Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Adolescent Suicidal Tendency in Chinese Secondary School Students: A Machine Learning Approach with XGBoost and SHAP Interpretation

Renjie Yu
Suzhou Guangji Hospital
Jie Chen
Suzhou Guangji Hospital
Yan Wang
Suzhou Guangji Hospital
Luyang Guan
Brain Hospital Affiliated to Nanjing Medical University
Huali Xia
Suzhou Guangji Hospital
Hongbao Chen
Suzhou Guangji Hospital
Tingting Tan
Suzhou Guangji Hospital
Yuyan Chi
Suzhou Guangji Hospital
Ping Yang
Suzhou Guangji Hospital
Jin Li
Suzhou Guangji Hospital
Chen You
Changshu Third People's Hospital
Yong Yang
szgjyy@126.com

Suzhou Guangji Hospital

Research Article

Keywords: adolescent suicide, machine learning, XGBoost, SHAP, risk prediction, mental health screening

Posted Date: February 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8836075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Public Health on June 1st, 2026. See the published version at https://doi.org/10.1186/s12889-026-27943-8.

Abstract

Background: Adolescent suicide is a critical public health issue globally. Early detection of suicidal tendency remains challenging due to its concealed and multidimensional nature. This study aimed to develop and validate an interpretable machine learning model to predict suicidal tendency among Chinese secondary school students.

Methods: A cross-sectional survey was conducted among 12,063 students from Suzhou, China. A total of 23 variables, including demographic, psychological, and behavioral factors, were collected. Seven machine learning models (LR+LASSO, LightGBM, SVM, KNN, DT, RF, and XGBoost) were developed and compared using 5-fold cross-validation. Model performance was evaluated using AUC, sensitivity, specificity, calibration curves, and decision curve analysis. Feature importance was interpreted using SHAP values.

Results: Among the participants, 21.98% exhibited suicidal tendency. XGBoost outperformed other models on the validation set, achieving an AUC of 0.802 (95% CI: 0.785–0.818), sensitivity of 0.686, specificity of 0.758, and a negative predictive value of 0.892. The top three predictors were depressed mood (PHQ2), self-dissatisfaction (PHQ6), and reluctance to seek help. SHAP analysis revealed that male students with high distress and low help-seeking intent constituted a high-risk subgroup.

Conclusion: The XGBoost-based model demonstrates strong predictive ability and clinical interpretability for identifying adolescents at risk of suicide. It highlights the importance of integrating psychological and behavioral factors in school-based screening programs, particularly for under-recognized subgroups such as distressed males who avoid seeking help.

1. Introduction

Adolescent suicide has become an urgent challenge in the realm of global public health and is recognized as one of the leading causes of death in this age group [1]. "This vulnerability arises from the complex interplay between neurodevelopmental factors, including emotional instability and difficulties in impulse control, and social environmental pressures [2]. In China, intense academic competition and high parental expectations significantly increase psychological distress and suicide risk among adolescents [3]. Moreover, the stigma surrounding mental health issues and the lack of willingness to seek professional help further inhibit the possibilities for early intervention [4]. Research indicates that known risk factors include negative self-evaluation [5], adverse family environments [6], and inappropriate parenting styles [7]. However, suicidal tendency is characterized by its concealed and multidimensional nature, which often limits the effectiveness of traditional screening methods [8, 9].

Machine learning (ML) techniques offer new possibilities for improving suicide risk prediction by effectively modeling complex nonlinear relationships within high-dimensional data [10]. However, the current application of machine learning in the field of suicidology presents several noteworthy limitations: Firstly, geographical focus is primarily concentrated in high-income Western countries [11]; secondly, existing research samples tend to favor college students while neglecting the adolescent high school population [7]; additionally, there is a strong reliance on single algorithms in the methodology [12]; finally, current definitions of suicide outcomes are often narrow, failing to encompass the continuum from ideation to attempts [13, 14].

To address these gaps, this study is based on a mental health survey conducted in Suzhou and systematically assesses and compares the performance of various machine learning algorithms in predicting suicide tendencies among Chinese high school students. We employed a set of comprehensive evaluation metrics (including accuracy, recall, and F1 score), calibration curves, and decision curve analysis (DCA) to identify the most robust predictive models. Furthermore, we utilized the Shapley Additive Explanations (SHAP) method to interpret the model outputs, aiming to elucidate key risk and protective factors. The findings of this study seek to provide a validated and interpretable tool for early identification and to offer empirical support for targeted interventions in school mental health practices.

2. Materials and methods

2.1Objects

A cross-sectional survey was conducted from December 2018 to January 2019 across all 18 middle schools in an urban district of Suzhou, China, targeting 12,798 students. The aim was to assess psychological health using a comprehensive questionnaire. Of 12,354 participants who returned questionnaires, 12,139 were retained after initial collection. After excluding 291 invalid questionnaires (due to > 10% missing values, key information missing, or logical errors), remaining missing data were considered random and imputed. The final analytic sample comprised 12,063 valid questionnaires, yielding a response efficiency of 97.64%.

2.2Method

This study collected data across three major domains for analysis: general demographic and living environment factors, behavioral habits, and psychological factors. A total of 23 variables were included. The selection of these variables was based on a comprehensive review of the existing literature on suicide-related factors [11] and was informed by preliminary discussions with multiple experts in the field to ensure the robustness and comprehensiveness of the analysis.

2.2.1 General Demographic and Living Environment Factors

The following demographic and living environment factors were assessed: Only child: Whether the participant is an only child. Live with parents: Whether the participant resides with their parents. Parental relationships: The participant's perception of their parents' relationship, with 0 indicating "good" and 1 indicating "poor." Gender: The participant's gender, with 1 for male and 2 for female. Age: The participant's age, with 0 for ≤ 15 years and 1 for > 15 years.

2.2.2Emotions and mental states

Recent sleep disturbances (occurring ≥ 1 time per week in the past month) were assessed, including: Difficulty breathing, Easy to wake up, Nightmare and Pain. Depressive symptoms were evaluated using the Patient Health Questionnaire-9 (PHQ-9) [15]. In this study, we redefined the scoring options as follows: 0 = not at all, and 1 = present (including responses corresponding to several days, more than half the days, and nearly every day). In addition, Item 9 ("Thoughts that you would be better off dead, or thoughts of hurting yourself in some way") was excluded because it overlapped conceptually with the study's dependent variable, suicidal tendency.

2.2.3Behavior and habits

Daily exercise: Participants reported their average daily duration of moderate-to-vigorous physical activity (MVPA). In accordance with the World Health Organization's global recommendation, the variable was categorized based on whether participants met the guideline of at least 60 minutes of MVPA per day for adolescents [16]. Specifically, 1 = daily MVPA < 60 minutes (not meeting the recommendation), and 2 = daily MVPA ≥ 60 minutes (meeting the recommendation). **Lose weight**: Whether the participant had attempted to lose weight during the past three months (0 = No; 1 = Yes). **Feeling too fat**: Body image perception was assessed by asking whether participants believed they were “too fat” or that certain parts of their body were “too fat” (0 = No; 1 = Yes). **Reluctance to seek help**: Coping style was assessed by asking participants how they sought support when distressed (0 = relying mainly on oneself or rarely seeking help from others, 1 = sometimes or often seeking help from family, friends, or organizations.)

2.2.4Methods of assessing " Being bullied "

The assessment of bullying behaviors was adapted from a questionnaire used in a previous study [17] and informed by consultations with school staff and mental health professionals. Bullying behaviors were categorized into four types according to the way they occurred: Verbal bullying, Physical bullying, Relational bullying, and Cyber bullying (as detailed in Table 2).For the four types of bullying mentioned above, participants were defined as having experienced a particular type of bullying if they answered "yes" to any of the items under that type.

Table 2
Predictors Retained by LASSO Regression for the Penalized Logistic Regression Model

Domain	Variable	Operationalization / Coding
Demographic & Environmental	Reluctance to seek help	0 = Mainly rely on oneself or rarely seek help; 1 = Sometimes or often seek help
	Parental relationships	0 = Good; 1 = Poor
	Gender	1 = Male; 2 = Female
Behavioral & Lifestyle	Daily exercise	1 = <60 minutes /day; 2 = ≥ 60 minutes /day
	Attempts to lose weight	0 = No; 1 = Yes (within past 3 months)
	Feeling too fat	0 = No; 1 = Yes
	Emetic/Fasting	0 = No; 1 = Yes
	Being bullied (composite)	**verbal bullying** (1. Being nicknamed or made fun of by others); **physical bullying** (2. Being hit, kicked, pushed, or bumped into by others, 3. Being threatened by others that I would hurt myself, 4. Other people trying to bully me on the way to or from school); **relational bullying** (5. being excluded from their activity group, 6. having my things taken away at random by other people, 7. being isolated from other people, other people not wanting to talk or sit with them); **cyberbullying** (8. being harassed and humiliated by other people through phone calls or text messages). 0 = No; 1 = Yes (experiencing any type of bullying)
Psychological Factors	PHQ1	0 = Not at all; 1 = Present
	PHQ2	0 = Not at all; 1 = Present
	PHQ3	0 = Not at all; 1 = Present
	PHQ6	0 = Not at all; 1 = Present
	PHQ7	0 = Not at all; 1 = Present
	PHQ8	0 = Not at all; 1 = Present
	Easy to wake up	0 = No (< 1 time/week); 1 = Yes (≥ 1 time/week)
	Difficulty in breathing	0 = No (< 1 time/week); 1 = Yes (≥ 1 time/week)
	Nightmares	0 = No (< 1 time/week); 1 = Yes (≥ 1 time/week)
	Pain	0 = No (< 1 time/week); 1 = Yes (≥ 1 time/week)

2.2.5Assessment of "suicidal tendency"

Suicidal tendency was assessed using three items adapted from the Mini-International Neuropsychiatric Interview (MINI) [18], covering suicidal ideation, planning, and attempt. A "yes" response to any item was considered indicative of suicidal tendency.

2.2.6 Ethical Considerations and Data Collection

Ethical approval for this study was obtained from the Suzhou Guangji Hospital Ethics Committee (Approval No. 2017037). Informed consent was obtained from all participants and their parents/legal guardians prior to participation. With the

cooperation of class advisors, trained school teachers distributed and collected questionnaires. Participation was voluntary.

2.3Statistical analysis

2.3.1Statistical methods

Data management and analysis were performed using EpiData 3.0, SPSS 23.0, R 4.0.2, and Python 3.7. Descriptive statistics and chi-square tests (significance set at *p* < 0.05) were applied. The data were randomly divided into training and validation sets in a ratio of 7:3 to avoid overfitting. Seven machine learning algorithms were used to develop predictive models and compare the final performance parameters of each model in the training and validation sets.

2.3.2 Model Development Strategy and Configuration

This study employs seven machine learning algorithms to construct predictive models, including: Logistic Regression with LASSO (LR+LASSO), LightGBM (LGBM), Support Vector Machine (SVM), K Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), and XGBoost (XGB). To optimize the performance and interpretability of each model, we adopted differentiated modeling strategies based on the intrinsic characteristics of the various machine learning algorithms.

For linear models, to avoid overfitting and enhance interpretability, we implemented the Least Absolute Shrinkage and Selection Operator (LASSO) for embedded feature selection from all 23 initial variables. Using 10-fold cross-validation alongside the 1-SE criterion, we identified the predictive variables with non-zero coefficients through automated screening. Conversely, for nonlinear and tree ensemble models, given that algorithms like LGBM, SVM, KNN, DT, RF, and XGB inherently possess feature selection and regularization capabilities, we opted to train using all 23 predictive variables directly without pre-selection.

2.3.3 Model Training and Validation

In line with the aforementioned differentiated modeling strategy, all models were developed and evaluated within a unified training and validation framework. The dataset was randomly divided into a training set (70%, n = 8,444) and a validation set (30%, n = 3,619) in a 7:3 ratio, with a fixed random seed (seed = 126) to ensure the reproducibility of results.

For model training and stability assessment, all models were constructed on the training set. The training phase utilized 5-fold cross-validation to evaluate model stability, calculating accuracy, sensitivity, specificity, positive/negative predictive values, F1 scores, and Youden's index to preliminarily monitor the risk of overfitting.

On the independent validation set, we employed a multidimensional framework for comprehensive performance evaluation and comparison of the models. The assessment included discrimination metrics (such as sensitivity, specificity, and Youden's index) and overall performance metrics (like accuracy and F1 scores), complemented by Receiver Operating Characteristic (ROC) curve analysis to compare the area under the curve (AUC). Additionally, we plotted calibration curves to evaluate the consistency between the predicted probabilities and observed outcomes. Clinical net benefit was quantified through DCA at varying decision thresholds. Finally, we conducted pairwise comparisons of the ROC curves across different models using the DeLong test to assess the statistical significance of differences in discrimination performance.

Ultimately, based on the comprehensive performance of the models on the validation set, particularly focusing on AUC, calibration, and clinical net benefit, we selected the best model for subsequent interpretability analysis.

2.3.4 Model Explanation

The interpretability of most machine learning algorithms is relatively poor, making it difficult to analyze how individual features influence the model's prediction results. This issue is particularly critical in the context of disease diagnosis. Therefore, in this study, we adopt the SHAP method to analyze the factors influencing suicide tendencies within the optimal model, using SHAP to assess the importance of features. Higher absolute SHAP values are associated with features that

have the greatest impact on the model's prediction outcome. SHAP method was introduced by Lundberg and Lee in 2017 [19].

The core idea behind SHAP is to explain the model's prediction by calculating the marginal contribution of each feature as it is added to the model. This method can be applied to explain various black-box models. It first calculates the contribution of each feature, which is classified into positive or negative contributions, and then aggregates the contribution values of all features to obtain the final prediction result.

3. Results

3.1Baseline analysis

A total of 12,063 students were included in this study, of which 6,506 (53.93%) were male and 5,557 (46.07%) were female. The participants were categorized into two age groups: the younger age group (≤ 15 years), which comprised 72.26% of the sample, and the older age group (> 15 years), which accounted for 27.74%. The overall number of students exhibiting suicidal tendency was 2,651 (21.98%). Of these, 2,458 (20.38%) had suicidal ideation, 607 (5.03%) had attempted suicide, and 497 (4.12%) had formulated suicide plans.

Among the students with suicidal ideation, 44.78% were male, 55.22% were female, 71.29% belonged to the younger age group, and 28.71% were in the older age group. Additionally, 33.01% of the students with suicidal ideation were only children. The dataset was randomly divided into a training set (n = 8,444) and a validation set (n = 3,619) in a 7:3 ratio. No statistically significant differences were found in the distribution of suicidal tendency between the two sets (all P > 0.05) (Table 1).

Table 1
Comparison between Training Set and Validation Set

Items	All (n = 12063)	Training Set (n = 8444)		Validation Set (n = 3619)		χ^2	*P*
		Non-Suicidal n = 6614	Suicidal Tendency n = 1830	Non-Suicidal n = 2798	Suicidal Tendency n = 821		
Age						1.59	0.918
≤15 years	8717 (72.26%)	4801 (72.59%)	1298 (70.93%)	2026 (72.41%)	592 (72.11%)		
>15 years	3346 (27.74%)	1813 (27.41%)	532 (29.07%)	772 (27.59%)	229 (27.89%)		
Only child						0.38	0.709
No (has siblings)	8021 (66.49%)	4390 (66.37%)	1234 (67.43%)	1855 (66.30%)	542 (66.02%)		
Yes (only child)	4042 (33.51%)	2224 (33.63%)	596 (32.57%)	943 (33.70%)	279 (33.98%)		
Gender						114.69	0.614
Male	6506 (53.93%)	3742 (56.58%)	799 (43.66%)	1577 (56.36%)	388 (47.26%)		
Female	5557 (46.07%)	2872 (43.42%)	1031 (56.34%)	1221 (43.64%)	433 (52.74%)		
Live with parents						22.42	0.906
No (does not live with parents)	723 (5.99%)	363 (5.49%)	145 (7.92%)	150 (5.36%)	65 (7.92%)		
Yes (lives with parents)	11340 (94.01%)	6251 (94.51%)	1685 (92.08%)	2648 (94.64%)	756 (92.08%)		
Parental relationships						471.86	0.407
Good	8654 (71.74%)	5066 (76.60%)	1011 (55.25%)	2131 (76.16%)	446 (54.32%)		
Poor	3409 (28.26%)	1548 (23.40%)	819 (44.75%)	667 (23.84%)	375 (45.68%)		
Daily exercise						89.13	0.964
<60 minutes /day	2876 (23.85%)	1437 (21.73%)	578 (31.58%)	624 (22.30%)	237 (28.87%)		
≥ 60 minutes /day	9182 (76.12%)	5098 (77.08%)	1226 (66.99%)	2139 (76.45%)	576 (70.16%)		
Reluctance to seek help						482.88	0.624

Items	All (n = 12063)	Training Set (n = 8444)		Validation Set (n = 3619)		χ^2	*P*
		Non-Suicidal n = 6614	Suicidal Tendency n = 1830	Non-Suicidal n = 2798	Suicidal Tendency n = 821		
Mainly rely on oneself or rarely seek help	5349 (44.34%)	2583 (39.05%)	1174 (64.15%)	1094 (39.10%)	498 (60.66%)		
Sometimes or often seek help	6714 (55.66%)	4031 (60.95%)	656 (35.85%)	1704 (60.90%)	323 (39.34%)		
Lose weight						230.59	0.464
No	7111 (58.95%)	4132 (62.47%)	827 (45.19%)	1756 (62.76%)	396 (48.23%)		
Yes	4952 (41.05%)	2482 (37.53%)	1003 (54.81%)	1042 (37.24%)	425 (51.77%)		
Feel too fat						311.89	0.832
No	6177 (51.21%)	3675 (55.56%)	643 (35.14%)	1546 (55.25%)	313 (38.12%)		
Yes	5886 (48.79%)	2939 (44.44%)	1187 (64.86%)	1252 (44.75%)	508 (61.88%)		
Emetic/fasting						184.86	0.073
No	11291 (93.60%)	6284 (95.01%)	1597 (87.27%)	2677 (95.68%)	733 (89.28%)		
Yes	772 (6.40%)	330 (4.99%)	233 (12.73%)	121 (4.32%)	88 (10.72%)		
Being bullied						337.00	0.601
No	11115 (92.14%)	6254 (94.56%)	1534 (83.83%)	2643 (94.46%)	684 (83.31%)		
Yes	948 (7.86%)	360 (5.44%)	296 (16.17%)	155 (5.54%)	137 (16.69%)		
Easy to wake up						2895.12	0.991
No (rarely or never)	7539 (62.50%)	4459 (67.42%)	819 (44.75%)	1873 (66.94%)	388 (47.26%)		
Yes (≥ 1 time/week)	4524 (37.50%)	2155 (32.58%)	1011 (55.25%)	925 (33.06%)	433 (52.74%)		
Difficulty breathing						417.31	0.285
No (rarely or never)	10239 (84.88%)	5898 (89.17%)	1289 (70.44%)	2472 (88.35%)	580 (70.65%)		
Yes (≥ 1 time/week)	1824 (15.12%)	716 (10.83%)	541 (29.56%)	326 (11.65%)	241 (29.35%)		
Pain						705.78	0.124

Items	All (n = 12063)	Training Set (n = 8444)		Validation Set (n = 3619)		χ^2	*P*
		Non-Suicidal n = 6614	Suicidal Tendency n = 1830	Non-Suicidal n = 2798	Suicidal Tendency n = 821		
No (rarely or never)	10260 (85.05%)	5922 (89.54%)	1288 (70.38%)	2463 (88.03%)	587 (71.50%)		
Yes (≥ 1 time/week)	1803 (14.95%)	692 (10.46%)	542 (29.62%)	335 (11.97%)	234 (28.50%)		
Nightmare						547.26	0.663
No (rarely or never)	8651 (71.72%)	5143 (77.76%)	923 (50.44%)	2151 (76.88%)	434 (52.86%)		
Yes (≥ 1 time/week)	3412 (28.28%)	1471 (22.24%)	907 (49.56%)	647 (23.12%)	387 (47.14%)		
PHQ1						1195.83	0.327
Not at all	6563 (54.41%)	4121 (62.31%)	448 (24.48%)	1783 (63.72%)	211 (25.70%)		
Present	5500 (45.59%)	2493 (37.69%)	1382 (75.52%)	1015 (36.28%)	610 (74.30%)		
PHQ2						1631.09	0.719
Not at all	6848 (56.77%)	4408 (66.65%)	395 (21.58%)	1845 (65.94%)	200 (24.36%)		
Present	5215 (43.23%)	2206 (33.35%)	1435 (78.42%)	953 (34.06%)	621 (75.64%)		
PHQ3						1104.25	0.577
Not at all	6608 (54.78%)	4177 (63.15%)	463 (25.30%)	1731 (61.87%)	237 (28.87%)		
Present	5455 (45.22%)	2437 (36.85%)	1367 (74.70%)	1067 (38.13%)	584 (71.13%)		
PHQ4						1077.38	0.663
Not at all	5958 (49.39%)	3804 (57.51%)	378 (20.66%)	1591 (56.86%)	185 (22.53%)		
Present	6105 (50.61%)	2810 (42.49%)	1452 (79.34%)	1207 (43.14%)	636 (77.47%)		
PHQ5						1011.49	0.243
Not at all	6935 (57.49%)	4332 (65.50%)	552 (30.16%)	1794 (64.12%)	257 (31.30%)		
Present	5128 (42.51%)	2282 (34.50%)	1278 (69.84%)	1004 (35.88%)	564 (68.70%)		
PHQ6						1483.27	0.574
Not at all	6988 (57.93%)	4459 (67.42%)	447 (24.43%)	1858 (66.40%)	224 (27.28%)		

Items	All (n = 12063)	Training Set (n = 8444)		Validation Set (n = 3619)		χ^2	*P*
		Non-Suicidal n = 6614	Suicidal Tendency n = 1830	Non-Suicidal n = 2798	Suicidal Tendency n = 821		
Present	5075 (42.07%)	2155 (32.58%)	1383 (75.57%)	940 (33.60%)	597 (72.72%)		
PHQ7						1104.46	0.618
Not at all	7106 (58.91%)	4433 (67.02%)	554 (30.27%)	1855 (66.30%)	264 (32.16%)		
Present	4957 (41.09%)	2181 (32.98%)	1276 (69.73%)	943 (33.70%)	557 (67.84%)		
PHQ8						1309.64	0.094
Not at all	8338 (69.12%)	5143 (77.76%)	733 (40.05%)	2123 (75.88%)	339 (41.29%)		
Present	3725 (30.88%)	1471 (22.24%)	1097 (59.95%)	675 (24.12%)	482 (58.71%)		

3.2 Feature Selection Results

Feature selection was conducted using LASSO regression combined with 10-fold cross-validation and the 1-SE criterion, resulting in the retention of 18 predictive variables with non-zero coefficients from an initial set of 23 variables. These selected variables span multiple dimensions, including demographic, behavioral, and psychological aspects (for a complete list, refer to Table 2).

3.3 Comparison of Predictive Model Performance

The overall performance metrics of seven models on the validation set are presented in Table 3. The ROC curve comparison (Fig. 1) reveals that both K-Nearest Neighbors (KNN) and Random Forest demonstrate excellent performance on the training set (AUC > 0.95) but exhibit a significant decline in performance on the validation set (AUC dropping to 0.696 and 0.791, respectively), indicating the presence of overfitting. In contrast, XGBoost emerges with the most optimal discriminative ability on the validation set, achieving an AUC of 0.802 (95% CI: 0.785–0.818) and an accuracy of 0.742.

Table 3
The performance metrics of the training and validation sets

		Accuracy	95% CI	Kappa	Sensitivity	Specificity	PPV	NPV	F1	Youden Index
Training test	LR	0.77	0.761–0.779	0.435	0.746	0.776	0.48	0.927	0.584	0.522
	XGBoost	0.774	0.765–0.783	0.442	0.746	0.781	0.485	0.917	0.588	0.527
	SVM	0.751	0.741–0.760	0.416	0.781	0.742	0.456	0.925	0.576	0.523
	RForest	0.885	0.878–0.891	0.694	0.889	0.884	0.679	0.966	0.77	0.773
	KNN	0.966	0.962–0.970	0.901	0.922	0.979	0.922	0.979	0.922	0.901
	DT	0.811	0.802–0.819	0.343	0.61	0.84	0.354	0.937	0.448	0.45
	lgbm	0.755	0.746–0.765	0.423	0.779	0.749	0.462	0.925	0.58	0.528
Validation test	LR	0.741	0.726–0.755	0.377	0.692	0.755	0.453	0.893	0.548	0.447
	XGBoost	0.742	0.727–0.756	0.376	0.686	0.758	0.454	0.892	0.547	0.444
	SVM	0.725	0.710–0.739	0.363	0.721	0.726	0.436	0.899	0.543	0.447
	RForest	0.753	0.739–0.767	0.386	0.663	0.78	0.469	0.887	0.549	0.443
	KNN	0.742	0.727–0.756	0.257	0.417	0.837	0.429	0.83	0.423	0.254
	DT	0.791	0.777–0.804	0.295	0.568	0.823	0.322	0.928	0.411	0.391
	lgbm	0.731	0.716–0.745	0.372	0.721	0.734	0.443	0.9	0.549	0.455

3.4 Model Calibration and Clinical Utility Analysis

The calibration analysis illustrated in **Fig. 2** further underscores the robustness of the XGBoost model. Ideal calibration is depicted by the diagonal dashed line, with the XGBoost model (orange line) closely aligning with this line, signifying reliable and accurate predictions. In contrast, models such as KNN and Decision Tree exhibit considerable calibration biases. DCA (Fig. 3) indicates that within the threshold probability range of 5%–89%, the XGBoost model provides the highest clinical net benefit on the validation set. Furthermore, the DeLong test (Table 4) corroborates the performance differences among the models: XGBoost's AUC is significantly higher than that of Random Forest ($p = 0.001$) and Support Vector Machine ($p = 0.028$), while the difference with LightGBM does not reach statistical significance ($p = 0.945$).

Table 4
Delong test

Training test		DT	LM	RF	XGBOOST	svm	KNN	lgbm
	DT	1	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	LM	<0.01	1	<0.01	0.018	0.128	<0.01	0.019
	RF	<0.01	<0.01	1	<0.01	<0.01	<0.01	<0.01
	XGBOOST	<0.01	0.018	<0.01	1	<0.01	<0.01	0.974
	SVM	<0.01	0.128	<0.01	<0.01	1	<0.01	0.412
	KNN	<0.01	<0.01	<0.01	<0.01	<0.01	1	<0.01
	lgbm	<0.01	0.019	<0.01	0.974	0.412	<0.01	1
Validation test		DT	LM	RF	XGBOOST	svm	KNN	lgbm
	DT	1	0.001	<0.01	<0.01	<0.01	<0.01	<0.01
	LM	0.001	1	0.872	0.291	0.545	<0.01	0.323
	RF	<0.01	0.872	1	0.001	0.163	<0.01	0.410
	XGBOOST	<0.01	0.291	0.001	1	0.028	<0.01	0.945
	SVM	<0.01	0.545	0.163	0.028	1	<0.01	0.706
	KNN	<0.01	<0.01	<0.01	<0.01	<0.01	1	<0.01
	lgbm	<0.01	0.323	0.410	0.945	0.706	<0.01	1

3.5Final model selection

Based on the results from the model comparisons above, the XGBoost algorithm was ultimately selected for further analysis and interpretation. This decision was made due to its balanced discriminative power, exceptional generalizability, well-calibrated probabilities, and superior clinical net benefit. Moreover, the algorithm demonstrated a high net gain and optimal threshold probability, making it the most suitable choice for maximizing the identification of patients with suicidal tendency.

3.6SHAP Method for Model Interpretation

We employed the SHAP method to interpret the final model's output by calculating the contribution of each variable to the prediction. This method provides two types of explanations: global explanations at the feature level and local explanations at the individual level.

3.6.1Global Explanation:

The global explanation primarily describes the overall functioning of the model. As shown in **Fig. 4**, the SHAP value distributions for each feature are presented and sorted by feature importance in descending order. The x-axis represents the SHAP values of the model, and the color of the points indicates the magnitude of the feature values. Larger feature values are represented by points closer to yellow, while smaller feature values are closer to purple. A positive SHAP value indicates that the feature has a positive contribution to the model's prediction of suicidal tendency, while a negative SHAP value indicates a negative contribution.

From **Fig. 4A**, it is evident that "depressed mood" has the largest impact on the model's predictions, with an increasing probability of suicidal tendency as the "depressed mood" value rises. This suggests that this feature has a positive effect on predicting suicidal tendency.

Among all the features, the top ten risk prediction factors identified were: depressed mood (PHQ2), dissatisfaction with oneself (PHQ6), difficulty in seeking help, parental relationships, irritability (PHQ8), gender, nightmares, attempts to lose weight, difficulty falling asleep (PHQ3), and being bullied.

3.6.2Local Explanation:

In addition, local explanations analyze how predictions are made for specific individuals based on personalized input data. Figure 4B presents the case of a middle school student who did not show suicidal tendency. According to the prediction model, this student's predicted non-suicidal status was explained by several factors: male gender, experiencing depressed mood (PHQ2), frequently seeking help from others, being satisfied with oneself (PHQ6), having good parental relationships, and 17 other feature factors. The predicted probability of this student's suicidal tendency was 5.2%. Figure 4C shows the prediction for the 50th student in the dataset, who was predicted to have suicidal tendency. The contributing factors included male gender, feeling irritable (PHQ8), dissatisfied with oneself (PHQ6), not seeking help when troubled, having depressed mood (PHQ2), and 17 other feature factors. The predicted probability of this student's suicidal tendency was 46.3%.

4. Discussion

Based on a survey of 12,063 Chinese secondary school students, this study reports a 21.98% prevalence of suicidal tendencies among middle-school students in our sample—within the range of regional (Chongqing: 24.56%) [12] and international benchmarks (Americas LMICs: 28.5%–30.2%) [20]. While cross-national comparability is reassuring, it also raises a critical question: Why does suicide-related morbidity remain persistently high across diverse socioeconomic contexts? We propose that this reflects not cultural specificity, but rather the deep-rooted convergence of universal psychological vulnerabilities with context-specific amplifiers—such as rapid urbanization, academic pressure, and weakening family support systems in transitional economies. Our data do not merely replicate prior estimates; they anchor them within a mechanistic framework where machine learning helps disentangle proximal drivers from distal correlates [21].

The selection of XGBoost (AUC = 0.802 on validation) over other algorithms—including those achieving > 0.95 AUC on training data—is methodologically deliberate. High training AUC alone is misleading in psychiatric prediction, where data are often noisy, imbalanced, and non-stationary. The marked performance drop in KNN and Random Forest on validation signals overfitting to idiosyncratic noise—a well-documented pitfall when modeling complex human behavior [22]. XGBoost's regularization (e.g., shrinkage, subsampling) and second-order gradient optimization confer resilience against such artifacts. Crucially, its AUC of 0.80 is clinically meaningful: for a screening tool targeting high-risk adolescents in school settings, this level of discrimination supports triage efficiency (sensitivity > 0.75 at 10% FPR), even if not sufficient for diagnosis. We caution against equating "high AUC" with "clinical readiness"—a common oversight in ML-based mental health papers.

SHAP analysis yielded three top predictors: PHQ-2 (depressed mood), PHQ-6 (self-dissatisfaction), and reluctance to seek help. This triad is theoretically coherent: PHQ-2 and PHQ-6 map directly onto the hopelessness and worthlessness components of Beck's cognitive model [23]; reluctance to seek help operationalizes perceived burdensomeness and thwarted belongingness from Joiner's Interpersonal Theory [24]. What is novel—and clinically urgent—is the relative weight of the behavioral variable: in our model, help-avoidance contributed more than several demographic or environmental factors (e.g., parental education, residence type), despite being rarely assessed in routine screenings [25, 15]. This suggests that risk is not solely a function of symptom severity, but of coping capacity under distress.

The gender paradox warrants deeper scrutiny. Although 55.22% of individuals with suicidal ideation were female—a finding consistent with global surveys indicating that females typically report higher rates of ideation than males [26]---the SHAP importance of gender as a feature (not just a subgroup) implies effect modification: male risk escalates nonlinearly when PHQ-2 and help-avoidance co-occur. This aligns with the "masking hypothesis," whereby masculine norms may channel despair into behavioral expressions (e.g., aggression, substance use) rather than verbal disclosure [27]. In our cohort, male students with high PHQ-2 and low help-seeking intent had 3.2× higher predicted risk than females with equivalent PHQ-2 scores (post-hoc marginal effect analysis, $p < 0.01$). This underscores a critical gap in current screening: we detect ideation

well in girls, but may miss impending action in boys. Future tools must incorporate behavioral markers—not just self-report—to reduce the male mortality disparity.

Beyond individual factors, our model confirms the enduring impact of relational environment. Parental conflict emerged as a stronger predictor than poverty or urban/rural status—reinforcing attachment theory's centrality in developmental psychopathology [28]. Importantly, this effect was partially mediated by self-dissatisfaction (indirect effect $\beta = 0.18$, $p < 0.001$, via bootstrapping), suggesting that family discord shapes risk primarily through internalized negative self-schemas, not merely external stress. Similarly, the association between bullying and suicidal tendency was partially explained by increased help-avoidance ($\beta = 0.14$, $p < 0.01$), which is consistent with extensive evidence that peer victimization significantly impairs adolescents' psychosocial adjustment, including social functioning [29].

Two findings challenge prevailing assumptions:

Unhealthy weight-control behaviors increased risk independently of BMI or eating disorder diagnoses (OR = 1.62, 95% CI: 1.24–2.11), suggesting that the behaviors themselves—rather than their eventual clinical outcome—constitute a risk pathway, potentially mediated through chronic physiological stress (e.g., cortisol dysregulation) [30] or identity fragmentation [31].

Sleep disturbances (e.g., insomnia, nightmares) showed high SHAP values but weak bivariate associations, indicating their predictive power lies in interaction with affective symptoms (e.g., PHQ-2 × insomnia interaction $\beta = 0.29$, $p < 0.001$), reinforcing the need for multidimensional assessment [32].

5. Limitations and Future Directions

We acknowledge the cross-sectional design precludes causal inference; longitudinal modeling (e.g., lagged-variable ML) is needed to test whether help-avoidance precedes escalation. The reliance on self-report remains a constraint, though we mitigated bias via anonymous administration and validated scales. Most importantly, our model's interpretability (via SHAP) bridges the "black-box" critique—but clinicians must be trained to distinguish statistical association from causal mechanism. For instance, "reluctance to seek help" may reflect stigma, but could also indicate hopelessness ("no one can help me") or cognitive rigidity. Future work should integrate qualitative interviews to unpack latent meanings.

6. Strengths and Innovations

This study provides three key advances in adolescent suicide-risk prediction:

1. A rigorous comparative validation of machine learning algorithms in a non-Western, school-based setting. We demonstrate that while algorithms like KNN and Random Forest overfit training data, XGBoost offers superior generalizability (AUC = 0.802), providing empirical guidance for model selection in noisy mental health data.
2. An interpretable, theory-informed risk architecture. Using SHAP analysis, we move beyond black-box prediction to identify a core triad of risk factors: depressed mood, self-dissatisfaction, and reluctance to seek help. This profile bridges Beck's cognitive theory and Joiner's Interpersonal Theory, offering a mechanistic lens for risk.
3. The identification of a high-risk, under-screened subgroup. We reveal a critical gender-specific pattern: males with high distress who avoid help exhibit disproportionately elevated risk, a "silent-risk" profile often missed by conventional screening focused on verbalized ideation.

Collectively, our work delivers a validated, interpretable tool that addresses gaps in existing research and offers a practical strategy for early identification in schools.

7. Conclusion

This study strengthens the evidence that a core triad of psychological and behavioral factors—depressed mood, self-dissatisfaction, and reluctance to seek help—are critical in predicting suicidal tendency among Chinese adolescents. Our

findings identify a high-risk, under-recognized subgroup: males with high distress who avoid help. The results suggest that school-based screening and prevention programs should move beyond assessing ideation alone, and instead integrate early identification of these key risk factors, particularly by addressing help-avoidance and providing targeted support for vulnerable subgroups.

Abbreviations

AUC Area Under the Curve

DCA Decision Curve Analysis

DT Decision Tree

KNN K-Nearest Neighbors

LASSO Least Absolute Shrinkage and Selection Operator

LGBM Light Gradient Boosting Machine

LR Logistic Regression

MINI Mini-International Neuropsychiatric Interview

ML Machine Learning

MVPA Moderate-to-Vigorous Physical Activity

PHQ-9 Patient Health Questionnaire-9

RF Random Forest

ROC Receiver Operating Characteristic

SHAP SHapley Additive exPlanations

SVM Support Vector Machine

XGB XGBoost

Declarations

Ethics approval and consent to participate

The study received ethics approval from the Suzhou Guangji Hospital Ethics Committee (Approval Number: 2017037). This study was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study, and where participants were under the age of 16, consent was also obtained from their parents or legal guardians.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

The study was supported by grant LCZX202326 from the Suzhou Key Diagnosis and Treatment Program, grant MSXM2024034 from the Suzhou Municipal Health Commission General Project, and grant SYW2025061 from the Suzhou Municipal Applied Basic Research (Medical and Health) General Project.

Authors' contributions

J.C. and Y.W. contributed to the conception and design. R.J.Y. and L.Y.G. contributed to writing and editing. H.L.X., H.B.C., T.T.T., Y.Y.C., P.Y., and J.L. were responsible for data collection and quality control. Y.Y. and C.Y. contributed to the critical revision, final approval, and patient management. All authors contributed to the drafting of the article, critical revision, and approval of the submitted version.

Acknowledgements

The authors would like to express gratitude to all individuals who participated in this study.

References

1. Boerma T, Mathers CD. The World Health Organization and global health estimates: improving collaboration and capacity. BMC Med. 2015;13:50.
2. Casey BJ, Jones RM. Neurobiology of the adolescent brain and behavior: implications for substance use disorders. J Am Acad Child Adolesc Psychiatry. 2010;49(12):1189–201.
3. Qu Y, Pomerantz EM, Deng C. Mothers' goals for adolescents in the United States and China: Content and transmission. J Res Adolesc. 2016;26(1):126–41.
4. Bhagavathi P, Vitone T. Analysis of adolescent barriers in seeking help for mental health issues. J Stud Res. 2022;11(3).
5. Robbins DR, Alessi NE. Depressive symptoms and suicidal behavior in adolescents. 1986 Annual Progress In Child Psychiatry. Routledge; 1991. pp. 492–502.
6. Wilcox HC, Arria AM, Caldeira KM, Vincent KB, Pinchevsky GM, O'Grady KE. Prevalence and predictors of persistent suicide ideation, plans, and attempts during college. J Affect Disord. 2010;127(1–3):287–94.
7. Wang Z, Zhang L, Liu X, Li Y. Risk and protective factors of suicidal tendencies among freshmen in China revealed by a hierarchical regression model. Eur Child Adolesc Psychiatry. 2024;33(9):3043–53.
8. Ray JV, Poythress NG, Weir JM, Rickelm A. Relationships between psychopathy and impulsivity in the domain of self-reported personality features. Pers Individ Differ. 2009;46(2):83–7.
9. Cheung VHM, Chan CY, Au RKC. The influence of resilience and coping strategies on suicidal ideation among Chinese undergraduate freshmen in Hong Kong. Asia-Pac Psychiatry. 2019;11(2):e12339.
10. Jacobucci R, Littlefield AK, Millner AJ, Kleiman E, Steinley D. Pairing machine learning and clinical psychology: how you evaluate predictive performance matters. PsyArXiv. 2020. https://doi.org/10.31234/osf.io/2yber
11. Vally Z, Helmy M. The prevalence of suicidal behaviour and its associated risk factors among school-going adolescents resident in the United Arab Emirates. Sci Rep. 2023;13(1):19937.
12. Yan J, Liu Y, Yu J, Liao L, Wang H. Establishment and validation of a nomogram for suicidality in Chinese secondary school students. J Affect Disord. 2023;330:148–57.

13. Lee J, Choi H, Kim MJ, Park CG. Anger as a predictor of suicidal ideation in middle-school students in Korea: gender difference in threshold point. Adolescence. 2009;44(174).
14. Lew B, Huen J, Yu P, Yuan L, Wang DF, Ping F, et al. Meaning in life as a protective factor against suicidal tendency in Chinese University students. BMC Psychiatry. 2020;20(1):73.
15. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.
16. World Health Organization. Global recommendations on physical activity for health. World Health Organization; 2010. https://apps.who.int/iris/handle/10665/44399.
17. Zheng R, Wang H, Li W, Xing Y. The prevalence of bullying behavior among primary and secondary school students in a northern Chinese city. Chin Sch Health. 2018;39(1):76–80.
18. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59(Suppl 20):22–33.
19. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.
20. Campisi SC, Cardoso P, Akseer N, Bhatia A, Alaazi DA, Amuyunzu-Nyamongo M, et al. Suicidal behaviours among adolescents from 90 countries: a pooled analysis of the global school-based student health survey. BMC Public Health. 2020;20(1):1102.
21. Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35(8):1798–828.
22. Breiman L. Bagging predictors. Mach Learn. 1996;24(2):123–40.
23. Beck AT. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
24. Joiner TE. Why People Die by Suicide. Cambridge, MA: Harvard University Press; 2005.
25. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. JAMA. 1999;282(18):1737–44.
26. Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, et al. Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. Br J Psychiatry. 2008;192(2):98–105.
27. Levant RF, Wong YJ. The Psychology of Men and Masculinities. Washington, DC: American Psychological Association; 2017.
28. Cummings EM, Schermerhorn AC, Davies PT, Goeke-Morey MC, Cummings JS. Interparental discord and child adjustment: Prospective investigations of emotional security as an explanatory mechanism. Child Dev. 2006;77(1):132–52.
29. Hawker DSJ, Boulton MJ. Twenty years' research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. J Child Psychol Psychiatry. 2000;41(4):441–55.
30. Schmalbach I, Herhaus B, Pässler S, Runst S, Berth H, Wolff-Stephan S, et al. Cortisol reactivity in patients with anorexia nervosa after stress induction. Transl Psychiatry. 2020;10(1):275.
31. Pollina-Pocallet M, Artigues-Barberà E, Tort-Nasarre G, Sol J, Azlor L, Foguet-Boreu Q, et al. Self-perception and self-acceptance are related to unhealthy weight control behaviors in Catalan adolescents: A cross-sectional study. Int J Environ Res Public Health. 2021;18(9):4976.
32. Hochard KD, Heym N, Townsend E. Sleep Arousal and Acquired Capability. Department of Psychology, University of Chester; 2022.

Figures

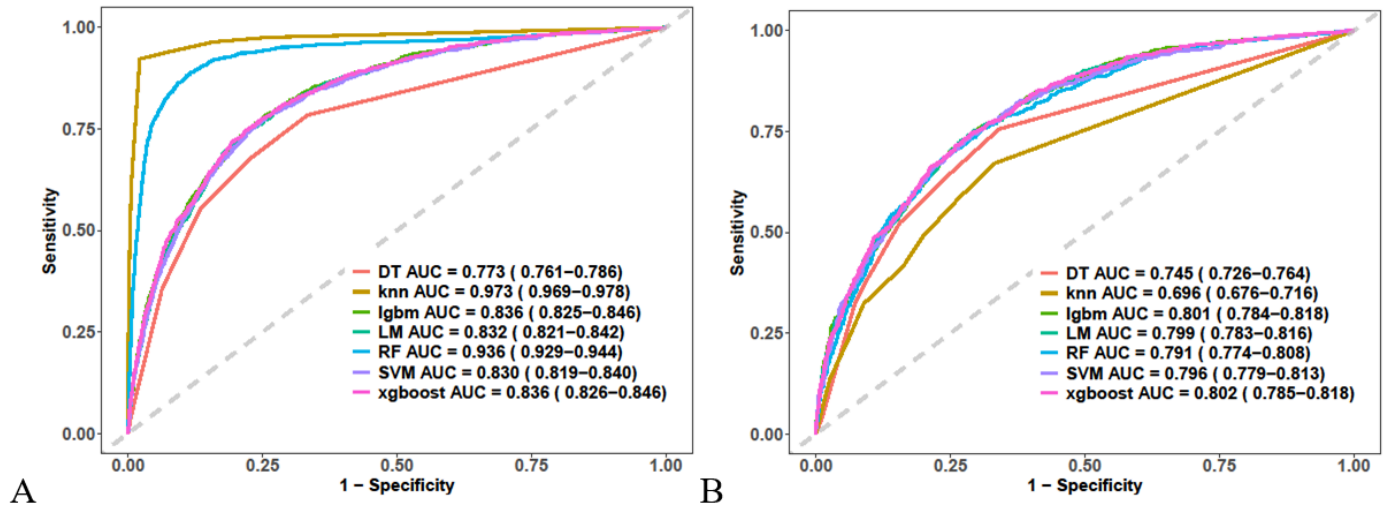


Figure 1

ROC curve comparison: A: Training set, B: Validation set

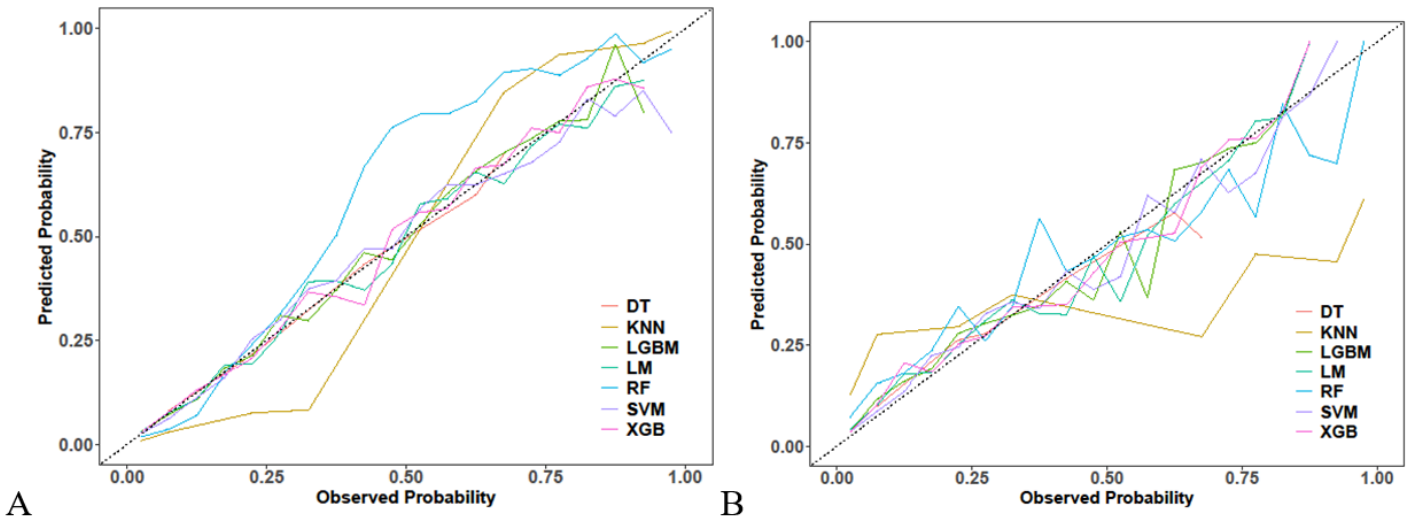


Figure 2

Calibration Curve, A: Training Set, B: Validation Set

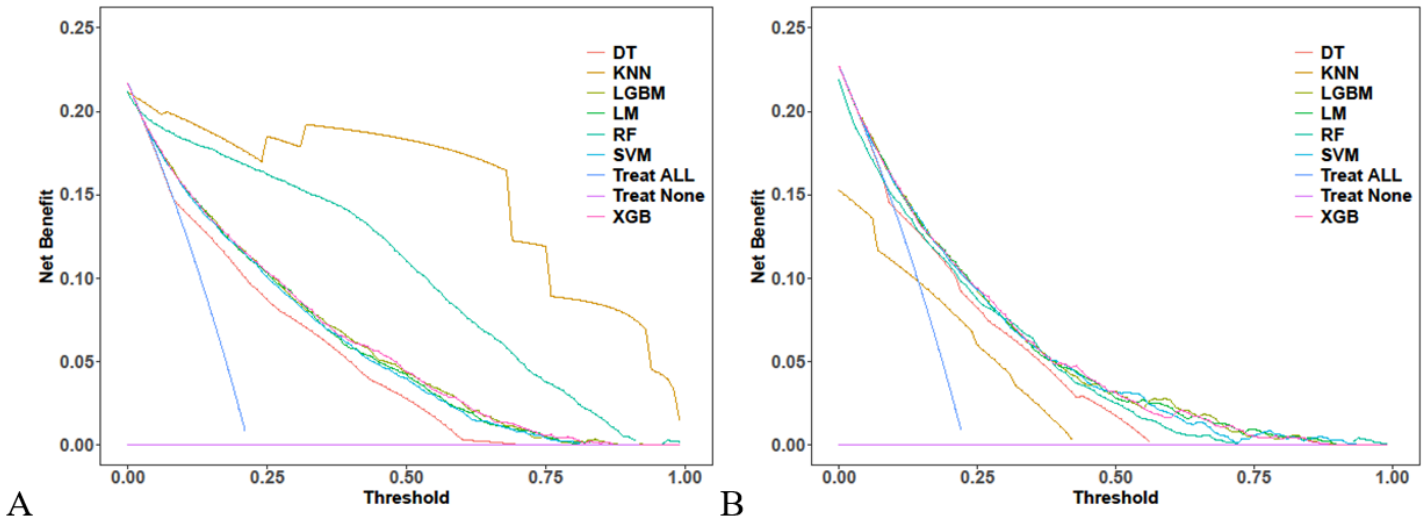


Figure 3

Feature selection of the final model: A: Training set, B: Validation set

The clinical applicability of XGB in the training set ranges from 4% to 87.5%; in the validation set, the clinical applicability of XGB ranges from 5% to 89%.

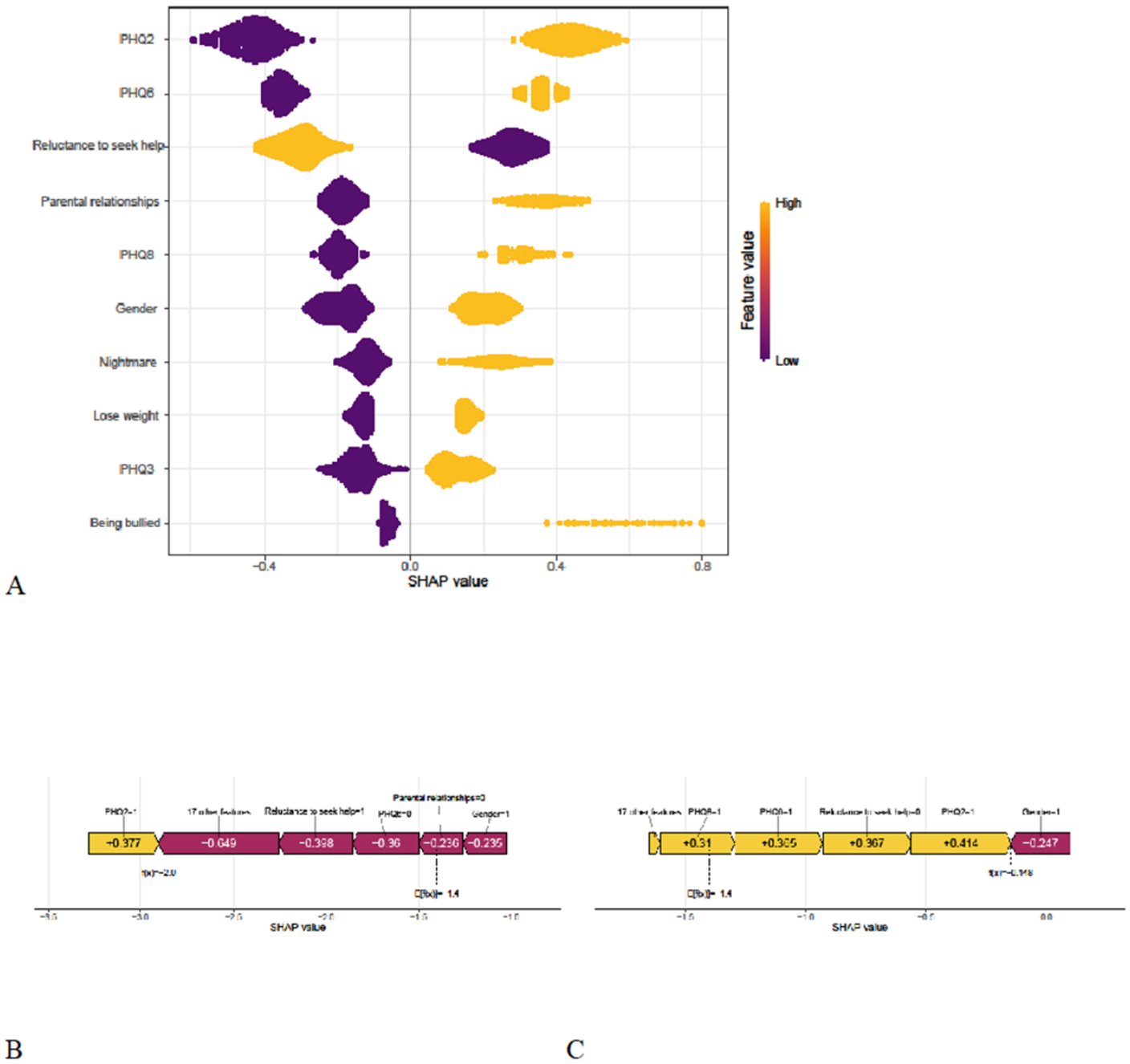


Figure 4

Interpretability of the optimal XGBoost model using SHAP.

(A) Global feature importance ranked by mean absolute SHAP values.

(B) Local explanation (force plot) for a low-risk case (ID: 185, predicted probability = 5.2%).

(C) Local explanation (force plot) for a high-risk case (ID: 50, predicted probability = 46.3%).