

A School–Family Framework for Initial Identification of Suicide-Related Vulnerability in Children and Adolescents: Development and Geographic External Validation

Jie Liu¹

Children's Hospital of Chongqing Medical University

Ning Lan¹

Children's Hospital of Chongqing Medical University

Chunbo Wang²

The First Affiliated Hospital of Chongqing Medical University

Zhijuan Li³

Xi'an Children's Hospital

Jiangling Su⁴

Yuxi Children's Hospital

Lu Tang

Children's Hospital of Chongqing Medical University

Gaofu Zhang

zhanggaofu19992004@163.com

Children's Hospital of Chongqing Medical University

Research Article

Keywords: Suicide-related thoughts and behaviors, adolescents, risk screening, machine learning, mental health

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10862310/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Suicide-related thoughts and behaviors (STBs) are important indicators of suicide risk among children and adolescents and represent a major public health concern. Although validated suicide-specific instruments are essential for formal risk assessment, large-scale implementation may be challenging in school, community, and primary care settings where specialized mental health resources are limited. This study aimed to develop and externally validate a school–family-based framework using routinely available, non-sensitive information to support the initial identification of suicide-related vulnerability.

Methods

This multicenter school-based cross-sectional study included 29,662 children and adolescents from Chongqing, Shaanxi, and Yunnan provinces in China. STBs were operationalized as a broad composite indicator encompassing self-harm behaviors, suicidal ideation, or suicidal behaviors. The Chongqing cohort was used for model development and internal validation, while the Shaanxi and Yunnan cohorts served as independent external validation datasets. Predictors were selected using least absolute shrinkage and selection operator regression. Logistic regression, random forest, extreme gradient boosting, and light gradient boosting machine models were evaluated for discrimination, calibration, and potential utility. Shapley additive explanations were used for model interpretation.

Results

Twenty-two predictors were retained. Models demonstrated stable performance across validation cohorts, with AUROC values ranging from 0.783 to 0.817 in internal validation, 0.795 to 0.849 in Shaanxi external validation, and 0.785 to 0.840 in Yunnan external validation. Logistic regression showed comparable discrimination with favorable calibration and interpretability and was selected as the final model. Major contributors included child anxiety, family functioning, sleep characteristics, emotional and behavioral difficulties, and developmental factors.

Conclusions

Routinely available, non-sensitive school–family information supported the identification of suicide-related vulnerability across two geographically independent external populations. The proposed framework may provide a low-barrier approach to initial risk stratification in settings with limited specialized mental health resources, helping prioritize potentially vulnerable children for subsequent suicide-specific assessment and professional evaluation. Prospective studies are needed to evaluate its real-world implementation and impact.

Introduction

Suicidal ideation, suicide attempts, and self-harm behaviors represent major and closely related public health concerns among children and adolescents. Suicide remains among the leading causes of death in young populations, ranking as the second leading cause of death among individuals aged 10–14 years and the third among adolescents and young adults aged 15–24 years(1). In China, suicide-related mortality among young populations has also shown an increasing trend in recent years, highlighting the urgent need for effective prevention strategies(2). Beyond suicide mortality, STBs represent important indicators of psychological vulnerability and are associated with subsequent suicidal behaviors, psychiatric morbidity, and adverse psychosocial outcomes(3, 4). A large meta-analysis involving more than 680,000 children and adolescents reported substantial lifetime prevalence of suicidal ideation, suicide attempts, and self-harm behaviors, further emphasizing the public health importance of timely risk identification and intervention(5).

Current approaches for identifying children and adolescents at risk of suicide commonly involve validated suicide-specific instruments, such as the Ask Suicide-Screening Questions (ASQ), Columbia-Suicide Severity Rating Scale (C-SSRS), and Patient Health Questionnaire-9 (PHQ-9)(6–8). These instruments remain essential for formal suicide risk assessment. However, their universal implementation across large school and community populations may be challenging, particularly in settings with limited trained personnel, time, referral capacity, or specialized mental health resources. In addition, some children and adolescents may be reluctant to disclose suicide-related experiences because of stigma, fear of negative consequences, or difficulties in recognizing and communicating emotional distress. A practical public health challenge is therefore how to identify children who may warrant further suicide-specific assessment when comprehensive assessment of an entire population is difficult to implement. Routinely available, non-sensitive information from school and family environments may provide a complementary first-stage approach to addressing this challenge.

STBs rarely occur in isolation but are associated with multiple psychological, behavioral, familial, and environmental vulnerabilities. Developmental perspectives suggest that suicide-related vulnerability reflects complex interactions between individual characteristics and contextual factors, including psychological difficulties, family environments, sleep problems, and social functioning. Previous studies have identified emotional symptoms, anxiety, depressive symptoms, sleep disturbances, family dysfunction, and social maladaptation as important correlates of suicidal outcomes among young people(9–15). Importantly, many of these indicators can be routinely assessed within school and family settings without requiring suicide-specific questioning. Integrating such multidomain information may therefore offer a low-barrier means of initial risk stratification and help prioritize children who may benefit from subsequent standardized suicide-specific assessment and professional evaluation.

Machine-learning approaches provide opportunities to integrate multidomain information and identify complex patterns across heterogeneous indicators(16, 17). However, methodological complexity alone does not ensure that a model will be useful for population-level risk identification. For public health

applications, models should demonstrate not only adequate discrimination but also calibration, transportability across independent populations, interpretability, and practical feasibility. Existing suicide-related prediction studies have frequently relied on clinical populations, psychiatric diagnoses, suicide-specific variables, or internal validation alone. Evidence regarding models based on routinely available, non-sensitive school–family information and evaluated across geographically independent populations remains limited. This gap is particularly relevant to school, community, and primary care settings where specialized mental health resources may be constrained.

Accordingly, this multicenter study aimed to develop and geographically validate a school–family-based framework for the initial identification of suicide-related vulnerability among children and adolescents using routinely available, non-sensitive information. We integrated psychological and behavioral characteristics, family environment, sleep-related factors, developmental indicators, and health-related information and compared interpretable statistical and machine-learning approaches across two independent external populations. Rather than replacing validated suicide-specific assessment, the proposed framework was intended to explore a low-barrier first-stage risk-stratification strategy for prioritizing potentially vulnerable children for subsequent suicide-specific screening and professional evaluation, particularly in settings with limited specialized mental health resources.

Methods

Study Design and Participants

This multicenter school-based cross-sectional study was conducted among children and adolescents from Chongqing, Shaanxi, and Yunnan provinces in China. The study was reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.

Participants were recruited through schools using a standardized questionnaire-based survey protocol. Before data collection, caregivers received standardized online training regarding questionnaire completion procedures and the interpretation of suicide-related items. Data quality control procedures included mandatory completion of key items, response-time monitoring, logical consistency checks, and telephone verification of potentially high-risk suicide-related thoughts and behaviors (STBs) responses by trained researchers.

After data cleaning and quality control, no missing values were present in the variables included in the analyses; therefore, no imputation was required. A total of 29,662 participants were included in the final analysis. The Chongqing cohort was used for model development and internal validation and was randomly divided into training (70%) and internal validation (30%) datasets using stratified sampling according to STBs status. The Shaanxi and Yunnan cohorts were reserved as geographically independent external validation datasets. The overall study workflow is presented in Fig. 1.

Outcome Definition

For the purposes of this study, STBs were operationalized as a broad composite indicator of suicide-related vulnerability encompassing caregiver-reported self-harm behaviors, suicidal ideation, or suicidal behaviors. Participants were classified as STBs-positive if any of these indicators were present and as STBs-negative otherwise.

Candidate Predictors

Candidate predictors were selected based on previous evidence regarding factors associated with STBs and their potential availability in school and family settings. Predictors were categorized into five theoretically informed domains: sociodemographic characteristics, psychological and behavioral characteristics, sleep-related characteristics, family environment, and developmental, perinatal, and health-related factors.

Sociodemographic characteristics included sex, age group, body mass index (BMI) category, household socioeconomic indicators, caregiver education, family structure, and academic performance.

Psychological and behavioral characteristics included separately assessed depressive symptoms, anxiety symptoms, temperament, and behavioral problems. In addition, child behavioral and emotional problems were assessed using the parent version of the Rutter Children's Behaviour Questionnaire(18).

Sleep-related characteristics included sleep duration and sleep problems assessed using the Children's Sleep Habits Questionnaire (CSHQ), a validated parent-reported instrument for pediatric sleep assessment(19). The CSHQ evaluates multiple sleep domains, including bedtime routine, sleep onset delay, sleep maintenance, sleep anxiety, night awakening, parasomnias, sleep-disordered breathing, and daytime sleepiness(20).

Family-related characteristics included family cohesion, family conflict, parental emotional status, family history indicators, parental marital status, caregiver characteristics, and environmental exposures.

Family cohesion and family conflict were assessed using the Cohesion and Conflict subscales of the Chinese version of the Family Environment Scale (FES-CV)(21), while parental anxiety and depression were evaluated using the Zung Self-Rating Anxiety Scale and Self-Rating Depression Scale, respectively(22).

Developmental, perinatal, and health-related characteristics included gestational age, birth weight, pregnancy-related factors, urinary symptoms, respiratory conditions, and ADHD-related symptoms. Urinary symptoms, including nocturnal enuresis, urinary incontinence, bladder bowel dysfunction, and overactive bladder, were assessed according to established clinical definitions(23–26). ADHD-related symptoms were evaluated using the Swanson, Nolan, and Pelham Rating Scale-IV (SNAP-IV)(27).

Data Preprocessing and Feature Selection

All data preprocessing procedures were performed exclusively within the Chongqing training dataset to minimize information leakage. Continuous variables were standardized when applicable, and categorical variables were converted into dummy variables based on the levels observed in the training dataset. The

same preprocessing parameters and variable-encoding scheme derived from the training dataset were subsequently applied to the internal and external validation datasets.

Four prediction models were developed using the predictors retained through LASSO: logistic regression, random forest (RF), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM). Logistic regression was included as an interpretable reference model, whereas the tree-based algorithms were used to capture potential nonlinear relationships and higher-order interactions among predictors. All model development procedures were conducted exclusively within the Chongqing training dataset. The resulting models were subsequently evaluated without model refitting in the internal validation dataset and the two external validation cohorts.

The complete R code used for model development is provided in the Supplementary Materials to facilitate reproducibility.

Model Evaluation

Model performance was evaluated in terms of discrimination, calibration, and potential utility. Discrimination was assessed using the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC). Given the relatively low prevalence of STBs, AUPRC was reported alongside AUROC to provide additional information regarding positive-case identification under class imbalance.

Calibration was evaluated using calibration curves, calibration-in-the-large (CITL), calibration slope, and the Brier score. These measures were used to assess agreement between model-estimated probabilities and observed outcomes.

Potential utility was explored using decision curve analysis (DCA), which estimates model-based net benefit across a range of threshold probabilities and compares model-guided strategies with reference strategies of treating all or no participants as positive.

Additional classification metrics, including sensitivity, specificity, accuracy, and F1-score, were calculated using a data-driven cutoff based on the Youden index. The Youden index was used to provide a standardized cutoff for model comparison rather than to define a definitive clinical or population-screening threshold.

Model selection considered discrimination, calibration, performance consistency across independent validation populations, interpretability, and potential implementation feasibility rather than relying solely on apparent performance in the training dataset.

External Validation

Independent geographic external validation was conducted using cohorts from Shaanxi and Yunnan provinces. Models developed using the Chongqing training dataset were directly applied to the two

external validation cohorts without model refitting, additional training, or recalibration.

Model performance in each external validation cohort was assessed independently using the same evaluation framework as that used for internal validation, including discrimination, calibration, classification metrics, and DCA. This approach was used to evaluate model performance across geographically distinct populations.

Model Interpretation Using SHAP

Model interpretability was assessed using Shapley Additive Explanations (SHAP). SHAP values were used to quantify the contribution of individual predictors to model predictions and to identify variables contributing most strongly to model-estimated STBs risk.

Global feature-importance and summary plots were generated to characterize overall predictor contributions across models. Individual-level SHAP visualizations were additionally generated for the selected final model to illustrate how individual predictor values contributed to model-estimated risk. SHAP results were interpreted as explanations of model predictions rather than as evidence of causal relationships between individual predictors and STBs.

Web-Based Implementation Prototype

To demonstrate potential implementation and facilitate accessibility of the final model, the selected logistic regression model was implemented as an interactive web-based prototype. The prototype generates model-estimated probabilities of STBs based on routinely available school- and family-related predictor information entered by the user.

The prototype was developed to illustrate a potential first-stage risk-stratification approach rather than to provide a standalone diagnostic or clinical decision-making tool. It is intended to complement, rather than replace, validated suicide-specific screening instruments and professional mental health assessment. Its usability, effectiveness, optimal implementation threshold, and impact on referral or mental health outcomes were not evaluated in the present study and require prospective evaluation before routine implementation.

Results

Participant Characteristics

A total of 29,662 children and adolescents from Chongqing, Shaanxi, and Yunnan provinces were included, of whom 934 (3.15%) were classified as having STBs and 28,728 (96.85%) as non-STBs. The Chongqing cohort (n = 16,686) was used for model development and internal validation, while the Shaanxi (n = 8,307) and Yunnan (n = 4,669) cohorts were used for independent external validation. The prevalence of STBs was 3.01% (250/8,307) in Shaanxi and 2.21% (103/4,669) in Yunnan.

Baseline characteristics according to STBs status are summarized in Table 1, with detailed distributions of candidate predictors presented in Supplementary Table S1. Compared with participants without STBs, those with STBs showed differences across sociodemographic, psychological and behavioral, family-related, sleep-related, and developmental characteristics. Overall, participants with STBs exhibited greater emotional and behavioral difficulties, more adverse family characteristics, and more sleep problems, highlighting the multidomain nature of STBs risk.

Table 1
Baseline characteristics of participants according to STBs status

Variable	STBs group	Non-STBs group	Test statistic	P
Gender				
male	387 (41.43%)	14666 (51.05%)		
female	547 (58.57%)	14062 (48.95%)	$\chi^2 = 33.086$	< 0.001
Age Group				
5–9	142 (15.2%)	11265 (39.21%)		
10–14	514 (55.03%)	13155 (45.79%)		
15–19	278 (29.76%)	4308 (15%)	H = 279.959	< 0.001
BMI				
Underweight	83 (8.89%)	3595 (12.51%)		
Normal	586 (62.74%)	16351 (56.92%)		
Overweight	111 (11.88%)	3497 (12.17%)		
obesity	154 (16.49%)	5285 (18.4%)	H = 16.831	< 0.001
Income				
< 30000	363 (38.87%)	10312 (35.9%)		
30000–80000	350 (37.47%)	10817 (37.65%)		
80000–150000	150 (16.06%)	5156 (17.95%)		
150000–300000	56 (6%)	2010 (7%)		
300000–1000000	13 (1.39%)	362 (1.26%)		
> 1000000	2 (0.21%)	71 (0.25%)	H = 5.492	0.359
Caregiver Education Level				
Junior high school and below	635 (67.99%)	17427 (60.66%)		
High School	189 (20.24%)	6775 (23.58%)		
Undergraduate	98 (10.49%)	4411 (15.35%)		
Master's degree or higher	12 (1.28%)	115 (0.4%)	H = 42.882	< 0.001
<i>λ Note: Data are presented as number (%) for categorical variables and median (interquartile range) for continuous variables. Group comparisons were performed using the chi-square test for categorical variables and the Kruskal–Wallis H test for continuous or ordinal variables, as appropriate. STBs, suicide-related thoughts and behaviors.</i>				

Variable	STBs group	Non-STBs group	Test statistic	P
Place of residence				
City	384 (60.09%)	13347 (66.95%)		
Non-urban	255 (39.91%)	6589 (33.05%)	$\chi^2 = 12.801$	< 0.001
Recent Child Depression				
Nil	612 (65.52%)	18856 (65.64%)		
Slight	287 (30.73%)	9154 (31.86%)		
Moderate	26 (2.78%)	484 (1.68%)		
Substantial	9 (0.96%)	234 (0.81%)	H = 6.968	0.073
Recent Child Anxiety				
Nil	154 (16.49%)	18321 (63.77%)		
Slight	478 (51.18%)	9773 (34.02%)		
Moderate	172 (18.42%)	458 (1.59%)		
Substantial	130 (13.92%)	176 (0.61%)	H = 3159.27	< 0.001
ADHD-I				
Normal	847 (90.69%)	28297 (98.5%)		
Abnormal	87 (9.31%)	431 (1.5%)	$\chi^2 = 317.402$	< 0.001
ADHD-HI				
Normal	864 (92.51%)	28138 (97.95%)		
Abnormal	70 (7.49%)	590 (2.05%)	$\chi^2 = 120.602$	< 0.001
ODD				
Normal	859 (91.97%)	28500 (99.21%)		
Abnormal	75 (8.03%)	228 (0.79%)	$\chi^2 = 461.366$	< 0.001
Antisocial Behavior (A)				
Normal	923 (98.82%)	28446 (99.02%)		
Abnormal	11 (1.18%)	282 (0.98%)	$\chi^2 = 0.183$	0.668
<i>λ Note: Data are presented as number (%) for categorical variables and median (interquartile range) for continuous variables. Group comparisons were performed using the chi-square test for categorical variables and the Kruskal–Wallis H test for continuous or ordinal variables, as appropriate. STBs, suicide-related thoughts and behaviors.</i>				

Variable	STBs group	Non-STBs group	Test statistic	P
Neurotic behavior (N)				
Normal	491 (52.57%)	23873 (83.1%)		
Abnormal	443 (47.43%)	4855 (16.9%)	$\chi^2 = 572.646$	< 0.001
Hybrid (M)				
Normal	886 (94.86%)	27573 (95.98%)		
Abnormal	48 (5.14%)	1155 (4.02%)	$\chi^2 = 2.629$	0.105
Urinary Incontinence				
No	853 (91.33%)	27360 (95.24%)		
Yes	81 (8.67%)	1368 (4.76%)	$\chi^2 = 28.935$	< 0.001
Bladder Bowel Dysfunction				
No	746 (79.87%)	26752 (93.12%)		
Yes	188 (20.13%)	1976 (6.88%)	$\chi^2 = 232.866$	< 0.001
Overactive Bladder				
No	826 (88.44%)	27111 (94.37%)		
Yes	108 (11.56%)	1617 (5.63%)	$\chi^2 = 57.086$	< 0.001
Overall Sleep				
Normal	136 (14.56%)	8931 (31.09%)		
Abnormal	798 (85.44%)	19797 (68.91%)	$\chi^2 = 115.641$	< 0.001
Family History of Enuresis				
Yes	144 (15.42%)	2278 (7.93%)		
No	790 (84.58%)	26450 (92.07%)	$\chi^2 = 66.645$	< 0.001
Family History of ADHD				
Yes	164 (17.56%)	2244 (7.81%)		
No	770 (82.44%)	26484 (92.19%)	$\chi^2 = 113.928$	< 0.001
Parental Anxiety				
<i>λ Note: Data are presented as number (%) for categorical variables and median (interquartile range) for continuous variables. Group comparisons were performed using the chi-square test for categorical variables and the Kruskal–Wallis H test for continuous or ordinal variables, as appropriate. STBs, suicide-related thoughts and behaviors.</i>				

Variable	STBs group	Non-STBs group	Test statistic	P
Normal	706 (75.59%)	26354 (91.74%)		
Mild anxiety	156 (16.7%)	2035 (7.08%)		
Moderate anxiety	52 (5.57%)	278 (0.97%)		
Severe anxiety	20 (2.14%)	61 (0.21%)	H = 434.435	< 0.001
Parental Depression				
Normal	458 (49.04%)	18531 (64.51%)		
Mild depression	265 (28.37%)	5911 (20.58%)		
Moderate depression	191 (20.45%)	4082 (14.21%)		
Severe depression	20 (2.14%)	204 (0.71%)	H = 109.209	< 0.001
Family Cohesion				
Low intimacy	137 (14.67%)	1717 (5.98%)		
Moderate level	543 (58.14%)	11811 (41.11%)		
High intimacy	254 (27.19%)	15200 (52.91%)	H = 287.072	< 0.001
Family Conflict				
Low conflict	585 (62.63%)	25077 (87.29%)		
Moderate level	329 (35.22%)	3596 (12.52%)		
High conflict	20 (2.14%)	55 (0.19%)	H = 552.064	< 0.001
<i>λ Note: Data are presented as number (%) for categorical variables and median (interquartile range) for continuous variables. Group comparisons were performed using the chi-square test for categorical variables and the Kruskal–Wallis H test for continuous or ordinal variables, as appropriate. STBs, suicide-related thoughts and behaviors.</i>				

Predictor Selection

LASSO logistic regression with ten-fold cross-validation was performed within the training dataset. The $\lambda_{\min}$ criterion retained 49 predictors with non-zero coefficients, whereas the more parsimonious λ_{1se} criterion retained 31 predictors. Considering model parsimony and implementation feasibility, λ_{1se} was selected for subsequent model development. After consolidation of dummy variables derived from categorical predictors, 22 variables spanning psychological, family-related, sleep-related, developmental, and perinatal domains were retained.

Model Development and Internal Validation

Four models—Logistic Regression, Random Forest, XGBoost, and LightGBM—were developed using the selected predictors. All models showed high apparent discrimination in the training dataset, with Random Forest achieving the highest performance (AUROC = 0.954; AUPRC = 0.868). (Table 2)

Table 2
Apparent performance of machine learning models in the training dataset

Model	AUROC	AUPRC	Best_Threshold	Accuracy	Sensitivity	Specificity	F1
Logistic	0.872	0.358	0.025	0.742	0.842	0.738	0.185
Random Forest	0.954	0.868	0.085	0.968	0.865	0.972	0.653
XGBoost	0.893	0.451	0.034	0.808	0.820	0.808	0.229
LightGBM	0.936	0.628	0.063	0.906	0.793	0.910	0.370
<i>λ Note: AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve. The optimal threshold was determined using the Youden index. Accuracy, sensitivity, specificity, and F1-score were calculated based on the optimal threshold.</i>							

In the independent internal validation cohort, model performance was more comparable, with AUROC values ranging from 0.783 to 0.817 and AUPRC values from 0.249 to 0.286. XGBoost (AUROC = 0.817) and Logistic Regression (AUROC = 0.816) showed similar discrimination (Table 3). Calibration analyses demonstrated overall agreement between predicted and observed risks, with Logistic Regression showing favorable calibration (calibration slope = 0.942; Brier score = 0.024). Decision curve analysis further indicated positive net benefit across clinically relevant threshold probabilities compared with the treat-all and treat-none strategies.

Table 3
Performance of machine learning models in the internal validation dataset

Model	AUROC	AUPRC	Best_Threshold	Accuracy	Sensitivity	Specificity	F1
Logistic	0.816	0.272	0.025	0.731	0.777	0.730	0.168
Random Forest	0.783	0.249	0.027	0.820	0.646	0.827	0.201
XGBoost	0.817	0.286	0.020	0.681	0.823	0.676	0.153
LightGBM	0.812	0.264	0.019	0.701	0.800	0.698	0.158
<i>λ Note: AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve. The optimal threshold was determined using the Youden index. Accuracy, sensitivity, specificity, and F1-score were calculated based on the optimal threshold.</i>							

External Validation and Generalizability

Model generalizability was evaluated in two independent external validation cohorts. In the Shaanxi cohort, all models maintained good discrimination, with AUROC values ranging from 0.795 to 0.849 and

AUPRC values from 0.238 to 0.276. XGBoost achieved the highest AUROC (0.849), closely followed by Logistic Regression (0.846) (Table 4).

Table 4 External validation performance of machine learning models in Shaanxi and Yunnan cohorts						
Region	Model	AUROC (95%CI)	AUPRC	Brier Score	Calibration -in- the-large	Calibration Slope
Shaanxi	Logistic	0.846 (0.820– 0.871)	0.276	0.025	-0.214	0.936
Shaanxi	Random Forest	0.795 (0.763– 0.827)	0.238	0.026	-2.446	0.070
Shaanxi	XGBoost	0.849 (0.824– 0.874)	0.272	0.025	-0.072	0.980
Shaanxi	LightGBM	0.842 (0.816– 0.868)	0.258	0.026	-0.319	0.874
Yunnan	Logistic	0.840 (0.798– 0.882)	0.210	0.020	-0.750	0.885
Yunnan	Random Forest	0.785 (0.734– 0.836)	0.194	0.020	-2.948	0.067
Yunnan	XGBoost	0.834 (0.790– 0.877)	0.227	0.019	-0.547	0.945
Yunnan	LightGBM	0.825 (0.780– 0.870)	0.207	0.020	-0.783	0.840
λ CI, confidence interval; AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; Brier score, calibration-in-the-large, and calibration slope were used to evaluate model calibration performance.						

Similar performance was observed in the Yunnan cohort, where AUROC values ranged from 0.785 to 0.840 and AUPRC values from 0.194 to 0.227. Logistic Regression achieved the highest AUROC (0.840) and showed favorable calibration, with a Brier score of 0.020 (Table 4). ROC and precision–recall curves showed broadly comparable discrimination across models in both external validation cohorts (Fig. 2).

Calibration analyses demonstrated overall agreement between predicted and observed risks across external validation datasets (Fig. 2). Decision curve analysis also showed positive net benefits relative to the treat-all and treat-none strategies across clinically relevant threshold probabilities (Fig. 3), supporting the potential utility of these models for population-level risk stratification.

Final Model Selection and Web-Based Implementation

No single algorithm consistently achieved the highest discrimination across internal and external validation cohorts. Although Random Forest showed the highest apparent performance in the training dataset, its discrimination decreased during validation. In contrast, Logistic Regression demonstrated consistently competitive discrimination, favorable calibration, positive clinical utility, and greater interpretability across validation datasets (Table 2; Table 3; Table 4; Figure 2; Figure 3).

Considering overall validation performance, calibration, interpretability, and implementation feasibility, Logistic Regression was selected as the final model. The selected model was subsequently implemented as a web-based risk screening tool using routinely available school- and family-based information to support individual risk estimation and initial stratification.

Factors Contributing to STBs Risk

SHAP analyses showed broadly consistent predictor contribution patterns across the four models, with psychological, family-related, sleep-related, and developmental characteristics contributing prominently to model predictions (Fig. 4).

Recent Child Anxiety showed the greatest overall contribution across models, while Age Group, Family Cohesion, Weekday Sleep Duration, Mood Disorders, and Family Conflict were also among the most influential predictors (Fig. 4). Although the relative ranking of individual predictors varied across algorithms, the overall pattern of contributions remained broadly consistent.

SHAP summary plots further indicated that greater anxiety symptoms, mood disorders, and family conflict generally contributed toward higher model-estimated STBs risk, whereas greater family cohesion contributed toward lower estimated risk (Fig. 4). Sleep-related characteristics, including sleep duration and sleep problems, also contributed substantially to model predictions.

Individual-level SHAP visualization based on the final Logistic Regression model demonstrated heterogeneous combinations of contributing factors among participants (Fig. 5). Individuals with similar estimated STBs risks could exhibit different psychological, family, sleep, and developmental contribution profiles, illustrating the multidimensional nature of STBs risk.

Discussion

Principal Findings

STBs among children and adolescents emerge from complex interactions among psychological characteristics, family environments, sleep conditions, physical health, and developmental vulnerabilities rather than from a single risk factor. In this multicenter study involving 29,662 children and adolescents from three regions in China, we developed and externally validated a school–family-based risk identification framework using routinely available non-sensitive information.

The major finding of this study was that children and adolescents with elevated STBs vulnerability could be identified without directly assessing suicidal thoughts or behaviors during initial screening. The developed models demonstrated stable performance across internal and independent external validation cohorts, supporting the feasibility of integrating routinely collected information into population-level risk stratification. Although complex ensemble algorithms demonstrated higher apparent performance during model development, no single algorithm consistently outperformed others across validation cohorts. Logistic Regression showed comparable discrimination with favorable calibration and interpretability, supporting its potential application in real-world screening contexts.

These findings extend previous approaches focused primarily on direct suicide assessment by suggesting that routinely observable characteristics within children's daily environments may provide complementary information for identifying individuals who may require further evaluation.

A School–Family Developmental Framework for Early STBs Identification

A major contribution of this study is the development of a school–family-based framework that integrates information from children's everyday environments to identify individuals with increased vulnerability to STBs. Current suicide screening approaches primarily rely on direct assessment of suicidal ideation, suicidal behaviors, or previous self-harm experiences. Although these approaches have demonstrated clinical utility, their application in universal school and community settings may be limited by the need for direct disclosure of sensitive experiences, trained professionals, referral pathways, and timely risk management.

In contrast, the present framework utilizes routinely available information across psychological, family, sleep, developmental, and health-related domains without requiring initial disclosure of suicide-related experiences. This approach is consistent with developmental ecological perspectives, in which children's mental health outcomes emerge from dynamic interactions between individual characteristics and surrounding environments.

Importantly, this framework should not replace established suicide-specific assessment tools. Instead, it may serve as an initial risk stratification approach to identify children who may benefit from subsequent standardized suicide assessment and professional evaluation. Such a two-stage strategy may improve the feasibility of population-based screening by allowing limited mental health resources to be directed toward children with greater potential vulnerability

Psychological and Family Vulnerability as Early Observable Signals

Psychological characteristics represented an important component of STBs vulnerability in this study. Recent child anxiety showed the greatest contribution across models, consistent with previous evidence linking anxiety symptoms, depressive symptoms, emotional dysregulation, and suicidal thoughts and behaviors among young populations(28, 29).

From a screening perspective, emotional symptoms represent accessible indicators that may be recognized before children disclose suicidal experiences. Persistent anxiety, emotional difficulties, somatic complaints, or behavioral changes may reflect underlying psychological vulnerability requiring further assessment. Previous population-based studies have shown that suicidal thoughts and behaviors frequently co-occur with broader emotional and behavioral difficulties, including mood, anxiety, and impulse-control problems(30, 31). Therefore, incorporating routinely assessed psychological and behavioral characteristics may facilitate earlier identification of children experiencing emerging psychological difficulties.

Family environments provide another important developmental context influencing emotional regulation, coping strategies, and resilience. Family-related indicators in this study, including family cohesion, family conflict, parental emotional status, family history of suicide, and caregiver characteristics, reflected broader psychosocial conditions surrounding children and adolescents. Previous studies have demonstrated that adverse family environments, including family conflict, reduced emotional support, and parental psychological difficulties, are associated with increased vulnerability to suicidal thoughts and behaviors(32, 33). Conversely, supportive family relationships may promote emotional security, adaptive coping, and resilience(33, 34).

Thus, family characteristics should be considered not only as predictive indicators but also as potentially modifiable intervention targets. Improving parent–child communication, reducing family conflict, and supporting parental mental health may provide opportunities for preventive intervention among children identified as vulnerable.

Sleep and Physical Health Indicators as Complementary Risk Signals

Sleep-related characteristics provided additional information beyond psychological and family domains. Sleep duration, sleep maintenance difficulties, night awakening, daytime sleepiness, and sleep-disordered breathing contributed to STBs identification, consistent with previous studies linking sleep disturbances with suicidal vulnerability among children and adolescents(35–37).

Sleep problems may represent broader psychosocial vulnerability rather than isolated symptoms. They may reflect interactions among emotional distress, family-related stressors, impaired daily functioning, and reduced emotion regulation capacity(36, 38, 39).

An important finding of this study was the contribution of bladder bowel dysfunction (BBD), a physical health-related factor that has been rarely considered in STBs prediction models. Previous studies have shown that children with lower urinary tract symptoms, including BBD, may experience psychosocial difficulties, behavioral problems, social impairment, and reduced quality of life(40–43).

Therefore, BBD may represent an accessible marker of broader psychosocial vulnerability rather than an independent determinant of STBs risk. The contribution of physical health indicators highlights the

importance of considering children's overall health conditions when developing comprehensive mental health screening strategies.

Developmental Vulnerability Across Childhood

Developmental and early-life characteristics, including age group, gestational age, and birth weight, provided additional information when integrated with current psychological and environmental factors. These characteristics should not be interpreted as independent causal determinants of STBs but may represent developmental vulnerability that interacts with later psychological experiences, family environments, and social functioning.

Age-related differences may reflect developmental changes during childhood and adolescence, including alterations in neurodevelopment, emotional regulation, social relationships, and stress responsiveness. Previous studies have suggested that developmental stage is associated with differences in vulnerability to psychological difficulties and suicidal thoughts and behaviors among young people(31, 44, 45)

Early-life factors, including gestational age and birth weight, may capture aspects of prenatal and perinatal developmental conditions that influence subsequent neurodevelopmental outcomes, emotional regulation, and behavioral adjustment(46, 47). Their contribution within a multidomain framework supports the concept that STBs vulnerability may accumulate across developmental stages through interactions between biological susceptibility and later psychosocial experiences.

Clinical Implications: From Risk Identification to Targeted Intervention

The findings of this study provide a potential pathway from early risk identification to targeted support. In practical settings, children may initially undergo risk stratification using routinely available school- and family-based information, including emotional symptoms, sleep characteristics, family functioning, developmental indicators, and physical health conditions. Individuals identified as potentially vulnerable could subsequently receive standardized suicide-specific assessment and professional psychological evaluation.

Importantly, risk screening should be considered an entry point rather than a diagnostic procedure. The purpose of this approach is not to replace comprehensive suicide assessment but to facilitate early identification and referral of children who may benefit from further evaluation and preventive support. Potential interventions may target modifiable vulnerability domains, including emotional difficulties, sleep problems, family functioning, and relevant physical health conditions.

To facilitate practical translation, the final Logistic Regression model was implemented as a web-based STBs risk screening tool. The tool enables individualized STBs risk estimation based on routinely available school- and family-based information and is intended to support early risk stratification and referral for professional assessment rather than replace comprehensive suicide-specific evaluation. The online prediction tool is available at: <https://stbs-predictor.shinyapps.io/stbs-predictor/>.

Future longitudinal studies should evaluate whether integrating risk stratification with targeted preventive strategies can improve referral efficiency and reduce the burden of STBs among children and adolescents.

Strengths and limitations

This study has several strengths. First, it included a large multicenter sample with independent external validation cohorts, strengthening confidence in model generalizability. Second, the framework was developed using routinely available non-sensitive information, supporting potential application in school- and family-based settings. Third, interpretable modeling approaches allowed identification of important contributing factors across multiple developmental domains.

Several limitations should also be acknowledged. First, the cross-sectional design prevents causal inference regarding associations between predictors and STBs. Second, STBs assessment relied on caregiver-reported information, which may not fully capture internal suicidal thoughts that are not disclosed. Future studies incorporating multi-informant assessments, including child self-report and clinician-rated measures, are warranted. Third, although the models demonstrated good performance, they should be considered screening tools rather than diagnostic instruments. Longitudinal studies are needed to evaluate temporal changes in risk and whether implementation of risk-based screening improves clinical and public health outcomes.

Abbreviations

Abbreviation	Full term
ADHD	Attention-deficit/hyperactivity disorder
AUPRC	Area under the precision–recall curve
AUROC	Area under the receiver operating characteristic curve
BBD	Bladder bowel dysfunction
CSHQ	Children’s Sleep Habits Questionnaire
DCA	Decision curve analysis
FES-CV	Chinese version of the Family Environment Scale
LASSO	Least absolute shrinkage and selection operator
LightGBM	Light Gradient Boosting Machine
PR	Precision–recall
RF	Random forest
SHAP	Shapley additive explanations
SNAP-IV	Swanson, Nolan, and Pelham Rating Scale-IV
STBs	Suicide-related thoughts and behaviors
XGBoost	Extreme Gradient Boosting

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Children's Hospital of Chongqing Medical University (Approval No. 2024-73) and was performed in accordance with the principles of the Declaration of Helsinki. Informed consent was obtained from all participants and their legal guardians. Guardians retained the right to withdraw their child from the study at any time without consequence.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form (including individual details, images, or videos).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. Data are not publicly available due to privacy restrictions and the terms of

the ethics approval, which limit data sharing to researchers who meet the criteria for access to confidential data.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Chongqing Science and Health Joint Medical Research Project (No. 2020FYYX18, No. 2024MSXM032); 2021 Chongqing Graduate Education Teaching Reform Research Project (yjg213057); 2021 Chongqing Medical University Graduate Education and Teaching Reform Research Project (xyjg210111). The funders had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Authors' contributions

JL: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. LN: Conceptualization, Data curation. CW: Conceptualization, Data curation, Resources, Investigation, Writing – review & editing. ZL: Conceptualization, Data curation, Resources, Investigation. JS: Resources, Investigation, Data curation. LT: Conceptualization, Methodology, Writing – review & editing. GZ: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

The authors express their sincere gratitude to Teacher Tu Peng from Chongqing Primary and Secondary School Health Care Center for her invaluable assistance in coordinating and completing the questionnaire survey. We also thank all the child guardians who participated in this study, as well as the teachers and school administrators who facilitated data collection across the three study regions.

References

1. Hua LL, Lee J, Rahmandar MH, Sigel EJ, Committee On A, Council O, Injury V et al. Suicide and Suicide Risk in Adolescents. *Pediatrics*. 2024;153(1).
2. Liu Y, Pan Y, Wang F, Li W, Gao L, Zhao G. Multivariate machine learning regression approaches to predict adolescent suicide risk. *Psychiatry Res*. 2026;361:117160.
3. Lewis CP, Klimes-Dougan B, Croarkin PE, Cullen KR. Understanding the emergence of suicidal thoughts and behaviors in adolescence from a brain and behavioral developmental perspective. *Neuropsychopharmacology*. 2026;51(1):259–72.
4. Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm. Systematic review. *Br J Psychiatry*. 2002;181:193–9.

5. Lim KS, Wong CH, McIntyre RS, Wang J, Zhang Z, Tran BX et al. Global Lifetime and 12-Month Prevalence of Suicidal Behavior, Deliberate Self-Harm and Non-Suicidal Self-Injury in Children and Adolescents between 1989 and 2018: A Meta-Analysis. *Int J Environ Res Public Health*. 2019;16(22).
6. Horowitz LM, Bridge JA, Teach SJ, Ballard E, Klima J, Rosenstein DL, et al. Ask Suicide-Screening Questions (ASQ): a brief instrument for the pediatric emergency department. *Arch Pediatr Adolesc Med*. 2012;166(12):1170–6.
7. Bjureberg J, Dahlin M, Carlborg A, Edberg H, Haglund A, Runeson B. Columbia-Suicide Severity Rating Scale Screen Version: initial screening for suicide risk in a psychiatric emergency department. *Psychol Med*. 2022;52(16):3904–12.
8. Fonseca-Pedrero E, Diez-Gomez A, Perez-Albeniz A, Al-Halabi S, Lucas-Molina B, Debbane M. Youth screening depression: Validation of the Patient Health Questionnaire-9 (PHQ-9) in a representative sample of adolescents. *Psychiatry Res*. 2023;328:115486.
9. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: A meta-analysis. *Sleep Med Rev*. 2018;38:3–16.
10. Kalin NH. Anxiety, Depression, and Suicide in Youth. *Am J Psychiatry*. 2021;178(4):275–9.
11. Serra G, De Crescenzo F, Maisto F, Galante JR, Iannoni ME, Trasolini M, et al. Suicidal behavior in juvenile bipolar disorder and major depressive disorder patients: Systematic review and meta-analysis. *J Affect Disord*. 2022;311:572–81.
12. Brent DA, Hur K, Gibbons JB, Porta G, Gibbons RD. Parental Diseases of Despair and Suicidal Events in Their Children. *JAMA Netw Open*. 2025;8(9):e2531442.
13. MacGregor EK, Grunebaum MF, Galfalvy HC, Melhem N, Burke AK, Brent DA, et al. Depressed parents' attachment: effects on offspring suicidal behavior in a longitudinal family study. *J Clin Psychiatry*. 2014;75(8):879–85.
14. Hammond NG, Semchishen SN, Geoffroy MC, Sikora L, Wafy G, Hsueh L, et al. Family dynamics and self-harm and suicidality in children and adolescents: a systematic review and meta-analysis. *Lancet Psychiatry*. 2025;12(9):660–72.
15. Miller AB, Esposito-Smythers C, Leichtweis RN. Role of social support in adolescent suicidal ideation and suicide attempts. *J Adolesc Health*. 2015;56(3):286–92.
16. Galatzer-Levy IR, Onnela JP. Machine Learning and the Digital Measurement of Psychological Health. *Annu Rev Clin Psychol*. 2023;19:133–54.
17. Kim J, Lee D, Park E. Machine Learning for Mental Health in Social Media: Bibliometric Study. *J Med Internet Res*. 2021;23(3):e24870.
18. Ekblad S. The children's behaviour questionnaire for completion by parents and teachers in a Chinese sample. *J Child Psychol Psychiatry*. 1990;31(5):775–91.
19. Owens JA, Spirito A, McGuinn M. The Children's Sleep Habits Questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep*. 2000;23(8):1043–51.

20. Li SH, Jin XM, Shen XM, Wu SH, Jiang F, Yan CH, et al. [Development and psychometric properties of the Chinese version of Children's Sleep Habits Questionnaire]. *Zhonghua Er Ke Za Zhi*. 2007;45(3):176–80.
21. Fei LPSQ, Zheng YP, Zhao JP, Jiang SA, Wang LW, Wang XD. Preliminary evaluation of the Family Cohesion and Adaptability Scale and the Family Environment Scale. *Chin Mental Health J*. 1991(5):198–202.
22. Zung WW. A rating instrument for anxiety disorders. *Psychosomatics*. 1971;12(6):371–9.
23. Neveus T, Eggert P, Evans J, Macedo A, Rittig S, Tekgul S, et al. Evaluation of and treatment for monosymptomatic enuresis: a standardization document from the International Children's Continence Society. *J Urol*. 2010;183(2):441–7.
24. Neveus T, Fonseca E, Franco I, Kawauchi A, Kovacevic L, Nieuwhof-Leppink A, et al. Management and treatment of nocturnal enuresis-an updated standardization document from the International Children's Continence Society. *J Pediatr Urol*. 2020;16(1):10–9.
25. Santos JD, Lopes RI, Koyle MA. Bladder and bowel dysfunction in children: An update on the diagnosis and treatment of a common, but underdiagnosed pediatric problem. *Can Urol Assoc J*. 2017;11(1–2Suppl1):S64–72.
26. Austin PF, Bauer SB, Bower W, Chase J, Franco I, Hoebeke P, et al. The standardization of terminology of lower urinary tract function in children and adolescents: update report from the Standardization Committee of the International Children's Continence Society. *J Urol*. 2014;191(6):1863–e513.
27. Bussing R, Fernandez M, Harwood M, Wei H, Garvan CW, Eyberg SM, et al. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: psychometric properties and normative ratings from a school district sample. *Assessment*. 2008;15(3):317–28.
28. Bentley KH, Franklin JC, Ribeiro JD, Kleiman EM, Fox KR, Nock MK. Anxiety and its disorders as risk factors for suicidal thoughts and behaviors: A meta-analytic review. *Clin Psychol Rev*. 2016;43:30–46.
29. Orri M, Galera C, Turecki G, Forte A, Renaud J, Boivin M, et al. Association of Childhood Irritability and Depressive/Anxious Mood Profiles With Adolescent Suicidal Ideation and Attempts. *JAMA Psychiatry*. 2018;75(5):465–73.
30. Stein DJ, Chiu WT, Hwang I, Kessler RC, Sampson N, Alonso J, et al. Cross-national analysis of the associations between traumatic events and suicidal behavior: findings from the WHO World Mental Health Surveys. *PLoS ONE*. 2010;5(5):e10574.
31. Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry*. 2013;70(3):300–10.
32. Zhang YY, Yang XF, Liu X, Jia CX. Longitudinal association of family conflict and suicidal behaviors among Chinese adolescents: The mediation effect of internalizing and externalizing problems. *J Affect Disord*. 2023;321:96–101.

33. Harris TL, Molock SD. Cultural orientation, family cohesion, and family support in suicide ideation and depression among African American college students. *Suicide Life Threat Behav.* 2000;30(4):341–53.
34. Buelga S, Cava MJ, Ortega-Baron J, Goossens L, Moreno Ruiz D. The Moderating Role of Family Communication in Suicidal Ideation and Victimization in Adolescents. *Psicothema.* 2024;36(4):369–79.
35. Miller BJ, McCall WV. Meta-analysis of insomnia, suicide, and psychopathology in schizophrenia. *Curr Opin Psychiatry.* 2023;36(3):156–65.
36. Asarnow LD, Mirchandaney R. Sleep and Mood Disorders Among Youth. *Psychiatr Clin North Am.* 2024;47(1):255–72.
37. Gowin JL, Stoddard J, Doykos TK, Sammel MD, Bernert RA. Sleep Disturbance and Subsequent Suicidal Behaviors in Preadolescence. *JAMA Netw Open.* 2024;7(9):e2433734.
38. Maratia F, Bacaro V, Crocetti E. Sleep Is a Family Affair: A Systematic Review and Meta-Analysis of Longitudinal Studies on the Interplay between Adolescents' Sleep and Family Factors. *Int J Environ Res Public Health.* 2023;20(5).
39. Gregory AM, Sadeh A. Sleep, emotional and behavioral difficulties in children and adolescents. *Sleep Med Rev.* 2012;16(2):129–36.
40. Marciano RC, Cardoso MGF, Vasconcelos MA, Paula JJ, Pinho NC, Oliveira AC, et al. Behavioral disorders and impairment of quality of life in children and adolescents with lower urinary tract dysfunction. *J Pediatr Urol.* 2018;14(6):568. e1- e7.
41. von Gontard A, Baeyens D, Van Hoecke E, Warzak WJ, Bachmann C. Psychological and psychiatric issues in urinary and fecal incontinence. *J Urol.* 2011;185(4):1432–6.
42. Ikeda H, Ono T, Oyake C, Oonuki Y, Watanabe Y, Watanabe T. Comparative analysis of health-related quality of life between children with bladder and bowel dysfunction versus lower urinary tract dysfunction and healthy controls. *Investig Clin Urol.* 2024;65(5):494–500.
43. Li T, Di X, Li Y, Wei J, Liao B, Wang K. The Association between Depression and Overactive Bladder: A Cross-Sectional Study of NHANES 2011–2018. *Int Urogynecol J.* 2025;36(2):373–80.
44. Bridge JA, Goldstein TR, Brent DA. Adolescent suicide and suicidal behavior. *J Child Psychol Psychiatry.* 2006;47(3–4):372–94.
45. Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? *J Child Psychol Psychiatry.* 2011;52(10):1015–25.
46. Arpino C, Compagnone E, Montanaro ML, Cacciato D, De Luca A, Cerulli A, et al. Preterm birth and neurodevelopmental outcome: a review. *Childs Nerv Syst.* 2010;26(9):1139–49.
47. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA.* 2002;288(6):728–37.

Figures

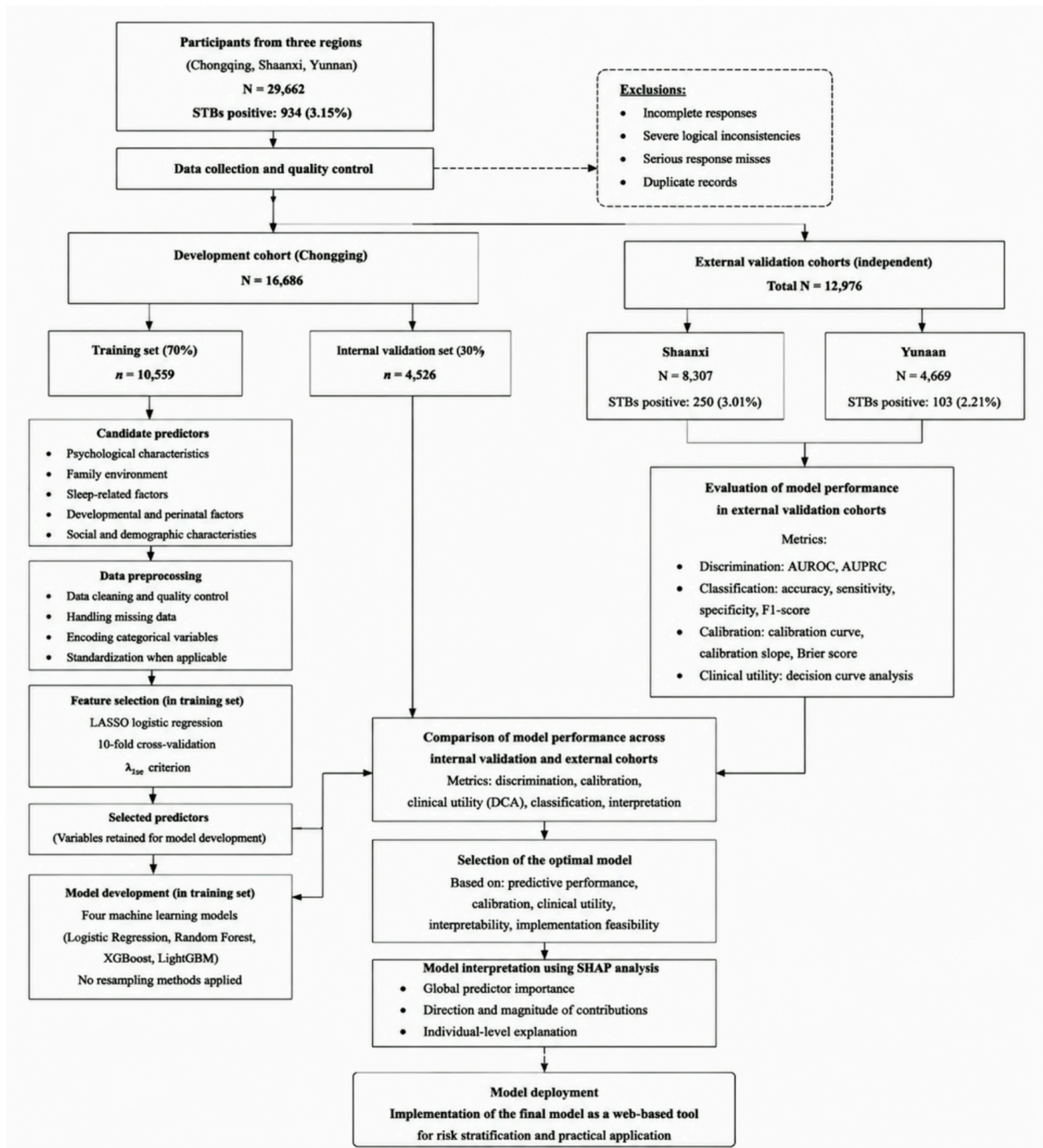


Figure 1

Flowchart

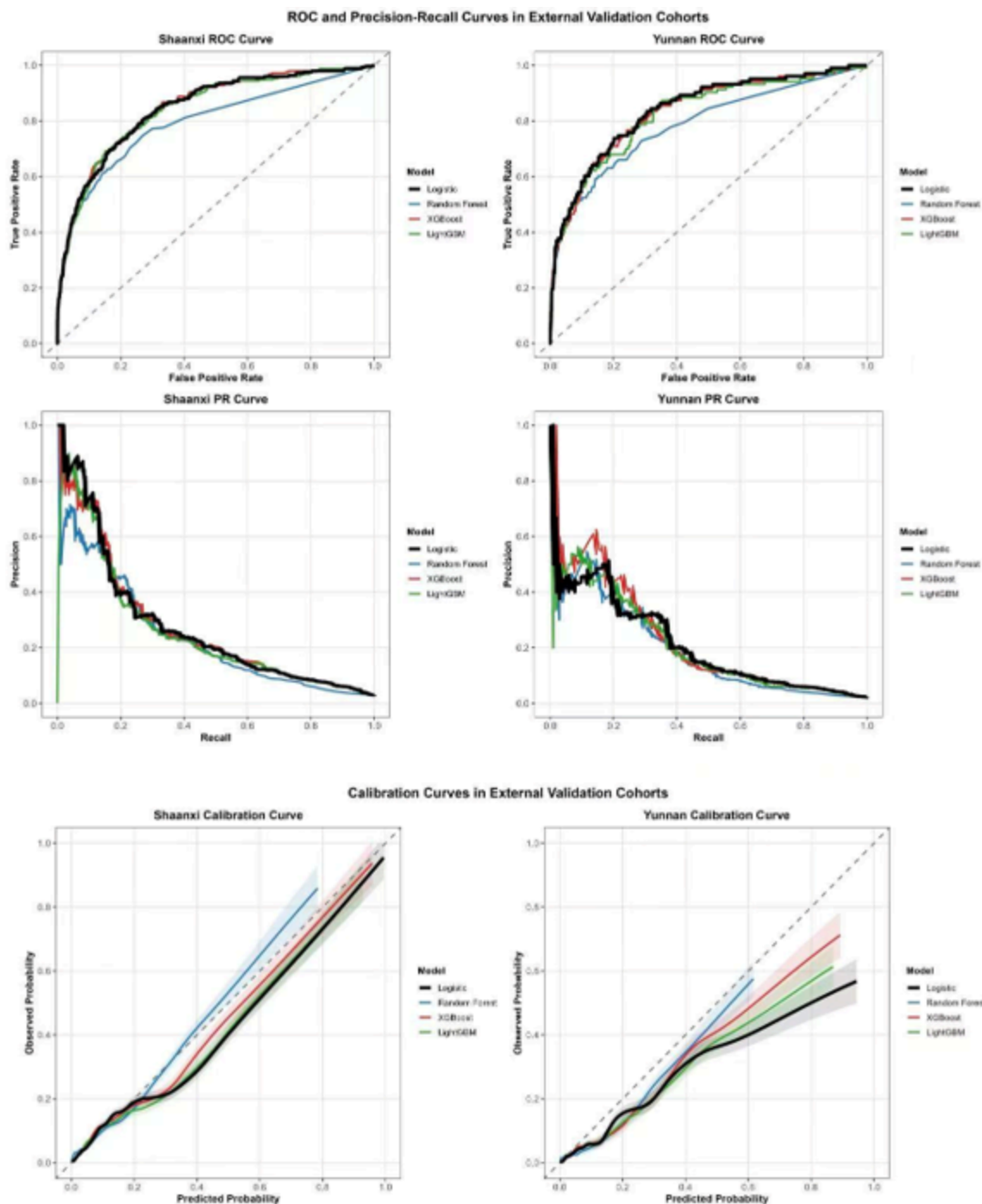


Figure 2

Performance of machine learning models in external validation cohorts

Note: ROC, receiver operating characteristic; PR, precision–recall. For ROC curves, the x-axis represents the false positive rate (1–specificity), and the y-axis represents the true positive rate (sensitivity). The area under the receiver operating characteristic curve (AUROC) reflects the overall discrimination ability of the model, with higher values indicating better discrimination. For PR curves, the x-axis represents recall, and the y-axis represents precision. The area under the precision–recall curve (AUPRC) reflects model performance in identifying positive cases under imbalanced outcomes, with higher values indicating better

performance. For calibration curves, the x-axis represents predicted probability, and the y-axis represents observed probability. Curves closer to the 45-degree reference line indicate better agreement between predicted and observed risks.

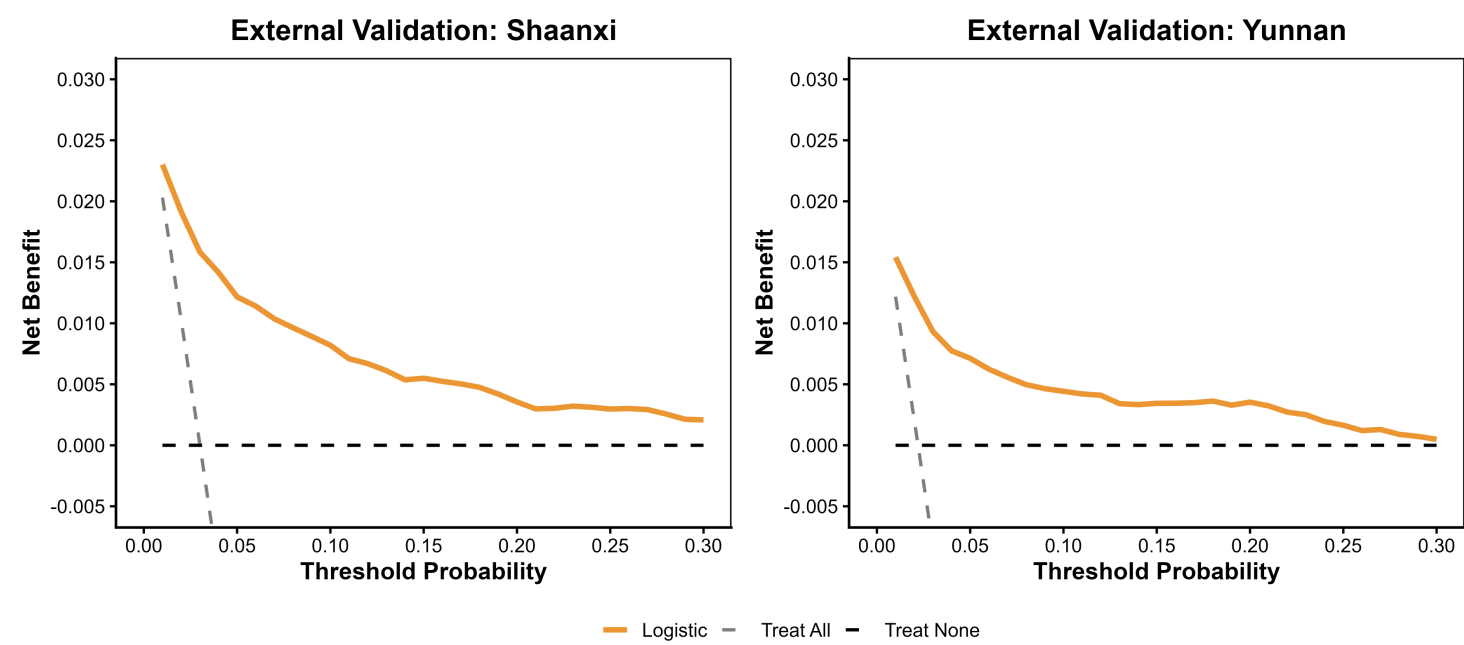


Figure 3

Decision curve analysis of the selected model in external validation cohorts

Note: DCA, decision curve analysis. The x-axis represents the threshold probability, and the y-axis represents the net benefit of the model. Net benefit reflects the clinical utility of the model by considering the benefits of correctly identifying individuals with STBs and the potential harms of unnecessary assessments. The model performance was compared with the treat-all and treat-none strategies. A higher net benefit indicates greater potential clinical usefulness across the corresponding threshold probability range.

to model-based STBs vulnerability identification. In SHAP summary plots (beeswarm plots), each point represents an individual participant. The x-axis represents the SHAP value, indicating the magnitude and direction of predictor contributions to model predictions. Positive SHAP values indicate increased model-estimated STBs vulnerability, whereas negative SHAP values indicate decreased predicted vulnerability. The color gradient represents the relative feature value, with red indicating higher values and blue indicating lower values.

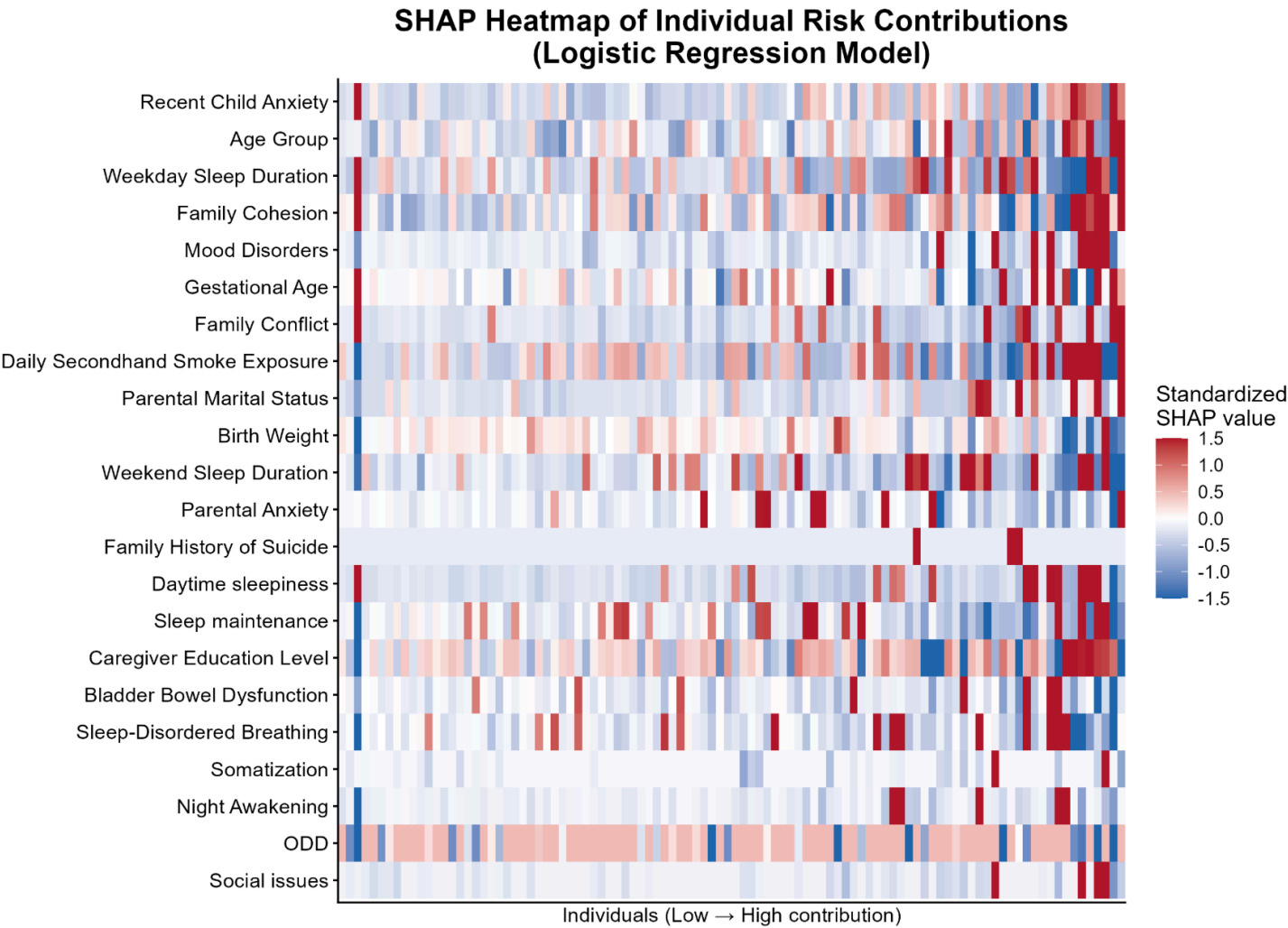


Figure 5

Individual-level SHAP heatmap illustrating heterogeneous risk contributions in the Logistic Regression model

Note: SHAP, Shapley additive explanations. The x-axis of SHAP feature importance plots represents the mean absolute SHAP value (Mean |SHAP value|), and the y-axis represents the included predictors. Higher mean absolute SHAP values indicate greater contributions to model predictions. In SHAP summary plots and individual-level heatmaps, positive SHAP values indicate increased model-estimated STBs vulnerability, whereas negative SHAP values indicate decreased predicted vulnerability. In the heatmap, rows represent

predictors and columns represent individual participants; color intensity represents standardized SHAP values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryCodeSTBsMachineLearning.r](#)
- [SupplementaryFile1.docx](#)
- [SupportingInformationLASSOFeatureSelection.docx](#)
- [TableS1.docx](#)