SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(3), 1477.
https://doi.org/10.62617/mcb1477

Article

College student mental health assessment: Predictive models based on machine learning and feature importance analysis

Cheng Liu[*]**, Lin Ji, Juncheng Lu, Jiaze Ma, Xinyuan Sui**

Department of Digital Business, Jiangsu Vocational Institute of Commerce, Nanjing 211168, China
*** Corresponding author:** Cheng Liu, 200022@jvic.edu.cn

CITATION

Liu C, Ji L, Lu J, et al. College student mental health assessment: Predictive models based on machine learning and feature importance analysis. Molecular & Cellular Biomechanics. 2025; 22(3): 1477. https://doi.org/10.62617/mcb1477

ARTICLE INFO

Received: 2 February 2025
Accepted: 13 February 2025
Available online: 21 February 2025

Abstract: To assess and forecast the mental health conditions of university students utilizing machine learning methodologies, focusing particularly on the influence of the nine psychological symptom dimensions encompassed by the Symptom Checklist-90 (SCL-90). The prevalence of mental health issues among college students is a significant concern. Traditional methods for assessing mental health may lack the precision required for early detection and intervention. Machine learning offers advanced tools to analyze complex data and predict outcomes based on multiple variables. The primary objective is to construct and evaluate predictive models for the mental health status of college students using various machine learning algorithms, optimize their performance, and identify the most impactful psychological symptom dimensions. Psychological health data from 11,943 college students were gathered via an online questionnaire platform. Multiple machine learning algorithms were utilized to develop predictive models. Hyperparameter optimization was achieved through K-fold cross-validation and the northern goshawk optimization algorithm. To tackle class imbalance, the synthetic minority over-sampling technique was employed to create synthetic samples for underrepresented classes. Model performance was assessed using metrics such as accuracy, recall, and f1 score. The light gradient boosting algorithm demonstrated superior performance, with only 6 misclassifications out of 2,388 test samples. Tree-based ensemble methods like random forest and extreme gradient boosting consistently outperformed non-ensemble methods such as k-nearest neighbors, multi-layer perceptron, and kernel discriminant analysis. A detailed analysis using Shapley additive explanations values indicated that features such as obsessive-compulsive symptoms and anxiety were the most influential in the model’s predictions. This study underscores the efficacy and potential of machine learning in mental health assessment. The results provide a robust scientific foundation for the development of early warning systems and targeted intervention strategies to enhance the mental well-being of college students.

Keywords: mental health assessment; light gradient boosting; machine learning; k-fold cross-validation; shapley additive explanations; northern goshawk optimization

1. Introduction

The mental health issues of college students are becoming increasingly prominent with the intensification of competition and the continuous increase of social pressure in modern society [1]. The transition from high school to college signifies not only a change in the learning environment but also the introduction of new social circles, increased academic pressure, and uncertainties about the future. Particularly for college students in China, the pressures of family expectations and social competition are more significant, making them more susceptible to mental health issues.

Numerous studies have indicated that students' mental health is influenced by a variety of factors, such as individual factors [2], environmental factors [3], and sociocultural factors [4]. Regarding environmental factors, aspects like campus atmosphere, teacher-student relationships, and peer relationships also significantly impact students' mental health [5]. On the social front, social support systems have a notable effect on college students' mental health. The lack of social support may increase the risk of psychological issues. Additionally, academic stress and economic factors [6] are also significant elements affecting college students' mental health. Mental health not only affects students' academic performance and personal development but may also have profound implications for their future social adaptability and quality of life.

Traditional mental health assessment methods typically rely on face-to-face interviews and questionnaires conducted by professional psychologists, which, while accurate, are time-consuming and difficult to implement on a large scale. And, they rely heavily on the subjective judgment of professionals, which may lead to inconsistent results. In contrast, machine learning methods can analyze a large amount of data objectively and quickly, potentially providing more precise and timely assessment results for early detection and intervention among college students. In recent years, with the advancement of machine learning [7] and artificial intelligence technologies, an increasing number of studies have begun to explore the use of machine learning methods to assess and predict mental health conditions. Machine learning models can automatically extract features from vast amounts of data, uncover underlying patterns, and thus providing more efficient and accurate assessment tools.

This study aims to utilize machine learning technology, in conjunction with the symptom checklist-90 (SCL-90) scale, to construct a model capable of effectively assessing and predicting the mental health status of college students. The SCL-90 scale is a widely used self-reporting questionnaire in the field of psychology, covering nine distinct psychological symptom dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, paranoid ideation, psychoticism, and others [8]. By collecting mental health data from college students through online questionnaires and employing various machine learning algorithms for model training and evaluation, we address the issue of class imbalance in the dataset using synthetic minority over-sampling technique (SMOTE) technology for data augmentation.

Through this research, we provide a new and effective tool for the assessment of college students' mental health, offering a scientific basis for early warning and intervention. The innovation of this study lies in the integration of machine learning technology with mental health assessment, proposing a novel evaluation method. By analyzing feature importance, this study reveals the key factors affecting college students' mental health, offering a new perspective for mental health education and intervention. Additionally, the study demonstrates the potential of machine learning models in handling imbalanced data and improving prediction accuracy. Specifically, the main objectives of this study include:

- Collecting and analyzing mental health data from college students, assessing the distribution characteristics of various psychological symptoms.

- Constructing multiple machine learning models, evaluating their performance in mental health prediction tasks, and using the northern goshawk optimization (NGO) algorithm to optimize algorithm hyperparameters, enhancing model performance.
- Utilizing shapley additive explanations (SHAP) value analysis to explore the significance and influence mechanisms of different features on model prediction outcomes.

The structure of this paper is as follows: the second section reviews relevant literature, discussing the current state of mental health assessment and the application of machine learning in this field. The third section provides a detailed introduction to the research methodology, including data process, model construction, and training. Section four presents the experimental results and conducts an in-depth analysis. Section five summarizes the study and proposes directions for future research.

2. Related literature

The factors influencing college students' mental health are multifaceted, encompassing personal, family, academic, and societal dimensions. Intense academic pressure is a common source of mental health issues among college students. As the academic load increases, many students may experience anxiety, depression, and excessive stress. Academic challenges, exam stress, and concerns about future careers can all contribute to the occurrence of mental health problems [9]. Family factors, such as socioeconomic status, parenting styles, and interpersonal relationships, also have a profound impact on college students' mental health [10]. Additionally, behaviors like alcohol abuse and internet addiction can increase students' psychological stress [11]. Personality traits, including introversion, loneliness, and emotional instability, are closely related to mental health issues among college students [12].

Machine Learning (ML), as an emerging technology, shows great potential in the field of mental health assessment. It can analyze vast amounts of data to help identify mental health issues, predict disease progression, and assist in clinical decision-making [13]. In detection and diagnosis, by analyzing patients' social data, ML models can aid in diagnosing mental health issues such as depression and anxiety [14]. In terms of prognosis, treatment, and support, ML can predict patients' responses to specific treatment plans, providing support for personalized therapy [15]. In the literature [16], researchers utilized machine learning models to improve the diagnostic accuracy of depression. The results showed that machine learning models could identify depressive symptoms more accurately. Despite challenges like data privacy, model interpretability, and algorithm generalization capabilities, the application of ML in mental health assessment is promising [17].

To enhance model accuracy, scholars have conducted research on feature selection and model optimization. Feature selection is a key step to improve model performance and reduce computational costs. The literature [18] provides a comprehensive review of feature selection techniques, including filter methods, wrapper methods, and embedded methods, which have been widely applied in areas like text classification. Hyperparameter optimization, such as particle swarm optimization and genetic algorithms, can perform the selection of optimal hyperparameter values in an autonomous manner [19]. The issue of data imbalance is prevalent in many practical applications, leading models to bias towards the majority class. The SMOTE method, a synthetic minority oversampling technique, is used to address data imbalance issues [20]. Furthermore, the literature [21] discusses the application of evolutionary computation in hyperparameter optimization, which simulates the process of natural selection to find the optimal hyperparameter settings. Concurrently, to improve model interpretability, scholars have proposed the SHAP method, offering fair and consistent explanations of feature importance for ML models. SHAP value analysis, as a method to interpret model predictions [22], can help researchers understand how models make decisions, which is crucial for enhancing model credibility and acceptance.

Although existing studies have made progress in using ML models to predict mental health status, most have not fully considered the model's interpretability. In addition, compared to traditional grid search and random search, incorporating intelligent optimization algorithms into ML models can more efficiently find the optimal hyperparameter combinations and enhance model performance. By integrating ML predictive models with SHAP value analysis, not only is predictive accuracy improved, but the transparency and interpretability of the models are also enhanced.

3. Methodology

3.1. Machine learning process for college student mental health assessment

The machine learning process for college student mental health assessment is a systematic approach designed to identify and address mental health issues among students using advanced computational techniques. The literature [23] used a simplified version of the SCL-90 to reduce the response burden on patients, and employed machine learning algorithms to build a classification model for differentiating between depression and anxiety. The literature [24] proposed an intelligent perception method for psychological stress among college students. It combines the SCL-90 with other factors (such as sleep, exercise, etc.), and after feature extraction, uses machine learning techniques to establish an evaluation model. In this process, we need to complete the collection and processing of data, as well as the training and optimization of the model. We collected data using a questionnaire, specifically the SCL-90 scale, and calculated the average scores for the nine dimensions of the scale. To address the imbalance in the data, we employed the SMOTE algorithm, and used the K-fold method to split the dataset into training and testing sets. We optimized the parameters with the northern goshawk optimization algorithm and made predictions, and finally, we interpreted the model using SHAP theory. The workflow is as follows:

- Data Collection, utilize an online platform to gather data through questionnaires, organize the results, remove invalid data, and ensure uniform data formatting.
- Feature Calculation, compute the average scores for the nine dimensions, standardize the data for each dimension, and then calculate the average score for

each sample in those dimensions. Use the SMOTE algorithm to generate synthetic samples for the minority class to balance the dataset.

- Apply the K-fold cross-validation method to divide the dataset into training and testing sets, splitting the data into K portions. For each iteration, use K-1 portions for training and the remaining portion for testing, repeating this process K times.
- Optimize the machine learning algorithms with the NGO algorithm, adjusting the model parameters. Identify the parameters that needed optimization within each classification algorithm and their value ranges, using the NGO algorithm to search the parameter space for the optimal combination. Assess the performance of the best parameter combination using cross-validation results.
- Make predictions using the test sets and output various metrics for each model, including accuracy, recall, and f1 score.
- Model interpretation, use SHAP theory to explain the model, calculating the contribution of each feature to the model's output with the SHAP library. Based on the SHAP interpretation results, analysis which features were most important for the model's predictions and how they influenced the outcomes.

3.2. SMOTE data augmentation

SMOTE is an oversampling method used to address imbalanced datasets [25]. Its principle involves generating new synthetic samples by interpolating between minority class samples to increase the number of minority class instances. The process of the SMOTE algorithm is as below:

Randomly select a sample from the minority class x, and select k samples from the majority class$\{x_1,x_2,...,x_k\}$; Calculate the distance between the minority class sample x and each majority class sample $x_i(\mathrm{i} = 1,2,\ldots,\mathrm{k})$. The Euclidean distance is commonly used as a measure of distance, defined as below:

$$\mathrm{d}(x,x_i) = \sqrt{\sum_{j=1}^{n}(x_j - x_{i,j})^2} \tag{1}$$

x_j is the *j*-th feature of sample x, $x_{i,j}$ is the *j*-th feature of sample x_i, and n is the total number of features.

Then, randomly select a point between x and each x_i to generate a new synthetic sample x'. This process can be represented as:

$$x' = x + \frac{r}{k}\sum_{i=1}^{k}(x_i - x) \tag{2}$$

r is a random number within the range [0,1], and k is set to 5.

Repeat the above steps until the number of minority class samples reaches the predetermined oversampling ratio. Merge the generated synthetic samples with the original dataset to form a new balanced dataset.

The advantage of SMOTE lies in the fact that it not only increases the number of minority class samples but also generates new samples through interpolation between minority class samples, which helps to maintain the diversity of minority class samples and thus enhances the generalization ability of the model. However, SMOTE may also introduce noise because it generates synthetic samples rather than real ones. Therefore, when applying SMOTE, it is necessary to make appropriate adjustments and validations based on the specific problem and the characteristics of the dataset.

3.3. K-fold cross-validation data division

K-fold cross-validation is a commonly used model evaluation technique widely applied in machine learning and statistics [26]. Its main purpose is to better assess model performance and avoid overfitting and underfitting.

K-fold cross-validation involves dividing the dataset into k equal-sized subsets and then conducting K rounds of training and testing. In each round, one subset is used as the test set, while the remaining the k-1 subsets serve as the training set. This way, each subset gets a chance to be the test set once. The final model performance is the average of the k test results, as shown in **Figure 1**.

K-fold cross-validation provides a more comprehensive assessment of the model's generalization capability by repeatedly training and testing the model on different data subsets. In K-fold cross-validation, each training and testing is conducted on a different subset of data, ensuring no overlap between training and test data. This helps prevent the model from learning noise or outliers specific to the training data, thereby reducing overfitting. The final performance estimate is based on the average of all these assessments. This averaging process reduces the chance of performance fluctuations due to specific characteristics of the dataset, providing a more robust performance estimate, in this paper, the value of K in K-fold cross-validation is set to 5.

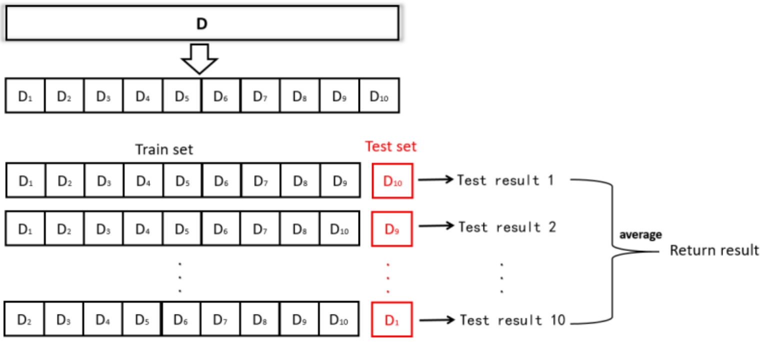
D
D₁ D₂ D₃ D₄ D₅ D₆ D₇ D₈ D₉ D₁₀
Train set
Test set
D₁ D₂ D₃ D₄ D₅ D₆ D₇ D₈ D₉ D₁₀
Test result 1
D₁ D₂ D₃ D₄ D₅ D₆ D₇ D₈ D₁₀ D₉
Test result 2
D₂ D₃ D₄ D₅ D₆ D₇ D₈ D₉ D₁₀ D₁
Test result 10
average
Return result

Figure 1. The principle of K-fold cross-validation.

3.4. NGO algorithm

The NGO algorithm is an optimization algorithm inspired by the predatory behavior of the northern goshawk [27]. It simulates the prey identification and attack, pursuit, and escape behaviors of goshawks during hunting.

During the prey identification phase, the northern goshawk randomly selects prey and then swiftly attacks it. Since the selection of prey in the search space is random, this phase enhances the exploration capability of NGO. This phase is a global search aimed at identifying the optimal area, and the model is as follows:

$$P_i = X_k, i = 1,2,\ldots,\mathrm{N} \tag{3}$$

$$x_{i,j}^{new,P1} = \begin{cases} x_{i,j} + r\left(p_{i,j} - x_{i,j}\right), F_{P_i} < F_i \\ x_{i,j} + r\left(x_{i,j} - x_{i,j}\right), F_i \le F_{P_i} \end{cases} \tag{4}$$

$$X_i = \begin{cases} X_i^{new,P1}, & F_i^{new,P1} < F_i \\ X_i, & F_i \le F_i^{new,P1} \end{cases} \tag{5}$$

here, P_i represents the position of the prey selected by the i-th northern goshawk, and F_{P_i} denotes the objective function value corresponding to that prey position. In the exploration phase of the algorithm, k serves as a randomly selected natural number that helps the algorithm to make random jumps among candidate solutions, enhancing the diversity and breadth of the search. $X_i^{new,P1}$ represents the new state of the i-th individual after being updated based on the search strategy of the first phase, while $x_{i,j}^{new,P1}$ is the specific manifestation of this new state in the *j*-th dimension. $F_i^{new,P1}$ is the objective function value of the solution calculated according to the exploration process of the first phase, reflecting the fitness of the model.

After the northern goshawk attacks its prey, the prey will attempt to escape. Therefore, during the pursuit, the northern goshawk continuously chases the prey. Due to the extremely high speed of the northern goshawk, they can almost chase and eventually capture the prey under any circumstances. Simulating this behavior enhances the algorithm's ability to explore the local search space.

In the proposed NGO algorithm, it is assumed that this hunting takes place within a range centered on the attack position with a radius R, and the model is as follows:

$$x_{i,j}^{new,P2} = x_{i,j} + R(2r - 1)x_{i,j} \tag{6}$$

$$R = 0.02\left(1 - \frac{t}{T}\right) \tag{7}$$

$$X_i = \begin{cases} X_i^{new,P2}, F_i^{new,P2} < F_i \\ X_i, F_i \le F_i^{new,P2} \end{cases} \tag{8}$$

In which t is the counter for the number of iterations, T is the maximum number of iterations, $X_i^{new,P2}$ is the new state of the i-th proposed solution, $x_{i,j}^{new,P2}$ represents the *j*-th dimension of that state, and $F_i^{new,P2}$ is the objective function value based on the second phase of the NGO algorithm.

At the beginning of the NGO algorithm, a series of parameters need to be set, including the population size, the initial learning rate, and the maximum number of iterations, to ensure that the algorithm does not run indefinitely. In each iteration, it is necessary to calculate the fitness of each individual. For the mental health assessment model in this study, the fitness function is based on the accuracy of the model on the test set. By calculating the fitness, the algorithm can evaluate the advantages and disadvantages of each individual in the current search state, thus guiding the subsequent search direction.

3.5. Light gradient boosting algorithm

The light gradient boosting (LGB) algorithm is a machine learning method based on the gradient boosting framework [28]. The algorithm iteratively builds multiple weak learners such as decision trees and combines them to form a powerful model. Gradient boosting, as an ensemble learning method, iteratively adds new models to correct the errors of existing models and thus improves the overall prediction performance.

In each step, the new model tries to minimize the loss function of all the models combined from the previous step. For a given dataset $\mathrm{D} = \{(x_i, y_i)\}_{i=1}^{n}$, where x_i is the feature vector and y_i is the target variable,the goal of the gradient boosting algorithm is to minimize the loss function in the following form:

$$\mathrm{L}(\theta) = \sum_{i=1}^{n} l\left(y_{i,}\, F(x_i)\right) \tag{9}$$

In which l stands for the loss function, typically using mean squared error or logarithmic loss, and $F(x_i)$ represents the current predictive model.

To enhance the efficiency of selecting feature split points and reduce computational complexity, LGB introduces a histogram algorithm to optimize the search for the best feature split points. The core idea is to discretize continuous floating-point feature values into a fixed number of k integer values and construct a histogram with a width of k. When iterating over the data of a certain feature, the discretized values serve as indices to accumulate statistics into the histogram. After a single complete pass through the data, the histogram has accumulated the necessary statistical information, allowing for the efficient determination of the optimal split point based on these aggregated statistics. Furthermore, to further reduce the volume of samples processed, LGB removes samples with smaller weights that contribute less to the model during training, focusing only on calculating the information gain for the remaining significant samples. This strategy not only reduces computational costs but also accelerates the speed of model training. Additionally, LGB employs a method of bundling mutually exclusive features, reducing the number of features while preserving the integrity of the original feature information. This approach not only speeds up the training process but also helps prevent overfitting issues that can arise from an excess of features.

LGB also adopts a leaf-wise growth strategy with depth limits, diverging from traditional level-wise growth methods. This strategy allows the algorithm to prioritize splitting leaves that could potentially bring the greatest gain, rather than expanding layer by layer in sequence. In this way, LGB can generate deeper yet more precise tree structures until it reaches the preset maximum depth or there is no more significant gain. This method typically results in more compact and efficient decision trees, thereby enhancing the overall performance and prediction accuracy of the model.

3.6. SHAP method

SHAP is a method used to explain the output of machine learning models [29]. It is based on the concept of shapley values from game theory, which are used to fairly distribute the gains or costs among the members of a coalition in a cooperative game.

The sum of all feature contributions equals the difference between the model's predicted value and the baseline value. If two features contribute equally to the prediction, their SHAP values are also equal. Features that have no impact on the prediction have a SHAP value of zero. The formula for calculating SHAP values is as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!}\,(\,v(S \cup \{i\}) - v(S)) \tag{10}$$

In which, ϕ_i is the SHAP value of feature i, N is the set of all features, $S \subseteq N \setminus \{i\}$ is a subset of features excluding feature i, $|S|$ denotes the size of subset S, $v(S)$ is the model's output for the subset of features S, and $v(S \cup \{i\}) - v(S)$ is the marginal contribution of feature i to subset S. The weight $\frac{|S|!(|N|-|S|-1)!}{|N|!}$ measures the importance weight of feature i across all possible subsets.

In practical applications, calculating all possible combinations of feature subsets is usually infeasible due to the number of combinations grows exponentially with the number of features, so approximation methods or sampling approaches are employed to estimate SHAP values. For instance, Kernel SHAP transforms the problem into a linear regression problem to approximate the SHAP values, while tree SHAP is an efficient algorithm specifically designed for tree models, capable of accurately computing SHAP values in polynomial time.

4. Result and discussion

4.1. Data collection and exploration

The SCL-90 scale comprises 90 items, covering nine distinct symptom dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, paranoid ideation, psychoticism, and others. This study employs a quantitative scoring system to assess participants' mental health status. Participants are asked to self-evaluate the severity of a range of psychological symptoms based on personal experience. The scoring criteria for each symptom are as follows: 1 indicates no symptoms, with no impact; 2 indicates mild symptoms, occurring occasionally, with minimal impact; 3 indicates moderate symptoms, occurring regularly, with noticeable impact; 4 indicates severe symptoms, occurring frequently, with significant impact; 5 indicates extremely severe symptoms, persistently present, severely affecting daily life. The average scores for different symptoms can be calculated as shown in **Table 1**.

Table 1. SCL-90 symptom checklist: Symptom categories and corresponding items.

Symptoms	**Item Numbers**
somatization	1, 4, 12, 27, 40, 42, 48, 49, 52, 53, 56, 58
obsessive-compulsive	3, 9, 10, 28, 38, 45, 46, 51, 55, 65
interpersonal sensitivity	6, 21, 34, 36, 37, 41, 61, 69, 73
depression	5, 14, 15, 20, 22, 26, 29, 30, 31, 32, 54, 71, 79
anxiety	2, 17, 23, 33, 39, 57, 72, 78, 80, 86
hostility	11, 24, 63, 67, 74, 81
phobic Anxiety	13, 25, 47, 50, 70, 75, 82
paranoid ideation	8, 18, 43, 68, 76, 83
psychoticism	7, 16, 35, 62, 77, 84, 85, 87, 88, 90
others	19, 44, 59, 60, 64, 66, 89

This study collected mental health data from college students through an online questionnaire format. During the data collection process, we strictly adhered to ethical standards, ensuring that all participants were involved in the study on the basis of informed consent, and that all data were anonymized before processing to protect the privacy of the participants. Participants who did not complete more than 95% of the questionnaire items were also excluded from the study to ensure the integrity of the data. Regarding missing data, we identified the proportion of missing values in each variable. If the missing rate of a variable was less than 10%, we used the mean value of that variable to impute the missing data. When the missing rate of a variable exceeded 10%, we excluded this variable from the analysis to avoid potential biases. A total of 11,943 valid entries were collected, including 3509 entries from first-year students, with 538 individuals reporting mental health issues; 3818 entries from second-year students, with 765 individuals reporting mental health issues; and 4616 entries from third-year students, with 2058 individuals having mental health concerns. For further analysis, we calculated the average score for each symptom for each student and processed the average scores of the nine symptoms using a binning method to facilitate visualization and subsequent statistical analysis.

As shown in **Figure 2**, the distribution of average scores across different psychological symptoms is displayed, with each subplot corresponding to a specific symptom. The *x*-axis represents the severity levels ranging from 1 to 5, and the *y*-axis shows the frequency at each level. The bar charts indicate that many students are concentrated at lower severity levels, which close to a score of 1, suggesting that the majority of students are relatively mentally healthy. However, there are significant differences among individual symptoms, including the peak frequency and the distribution across the range of severity levels. For instance, obsessive-compulsive and interpersonal sensitivity have a notably higher proportion of moderate to severe intensity compared to other symptoms.

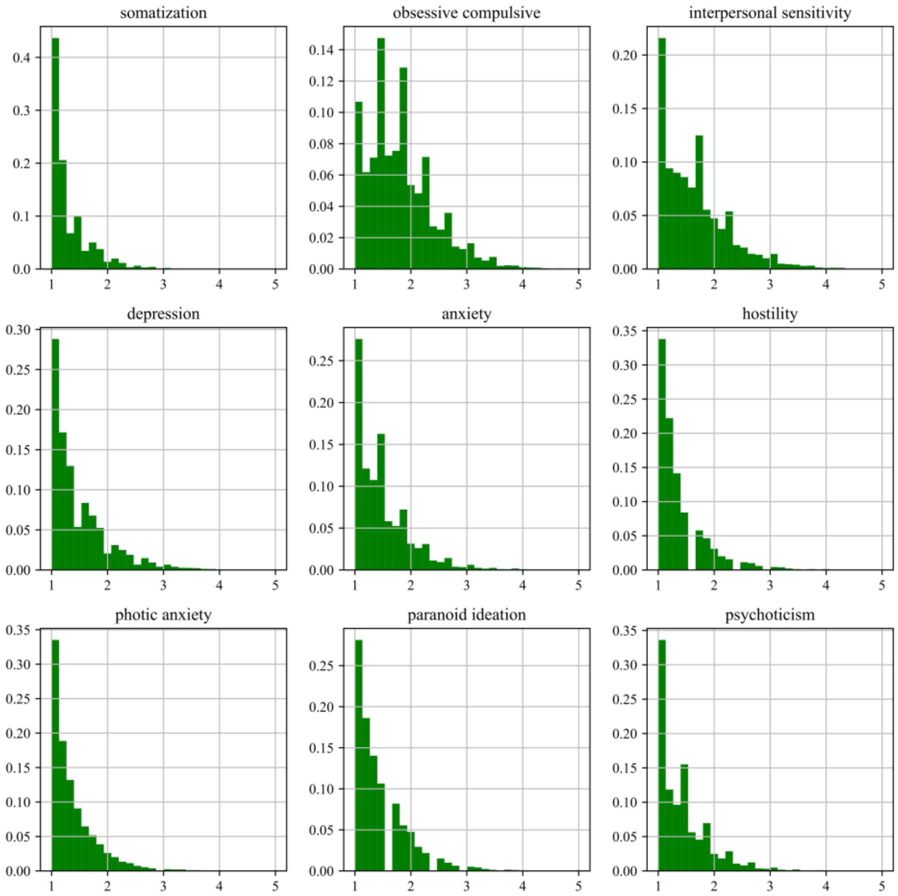

Figure 2. Comparison of distribution characteristics across nine symptom dimensions.

As depicted in **Figure 3**, each subplot illustrates a scatter plot between two dimensions along with their correlation coefficient, where the color encoding signifies the magnitude and direction of the correlation coefficient. Three asterisks denote that the correlation is statistically significant ($p < 0.01$). For instance, there is a very strong correlation between depression and anxiety ($r = 0.83$, $p < 0.01$), which likely reflects the frequent co-occurrence of depression and anxiety in mental health issues. The correlation between interpersonal sensitivity and depression is also notably strong ($r = 0.82$, $p < 0.01$), suggesting a close link between discomfort in social interactions and depressive symptoms. A high correlation is observed between paranoid ideation and psychoticism ($r = 0.79$, $p < 0.01$), indicating that these two dimensions often manifest concurrently when assessing mental health. Other dimensions, such as hostility and phobic anxiety, exhibit lower correlations with other dimensions but still demonstrate a degree of statistical significance.

These correlation results reveal the intricate relationships among mental health dimensions, which are of significant importance for understanding mental health issues among college students. This insight can guide targeted interventions and support strategies to address the multifaceted nature of mental health challenges faced by this population.

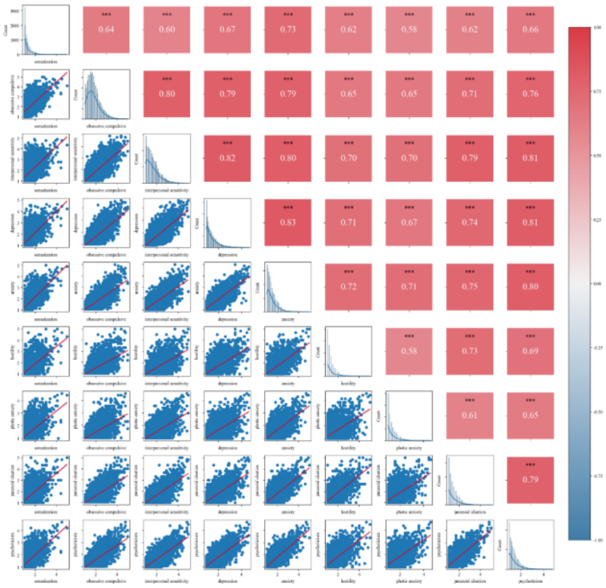

Figure 3. Analysis of correlations among nine symptom dimensions.

4.2. Machine learning-based model assessment

To comprehensively evaluate the performance of the model we proposed, this study employed three widely utilized classification metrics: accuracy, recall, and the f1 score. These metrics provide a multifaceted reflection of the model’s classification capabilities, ensuring a thorough and objective assessment.

Accuracy, measures the proportion of correctly predicted instances out of the total number of instances, offering a general overview of the model’s predictive power. The formula for its calculation is as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + FP + FN} \tag{11}$$

In which, TP (True Positives) denotes the number of samples that are actually positive and are correctly predicted as positive; TN (True Negatives) represents the number of samples that are actually negative and are correctly predicted as negative; FP (False Positives) indicates the number of samples that are actually negative but are incorrectly predicted as positive; FN (False Negatives) refers to the number of samples that are actually positive but are incorrectly predicted as negative.

Recall, indicates the ability of the model to identify all relevant instances within the dataset, crucial for understanding its effectiveness in capturing positive cases. The formula for its calculation is as follows:

$$\text{recall} = \frac{\text{TP}}{\text{TP} + FN} \tag{12}$$

F1 score, a harmonic mean of precision and recall, providing a balanced measure that considers both the false positives and false negatives, the calculation method is as follows:

$$\text{f1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

In which, precision is defined as the ratio of the number of samples correctly predicted as positive to the total number of samples predicted as positive, calculated by the formula:

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + FP} \tag{14}$$

The NGO algorithm is particularly effective in navigating the parameter space to find the most efficient combination that enhances the model's predictive accuracy. When using the NGO algorithm to optimize the model parameters, the maximum number of iterations of the model parameters is set to 500, and the specific population size is set to 30. A larger population size can usually explore the search space more comprehensively, but it will also increase the computational cost. The initial learning rate is set to 0.01. A suitable initial learning rate can balance the convergence speed and accuracy of the algorithm. By employing the NGO algorithm, we have optimized the hyperparameters of the following machine learning algorithms: Random Forest (RF), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), KNeighbors (KNN), Decision Tree (DT), eXtreme Gradient Boosting (XGB), Light Gradient Boosting (LGB), Support Vector Classifier (SVC), and Multi-layer Perceptron Classifier (MLP). The optimization algorithm and the parameters to be optimized are presented in **Table 2**. Then, these optimal parameters are then applied to the models for analysis of the test samples.

Table 2. Algorithm parameter optimization search space and optimal values.

Algorithm	Parameter	Search Space	Optimal Values
RF	n_estimators	[10, 100]	93
	max_depth	[2, 20]	15
	min_samples_split	[2, 10]	3
	min_samples_leaf	[1, 4]	2
LDA	solver	['svd', 'lsqr', 'eigen']	'svd'
QDA	reg_param	[0.01, 0.1, 1.0]	0.1
KNN	n_neighbors	[1, 20]	5
DT	max_depth	[2, 20]	10
	min_samples_split	[2,10]	3
	min_samples_leaf	[1,4]	3
XGB	max_depth	[2, 20]	12
	learning_rate	[0.01, 1]	0.16
	n_estimators	[10, 100]	82
LGB	boosting_type	['gbdt', 'dart', 'goss']	'gbdt'
	learning_rate	[0.01, 1]	0.18
	n_estimators	[10, 100]	76
SVC	n_estimators	[10, 100]	47
MLP	hidden_layer_sizes	[5,20]	8
	learning_rate	[0.01, 1]	0.14

From **Table 3**, it can be observed that almost all models experienced a significant improvement in recall after the application of SMOTE. For instance, the recall of LDA increased from 0.923 to 0.978, a notable enhancement. The recall for QDA and MLP also showed a marked improvement. This phenomenon suggests that SMOTE can significantly enhance the models' ability to recall when dealing with imbalanced datasets.

In classification tasks, the LGB and XGB algorithms performed the best on both SMOTE and non-SMOTE datasets. Previous studies, have also applied machine learning techniques to predict mental health. Vaishnavi used a stacking technique, and achieved an accuracy of 0.8175 [30]. Cheng employed random forest model, and reported an accuracy of 0.8323 [31]. In comparison, our LGB model, with its leaf-wise growth strategy and histogram based optimization, is more efficient in handling large scale datasets and can better capture non-linear relationships among features, achieved an accuracy of 0.9980 and a recall rate of 0.9983. Despite SMOTE's effectiveness in improving recall, certain algorithms, such as KNN, experienced a decrease in accuracy on balanced datasets. This indicates that SMOTE might introduce some noisy samples, which could slightly affect the overall accuracy of the classifiers.

Furthermore, the impact of SMOTE varies significantly across different types of models. Linear models like LDA and QDA benefited the most, with performance substantially enhanced across all metrics; whereas ensemble models like LGB and XGB, which already perform near optimally on imbalanced datasets, show a limited improvement with SMOTE.

Table 3. Comparison of the impact of SMOTE with different algorithms.

Algorithms	**no-smote**			**smote**		
	accuracy	**f1 score**	**recall**	**accuracy**	**f1 score**	**recall**
RF	0.9916	0.9916	0.9918	0.9954	0.9968	0.9982
LDA	0.9187	0.9191	0.9231	0.9263	0.9502	0.9784
QDA	0.8558	0.8701	0.9662	0.8853	0.9252	0.9878
KNN	0.9714	0.9711	0.9610	0.9489	0.9653	0.9901
DT	0.9875	0.9876	0.9959	0.9937	0.9956	0.9994
XGB	0.9950	0.9951	0.9959	0.9946	0.9962	0.9977
LGB	0.9966	0.9977	0.9983	0.9980	0.9980	0.9983
SVC	0.9432	0.9426	0.9324	0.9552	0.9690	0.9755
MLP	0.9196	0.9427	0.9196	0.9508	0.9497	0.9301
LR	0.9303	0.9219	0.9301	0.9321	0.9346	0.9531

The confusion matrix, as shown in **Figure 4**, illustrates that RF and LGB exhibit the best overall performance in the classification task, maintaining the lowest number of misclassified samples for both the positive and negative classes. In contrast, the recall rate of MLP is relatively low, with a higher number of false negatives, which is associated with its insufficient ability to learn the boundary features of positive samples effectively. Meanwhile, QDA and LDA exhibit a higher false positive rate in the classification of the negative class, indicating a weaker robustness to imbalanced data.

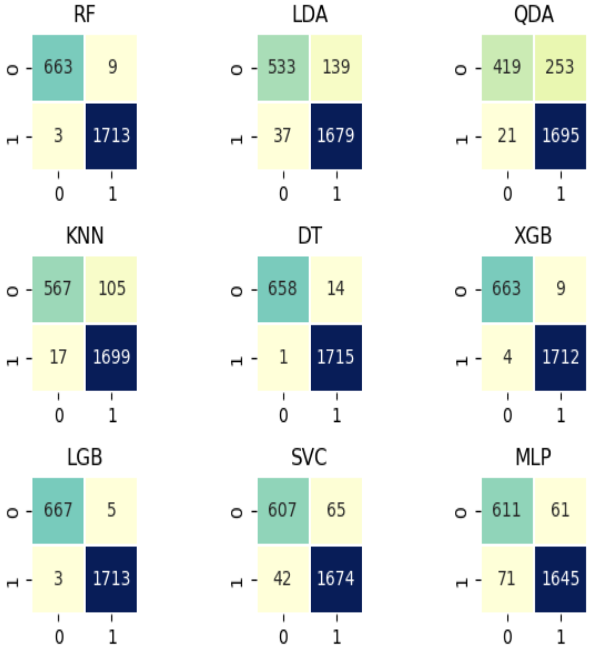

Figure 4. Confusion matrix with different algorithms.

As depicted in **Figure 5**, an analysis of the receiver operating characteristic (ROC) curves and area under the curve (AUC) values reveals that LGB achieves the optimal performance in classification tasks, with an AUC of 1.000, demonstrating a perfect discrimination capability between positive and negative samples. Other ensemble models such as XGB and RF follow closely, also attaining high values, indicating their strong performance. The shape of the ROC curves indicates that most models maintain a curve close to the top-left corner, signifying a low rate of misclassification. However, the ROC curve for LDA deviates from the top-left corner, with an AUC of 0.978, suggesting a higher risk of misclassification. This suggests that the LDA model is less robust compared to other models when applied to this dataset, potentially due to its reduced capacity to handle the complexity and imbalance present in the data. The superior performance of LGB and the competitive performance of RF underscore their efficacy in providing accurate and reliable classification outcomes.

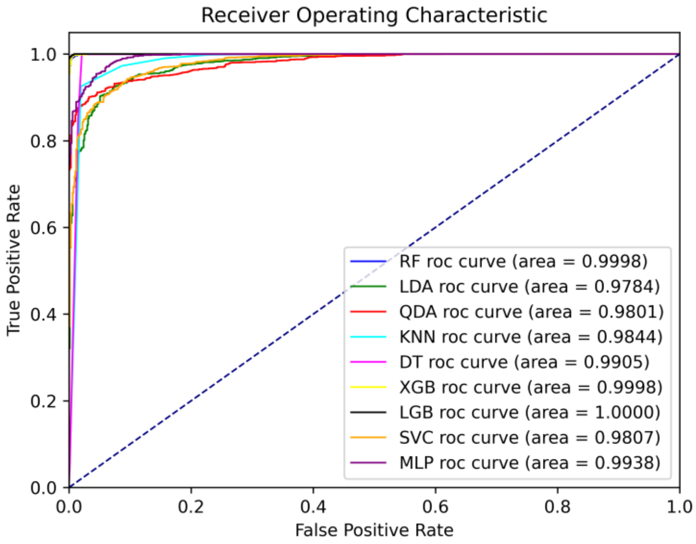

Figure 5. Comparison of classification models using ROC Curves and AUC metrics.

4.3. Analysis based on SHAP model interpretation

By utilizing SHAP model interpretation, we elucidate the impact of various features on the LGB model. As depicted in **Figure 6**, there are notable differences in the contributions of distinct mental health characteristics, such as obsessive-compulsive symptoms, depression, and anxiety, to the model’s predictions.

The distribution of SHAP values for obsessive-compulsive and anxiety indicates a broader range and a clear upward trend, signifying their higher weights within the model and their status as pivotal predictive factors. In contrast, hostility and psychoticism exhibit a narrower range of SHAP values, suggesting a more limited influence on predictions, manifesting significantly only under extreme conditions. This outcome may correlate with the clinical presentation of these psychological traits. For instance, obsessive-compulsive symptoms are typically more pronounced and easier to detect, whereas hostility and psychoticism are less likely to be reflected through straightforward questionnaires or specific scales. The distribution of SHAP values reveals that obsessive-compulsive symptoms and anxiety are the most influential mental health characteristics within the model, with their broad distribution and upward trend underscoring their critical role in predicting mental health status.

Additionally, the red line in each plot illustrates the model’s response trend to variations in feature values. For instance, depression and anxiety show a linear increase in SHAP values as their feature values rise, indicating a continuously enhancing contribution to predictions. In the case of somatization and interpersonal sensitivity, there is a notable increase in SHAP values beyond specific thresholds (e.g., feature values > 2), indicating heightened sensitivity of these characteristics to model predictions within certain ranges, aligning with the distribution characteristics of the relevant clinical mental health features.

The trend in SHAP value variations suggests that the model is highly sensitive to diagnostic thresholds of mental health traits. When feature values exceed certain thresholds, the model’s reliance on these features for predictions significantly increases. This indicates the model’s capability to effectively capture key points of change in mental health status. The global and local analysis of SHAP value distribution further unveils the model’s dependency on mental health characteristics. The global trend indicates that anxiety and obsessive-compulsive symptoms significantly influence overall prediction outcomes, while the local distribution reflects individual differences, particularly the notable variance in the contribution of obsessive-compulsive symptoms across samples. The SHAP analysis provides a scientific basis for feature selection and model optimization. Future research could focus on refining the modeling process for less sensitive traits like hostility and further exploring the model’s segmented predictive capabilities for key characteristics such as anxiety and obsessive-compulsive symptoms to enhance early warning effects for high-risk groups.

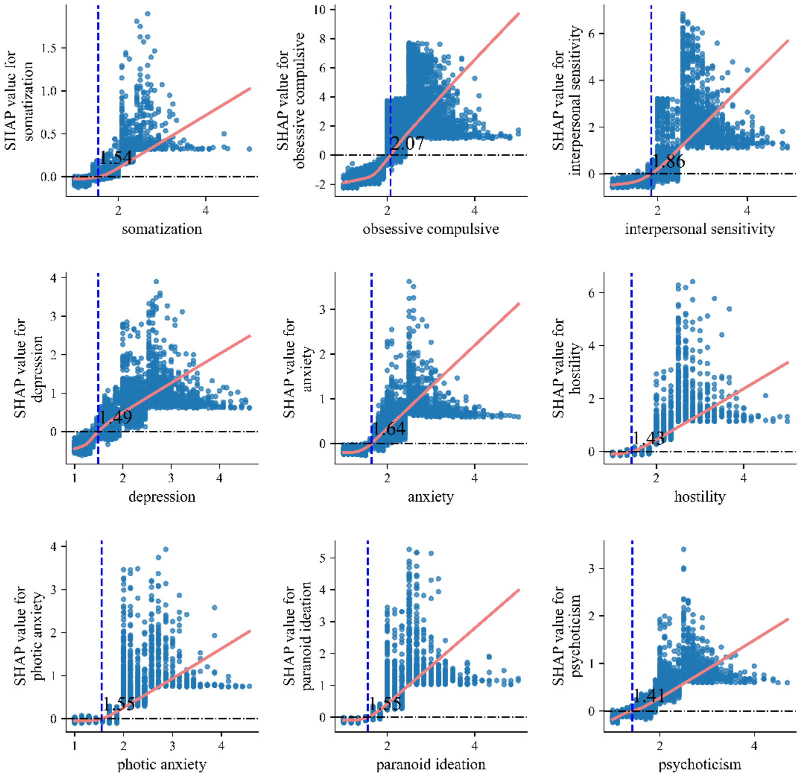

Figure 6. Analysis of SHAP values and diagnostic thresholds.

As shown in **Figure 7**, obsessive-compulsive symptoms have the greatest influence on the model’s output, indicating that this feature is crucial to the decision-making process of the model. Grade level follows closely, also being one of the significant factors affecting the model’s output. Interpersonal sensitivity, depression, and anxiety, these psychological state features also show relatively high importance, although their impact is lower than the first two features. The features of paranoid ideation, hostility, others, psychoticism, phobic anxiety, and somatization have a lesser impact on the model’s output.

In specific cases as shown in **Figure 8**, when the value of obsessive-compulsive is 2.6, it becomes the most influential positive feature with a SHAP value of +4.76, indicating a significant positive impact on the model’s predictive outcome. Conversely, when the the grade is freshman corresponds to a SHAP value of −0.91, marking it as the most influential negative feature, which implies a substantial negative effect on the prediction. This is primarily because, in the dataset, the proportion of freshmen with mental health issues is the lowest.

The phobic anxiety feature holds significance as the second most impactful positive contributor, with a SHAP value of +2.69. This indicates that phobic anxiety plays a potentially crucial role in predicting mental health problems. The interpersonal sensitivity feature also makes a positive contribution to the model's predictions, with a SHAP value of +0.41. Although its influence is relatively less compared to phobic anxiety, it still cannot be overlooked. Meanwhile, featureS like hostility and photic anxiety have relatively small SHAP values. Nevertheless, they do have a minor negative impact on the predictive outcomes, suggesting that they might be factors reducing the likelihood of certain mental health issue predictions.

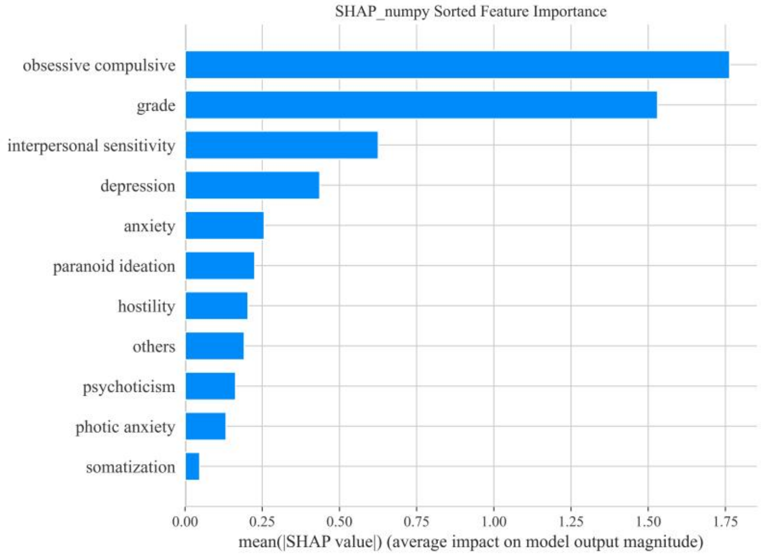

Figure 7. Feature importance analysis based on SHAP.

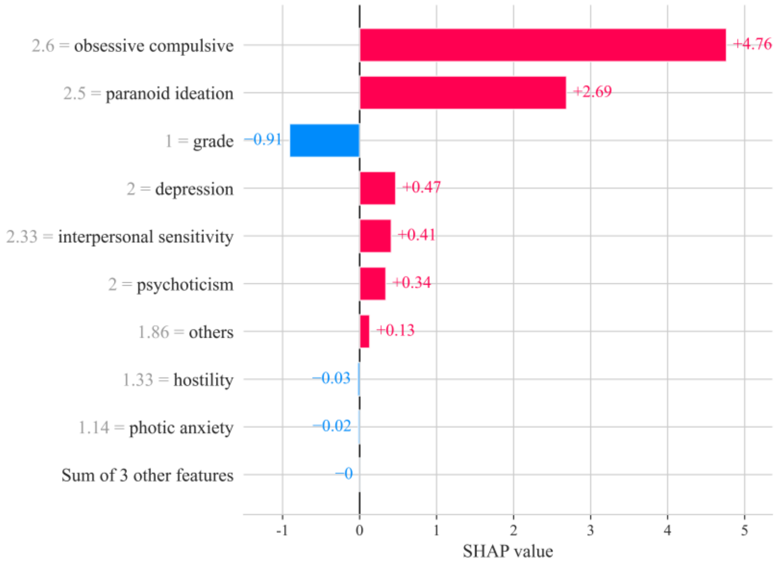

Figure 8. SHAP explanation for instance.

5. Conclusion

This study successfully applied machine learning techniques to the assessment and prediction of college students' mental health, leveraging the symptom checklist-90 scale. The findings underscore the efficacy of machine learning models, particularly LGB, in accurately predicting mental health outcomes. The integration of NGO algorithm for hyperparameter optimization and the SMOTE for addressing class imbalance significantly enhanced model performance. The SHAP value analysis revealed that obsessive- compulsive symptoms and anxiety were the most influential features in the model's predictions, providing valuable insights into the psychological dimensions that significantly affect mental health outcomes. These findings offer a scientific basis for developing targeted intervention strategies and early warning systems to support the mental well-being of college students. However, participants' responses may be influenced by various factors, some students may overreport symptoms, perhaps seeking more attention or assistance. These biases could potentially distort the real distribution of mental health issues among college students and affect the accuracy of our model predictions.

The study's contributions lie in its novel approach to mental health assessment using machine learning, its proposal of a new evaluation method, and its demonstration of the potential of machine learning models in handling imbalanced data and improving prediction accuracy. To further improve the accuracy of the model, other factors such as sleep and exercise can be incorporated [27]. Meanwhile, the SCL-90 scale can be further streamlined to reduce the assessment time [28]. Future research should focus on refining the modeling process for less sensitive traits and further exploring the model's segmented predictive capabilities for key characteristics to enhance early warning effects for high-risk groups.

The results of this study not only advance the understanding of college students' mental health but also pave the way for the development of more effective mental health assessment tools and intervention strategies. By leveraging machine learning technology, we can better address the complex and multifaceted challenges of mental health in the college student population.

Author contributions: Conceptualization, CL and LJ; methodology, CL; software, CL; validation, CL; formal analysis, JL; data curation, JM and XS; writing—original draft preparation, CL; writing—review and editing, CL; visualization, LJ; supervision, CL; project administration, CL; funding acquisition, CL. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by "Research on Monitoring and Early Warning of College Students' Mental Health from the Perspective of Digital Intelligence Integration (Grant number: B-a/2024/09)" and Leading talents of Jiangsu Vocational Institute of Commerce.

Acknowledgments: We would like to express gratitude to the team from Jiangsu Vocational Institute of Commerce (Multi-dimensional Heterogeneous Elastic Network Engineering Technology and Application).

Ethical approval: This study was conducted in accordance with the Declaration of Helsinki and was approved and ratified by the Ethics Committee of Jiangsu Vocational Institute of Commerce. Informed consent was obtained from all participating subjects before the study was conducted.

Conflict of interest: The authors declare no conflict of interest.

References

1. Cheng S, An D, Yao Z. Association between mental health knowledge level and depressive symptoms among Chinese college students. International Journal of Environmental Research and Public Health. 2021; 18(4): 1850. doi: 10.3390/ijerph18041850
2. Costa PT, McCrae RR. Four ways five factors are basic. Personality and Individual Differences. 1992; 13(6): 653–665. doi: 10.1016/0191-8869(92)90236-I
3. Conger RD, Donnellan B. An interactionist perspective on the socioeconomic context of human development. Annual Review of Psychology. 2007; 58: 175–199. doi: 10.1146/annurev.psych.58.110405.085551
4. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychological Bulletin. 1985; 98(2): 310–357. doi: 10.1037/0033-2909.98.2.310
5. Ng ZJ, Huebner S, Hills KJ. Life satisfaction and academic performance in early adolescents: Evidence for reciprocal association. Journal of School Psychology. 2015; 53(6): 479–491. doi: 10.1016/j.jsp.2015.09.004
6. Sharma A, Blakemore A, Byrne M. Oral health primary preventive interventions for individuals with serious mental illness in low- and middle-income nations: Scoping review. Global Public Health. 2024; 19(1). doi: 10.1080/17441692.2024.2408597
7. Ahmed NN, Bhat TK, Powar S. Stacked ensemble machine learning approach for electroencephalography-based major depressive disorder classification using temporal statistics. Systems Science & Control Engineering. 2024; 12(1). doi: 10.1080/21642583.2024.2427028
8. Derogatis LR, Cleary PA. Factorial invariance across gender for the primary symptom dimensions of the SCL-90. British Journal of Social and Clinical Psychology. 1977; 16(4): 347–356. doi: 10.1111/j.2044-8260.1977.tb00241.x
9. Hamaideh SH. Stressors and reactions to stressors among university students. International Journal of Social Psychiatry. 2011; 57(1): 69–80. doi: 10.1177/0020764009348442
10. Wu J, Shen H, Shen Y, et al. The influence of family socioeconomic status on college students' mental health literacy: The chain mediating effect of parenting styles and interpersonal relationships. Frontiers in Psychology. 2024; 15: 1477221. doi: 10.3389/fpsyg.2024.1477221
11. Tavolacci MP, Ladner J, Grigioni S, et al. Prevalence and association of perceived stress, substance use and behavioral addictions: A cross-sectional study among university students in France, 2009–2011. BMC Public Health. 2013; 13: 724. doi: 10.1186/1471-2458-13-724
12. Lucas RE, Diener E, Suh E. Discriminant validity of well-being measures. Journal of Personality and Social Psychology. 1996; 71(3): 616. doi: 10.1037/0022-3514.71.3.616
13. Dehghan P, Alashwal H, Moustafa AA. Applications of machine learning to behavioral sciences: Focus on categorical data. Discoveries in Psychology. 2022; 2: 22. doi: 10.1007/s44202-022-00027-5
14. Xin C, Zakaria LQ. Integrating BERT with CNN and BiLSTM for explainable detection of depression in social media contents. IEEE Access. 2024; 12: 161203–161212. doi: 10.1109/ACCESS.2024.3488081
15. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: A scoping review of methods and applications. Psychological Medicine. 2019; 49(9): 1426–1448. doi: 10.1017/S0033291719000151
16. Osman AB, Tabassum F, Patwary MJA, et al. Examining Mental Disorder/Psychological Chaos through Various ML and DL Techniques: A Critical Review. Annals of Emerging Technologies in Computing. 2022; 6(2): 61–71. doi: 10.33166/AETiC.2022.02.005
17. Zheng BHY. The application of machine learning in mental health. Frontiers in Social Sciences. 2022; 11(11): 4814–4818. doi: 10.12677/ASS.2022.1111656
18. Deng X, Li Y, Weng J. Feature selection for text classification: A review. Multimedia Tools and Applications. 2019; 78(4): 3797–3816. doi: 10.1007/s11042-018-6083-5

19. Tani L, Rand D, Veelken C. Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics. European Physical Journal C. 2021; 81: 170. doi: 10.1140/epjc/s10052-021-08950-y
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002; 16: 321–357. doi: 10.1613/jair.953
21. Vincent AM, Jidesh P. An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms. Scientific Reports. 2023; 13(1): 4737. doi: 10.1038/s41598-023-32027-3
22. Zhang C, Cho S, Vasarhelyi M. Explainable artificial intelligence (XAI) in auditing. International Journal of Accounting Information Systems. 2022; 46: 100572. doi: 10.1016/j.accinf.2022.100572
23. Wang T, Xue C, Zhang Z, et al. Unraveling the distinction between depression and anxiety: A machine learning exploration of causal relationships. Computers in Biology and Medicine. 2024; 174: 108446. doi: 10.1016/j.compbiomed.2024.108446
24. Liao Z, Fan X, Ma W, Shen Y. An Examination of Mental Stress in College Students: Utilizing Intelligent Perception Data and the Mental Stress Scale. Mathematics. 2024; 12(10): 1501. doi: 10.3390/math12101501
25. Tiwari S, Vats S, Bhardwaj B, et al. Enhanced SMOTE strategy for handling imbalanced data in machine learning classification. In: Proceedings of the 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE); 1–2 November 2023; Chennai, India. doi: 10.1109/rmkmate59243.2023.10369381
26. Aghbalou A, Sabourin A, Portier F. On the bias of K-fold cross-validation with stable learners. Proceedings of Machine Learning Research. 2023; 206: 3775–3794.
27. Dehghani M, Hubálovský Š, Trojovský P. Northern goshawk optimization: A new swarm-based algorithm for solving optimization problems. IEEE Access. 2021; 9: 162059–162080. doi: 10.1109/ACCESS.2021.3133286
28. Fan J, Ma X, Wu L, et al. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agricultural Water Management. 2019; 225: 105758. doi: 10.1016/j.agwat.2019.105758
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 4-9 December 2017; Long Beach, CA, US. doi: 10.48550/arXiv.1705.07874
30. Vaishnavi K, Nikhitha Kamath U, Ashwath Rao B, Subba Reddy NV. Predicting Mental Health Illness using Machine Learning Algorithms. Journal of Physics: Conference Series. 2022; 2161(1): 012021. doi: 10.1088/1742-6596/2161/1/012021
31. Cheng JP, Haw SC. Mental Health Problems Prediction Using Machine Learning Techniques. International Journal on Robotics, Automation and Sciences. 2023; 5(2): 59–72. doi: 10.33093/ijoras.2023.5.2.7