Validation

To assess the prediction performance of our models, we employed highly recognized and commonly used metrics. These metrics included the AUC, accuracy, precision, recall, specificity, F1 score, Brier score, log loss, discrimination slope, calibration slope, and intercept [28, 29]. The AUC was calculated by applying 100 bootstraps and represents the overall performance of a model, as it measures the area under the receiver operating characteristic (ROC) curve. A higher AUC indicates a better discrimination ability of the model, and a value above 0.90 is typically indicative of excellent prediction performance. Accuracy, precision, recall, and specificity were evaluated using confusion matrix [29]. Accuracy is a fundamental metric that quantifies the ability of a classification model to correctly classify instances. It was calculated by dividing the number of correctly classified instances (true positives and true negatives) by the

total number of instances. Precision, on the other hand, focuses on the proportion of instances that were accurately predicted as positive out of all instances predicted as positive. It was calculated by dividing the number of true positive predictions by the sum of true positive and false positive predictions. Recall, also known as sensitivity or the true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances. It was calculated by dividing the number of true positive predictions by the sum of true positive and false negative predictions.

The Brier score is a commonly used metric for assessing the accuracy and calibration of probabilistic predictions [28]. It calculates the mean squared difference between predicted probabilities and the actual outcomes. A lower Brier score suggests that the probabilistic predictions are more accurate and well-calibrated. The Brier score can be calculated using the following formula [28]:

$$\text{Brier Score} = \frac{1}{N} \sum_{i=1}^n (p_i - o_i)^2$$

Where, N represents the total sample, p_i represents the predicted probability of severe mental distress for the i -th instance, and o_i represents the actual probability of severe mental distress for the i -th instance.

Log loss, commonly referred to as cross-entropy loss, is a widely employed metric in classification tasks [30]. It computes the average negative logarithm of the predicted probabilities for the correct class. This metric assesses the disparity between predicted probabilities and the actual class labels. A lower log loss signifies superior performance of the classification model. The log loss is determined by the following equation [11]:

$$\text{Log Loss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

Where, N represents the number of samples, M represents the number of classes, y_{ij} represents the true label of sample i for class j (0 or 1), and p_{ij} represents the predicted probability of sample i belonging to class j .

In addition to log loss, we utilized the discrimination slope to assess the model's ability to rank individuals based on their predicted probabilities. The discrimination slope measures how well the model distinguishes between high-risk and low-risk individuals [11, 28]. The calibration slope, on the other hand, evaluates the alignment between the model's predicted probabilities and the observed probabilities. A calibration slope value of 1 indicates perfect calibration. Both the calibration slope and the intercept-in-large were obtained from

the calibration curve, which provides insights into the model's calibration performance.

Furthermore, we conducted a decision curve analysis to evaluate the clinical net benefit of each model. This analysis assesses the net benefit of using the model's predictions compared to other strategies, considering the potential risks and benefits. To provide a comprehensive assessment of the predictive performance, we developed a scoring system based on previous studies [10–13]. This scoring system incorporates the 11 metrics mentioned above, assigning each metric a rating from one to six. Higher scores indicate better predictive performance, and the scoring system encompasses a range from 0 to 66.

Feature importance

This study employed the Shapley Additive Explanation (SHAP) method to assess the significance of each feature in order to enhance interpretability in clinical settings [31]. This method assigns a numerical value to each feature reflecting its influence on the model's output. Higher SHAP values indicate a stronger feature impact. We derived individual outcome predictions using the SHAP method. The feature importance can be elucidated through the following formula [26]:

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j Z'_j$$

Where, the output of the interpretation model is denoted by g , the total number of input parameters is represented by M , ϕ_0 stands for a constant term, ϕ_j signifies the attribution value (Shapley value) assigned to each model parameter, and Z'_j corresponds to the value of the j -th feature for the specific under examination.

Within the coalition vectors, a value of "1" denotes the presence of respective feature that aligns with the features of the case being analyzed. Conversely, a value of "0" indicates the absence of that feature in the current case. By setting all simplified features to "1" in a hypothetical scenario, the SHAP expression can be streamlined for a more concise depiction of feature importance based on SHAP values, and the equation is shown as follows.

$$g(x') = \phi_0 + \sum_{j=1}^M \phi_j$$

Deployment of the AI tool

The web-based AI tool created with the best model in our study was launched to offer a user-friendly platform for researchers, clinicians, and healthcare professionals.

GitHub was utilized as the code hosting platform for effective storage and version control of the codebase. Streamlit, a cloud infrastructure platform, was employed to host the online calculator, ensuring consistent and scalable performance. The tool's user interface was crafted to allow users to effortlessly input university student's information and promptly receive the predicted likelihood of the severe mental distress. It featured user-friendly panels for selecting model parameters, conducting probability calculations, and accessing details about the model. The interface aimed to deliver a smooth and engaging user experience, empowering users to interpret and evaluate the probabilities of severe mental distress in college students.

Statistical analysis

In our analysis, we summarized continuous variables by calculating the average and standard deviation (SD) of the data. Categorical variables were expressed as percentages. To assess the distribution of categorical variables, we utilized the Chi-square test. When comparing continuous variables, either the student t-test or Wilcoxon rank test was applied depending on the characteristics of the data. All statistical analyses were performed using the R programming language (version 4.1.2). Statistical significance was considered present when the two-tailed P value was below 0.05.

Results

Participant characteristics

A total of 2088 participants were included in the study. The mean age of the participants was 19.84 years (SD: 2.12 years), with a majority of 55.9% being female. Among the participants, 46.4% were in their sophomore year. The majority of participants were single, accounting for 76.4% of the sample. Notably, a significant proportion of participants reported not drinking (80.8%) or smoking (91.7%). Details of eating and physical activity habits are summarized in Table 1. Results revealed that 26.1% of participants had a preference for consuming fatty foods, whereas 29.8% preferred barbecue. Only 48.9% and 58.1% of participants reported a preference for consuming vegetables and fruits, respectively. Sedentary behavior was prevalent among the participants, with 42.4% reporting a daily sedentary time of 6 h or more. In terms of comorbidity burden, participants had relatively low rates of chronic diseases, with only 4.4% reporting a chronic condition. The participants had an average PSQI score of 5.57 (SD: 2.88), and the prevalence of severe mental distress was 4.07% (85/2088) among these participants.

Subgroup analysis of participants stratified by severe mental distress

Subgroup analysis revealed that participants with severe mental distress exhibited certain distinct characteristics. They tended to be older ($P=0.008$) and in higher grades ($P=0.016$) compared to those without severe mental distress. Additionally, they had a higher rate of smoking ($P=0.044$), a higher preference for consuming fat food ($P=0.037$), a higher monthly expense ($P<0.001$), a higher rate of chronic disease ($P=0.043$), and a higher PSQI score ($P<0.001$) (Table 1). In the study, we did not find a preference for eating vegetables ($P=0.648$) to be a protective factor against severe mental distress. Additionally, while participants without severe mental distress showed a relatively higher rate of fruit consumption, this difference did not reach statistical significance ($P=0.077$).

Furthermore, participants who engaged in physical activities frequently were found to have a lower likelihood of experiencing severe mental distress, although this association did not reach statistical significance ($P=0.057$).

Prediction performance of models

Among the developed models, the eXGBM model exhibited the highest AUC value of 0.932 (95% CI: 0.911–0.949), closely followed by the RF model with an AUC of 0.927 (95% CI: 0.905–0.943) (Fig. 2). The calibration curve demonstrated that most models, particularly eXGBM, RF, and KNN, displayed favorable calibration ability (Fig. 3). Further assessment of the calibration slope and intercept-in-large confirmed the good calibration of these models, with calibration slopes close to 1 and intercept-in-large values close to 0 (Supplementary Fig. 1). The probability density curve revealed the ability of the eXGBM, RF, and KNN models to effectively distinguish participants with and without severe mental distress. This was indicated by the leftward shift of the peak of the blue curve (participants without severe mental distress) and the rightward shift of the peak of the red curve (participants with severe mental distress) (Fig. 4). Violin plots supported this trend, with the eXGBM model exhibiting the highest discrimination slope (0.598), followed by the KNN model (0.594) and the RF model (0.553) (Fig. 5). In terms of performance measures, the eXGBM model demonstrated superior accuracy (0.850), precision (0.824), recall (0.890), specificity (0.810), F1 score (0.856), Brier score (0.103), and log loss (0.326) (Fig. 6; Table 2). The decision curve analysis for each model (Supplementary Fig. 2) indicated that the eXGBM model provided favorable clinical net benefit compared to the other models (Fig. 7). Based on the comprehensive evaluation scoring system, the eXGBM model received the highest score of 60, while RF achieved a score of 49 (Fig. 8). The scores for LR, DT, and SVM were only 19, 32, and 36, respectively.

Table 1 Participant's baseline characteristics

Characteristics	Overall	Severe mental distress		P value
		No	Yes	
n	2088	2003	85	
Age (years, mean [SD])	19.84 (2.12)	19.81 (2.06)	20.44 (3.12)	0.008
Gender (male/female, %)	921/1167 (44.1/55.9)	883/1120 (44.1/55.9)	38/47 (44.7/55.3)	0.999
Grade (%)				0.016
First	520 (24.9)	510 (25.5)	10 (11.8)	
Second	969 (46.4)	927 (46.3)	42 (49.4)	
Third	352 (16.9)	336 (16.8)	16 (18.8)	
Fourth	228 (10.9)	213 (10.6)	15 (17.6)	
Delayed	19 (0.9)	17 (0.8)	2 (2.4)	
Marital status (%)				0.233
Single	1596 (76.4)	1528 (76.3)	68 (80.0)	
Dating	472 (22.6)	457 (22.8)	15 (17.6)	
Married	20 (1.0)	18 (0.9)	2 (2.4)	
Drinking (%)				0.191
No	1688 (80.8)	1622 (81.0)	66 (77.6)	
Quitted	102 (4.9)	100 (5.0)	2 (2.4)	
Yes	298 (14.3)	281 (14.0)	17 (20.0)	
Smoking (%)				0.044
No	1914 (91.7)	1839 (91.8)	75 (88.2)	
Quitted	66 (3.2)	65 (3.2)	1 (1.2)	
Yes	108 (5.2)	99 (4.9)	9 (10.6)	
Low salt oil food (no/yes, %)	1460/628 (69.9/30.1)	1405/598 (70.1/29.9)	55/30 (64.7/35.3)	0.342
Fat food (no/yes, %)	1542/546 (73.9/26.1)	1488/515 (74.3/25.7)	54/31 (63.5/36.5)	0.037
Red meat (no/yes, %)	694/1394 (33.2/66.8)	669/1334 (33.4/66.6)	25/60 (29.4/70.6)	0.518
Barbecue (no/yes, %)	1466/622 (70.2/29.8)	1411/592 (70.4/29.6)	55/30 (64.7/35.3)	0.312
Vegetable (no/yes, %)	1067/1021 (51.1/48.9)	1021/982 (51.0/49.0)	46/39 (54.1/45.9)	0.648
Fruit (no/yes, %)	875/1213 (41.9/58.1)	831/1172 (41.5/58.5)	44/41 (51.8/48.2)	0.077
Monthly expense (¥, %)				<0.001
<2000	1637 (78.4)	1572 (78.5)	65 (76.5)	
≥2000 and <5000	430 (20.6)	416 (20.8)	14 (16.5)	
≥5000 and <10,000	12 (0.6)	9 (0.4)	3 (3.5)	
≥10,000	9 (0.4)	6 (0.3)	3 (3.5)	
Daily sedentary time (hours, %)				0.163
<1	110 (5.3)	107 (5.3)	3 (3.5)	
≥1 and <3	391 (18.7)	379 (18.9)	12 (14.1)	
≥3 and <6	701 (33.6)	677 (33.8)	24 (28.2)	
≥6	886 (42.4)	840 (41.9)	46 (54.1)	
Sport frequency every week (%)				0.057
0	439 (21.0)	413 (20.6)	26 (30.6)	
1 to 2	786 (37.6)	759 (37.9)	27 (31.8)	
3 to 4	441 (21.1)	420 (21.0)	21 (24.7)	
≥5	422 (20.2)	411 (20.5)	11 (12.9)	
Chronic disease (yes/no, %)	92/1996 (4.4/95.6)	84/1919 (4.2/95.8)	8/77 (9.4/90.6)	0.043
PSQI (mean [SD])	5.57 (2.88)	5.42 (2.77)	9.12 (3.24)	<0.001

SD, Standard deviation; PSQI, Pittsburgh sleep quality index

These results suggested that the eXGBM model was the optimal one.

External validation

External validation of the eXGBM model was conducted using a separate cohort of 751 participants. The baseline

characteristics of these participants are summarized in Supplementary Table 2. The external validation yielded an AUC value of 0.918 (95% CI: 0.904–0.933) (Supplementary Fig. 3). In terms of performance measures, the eXGBM model demonstrated an accuracy of 0.849, precision of 0.886, recall of 0.801, F1 score of 0.841, Brier

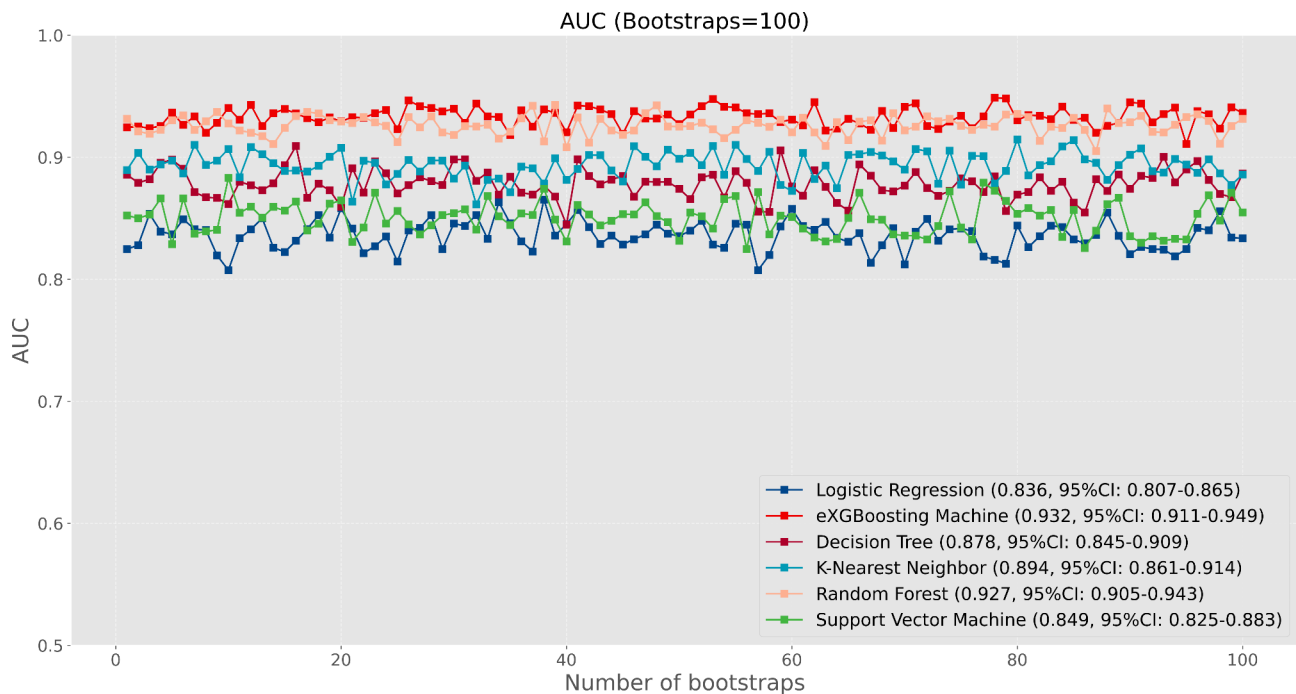


Fig. 2 The area under the curve value after conducting 100 bootstraps for each model

score of 0.115, and log loss of 0.408. The probability density curve showed that the eXGBM model had favorable discrimination (Supplementary Fig. 4), supported by a discrimination slope of 0.594 (Supplementary Fig. 5). The calibration slope was found to be 0.739, and the intercept-in-large value was 0.637 (Supplementary Fig. 6). Decision curve analysis further demonstrated that the eXGBM model provided favorable clinical net benefits (Supplementary Fig. 7). Additionally, it was observed that the PSQI was positively associated with the SHAP value for PSQI. Overall, the external validation results confirmed the robustness and generalizability of the eXGBM model in predicting the outcome in an independent cohort.

Feature importance and individual prediction

The SHAP analysis revealed that the three most important features for predicting the outcome were PSQI, age, and grade, as evidenced in both the training (Fig. 9A) and validation (Fig. 9B) groups. The relationship between continuous features, such as age and PSQI, and their corresponding SHAP values is depicted in Supplementary Fig. 8. The absolute value of the SHAP value for PSQI indicates its contribution to the outcome. A larger absolute SHAP value suggests a greater contribution, while a negative value represents a protective factor, and a positive value represents a promoting factor. Supplementary Fig. 9 illustrates a true positive case, where features such as PSQI, grade, monthly expense, age, smoking, and fat food were identified as risk factors, while chronic disease acted as a protective factor. Each feature had a

corresponding SHAP value, with larger values indicating a greater contribution to the outcome. The sum of the SHAP values in this case was 1.737, significantly larger than the base value of -0.010, indicating a positive prediction. On the other hand, Supplementary Fig. 10 depicts a true negative case.

Deployment of the AI tool

The web-based AI tool has been successfully developed by uploading the highly optimized eXGBM model and its code to a GitHub repository. This model has been integrated into the Streamlit platform, ensuring easy and user-friendly access to the AI tool. The code of the model can be available at: <https://github.com/Starxue-shu/predictionofmentaldistress>. Once accessed, users can generate highly personalized risk assessments for severe mental distress among college students (Fig. 10). By selecting their desired model parameters and simply clicking the “submit” button, the tool will provide individual risk assessments based on the powerful eXGBM model. Additionally, the tool provides a stratification of university students into high-risk and low-risk groups, enabling tailored recommendations for early prevention of mental distress. To ensure uninterrupted access and a smooth user experience, the platform includes a reactivation feature. In the event of platform inactivity or shutdown, users can reactivate it effortlessly by clicking on the “Yes, get this app back up!” option. Within a short span of approximately 30 s, the platform will be up and

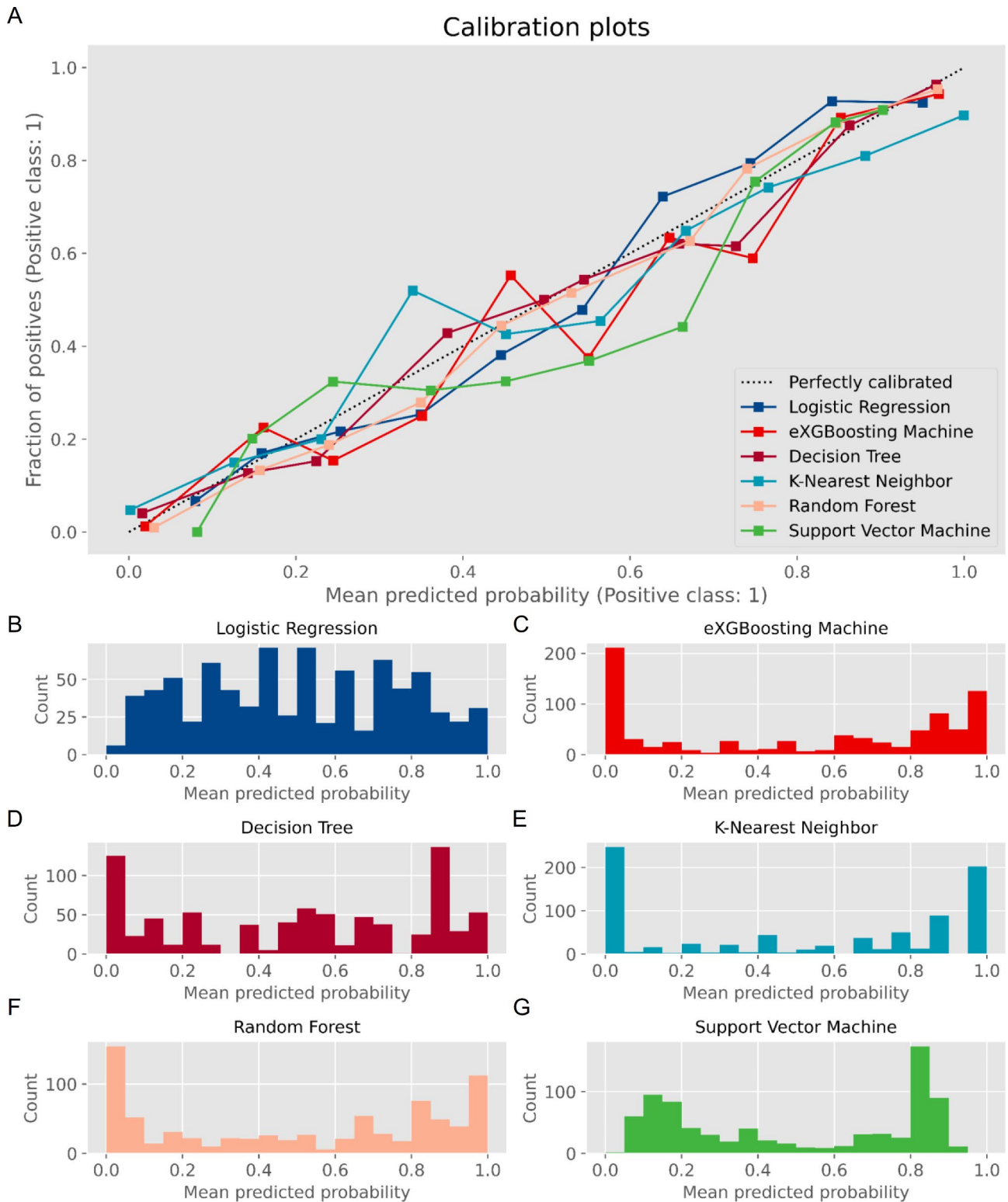


Fig. 3 Calibration curves and histogram plots of predicted probability for each model. **A** Calibration curves; **B**. Histogram plot of predicted probability for logistic regression; **C**. Histogram plot of predicted probability for eXGBoosting machine; **D**. Histogram plot of predicted probability for decision tree; **E**. Histogram plot of predicted probability for K-nearest neighbor; **F**. Histogram plot of predicted probability for random forest; **G**. Histogram plot of predicted probability for support vector machine

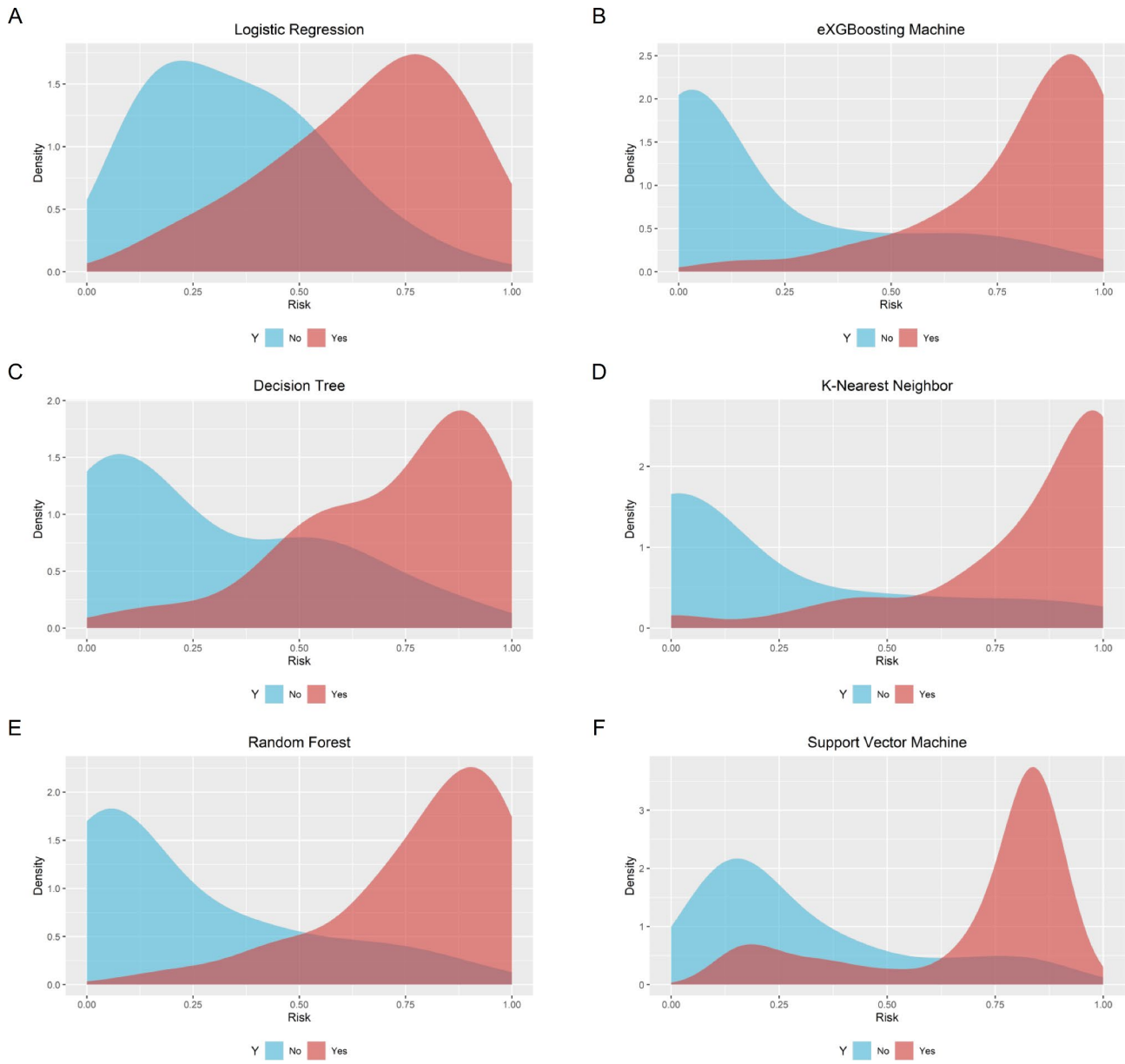


Fig. 4 Density curves for each model. (A) logistic regression; (B) eXGBoosting machine; (C) decision tree; (D) K-nearest neighbor; (E) random forest; (F) support vector machine

running again, allowing users to continue utilizing the online application without any inconvenience.

Discussion

Main findings

The main finding of this study is that the developed AI tool demonstrates promising predictive performance for identifying college students at risk of severe mental distress. Among the various machine learning models evaluated, the eXGBM model achieved the highest performance with an AUC value of 0.932. This indicates the model’s ability to accurately discriminate between individuals with and without severe mental distress. In

addition, external validation of the AI tool further supported its effectiveness, yielding an impressive AUC value of 0.918. This validates the generalizability and robustness of the tool’s predictive capabilities across different university populations. Thus, the AI tool could effectively stratify college students into high-risk and low-risk groups, enabling personalized recommendations for preventive interventions. By stratifying students into risk groups, the tool could facilitate targeted interventions and preventive measures, ultimately improving mental health outcomes and overall well-being in the college population.

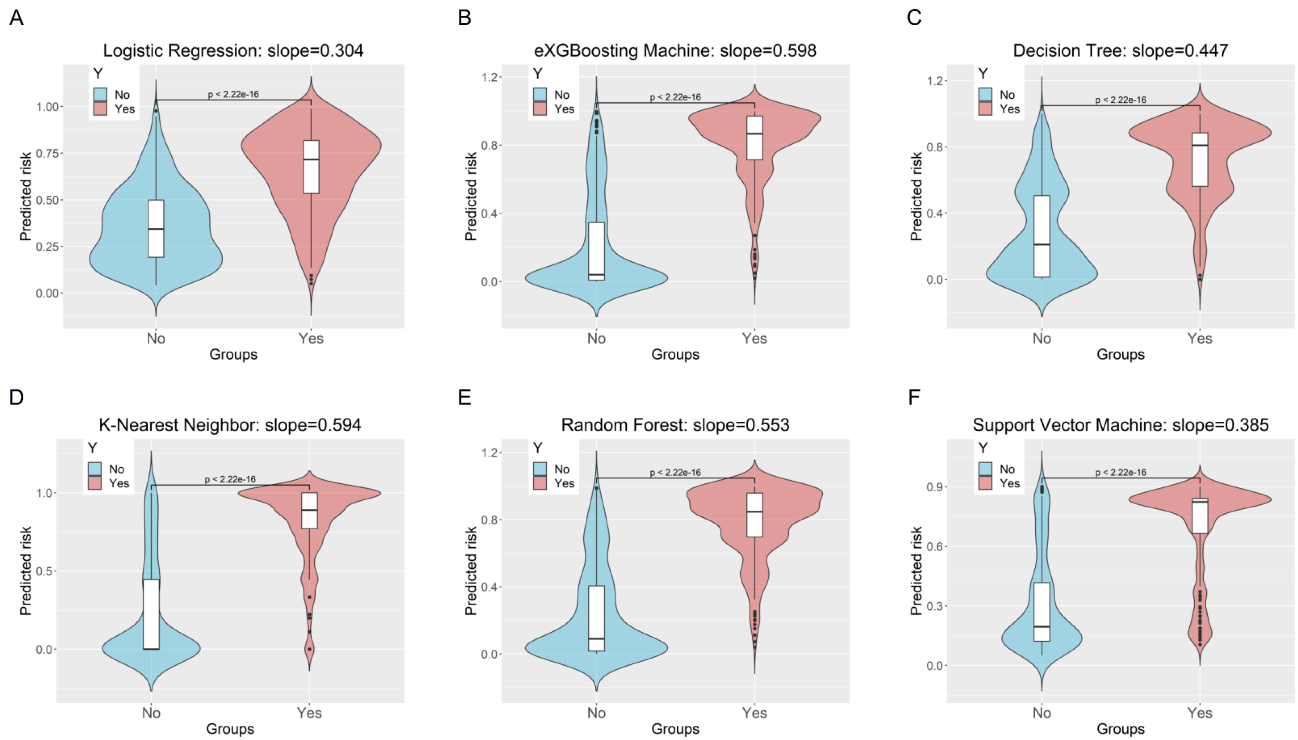


Fig. 5 Discrimination slope for each model. (A) logistic regression; (B) eXGBoosting machine; (C) decision tree; (D) K-nearest neighbor; (E) random forest; (F) support vector machine

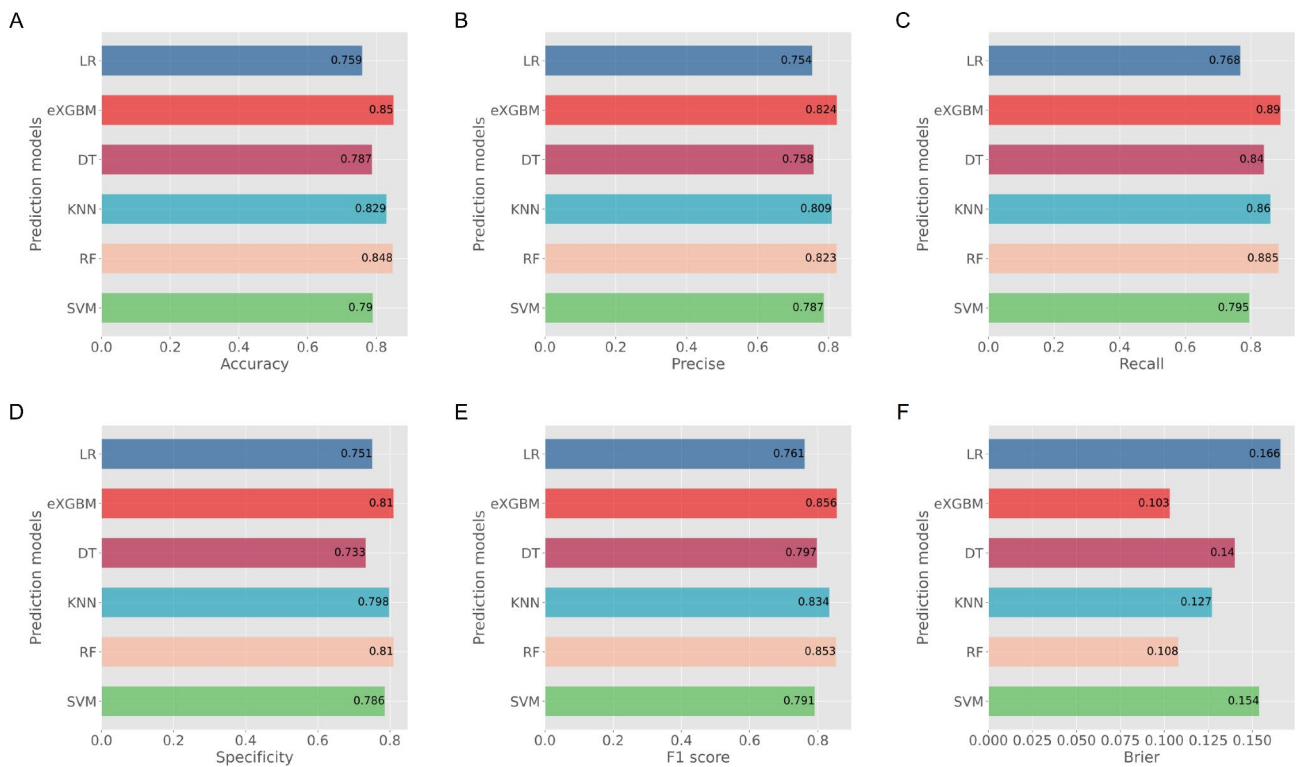


Fig. 6 Prediction performance for each model. (A) Accuracy; (B) Precise; (C) Recall; (D) Specificity; (E) F1 score; (F) Brier score

Table 2 Prediction performance of machine learning-based and traditional models in the internal validation set

Evaluation metrics	Models					
	LR	eXGBM	DT	KNN	RF	SVM
AUC (95%CI)	0.836 (95%CI: 0.807–0.865)	0.932 (95%CI: 0.911–0.949)	0.878 (95%CI: 0.845–0.909)	0.894 (95%CI: 0.861–0.914)	0.927 (95%CI: 0.905–0.943)	0.849 (95%CI: 0.825– 0.883)
Accuracy	0.759	0.850	0.787	0.829	0.848	0.790
Precise	0.754	0.824	0.758	0.809	0.823	0.787
Recall	0.768	0.890	0.840	0.860	0.885	0.795
Specificity	0.751	0.810	0.733	0.798	0.810	0.786
F1 score	0.761	0.856	0.797	0.834	0.853	0.791
Brier score	0.166	0.103	0.140	0.127	0.108	0.154
Log loss	0.507	0.326	0.544	1.676	0.345	0.478
Discrimination slope	0.304	0.598	0.447	0.594	0.553	0.385
Calibration slope	1.183	0.962	1.008	0.827	1.092	1.007
Intercept-in-large	-0.033	-0.104	-0.05	-0.065	-0.136	-0.006
Total score	19	60	32	40	49	36

AUC, area under the curve; CI, confident interval; LR, logistic regression; eXGBM, eXGBoosting Machine

AI prediction of mental problems in college students

Machine learning and artificial intelligence techniques have been utilized for early detection, prognostication, and prediction of negative psychological well-being states [7–9]. For instance, Rahman et al. [7] discovered that machine learning algorithms can effectively assess mental well-being, with random forest and adaptive boosting algorithms achieving the highest accuracy in identifying negative mental well-being traits. The key predictors of poor mental well-being included the frequency of sports activities per week, body mass index, grade point average, sedentary hours, and age. The study proposes that these findings could be utilized to offer cost-effective support and enhance mental well-being assessment and monitoring at both the individual and university levels. Baba et al. [8] developed a machine learning model to predict students' mental health problems using health survey data and response time metrics. The LightGBM model was found to be the most effective, with high predictive performance (AUC=0.857). Responses to questions about campus life were key predictors of mental health issues based on the SHAP analysis. While the inclusion of response time-related variables did not significantly improve predictions, certain derived variables based on response times enhanced prediction accuracy. The findings suggest the potential of using machine learning to predict mental health issues over time and highlight the importance of incorporating behavioral data in mental health assessments. Meda et al. [9] focused on the mental health of university students, revealing high levels of severe depressive symptoms and suicidal ideation. Economic worry was associated with depression, and demographic factors were found to be poor predictors of mental health outcomes. The random forest algorithm showed high accuracy in predicting students maintaining

well-being (Accuracy: 0.85) but had limitations in predicting symptom worsening (Accuracy: 0.49). Anbarasi et al. [14] used the RF to established a model to assess quality of life, and the study found a positive correlation between sleep quality and anxiety levels. Students were identified as being highly susceptible to mental health disturbances during the COVID-19 pandemic, particularly due to factors such as online learning challenges, parental involvement, and workload stress. In the present study, we also found that quality of sleep was an important contributor to severe mental distress, as it ranked first based on feature importance analysis. Ratul et al. [15] developed a reliable machine learning-based prediction model using the multilayer perceptron algorithm for perceived stress and achieved high accuracy (Accuracy: 0.805), precision, F1 score, and recall values. However, the convenience sampling technique used in the study may have biased results and lack generalizability. In addition, Rois et al. [16] used advanced machine learning approaches to predict the prevalence of stress among Bangladeshi university students ($n=355$), and identified important risk factors for stress, including pulse rate, blood pressure, sleep and smoking status, and academic background. The RF model showed the highest performance in predicting stress (AUC: 0.897), outperforming logistic regression and support vector machine models. The outcome indicator of our study was severe mental distress, while the outcome indicators in the above other studies included negative mental well-being traits, mental health problems, severe depressive symptoms, suicidal ideation, and perceived stress. Although the model variables differ, they generally cover similar aspects such as exercise habits and sleep habits. In addition, by comparing, it can be seen that although these studies all utilize machine learning and artificial intelligence techniques

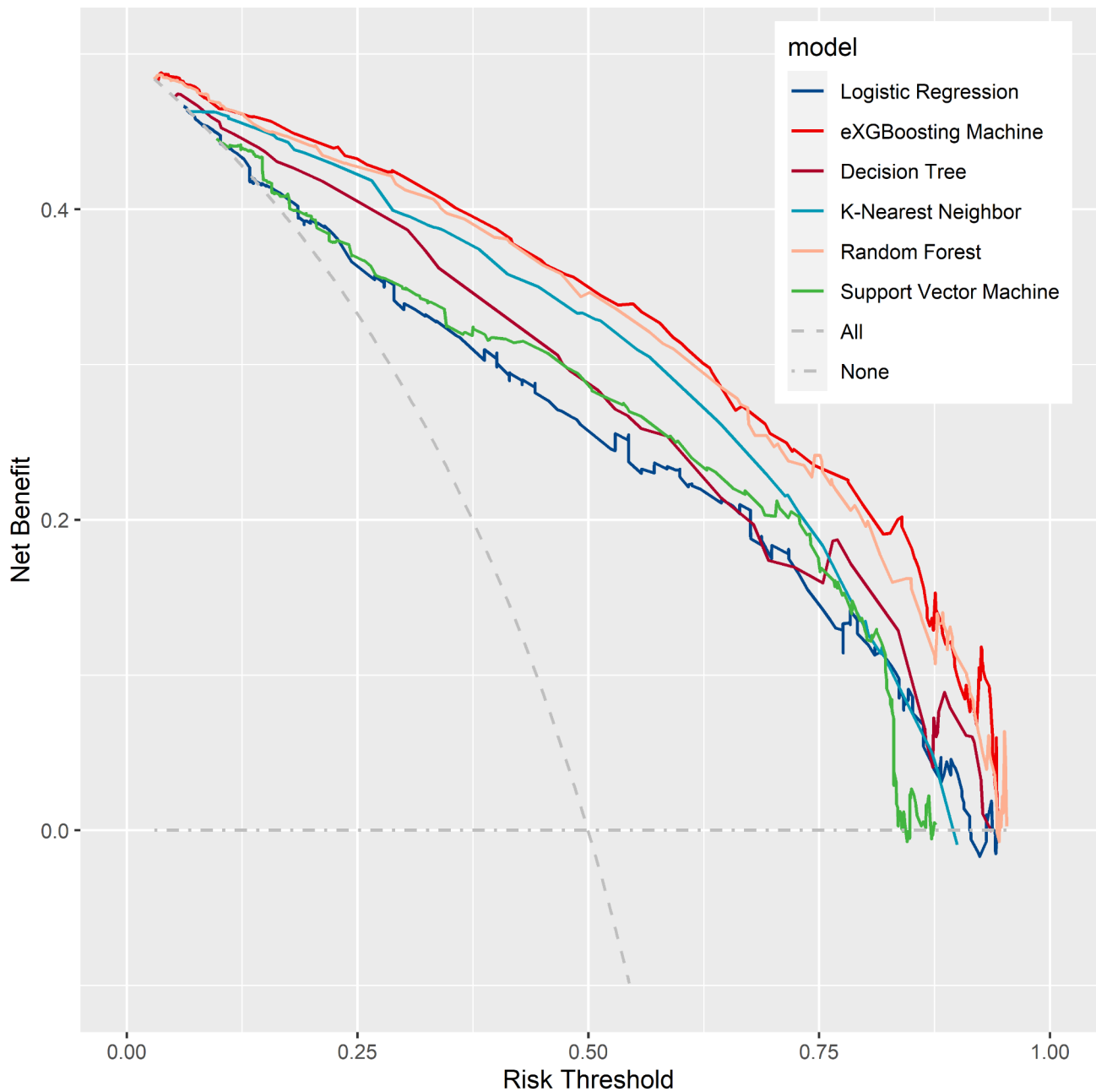


Fig. 7 Decision curve analysis for each model

to predict and assess the mental health status of college students, their research content and focus are different, exploring and studying different aspects of mental health issues.

Individualized intervention under AI guidance

For university students identified as high-risk individuals with severe mental distress, a comprehensive management approach is imperative to address their specific needs. Firstly, a multidisciplinary team comprising mental health professionals, counselors, and medical

practitioners should be involved in their care. This team can collaborate to develop personalized treatment plans tailored to the individual’s condition. Intensive therapy sessions, such as cognitive-behavioral therapy [32] or dialectical behavior therapy [33], can be implemented to help these students develop coping mechanisms and improve their emotional well-being. Additionally, pharmacological interventions, under the guidance of a psychiatrist [34], may be considered to alleviate symptoms and stabilize their mental health. Regular follow-up appointments and close monitoring of their progress are



Fig. 8 Heatmap to comprehensively present the prediction performance for each model. LR, logistic regression; eXGBM, extreme gradient boosting machine; DT, decision tree; KNN, K-nearest neighbor; RF, random forest; SVM, support vector machine

crucial to ensure the effectiveness of the management plan. It is crucial to acknowledge that although the AI application offers risk estimates and recommendations, clinical decision-making should encompass the expertise of healthcare providers and take into account the unique context of each student.

University students identified as low-risk individuals in terms of severe anxiety and depression require a different approach to management. While their mental health concerns may be less severe, proactive measures should still be taken to promote their overall well-being and prevent the development of more significant issues. One key aspect is the provision of mental health education and exercise or mindfulness-based programs on campus [35, 36]. These initiatives can help students recognize the signs of mental distress and equip them with self-help strategies to manage mental distress and maintain good mental health. Additionally, establishing a supportive environment through peer support groups or mentoring programs can foster a sense of belonging and provide a platform for students to share their experiences and seek guidance [37]. By implementing these preventive measures, the university can create a nurturing

environment that supports the mental well-being of all students, including those at low risk for severe mental distress. It is crucial to acknowledge that although the AI application offers risk estimates and recommendations, clinical decision-making should encompass the expertise of healthcare providers and take into account the unique context of each student. Notably, a comprehensive support mechanism was implemented during the study. This included providing participants with access to mental health professionals, offering counseling services, and ensuring that participants were informed about these resources prior to their involvement. Besides, this study established a clear protocol for managing distress during and after participation, ensuring participants had immediate support if needed.

Limitations

There are several limitations in this study. Firstly, the sample size was limited to 2088 college students from five universities, which may constrain the generalizability of the model. The results may not be applicable to populations with different culture in other nations. Secondly, although a favorable AUC value was achieved in

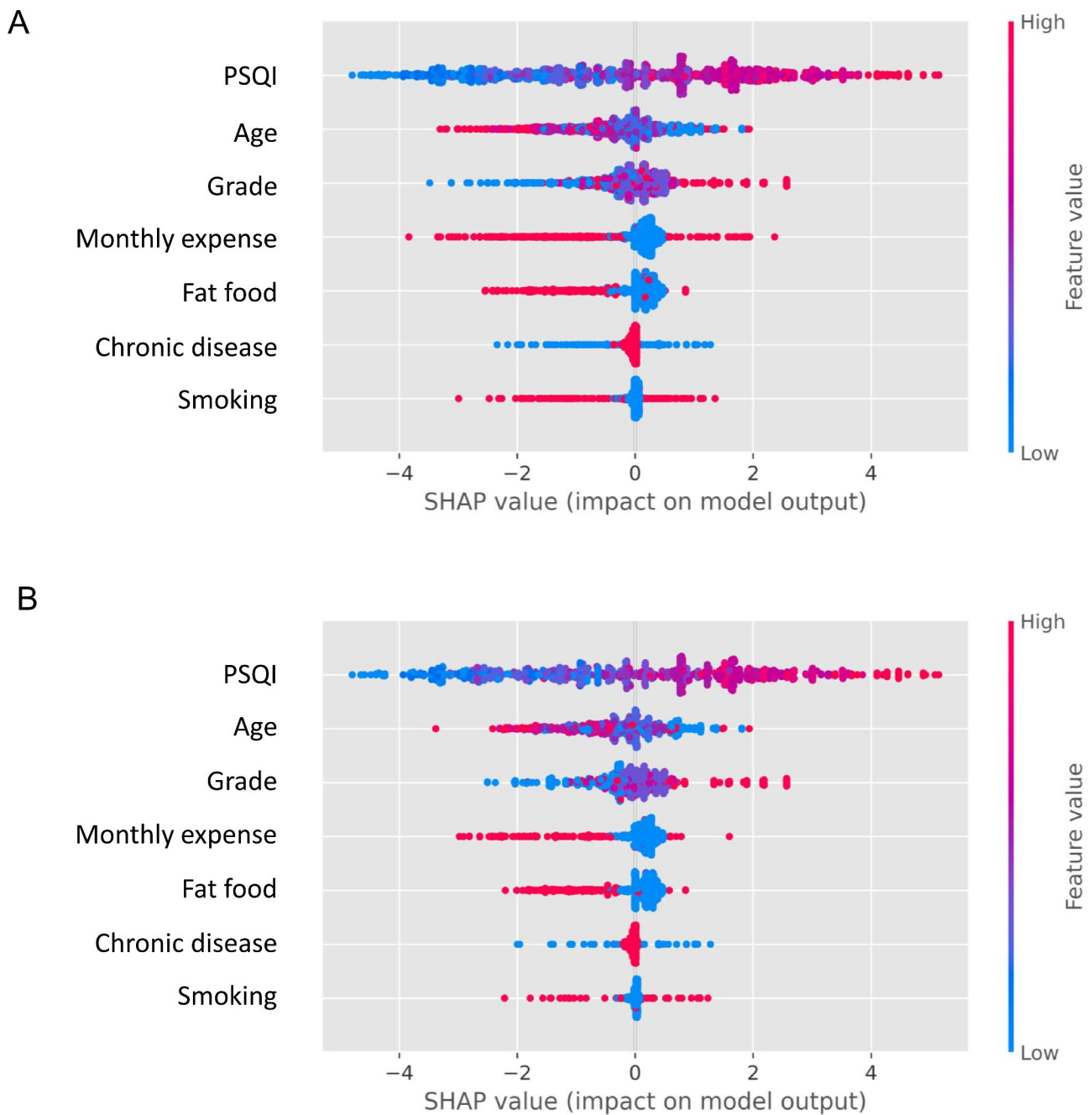


Fig. 9 Feature importance analysis using the SHAP method. (A) Training group; (B) Validation group

external validation, further extensive external validation is needed to ensure the robustness and reliability of the model. Additionally, while the machine learning models used in the study performed well on the training set, they may be influenced by data quality and feature selection in real-world applications, necessitating further optimization and improvement. What's more, some of the important confounding variables, such as social support, academic stress, financial stress, interpersonal relationships and exposure to digital media, were not included

for analysis. Incorporating these factors into the prediction model might further improve the prediction performance and impact of the model. Lastly, while the AI tool showed promising performance in predicting severe mental distress in college students, mental health issues are complex and diverse, and a simple prediction may not comprehensively assess an individual's mental health status. Therefore, a comprehensive evaluation and intervention combining other factors are still required. Further

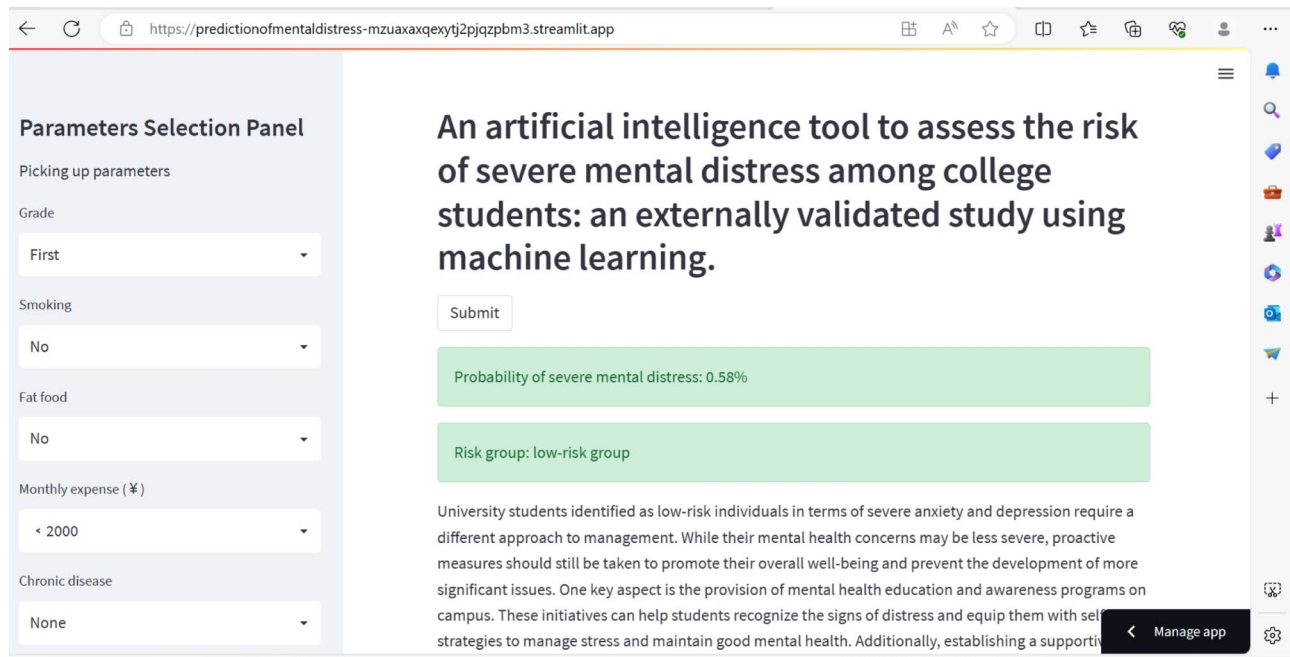


Fig. 10 The AI application to predict the risk of severe mental distress among university students

in-depth research and improvement are needed before applying the AI tool in practical clinical practice.

Conclusions

In conclusion, the developed AI tool demonstrates promising predictive performance for identifying college students at risk of severe mental distress. Its high accuracy and reliability highlight its potential to guide intervention strategies and support early identification and preventive measures. The tool's accessibility and ability to provide personalized recommendations make it a valuable resource for improving mental health outcomes among college students.

Abbreviations

AI	artificial intelligence
AUC	Area under the curve
CI	Confident interval
GAD-7	Generalized anxiety disorder-7
PHQ-9	Patient health questionnaire
PSQI	Pittsburgh Sleep Quality Index
SD	Standard deviation

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12888-024-06017-2>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

Not applicable.

Author contributions

LZ, SZ, ZY, HZ, and ML conceived and designed this study together. ML and LZ undertook the data analysis, results interpretation and manuscript preparation. LZ, SZ, and HZ performed supervision. All authors read and approved the final manuscript.

Funding

This study was funded by the Chongqing Social Science Planning Project (No.2023NDYB110) and Ministry of Education's Industry-Education Collaboration and Synergy Education Project (No.202102100024), and this study was also funded by the Study of the Mechanism and Pathways of Artificial Intelligence Empowering Sports to Prevent and Control Myopia among Chinese Adolescents (No. SKHF24029) and Military Training Injury Prevention and Control Special Project (No. 21XLS38).

Data availability

The data are available under reasonable request to the corresponding author.

Declarations

Ethics approval and consent to participant

The study was approved by the Academic Committee and Ethics Board of the Xiamen University of Technology, and informed consent was obtained from all subjects or legal guardians before filling the questions in the survey. Participants were all informed that their personal information was not identified and collected, and all data were anonymous. The study was abided by the Declaration of Helsinki. All experiments were performed in accordance with relevant guidelines and regulations.

Consent to participate

The informed consent was obtained from all subjects or legal guardians before filling the questions in the survey. Participants were all informed that their personal information was not identified and collected, and all data were anonymous.

Consent for publication

Not Applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Physical Education, Xiamen University of Technology, No. 600, Ligong Road, Jimei District, Xiamen, Fujian 361024, People's Republic of China

²School of Physical Education, Guizhou Normal University, Guizhou 550025, People's Republic of China

³College of Physical Education and Health Sciences, Chongqing Normal University, No. 37, Middle Road, University Town, Shapingba District, Chongqing 401331, People's Republic of China

⁴National Clinical Research Center for Orthopedics, Sports Medicine & Rehabilitation, Beijing 100039, People's Republic of China

⁵Department of Orthopedic Surgery, Hainan Hospital of Chinese PLA General Hospital, Beijing 100039, People's Republic of China

Received: 28 April 2024 / Accepted: 12 August 2024

Published online: 27 August 2024

References

1. Knapstad M, Sivertsen B, Knudsen AK, Smith ORF, Aarø LE, Lønning KJ, Skogen JC. Trends in self-reported psychological distress among college and university students from 2010 to 2018. *Psychol Med*. 2021;51(3):470–8.
2. Siraji A, Molla A, Ayele WM, Kebede N. Mental distress and associated factors among college students in Kemisie district, Ethiopia. *Sci Rep*. 2022;12(1):17541.
3. Dachew BA, Biftu BB, Tiruneh BT, Anlay DZ, Wassie MA. Prevalence of mental distress and associated factors among university students in Ethiopia: a meta-analysis. *J Ment Health*. 2022;31(6):851–8.
4. Stocker R, Tran T, Hammarberg K, Nguyen H, Rowe H, Fisher J. Patient Health Questionnaire 9 (PHQ-9) and general anxiety disorder 7 (GAD-7) data contributed by 13,829 respondents to a national survey about COVID-19 restrictions in Australia. *Psychiatry Res*. 2021;298:113792.
5. Jeffries V, Salzer MS. Mental health symptoms and academic achievement factors. *J Am Coll Health*. 2022;70(8):2262–5.
6. Grasdalsmoen M, Eriksen HR, Lønning KJ, Sivertsen B. Physical exercise, mental health problems, and suicide attempts in university students. *BMC Psychiatry*. 2020;20(1):175.
7. Abdul Rahman H, Kwikkis M, Ottom M, Amornsriwatanakul A, K HA-M, Rosenberg M, Dinov ID. Machine learning-based prediction of Mental Well-being using Health Behavior Data from University students. *Bioeng (Basel)* 2023, 10(5).
8. Baba A, Bunji K. Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach. *JMIR Ment Health*. 2023;10:e42420.
9. Meda N, Pardini S, Rigobello P, Visioli F, Novara C. Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students. *Epidemiol Psychiatr Sci*. 2023;32:e42.
10. Lei M, Wu B, Zhang Z, Qin Y, Cao X, Cao Y, Liu B, Su X, Liu Y. A web-based calculator to predict early death among patients with bone metastasis using machine learning techniques: Development and Validation Study. *J Med Internet Res*. 2023;25:e47590.
11. Shi X, Cui Y, Wang S, Pan Y, Wang B, Lei M. Development and validation of a web-based artificial intelligence prediction model to assess massive intra-operative blood loss for metastatic spinal disease using machine learning techniques. *Spine J*. 2024;24(1):146–60.
12. Han T, Xiong F, Sun B, Zhong L, Han Z, Lei M. Development and validation of an artificial intelligence mobile application for predicting 30-day mortality in critically ill patients with orthopaedic trauma. *Int J Med Inf*. 2024;184:105383.
13. Cui Y, Shi X, Qin Y, Wan Q, Cao X, Che X, Pan Y, Wang B, Lei M, Liu Y. Establishment and validation of an interactive artificial intelligence platform to predict postoperative ambulatory status for patients with metastatic spinal disease: a multicenter analysis. *Int J Surg* 2024.
14. Anbarasi LJ, Jawahar M, Ravi V, Cherian SM, Shreenidhi S, Sharen H. Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown. *Health Technol (Berl)*. 2022;12(4):825–38.
15. Ratul IJ, Nishat MM, Faisal F, Sultana S, Ahmed A, Al Mamun MA. Analyzing perceived psychological and Social Stress of University students: a Machine Learning Approach. *Heliyon*. 2023;9(6):e17307.
16. Rois R, Ray M, Rahman A, Roy SK. Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. *J Health Popul Nutr*. 2021;40(1):50.
17. Zhang L, Zhao S, Yang Z, Zheng H, Lei M. An Artificial Intelligence Platform to Stratify the risk of experiencing sleep disturbance in University Students after Analyzing Psychological Health, Lifestyle, and sports: a Multicenter externally validated study. *Psychol Res Behav Manag*. 2024;17:1057–71.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD). *Ann Intern Med*. 2015;162(10):735–6.
19. Hu B, Wu Q, Wang Y, Zhou H, Yin D. Factors associated with sleep disorders among university students in Jiangsu Province: a cross-sectional study. *Front Psychiatry* 2024, 15.
20. Guo S, Sun W, Liu C, Wu S. Structural validity of the Pittsburgh Sleep Quality Index in Chinese Undergraduate Students. *Front Psychol*. 2016;7:1126.
21. Zhou Y, Xu J, Rief W. Are comparisons of mental disorders between Chinese and German students possible? An examination of measurement invariance for the PHQ-15, PHQ-9 and GAD-7. *BMC Psychiatry*. 2020;20(1):480.
22. Eleftheriou A, Rokou A, Arvaniti A, Nena E, Steiropoulos P. Sleep Quality and Mental Health of Medical Students in Greece during the COVID-19 pandemic. *Front Public Health* 2021, 9.
23. Zhang C, Wang T, Zeng P, Zhao M, Zhang G, Zhai S, Meng L, Wang Y, Liu D. Reliability, validity, and Measurement Invariance of the General anxiety disorder scale among Chinese Medical University students. *Front Psychiatry*. 2021;12:648755.
24. Du N, Yu K, Ye Y, Chen S. Validity study of Patient Health Questionnaire-9 items for internet screening in depression among Chinese university students. *Asia Pac Psychiatry* 2017, 9(3).
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16(1):321–57.
26. Zhu C, Xu Z, Gu Y, Zheng S, Sun X, Cao J, Song B, Jin J, Liu Y, Wen X, et al. Prediction of post-stroke urinary tract infection risk in immobile patients using machine learning: an observational cohort study. *J Hosp Infect*. 2022;122:96–107.
27. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, Xu S, Stub D, Smith K, Tacey M et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: a retrospective international registry study. *PLoS Med* 2018, 15(11).
28. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38.
29. Cabot JH, Ross EG. Evaluating prediction model performance. *Surgery*. 2023;174(3):723–6.
30. Wang H, Wu W, Han C, Zheng J, Cai X, Chang S, Shi J, Xu N, Ai Z. Prediction model of osteonecrosis of the femoral Head after femoral Neck fracture: machine learning-based development and validation study. *JMIR Med Inf*. 2021;9(11):e30079.
31. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OK. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24(1):478.
32. Lee SH, Cho SJ. Cognitive behavioral therapy and mindfulness-based cognitive therapy for Depressive disorders. *Adv Exp Med Biol*. 2021;1305:295–310.
33. Kothgassner OD, Goreis A, Robinson K, Huscsava MM, Schmahl C, Plener PL. Efficacy of dialectical behavior therapy for adolescent self-harm and suicidal ideation: a systematic review and meta-analysis. *Psychol Med*. 2021;51(7):1057–67.
34. Marwaha S, Palmer E, Suppes T, Cons E, Young AH, Upthegrove R. Novel and emerging treatments for major depression. *Lancet*. 2023;401(10371):141–53.
35. Herbert C. Enhancing Mental Health, well-being and active lifestyles of University students by means of physical activity and Exercise Research Programs. *Front Public Health*. 2022;10:849093.
36. Galante J, Dufour G, Vainre M, Wagner AP, Stochl J, Benton A, Lathia N, Howarth E, Jones PB. A mindfulness-based intervention to increase resilience to stress in university students (the Mindful Student Study): a pragmatic randomised controlled trial. *Lancet Public Health*. 2018;3(2):e72–81.

37. Ning X, Wong JP, Huang S, Fu Y, Gong X, Zhang L, Hilario C, Fung KP, Yu M, Poon MK et al. Chinese University Students' Perspectives on Help-Seeking and Mental Health Counseling. *Int J Environ Res Public Health* 2022, 19(14).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.