

Original Article

Cite this article: Wu Y, Zhao Y, Zhong P, Chen C, Wu Y, Zheng X (2025) Age-specific prevalence and predictors of lifetime suicide attempts using machine learning in Chinese adults: a nationwide multi-centre survey. *Epidemiology and Psychiatric Sciences* **34**, e52, 1–16. <https://doi.org/10.1017/S2045796025100231>

Received: 26 June 2025
Revised: 23 August 2025
Accepted: 9 September 2025

Keywords:

age differences; associated factors; China; machine learning; prevalence; suicide attempt; suicidal behavior

Corresponding author: Xiaoying Zheng;
Email: zhengxiaoying@sph.pumc.edu.cn

Age-specific prevalence and predictors of lifetime suicide attempts using machine learning in Chinese adults: a nationwide multi-centre survey

Yu Wu¹ , Yihao Zhao¹, Panliang Zhong¹, Chen Chen¹, Yibo Wu² and Xiaoying Zheng³

¹Department of Population Health and Aging Science, School of Population Medicine and Public Health, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; ²School of Public Health, Peking University, Beijing, China and ³China Center for Environmental and Energy Economics (C2E3), Peking University, Chengze Yuan, Beijing, China

Abstract

Aims. The epidemiology and age-specific patterns of lifetime suicide attempts (LSA) in China remain unclear. We aimed to examine age-specific prevalence and predictors of LSA among Chinese adults using machine learning (ML).

Methods. We analyzed 25,047 adults in the 2024 Psychology and Behavior Investigation of Chinese Residents (PBICR-2024), stratified into three age groups (18–24, 25–44, ≥ 45 years). Thirty-seven candidate predictors across six domains—sociodemographic, physical health, mental health, lifestyle, social environment, and self-injury/suicide history—were assessed. Five ML models—random forest, logistic regression, support vector machine (SVM), Extreme Gradient Boosting (XGBoost), and Naive Bayes—were compared. SHapley Additive exPlanations (SHAP) were used to quantify feature importance.

Results. The overall prevalence of LSA was 4.57% (1,145/25,047), with significant age differences: 8.10% in young adults (18–24), 4.67% in adults aged 25–44, and 2.67% in older adults (≥45). SVM achieved the best test-set performance across all ages [area under the curve (AUC) 0.88–0.94, sensitivity 0.79–0.87, specificity 0.81–0.88], showing superior calibration and net clinical benefit. SHAP analysis identified both shared and age-specific predictors. Suicidal ideation, adverse childhood experiences, and suicide disclosure were consistent top predictors across all ages. Sleep disturbances and anxiety symptoms stood out in young adults; marital status, living alone, and perceived stress in mid-life; and functional limitations, poor sleep, and depressive symptoms in older adults.

Conclusions. LSA prevalence in Chinese adults is relatively high, with a clear age gradient peaking in young adulthood. Risk profiles revealed both shared and age-specific predictors, reflecting distinct life-stage vulnerabilities. These findings support age-tailored suicide prevention strategies in China.

Introduction

Suicide remains a major global public health challenge, claiming over 700,000 lives annually (WHO, 2021). In China, approximately 116,000 suicide deaths were reported in 2019, accounting for about 17% of the global total (WHO, 2021). The age-standardized rate was 6.7 per 100,000, lower than the global average of 9.0 per 100,000 (WHO, 2021). Although China has seen a substantial decline in suicide rates over the past three decades (Wu *et al.*, 2024b), the country's large population keeps the absolute burden high. Suicide attempt (SA) is among the strongest predictors of suicide (Bostwick *et al.*, 2016). An estimated 8.5% to 13% of attempts are fatal, with the first-attempt fatality rate around 3% (Bostwick *et al.*, 2016; Conner *et al.*, 2019; Miller *et al.*, 2012). Nearly half of suicide deaths occur following a prior attempt rather than on the first attempt (Anestis, 2016; Bostwick *et al.*, 2016). Preventing nonfatal SA provides a critical window for early intervention, especially in China.

Age is a well-established factor influencing suicidal behaviours (Yazdi-Ravandi *et al.*, 2023). Although no uniform age pattern has been identified, substantial epidemiological differences across age groups have been consistently reported (Franklin *et al.*, 2017; Nock *et al.*, 2008b). Globally, adolescents and young adults are widely recognized as high-risk groups for both SA and suicide deaths (Franklin *et al.*, 2017; Nock *et al.*, 2008b). In China, a nationwide analysis

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

of cause-of-death surveillance data revealed a bimodal age distribution in suicide mortality, with peaks among those aged 15–24 and older adults (Wu *et al.*, 2024b). However, nationally representative data on SA remain scarce, possibly due to suicide-related stigma, stringent ethical oversight, and the challenges of diagnostic complexity, participant recruitment, disclosure, and data collection in population-based surveys on suicidal behaviours (Lee *et al.*, 2007; Wu *et al.*, 2024b). Most available evidence comes from small, high-risk samples—such as individuals with mental disorders (Chen *et al.*, 2022; Song *et al.*, 2023), sexual and gender minorities (Chen *et al.*, 2019), children and adolescents (Su *et al.*, 2023; Wu *et al.*, 2021), and rural or metropolitan residents (Lee *et al.*, 2007; Lyu *et al.*, 2018)—and shows wide variability across subgroups. As a result, the overall prevalence and age-specific patterns of SA in the general population remain unclear. Given China's age-patterned suicide mortality across life stages, we hypothesize that SA similarly displays distinct age-specific profiles and predictors. Clarifying these patterns is critical for informing age-tailored suicide prevention strategies.

SA arises from multiple factors that often interact in complex, nonlinear ways (Franklin *et al.*, 2017). Most prediction studies have relied on linear models, yielding modest accuracy (Wu *et al.*, 2023). Although traditional approaches such as generalized additive models can model certain nonlinearities, they require pre-specifying interactions and are less suited to high-dimensional, mixed-type data. A meta-analysis of 365 studies found that models based solely on conventional risk factors performed only slightly better than chance, highlighting the need for machine-learning (ML) risk algorithms (Franklin *et al.*, 2017). ML can flexibly capture nonlinearities and higher order interactions, handle heterogeneous predictors and class imbalance, and do so without strict parametric assumptions (García de la Garza *et al.*, 2021). Given the likely age-related heterogeneity of SA in the general population, identifying age-specific predictors in non-clinical samples is essential (García de la Garza *et al.*, 2021; Gordon *et al.*, 2020). To date, no study in China has used ML for this purpose, leaving a key evidence gap.

In this nationally representative, multicentre survey, we aimed to (1) estimate the age-specific prevalence of lifetime suicide attempts (LSA) among Chinese adults and (2) identify key age-specific predictors using five ML algorithms. These findings may help inform tailored suicide prevention strategies across the life course.

Methods

Data source and study population

Data came from the Psychology and Behavior Investigation of Chinese Residents (PBICR), a nationally representative, multicentre cross-sectional survey led by Peking University from June to September 2024. Using stratified and quota sampling, the survey randomly selected 150 cities, 202 districts/counties, and 390 townships across 31 provincial-level regions (including Hong Kong and Macao), covering 800 communities or villages. Trained interviewers recruited participants online and onsite at survey locations, verified eligibility, and administered standardized electronic questionnaires in one-to-one, face-to-face interviews. Eligible participants were adults (≥ 18 years) who were Chinese nationals permanently residing in China and able to clearly understand and complete the questionnaire. We excluded individuals with diagnosed psychiatric or cognitive disorders, those concurrently enrolled in similar surveys, those who declined or did not sign informed consent,

and responses with completion times < 5 min. For participants with intact decisional capacity but insufficient motor ability, interviewers conducted one-to-one interviews and recorded responses on their behalf. All procedures complied with the Declaration of Helsinki and received institutional ethics approval (see Ethics Statement).

Of 38,793 invited residents, 38,424 responded (response rate: 99.05%). We excluded 2,563 invalid questionnaires due to age < 18 , non-Chinese nationality, no consent, response time < 5 min, or missing age and LSA, leaving 35,861 fully eligible participants. To improve national representativeness, post-stratification weights based on sex, age, and urban–rural distribution as reported in the *China Statistical Yearbook 2023* (National Bureau of Statistics of China, 2023) were applied, yielding a final analytic sample of 25,047. Missing covariate values ($n = 136$) were imputed using MissForest, a non-parametric algorithm based on random forests suitable for both continuous and categorical data (Stekhoven and Buhlmann, 2012). Based on prior suicide-related research, mortality surveillance practices, and developmental distinctions across life stages (Miller *et al.*, 2004; Wu *et al.*, 2024b), age was categorized into three groups (18–24, 25–44, and ≥ 45 years), reflecting distinct patterns of psychosocial development, health risk exposures, and suicide-related vulnerabilities. The sample selection flowchart is shown in Fig. 1.

Outcome assessment

LSA was defined as any self-inflicted, potentially injurious act at any point in life, carried out with at least some intent to die, regardless of the outcome (Nock *et al.*, 2008b). Consistent with prior suicide-related research (Lee *et al.*, 2007; Oquendo *et al.*, 2024), LSA was assessed using a dichotomous self-report item: “Have you ever attempted suicide?”. Participants who answered “yes” were classified as lifetime suicide attempters, and those who answered “no” as non-attempters.

Predictors

Based on previous studies (Su *et al.*, 2023; Wu *et al.*, 2021), we included 37 candidate predictors across six domains (details in Table S1):

- (1) Sociodemographic: gender (male/female), marital status (married vs. unmarried—never married, divorced, widowed, or separated), residence (urban/rural), ethnicity (Han/ethnic minority), educational level (low: junior high school or below; medium: senior high school or vocational/technical school; high: college degree or above), job status (in education or employment vs. not in education or employment), economic status (monthly household income: low < 3000 RMB; middle 3000–5999 RMB; high ≥ 6000 RMB), and medical insurance (yes/no).
- (2) Physical health: body mass index (BMI, underweight < 18.5 , normal 18.5–24.9, overweight/obese ≥ 25); chronic disease (yes/no); activities of daily living (ADL) limitation (yes/no), defined as difficulty with basic self-care tasks (walking across a room, bathing/showering, or dressing); and self-rated health (0–100, higher scores represent better status).
- (3) Lifestyle: smoking (yes/no), drinking (yes/no), living alone (yes/no), sleep quality (five-item sleep health score: chronotype, duration, insomnia, snoring, daytime sleepiness; range 0–5; higher scores indicate better sleep) (Fan *et al.*, 2020),

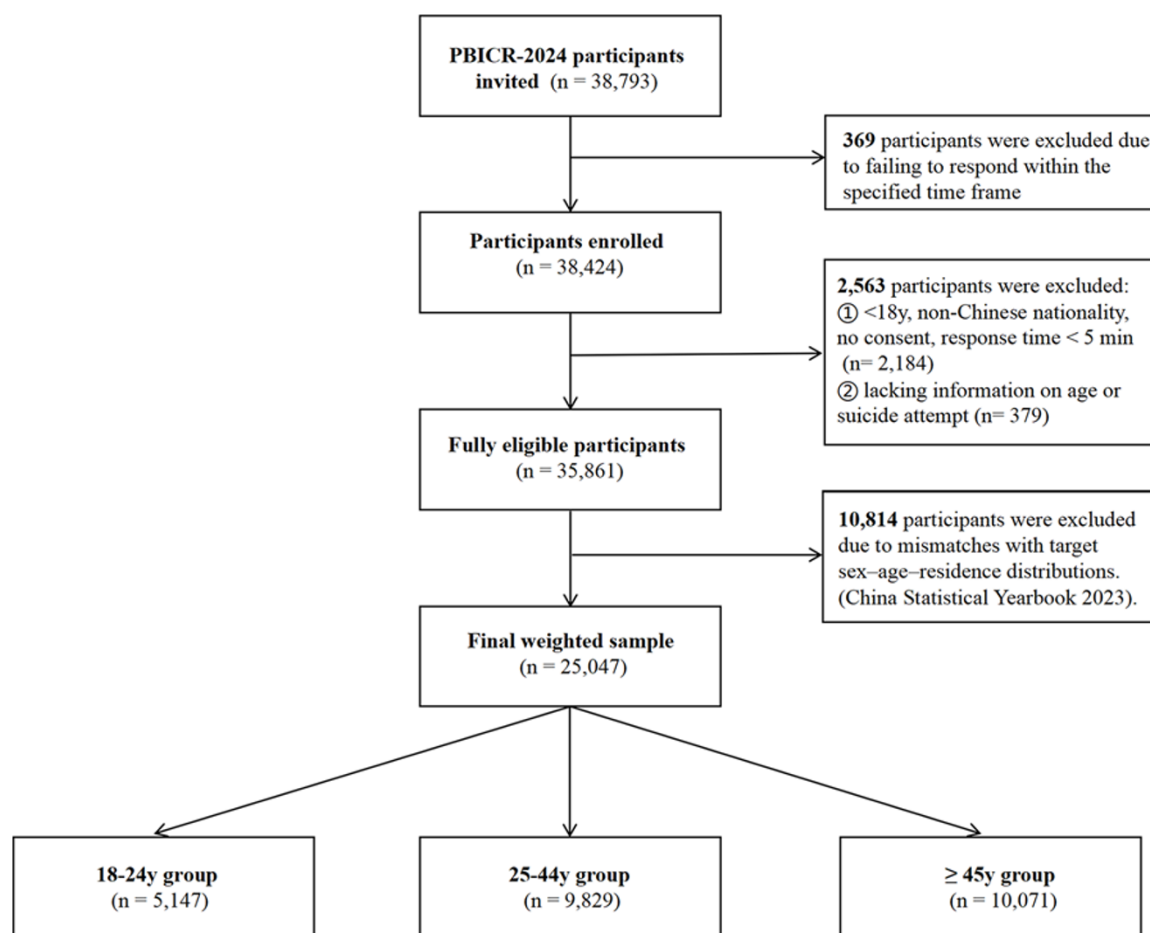


Fig. 1. Flow diagram of sample selection.

social media addiction (SMA, assessed with the Bergen Social Media Addiction Scale; total score 6–30; scores ≥ 19 indicate at-risk status; Cronbach's $\alpha = 0.926$ overall) (Leung *et al.*, 2020), and moderate- and vigorous-intensity physical activity (MPA, VPA) (measured as the number of self-reported days per week).

- (4) Social environment: social support (three-item Perceived Social Support Scale; total score 0–18, higher scores indicate stronger perceived support; Cronbach's $\alpha = 0.904$ overall) (Wu *et al.*, 2025); self-rated social status (a self-assessed item, 1–7; higher scores indicate higher perceived status); neighbourhood relations (a self-assessed item, 1–7; higher scores indicate better relations); and social participation (assessed by the five-item Social Connection Scale; total score 0–5; higher scores indicate greater engagement in shared activities and stronger social connectedness; Cronbach's $\alpha = 0.805$ overall) (Foster *et al.*, 2023).
- (5) Mental health: depressive symptoms (DS, assessed using Patient Health Questionnaire-9; total score 0–27, scores ≥ 10 indicating having DS; Cronbach's $\alpha = 0.940$ overall) (Levis *et al.*, 2019); anxiety symptoms (AS, assessed using Generalized Anxiety Disorder-3 scale; total score 0–9, scores ≥ 3 indicating AS; Cronbach's $\alpha = 0.901$ overall) (Wang *et al.*, 2024); negative life events (NLE, any major stressor in the past year; yes/no); loneliness (a self-rated item, 1–7; higher scores indicate greater loneliness); perceived stress (assessed

by the four-item Perceived Stress Scale; total score 0–16; higher scores indicate higher perceived stress; Cronbach's $\alpha = 0.934$ overall) (Huang *et al.*, 2020); burnout (measured with the seven-item Copenhagen Burnout Inventory; total score 0–28; higher scores indicate greater exhaustion; Cronbach's $\alpha = 0.851$ overall) (Borritz *et al.*, 2005); self-efficacy (assessed with the three-item New Generalized Self-Efficacy Scale; total score 0–12; higher scores indicate greater self-efficacy; Cronbach's $\alpha = 0.929$ overall) (Feng and Chen, 2012); intimate partner violence (IPV, measured with the five-item scale covering physical, emotional, and controlling behaviours; total score 0–20; higher scores indicate greater severity; Cronbach's $\alpha = 0.921$ overall) (Yount *et al.*, 2022); adverse childhood experiences (ACEs, cumulative score from 30 binary items covering household dysfunction, abuse, neglect, bereavement, and exposure to violence before age 18; total score 0–30; higher scores indicate greater adversity; Cronbach's $\alpha = 0.901$ overall) (Lin *et al.*, 2021).

- (6) Self-injury and suicide-related history (all variables refer to lifetime history unless otherwise specified) (Nock *et al.*, 2008b): Non-suicidal self-injury (NSSI): deliberate self-inflicted damage to the body's surface without suicidal intent (yes/no); NSSI-medical: history of NSSI that required medical treatment (yes/no); Suicidal ideation (SI): thoughts about engaging in behaviour intended to end one's life (yes/no); Suicide plan (SP): formulation of a specific method,

means, timing, or place for a suicide attempt (yes/no); Suicide disclosure: disclosure of suicide plans to others (yes/no).

Feature screening

Feature selection proceeded in two steps. First, univariable logistic regressions for LSA were run to prefilter predictors (retain if $p < 0.05$). Second, least absolute shrinkage and selection operator (LASSO) logistic regression with five-fold cross-validation was used to tune λ and apply an L1 penalty that shrinks irrelevant coefficients to zero, addressing high dimensionality and multicollinearity (Lee *et al.*, 2018). Only predictors retained by both steps were carried forward. This intersection rule was chosen to enhance model stability, interpretability, and out-of-sample generalizability.

Machine learning algorithms

To identify the most accurate classifier, we compared five supervised algorithms on each age-specific dataset. All models used the same feature set obtained from the *Predictors* section. Hyperparameters were tuned by grid search with stratified five-fold cross-validation in the training set, and performance was evaluated on a held-out 30% test set (see Table S2).

- (1) Random forest (RF) (Breiman, 2001): RF is an ensemble of decision trees grown on bootstrap samples with random feature subsampling. Predictions are aggregated by majority vote. This setup captures nonlinearities and interactions, is relatively robust to noise and overfitting, and provides internal variable-importance measures.
- (2) Logistic regression (LR) (Pan and Xu, 2022): LR models the log-odds of a binary outcome as a linear function of predictors. Coefficients are directly interpretable and estimation is efficient, suiting problems with modest feature sets or an emphasis on describing predictor–outcome associations.
- (3) Support Vector Machine (SVM) (Rezvani and Wu, 2023): SVM is a margin-based classifier that finds a separating hyperplane with maximal margin. Kernel functions allow nonlinear decision boundaries. It performs well in high-dimensional or sparse settings, particularly when the number of features exceeds the number of observations or when class boundaries are approximately linearly separable.
- (4) Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016): XGBoost is a gradient-boosted ensemble of decision trees fitted stage-wise to minimize a differentiable loss, with shrinkage and L1/L2 regularization to curb overfitting. It delivers strong accuracy on structured, high-dimensional data and supports efficient parallel training, making it well suited to complex health datasets.
- (5) Naive Bayes (Domingos and Pazzani, 1997): Naive Bayes is a probabilistic classifier grounded in Bayes' theorem that assumes conditional independence among features. Despite its simplicity, it performs well in many practical settings, particularly for high-dimensional, text-based, or weakly correlated data, and when computational efficiency is essential.

Statistical analysis

We estimated and compared the prevalence of LSA across age groups using chi-square tests. Within each age group, we compared participant characteristics between those with and without

LSA using independent-samples *t* tests (or Mann–Whitney *U* tests for non-normal data) for continuous variables and chi-square tests for categorical variables. Continuous variables are reported as mean $\pm$ standard deviation (SD) if approximately normal, and as median (interquartile range, IQR) otherwise; categorical variables are reported as number and percentage.

To mitigate class imbalance in the binary classification task, we applied Adaptive Synthetic (ADASYN) oversampling to the training set after feature selection (Haibo *et al.*, 2008). This method generates synthetic samples adaptively based on local feature density, improving classifier performance in imbalanced settings.

Each of the three age-specific datasets was randomly split into training (70%) and test (30%). Five-fold stratified cross-validation was performed within the training set for hyperparameter tuning and to reduce overfitting. Final performance was evaluated on the held-out test set using multiple metrics. The area under the receiver operating characteristic curve (AUC) assessed overall discrimination (0.5 indicating no discrimination and 1.0 perfect classification). Accuracy measured the proportion of correctly classified cases. Sensitivity reflected the proportion of true positives correctly identified, while specificity measured the proportion of true negatives. Precision (positive predictive value, PPV) and negative predictive value (NPV) captured the correctness of positive and negative predictions, respectively. The F1 score, defined as the harmonic mean of precision and sensitivity, offered a balanced metric under class imbalance. Beyond discrimination, model calibration was evaluated using calibration curves, and clinical utility was assessed with decision curve analysis (DCA). For the best-performing model in each dataset, we computed SHAP (SHapley Additive exPlanations) values to interpret feature contributions, offering insights into both global importance and individual-level predictions.

Sensitivity analyses were conducted to test the robustness of results: (1) reanalysis using complete cases to assess the impact of imputation; (2) 10-fold stratified cross-validation to test resampling sensitivity; (3) a 60:40 development/validation split to examine split-ratio effects; and (4) comparing feature-importance rankings across similarly performing models—to gauge predictor stability.

All analyses were performed using Python (version 3.7). Statistical significance was defined as two-sided $p < 0.05$.

Results

Prevalence of LSA and participant characteristics across age groups

As shown in Fig. 2, 1,145 of 25,047 participants reported LSA, including 417, 459, and 269 cases in the 18–24, 25–44, and ≥ 45 age groups, respectively. The overall prevalence of LSA was 4.57%, slightly higher in males (4.69%) than in females (4.45%), and similar between urban (4.49%) and rural (4.66%) residents. Prevalence varied significantly across age groups—peaking at 18–24 years (8.10%), then 25–44 years (4.67%), and lowest at ≥ 45 years (2.67%)—with all pairwise differences significant (all $p < 0.001$). Table 1 summarizes characteristics by LSA status within each age group across six domains. Age-related heterogeneity was most evident for sociodemographic factors (gender, marital status, education, income, insurance), whereas differences across lifestyle, social environment, mental health, and self-injury/suicide-history domains were broadly consistent across age groups.

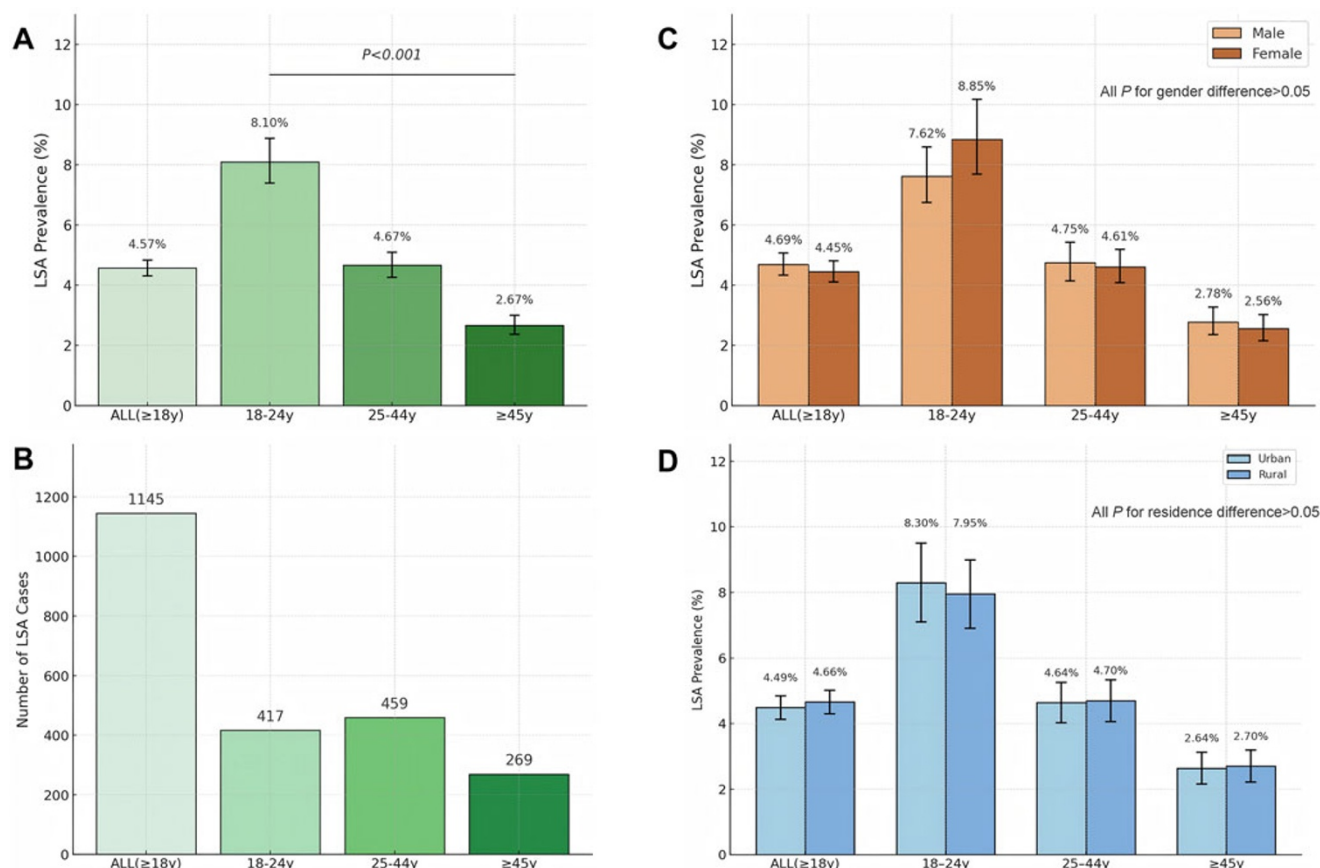


Fig. 2. Age-specific prevalence and subgroup differences in lifetime suicide attempts (LSA) among adults in China.

Note: (A) LSA prevalence by age group in the overall population; (B) Number of LSA cases by age group; (C) LSA prevalence by age group and sex; (D) LSA prevalence by age group and residential location (urban vs. rural). LSA prevalence differed by age, with all pairwise comparisons significant (all $p < 0.001$)

Predictor screening

We applied the predefined two-step feature selection separately within each age group. Figures S1–S3 show the univariable logistic-regression rankings, LASSO cross-validation error curves, and coefficient shrinkage paths; Table S3 shows the overlap in selected predictors across age groups. Eighteen predictors were retained for ages 18–24, 22 for 25–44, and 20 for ≥45. Several key features overlapped across groups, but age-specific differences were evident, consistent with distinct life-stage risk profiles. This combined approach improved the robustness of subsequent model development.

Model performance

We evaluated the predictive performance of five ML models on the test set across three age groups (Table 2). Discriminatory power and calibration were assessed using ROC curves (Fig. 3A–C) and calibration plots (Fig. 3D–F). Based on overall performance, SVM was identified as the optimal model in all groups and was further evaluated using DCA (Fig. 3G–I).

In the 18–24 group, SVM demonstrated the most balanced performance, achieving high AUC (0.88), accuracy (0.805), sensitivity (0.794), specificity (0.815), PPV (0.803), and NPV (0.861). Although XGBoost yielded a slightly higher AUC (0.89), SVM showed more consistent performance across all metrics. Its ROC curve approached the upper-left corner, and its calibration plot

showed strong agreement with the ideal reference line. DCA indicated greater net benefit for SVM across a wide range of clinically relevant threshold probabilities compared to “treat all” or “treat none” strategies.

In the 25–44 group, SVM outperformed all other models, with the highest AUC (0.94), accuracy (0.874), sensitivity (0.868), and NPV (0.870), along with strong specificity (0.880) and PPV (0.871). ROC and calibration curves confirmed its excellent discrimination and calibration. DCA further supported its clinical utility.

In the ≥45 group, both SVM and LR achieved the highest AUC (0.94). However, SVM showed slightly better performance in accuracy (0.858 vs. 0.857), specificity (0.885 vs. 0.879), and PPV (0.875 vs. 0.873), while maintaining comparable sensitivity and NPV. It also showed calibration closer to the 45° line. DCA consistently demonstrated higher net clinical benefit for SVM.

SHAP-based model interpretability analysis

To interpret the best-performing SVM models across age groups, we applied SHAP analysis to quantify each predictor's contribution to the predicted probability of LSA. For each group, the SHAP bar plot (left) shows mean absolute SHAP values (global importance), and the summary plot (right) illustrates the direction and magnitude of feature effects (Figs. 4, 5, and 6; full results in Figures S4–S6). For accessibility, SHAP values > 0 indicate an

Table 1. Basic characteristics of participants with lifetime suicide attempts across age groups

Features	18-24y (N = 5147)			25-44y (N = 9829)			≥45y (N = 10071)		
	Non-LSA	LSA	<i>P</i>	Non-LSA	LSA	<i>P</i>	Non-LSA	LSA	<i>P</i>
Sociodemographic									
Gender			0.115			0.744			0.487
Male	2908(61.5)	240(57.6)		4092(43.7)	204(44.4)		4818(49.2)	138(51.3)	
Female	1822(38.5)	177(42.4)		5278(56.3)	255(55.6)		4984(50.9)	131(48.7)	
Residence			0.648			0.892			0.858
Urban	2089(44.2)	189(45.3)		4848(51.7)	236(51.4)		5010(51.1)	136(50.6)	
Rural	2641(55.8)	228(54.7)		4522(48.3)	223(48.6)		4792(48.9)	133(49.4)	
Marriage			0.602			<0.001			<0.001
Unmarried	4624(97.8)	406(97.4)		3466(37.0)	252(54.9)		1521(15.5)	88(32.7)	
Married	106(2.2)	11(2.6)		5904(63.0)	207(45.1)		8281(84.5)	181(67.3)	
Ethnicity			0.456			0.004			0.164
Han	4143(87.6)	360(86.3)		8639(92.2)	406(88.5)		9297(94.9)	250(92.9)	
Ethnic minority	587(12.4)	57(13.7)		731(7.8)	53(11.5)		505(5.2)	19(7.1)	
Educational level			0.066			0.042			0.007
Low	118(2.49)	15(3.6)		1608(17.2)	59(12.9)		4812(49.1)	112(41.6)	
Middle	1920(40.6)	187(44.8)		3287(35.1)	162(35.3)		3260(33.3)	91(33.8)	
High	2692(56.9)	215(51.6)		4475(47.8)	238(51.9)		1730(17.7)	66(24.5)	
Job status			0.985			0.452			0.183
No	171(3.6)	15(3.6)		534(5.7)	30(6.5)		3797(38.7)	115(42.8)	
Yes	4559(96.4)	402(96.4)		8836(94.3)	429(93.5)		6005(61.3)	154(57.3)	
Economic status			0.341			0.005			0.116
Low	1196(25.3)	119(28.5)		1763(18.8)	114(24.8)		2913(29.7)	94(34.9)	
Middle	1914(40.5)	160(38.4)		3955(42.2)	185(40.3)		4266(43.5)	102(37.9)	
High	1620(34.3)	138(33.1)		3652(39.0)	160(34.9)		2623(26.8)	73(27.1)	
Medical insurance			0.099			0.298			<0.001
No	429(9.1)	48(11.5)		488(5.2)	29(6.3)		341(3.5)	23(8.6)	
Yes	4301(90.9)	369(88.5)		8882(94.8)	430(93.7)		9461(96.5)	246(91.5)	

(Continued)

Table 1. (Continued.)

Features	18-24y (N = 5147)			25-44y (N = 9829)			≥45y (N = 10071)		
	Non-LSA	LSA	P	Non-LSA	LSA	P	Non-LSA	LSA	P
Physical health									
BMI			0.422			0.362			0.533
Normal	3258(68.9)	293(70.3)		6458(68.9)	328(71.5)		6677(68.1)	178(66.2)	
Underweight	771(16.3)	58(13.9)		929(9.9)	37(8.1)		594(6.1)	14(5.2)	
Overweight/obese	701(14.8)	66(15.8)		1983(21.2)	94(20.5)		2531(25.8)	77(28.6)	
Chronic disease			<0.001			<0.001			<0.001
No	4442(93.9)	340(81.5)		8118(86.6)	328(71.5)		5497(56.1)	111(41.3)	
Yes	288(6.1)	77(18.45)		1252(13.4)	131(28.5)		4305(43.9)	158(58.7)	
ADL limitation			<0.001			<0.001			<0.001
No	4500(95.1)	337(80.8)		8604(91.8)	328(71.5)		7861(80.2)	150(55.8)	
Yes	230(4.9)	80(19.2)		766(8.2)	131(28.5)		1941(19.8)	119(44.2)	
Self-rated health	79.14(18.1)	68.23(21.3)	<0.001	77.80(18.4)	63.83(23.0)	<0.001	74.09(19.3)	63.54(23.0)	<0.001
Lifestyle									
Smoking			<0.001			<0.001			<0.001
No	4146(87.7)	329(78.9)		7735(82.6)	289(63.0)		7470(76.2)	164(61.0)	
Yes	584(12.3)	88(21.1)		1635(17.5)	170(37.0)		2332(23.8)	105(39.0)	
Drinking			<0.001			<0.001			<0.001
No	2980(63.0)	207(49.6)		6267(66.9)	203(44.2)		6685(68.2)	132(49.1)	
Yes	1750(37.0)	210(50.4)		3103(33.1)	256(55.8)		3117(31.8)	137(50.9)	
Living alone			0.063			<0.001			<0.001
No	3434(72.6)	285(68.4)		7166(76.5)	296(64.5)		8278(84.5)	190(70.6)	
Yes	1296(27.4)	132(31.7)		2204(23.5)	163(35.5)		1524(15.6)	79(29.4)	
Sleep quality	3.69(1.4)	2.87(1.3)	<0.001	3.73(1.4)	2.65(1.4)	<0.001	3.72(1.4)	2.82(1.3)	<0.001
Social media addiction score	10.16(5.6)	11.65(5.8)	<0.001	9.09(5.6)	12.10(5.5)	<0.001	6.92(5.6)	9.94(6.0)	<0.001
MPA	2.79(2.0)	2.61(1.9)	0.091	2.72(2.0)	2.89(2.0)	0.069	2.60(2.1)	2.94(2.1)	0.007
VPA	2.69(2.0)	2.42(1.9)	0.007	2.56(2.0)	2.75(2.0)	0.040	2.26(2.0)	2.59(1.9)	0.006

(Continued)

Table 1. (Continued.)

Features	18-24y (N = 5147)			25-44y (N = 9829)			≥45y (N = 10071)		
	Non-LSA	LSA	P	Non-LSA	LSA	P	Non-LSA	LSA	P
Social environment									
Social support	12.43(4.1)	11.09(3.7)	<0.001	12.39(4.1)	10.68(3.9)	<0.001	12.39(4.0)	10.50(4.1)	<0.001
Self-rated social status	3.84(1.3)	3.54(1.4)	<0.001	4.08(1.4)	3.74(1.5)	<0.001	4.17(1.4)	4.09(1.5)	0.345
Neighborhood relations	5.22(1.3)	4.73(1.4)	<0.001	5.32(1.3)	4.71(1.5)	<0.001	5.43(1.3)	5.04(1.4)	<0.001
Social participation	3.15(1.6)	3.09(1.7)	0.446	2.79(1.8)	3.35(1.8)	<0.001	1.97(1.8)	2.64(2.1)	<0.001
Mental health									
DS			<0.001			<0.001			<0.001
No	3623(76.6)	200(48.0)		7665(81.8)	177(38.6)		8291(84.6)	130(48.3)	
Yes	1107(23.4)	217(52.0)		1705(18.2)	282(61.4)		1511(15.4)	139(51.7)	
AS			<0.001			<0.001			<0.001
No	2800(59.2)	129(30.9)		6135(65.5)	106(23.1)		7132(72.8)	88(32.7)	
Yes	1930(40.8)	288(69.1)		3235(34.5)	353(76.9)		2670(27.2)	181(67.3)	
NLEs			<0.001			<0.001			<0.001
No	2748(58.1)	135(32.4)		6302(67.3)	172(37.5)		7299(74.5)	114(42.4)	
Yes	1982(41.9)	282(67.6)		3068(32.7)	287(62.5)		2503(25.5)	155(57.6)	
Loneliness	2.52(1.6)	3.67(1.6)	<0.001	2.22(1.5)	3.45(1.6)	<0.001	2.00(1.4)	3.27(1.6)	<0.001
Perceived stress	4.46(4.1)	7.66(4.2)	<0.001	4.01(3.7)	7.96(4.0)	<0.001	3.50(3.5)	7.27(4.3)	<0.001
Burnout	12.58(5.8)	16.28(5.4)	<0.001	12.02(5.8)	15.62(5.4)	<0.001	10.34(5.6)	14.69(5.6)	<0.001
Self-efficacy	8.13(2.6)	7.13(2.5)	<0.001	8.13(2.6)	6.87(2.5)	<0.001	7.87(2.6)	6.65(2.8)	<0.001
IPV	1.64(3.7)	3.55(5.1)	<0.001	2.21(3.9)	6.15(5.5)	<0.001	1.99(3.6)	6.03(5.1)	<0.001
ACEs	2.27(2.2)	5.09(4.3)	<0.001	2.29(2.5)	6.27(5.3)	<0.001	2.65(2.7)	6.65(5.1)	<0.001

(Continued)

Table 1. (Continued.)

Features	18-24y (N = 5147)			25-44y (N = 9829)			≥45y (N = 10071)		
	Non-LSA	LSA	P	Non-LSA	LSA	P	Non-LSA	LSA	P
Self-injury and suicide-related history									
NSSI			<0.001			<0.001			<0.001
No	4513(95.4)	272(65.2)		9066(96.8)	307(66.9)		9681(98.8)	193(71.8)	
Yes	217(4.6)	145(34.8)		304(3.2)	152(33.1)		121(1.2)	76(28.3)	
NSSI-medical			<0.001			<0.001			<0.001
No	4687(99.1)	378(90.7)		9255(98.8)	388(84.5)		9746(99.4)	230(85.5)	
Yes	43(0.9)	39(9.3)		115(1.2)	71(15.5)		56(0.6)	39(14.5)	
SI			<0.001			<0.001			<0.001
No	4282(90.5)	114(27.3)		8939(95.4)	120(26.1)		9608(98.0)	72(26.8)	
Yes	448(9.5)	303(72.7)		431(4.6)	339(73.9)		194(2.0)	197(73.2)	
Suicide plan			<0.001			<0.001			<0.001
No	4604(97.3)	262(62.8)		9254(98.8)	309(67.3)		9763(99.6)	199(74.0)	
Yes	126(2.7)	155(37.2)		116(1.2)	150(32.7)		39(0.4)	70(26.0)	
Suicide disclosure			<0.001			<0.001			<0.001
No	4468(94.5)	171(41.0)		9110(97.2)	175(38.1)		9706(99.0)	120(44.6)	
Yes	262(5.5)	246(59.0)		260(2.8)	284(61.9)		96(1.0)	149(55.4)	

Notes: Abbreviations: LSA = lifetime suicide attempts; BMI = body mass index; ADL = activities of daily living; MPA = moderate-intensity physical activity, days/week; VPA = vigorous-intensity physical activity, days/week; DS = depressive symptoms; AS = anxiety symptoms; NLE = negative life events; IPV = intimate partner violence; ACEs = adverse childhood experiences; SI = suicidal ideation; NSSI = non-suicidal self-injury; NSSI-medical = NSSI requiring medical treatment. BMI: underweight < 18.5; normal weight 18.5–24.9; overweight/obesity ≥ 25.0 kg/m². Educational level: low (junior high school or below), medium (senior high school or vocational/technical school), high (college degree or above). Job status: in education or employment (Yes) vs not in education or employment (No). Marital status: married or unmarried (never married, divorced, widowed, or separated).

Table 2. Comparison of model performance in predicting lifetime suicide attempts (LSA) on the test set data

Performance metrics	18-24y					25-44y					≥45y				
	RF	LR	SVM	XGBoost	Naive Bayes	RF	LR	SVM	XGBoost	Naive Bayes	RF	LR	SVM	XGBoost	Naive Bayes
Accuracy	0.799	0.764	0.805	0.818	0.787	0.820	0.819	0.874	0.843	0.806	0.805	0.857	0.858	0.823	0.8060
Sensitivity	0.778	0.704	0.794	0.873	0.726	0.790	0.740	0.868	0.803	0.722	0.872	0.835	0.831	0.818	0.699
Specificity	0.820	0.822	0.815	0.764	0.847	0.851	0.898	0.880	0.882	0.890	0.740	0.879	0.885	0.828	0.910
Positive Predictive Value	0.809	0.794	0.808	0.783	0.822	0.840	0.878	0.878	0.871	0.866	0.766	0.870	0.875	0.822	0.883
Negative Predictive Value	0.791	0.740	0.802	0.861	0.760	0.803	0.777	0.870	0.819	0.763	0.856	0.845	0.843	0.824	0.756
ROC Area	0.86	0.85	0.88	0.89	0.86	0.90	0.91	0.94	0.92	0.89	0.91	0.94	0.94	0.91	0.90

Notes: RF = Random Forest; LR = Logistic Regression; SVM = Support Vector Machine; XGBoost = Extreme Gradient Boosting; ROC = Receiver Operating Characteristic curve.

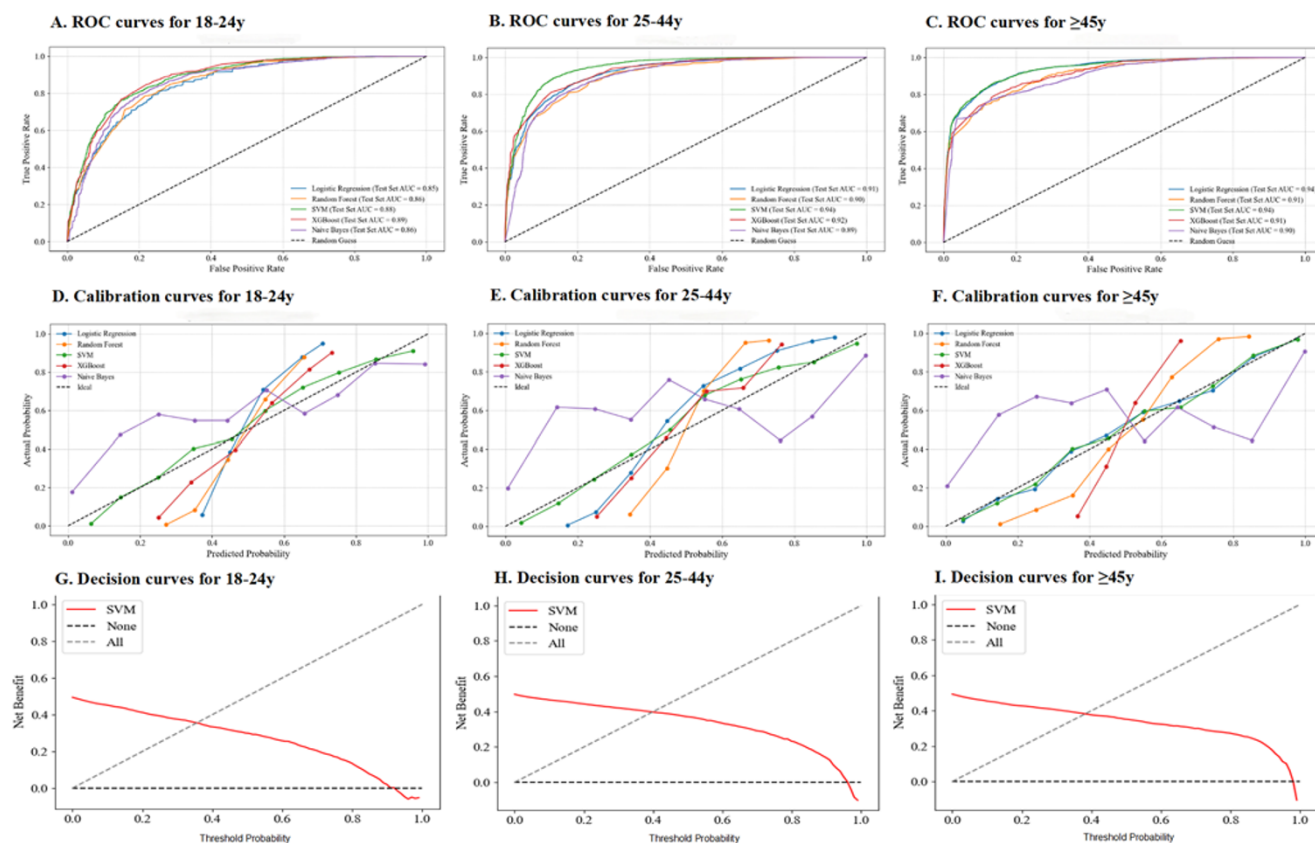


Fig. 3. Model performance in predicting lifetime suicide attempts across age groups.

Note: (A–C) Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) for 18–24y, 25–44y, and ≥ 45 y age groups; (D–F) Calibration curves comparing predicted versus observed probabilities for each model across the three age groups; (G–I) Decision Curve Analyses showing the net benefit of using Support Vector Machine (SVM) models across a range of threshold probabilities. Models included: Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Naive Bayes

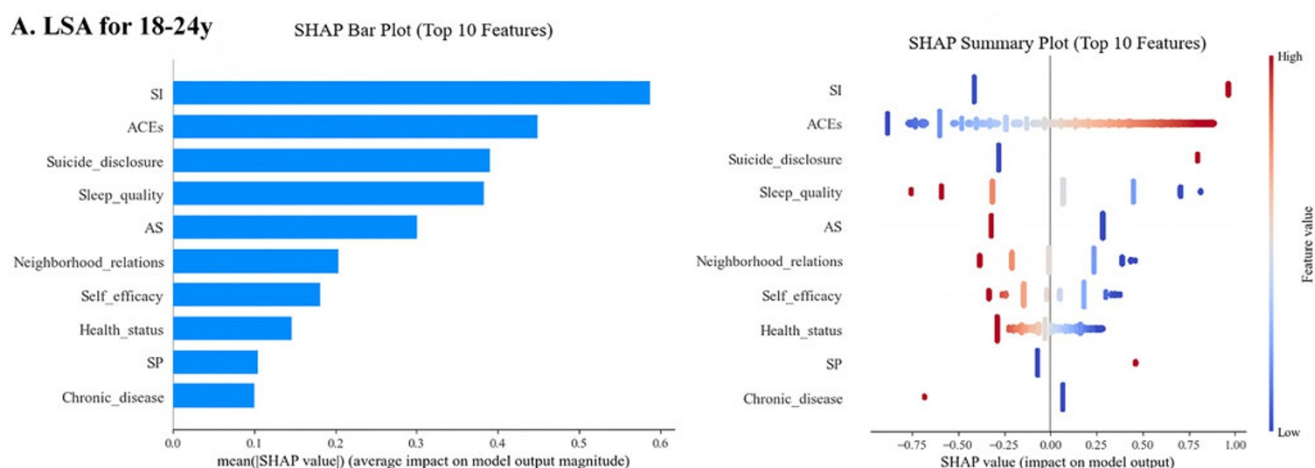


Fig. 4a. Top 10 features identified by SHAP using the best-performing model (SVM) in the 18–24y group.

Note: 1. Left panel: SHAP bar plot of mean absolute SHAP values (global importance; features ordered by mean SHAP). Right panel: SHAP summary plot (each dot = one participant). Colors encode feature values (red = higher, blue = lower). Positive SHAP values indicate an increase, and negative values a decrease, in the model-predicted probability of LSA. SHAP values were computed on the test set; 2. Abbreviations: SHAP = SHapley Additive exPlanations; SVM = support vector machine; LSA = lifetime suicide attempts; SI = suicidal ideation; ACEs = adverse childhood experiences; AS = anxiety symptoms; SP = suicide plan

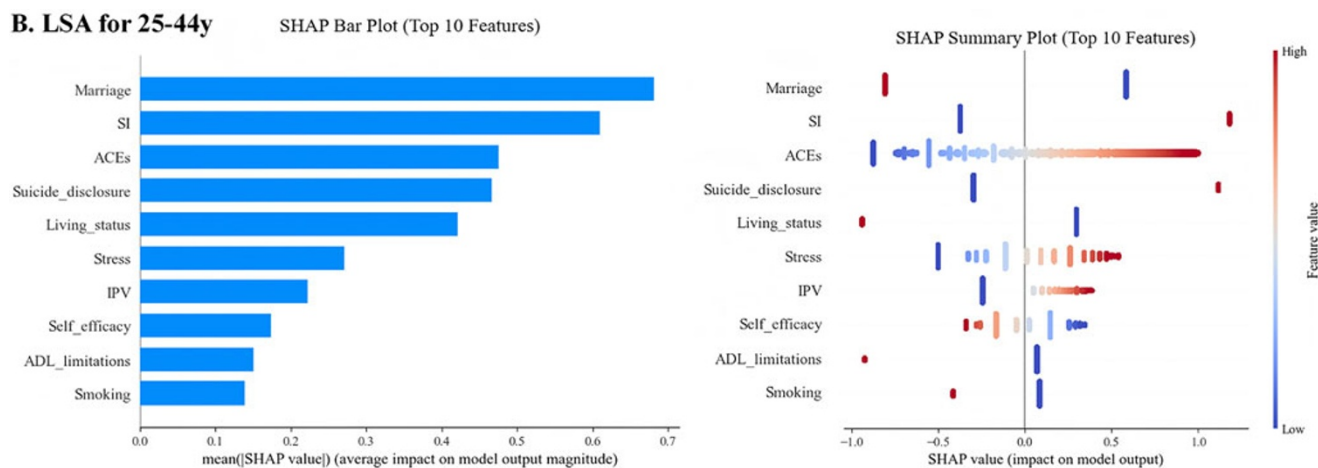


Fig. 4b. Top 10 features identified by SHAP using the best-performing model (SVM) in the 25-44y group.

Note: 1. Left panel: SHAP bar plot of mean absolute SHAP values (global importance; features ordered by mean SHAP). Right panel: SHAP summary plot (each dot = one participant). Colors encode feature values (red = higher, blue = lower). Positive SHAP values indicate an increase, and negative values a decrease, in the model-predicted probability of LSA. SHAP values were computed on the test set; 2. Abbreviations: SHAP = SHapley Additive exPlanations; SVM = support vector machine; LSA = lifetime suicide attempts; SI = suicidal ideation; ACEs = adverse childhood experiences; IPV = intimate partner violence; ADL = activities of daily living

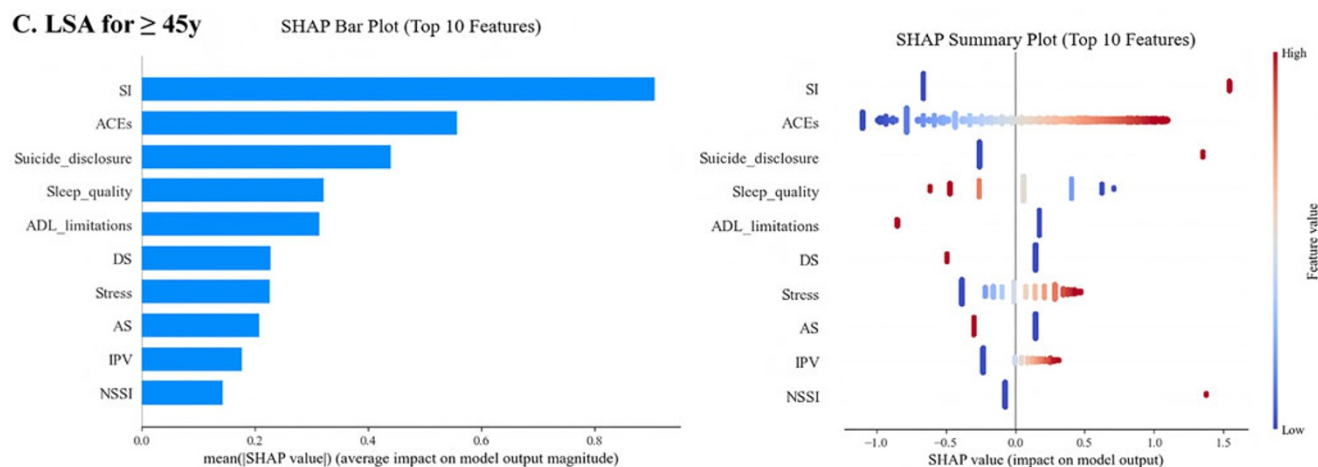


Fig. 4c. Top 10 features identified by SHAP using the best-performing model (SVM) in the $\geq 45y$ group.

Note: 1. Left panel: SHAP bar plot of mean absolute SHAP values (global importance; features ordered by mean SHAP). Right panel: SHAP summary plot (each dot = one participant). Colors encode feature values (red = higher, blue = lower). Positive SHAP values indicate an increase, and negative values a decrease, in the model-predicted probability of LSA. SHAP values were computed on the test set; 2. Abbreviations: SHAP = SHapley Additive exPlanations; SVM = support vector machine; LSA = lifetime suicide attempts; SI = suicidal ideation; ACEs = adverse childhood experiences; ADL = activities of daily living; DS = depressive symptoms; AS = anxiety symptoms; IPV = intimate partner violence; NSSI = non-suicidal self-injury

increase in the model-predicted probability of LSA (values < 0 indicate a decrease); the absolute value reflects contribution size, and point colour encodes the feature value (red = higher, blue = lower).

In the 18–24 group, the top predictors were SI, ACEs, suicide disclosure, sleep quality, and AS, followed by neighbourhood relations, self-efficacy, and self-rated health. Summary plots indicated that higher ACEs and SI were associated with a higher predicted probability of LSA, whereas better sleep quality and stronger neighbourhood relations were associated with a lower predicted probability.

In the 25–44 group, leading contributors included marriage, SI, ACEs, suicide disclosure, and living status. IPV, perceived stress, and self-efficacy also featured prominently. Being unmarried, exposure to IPV, and higher perceived stress were associated with higher predicted probability of LSA.

In the ≥ 45 group, SI, ACEs, and suicide disclosure again ranked highest, followed by sleep quality, ADL limitation, and DS. Functional limitations and psychological distress (e.g., perceived stress, AS, DS) contributed substantially to risk in this older cohort.

Across age groups, SI, ACEs, and suicide disclosure consistently emerged as core predictors. In contrast, other factors varied by life stage: sleep and AS weighed more in younger adults; relationship/structural factors (e.g., marital status, living status) dominated in midlife; and functional status and mental-health measures (e.g., ADL limitation, DS) were most salient in older adults.

Sensitive analysis

SHAP-identified top predictors of LSA remained largely consistent across age groups under all sensitivity checks, including exclusion of imputed data (Table S4; Figures S7–S10), application of

10-fold stratified cross-validation (Table S5; Figures S11–S14), use of a 60:40 train-test split (Table S6; Figures S15–S18), and comparison of feature rankings between similarly performing models (LR vs. SVM in the ≥ 45 y group; Figure S19), further supporting the robustness of the findings.

Discussion

To our knowledge, this is the first study to apply ML approaches to examine age-stratified prevalence and predictors of LSA among Chinese adults. Using a nationally representative cross-sectional survey, we identified three principal findings: First, the overall prevalence of LSA was 4.57%, indicating a relatively high burden with a pronounced age gradient—prevalence declined with increasing age, and young adults represented the highest-risk group. Second, SI, ACEs, and suicide disclosure consistently emerged as the most robust predictors across all age groups. Third, risk profiles varied by age: psychological distress and sleep-related problems predominated in young adults; marital and living status, IPV, and perceived stress were more salient in mid-life; and in older adults, poor sleep, functional limitations, and DS were the primary contributors. Collectively, these findings provide empirical evidence to guide the development of age-tailored suicide prevention strategies in China.

SA is among the strongest predictors of death by suicide (Bostwick *et al.*, 2016). In this nationally representative study, the prevalence of LSA among Chinese adults was 4.57%, with a marked age gradient: 8.10% in young adults, 4.67% in mid-life, and 2.67% in older adults. National epidemiological data on SA remain limited; most prior studies have focused on local settings or high-risk clinical groups, yielding widely varying estimates. For example, a 2001–2002 survey of community-dwelling adults in metropolitan China reported a LSA prevalence of 1.0% (Lee *et al.*, 2007). By contrast, a meta-analysis of 29 studies in Chinese college students estimated a pooled LSA prevalence of 2.8% (range 0.4%–10.5%) (Yang *et al.*, 2015). Internationally, nationally representative World Mental Health surveys across 17 countries reported adult LSA prevalence ranging from 0.5% in Italy to 5.0% in the United States (Nock *et al.*, 2008a). Age-specific patterns show broad cross-national regularities: most countries report higher LSA in younger adults, whereas Japan is a notable exception with relatively elevated risk in midlife. Our study indicates that the prevalence of LSA among Chinese adults is relatively high. This may reflect both substantive and methodological factors: rapid social change with rising depression and stress exposure; limited access and continuity of mental-health care alongside stigma that suppresses help-seeking and disclosure (Wu *et al.*, 2023, 2024b, 2021); and the use of self-reported recall measures—which typically yield higher estimates than diagnostic or registry data—plus differences in survey timing.

Notably, young adults show the highest LSA risk, plausibly reflecting heavier psychosocial stressors (academic and career pressures, identity formation), greater exposure to self-injury content online, and fewer coping resources (Wu *et al.*, 2021). This pattern should not be equated with the bimodal distribution of suicide mortality (youth and late-life peaks) (Wu *et al.*, 2024b): mortality reflects attempt incidence and case-fatality, and case-fatality increases with age (more lethal methods, greater frailty/comorbidity, lower rescue), yielding a late-life peak despite fewer attempts. Underreporting of lifetime attempts and survivorship may also depress observed LSA in older adults. These findings provide national epidemiological evidence for China and underscore the need for improved surveillance and age-targeted prevention.

Guided by the age-specific pattern of LSA, we applied multiple ML algorithms to identify key predictors across the lifespan. SI, ACEs, and suicide disclosure consistently emerged as the most robust predictors. SI was the strongest predictor, reaffirming its pivotal role in the suicidal process (Klonsky *et al.*, 2016). According to the Interpersonal Theory of Suicide, SI arises from perceived burdensomeness, thwarted belongingness, and psychological pain (Van Orden *et al.*, 2010), while the Three-Step Theory conceptualizes it as a necessary, though not sufficient, precursor, triggered by intolerable distress and hopelessness (Klonsky *et al.*, 2016). Although not all individuals with SI progress to attempts, SI substantially elevates risk and remains a necessary antecedent (Klonsky *et al.*, 2016). These findings support integrating brief SI screening into frontline settings (e.g., primary care, schools, crisis services), alongside the standardized response protocols and inclusion of SI indicators in public health surveillance to inform timely intervention and resource allocation.

ACEs represent cumulative early-life adversity with lasting impacts on psychological development (Norman *et al.*, 2012). Extensive evidence links childhood abuse, neglect, and interpersonal violence to elevated risk of suicidal behaviour in adulthood (Angelakis *et al.*, 2019; Norman *et al.*, 2012). One possible mechanism is that ACEs foster negative cognitive patterns—such as powerlessness, defeat, and entrapment—that impair emotional regulation under stress (Angelakis *et al.*, 2019). ACEs are also linked to post-traumatic stress disorder, where hopelessness and psychological disengagement may trigger SI and accelerate its progression to SA (Angelakis *et al.*, 2019). Their consistent predictive value underscores the need for upstream, trauma-informed policies that go beyond clinical screening—such as parenting support programs, school-based violence prevention, and cross-sector data integration to identify and intervene in at-risk environments before patterns of harm are established.

Suicide disclosure similarly predicted LSA across age groups. It often reflects a critical threshold of psychological distress and a strong need for help. From a cognitive-behavioural perspective, disclosure signals the externalization of suicidality—a shift from internal struggle to overt expression—and thus heightened risk (Rudd, 2000). While concealment may reflect fear of stigma, unsupported disclosure can intensify hopelessness and isolation, increasing the likelihood of escalation. These findings call for dedicated post-disclosure protocols across sectors, including hotline escalation pathways, school-based rapid response teams, and mandated training for frontline staff to recognize, triage, and support individuals who disclose suicidal thoughts or intent—bridging the gap between expression and action.

In addition to shared predictors, our SHAP analysis revealed distinct age-specific patterns of LSA risk. Among young adults, the pattern was primarily emotion–relationship driven. Poor sleep, high anxiety, weak neighbourhood ties, and low self-efficacy emerged as salient predictors, reflecting developmental challenges of identity formation, autonomy, and early academic or career stress. Sleep disturbances and anxiety may indicate emotional dysregulation and vulnerability to affective instability, increasing the risk of impulsive suicidal behaviours (Kearns *et al.*, 2020). Weak community ties and low self-efficacy may further diminish perceived agency and access to support, reinforcing powerlessness. These findings underscore the value of youth-centred prevention strategies emphasizing emotion regulation, connectedness, and empowerment.

In mid-aged adults, the pattern was largely role-stress driven. Being unmarried, living alone, perceived stress, and IPV ranked

highest. This stage is marked by heavier work and family responsibilities alongside shrinking informal support. Emotional isolation from marital disruption or living alone may heighten perceived burdensomeness and thwarted belongingness—key drivers of SI (Van Orden *et al.*, 2010). Chronic stress and IPV may also dysregulate the stress-response system, leading to emotional exhaustion and impaired coping (Vidal *et al.*, 2024). These findings highlight the need for integrated psychosocial strategies, including family therapy, IPV screening, and workplace stress reduction.

In older adults, the pattern was health–decline driven. DS, sleep problems, ADL limitations, and NSSI were top predictors. This stage is often accompanied by functional decline, bereavement, and shrinking social roles from retirement or caregiving burdens. ADL impairments undermine autonomy and increase dependency, which in collectivist contexts may be experienced as shame or perceived burden on others. Coexisting depression and sleep disturbances deepen emotional pain and reinforce cognitive distortions, while late-life NSSI may reflect chronic psychological distress or entrenched maladaptive coping. Prevention in later life thus requires a multidisciplinary approach integrating chronic disease management, ADL rehabilitation, and geriatric mental health services tailored to the loss and meaning reconstruction. Together, these findings highlight the need for age-tailored LSA screening and prevention strategies that are developmentally sensitive to the psychosocial vulnerabilities of each life stage.

Finally, our study demonstrates the potential of ML to improve identification of individuals at risk for LSA. Given the rarity and complexity of suicidal behaviour, accurate prediction of SA remains a major challenge (Su *et al.*, 2023). Traditional risk assessment tools have shown limited value, with meta-analyses reporting low sensitivity and poor positive predictive values across populations (Franklin *et al.*, 2017; Su *et al.*, 2023). In contrast, ML can capture complex, non-linear interactions among multiple factors (Su *et al.*, 2023; Wu *et al.*, 2023). In our analysis, SVM consistently outperformed other models across age groups, yielding balanced predictive performance—critical for reducing both false negatives and positives in suicide risk screening. These findings support the feasibility and added value of ML-based prediction. However, concerns remain regarding potential bias and inflated performance estimates (Jacobucci *et al.*, 2021). Thus, rigorous methodological standards and cautious interpretation are essential to ensure clinical applicability.

Strengths and limitations

This study has several strengths. First, it is the first nationwide investigation in China to assess LSA across age groups, providing valuable epidemiological data for future suicide-related research. Second, the inclusion of 37 variables across six domains enabled a comprehensive, multidimensional analysis of LSA predictors. Third, the application and comparison of multiple ML models provided a more robust assessment of predictive performance and model reliability.

However, several limitations should be acknowledged. First, the cross-sectional design precludes causal and temporal inferences. Time-window misalignment between current exposures and lifetime outcomes may introduce reverse causality. Prospective studies are needed to clarify temporal sequence and strengthen causal interpretation. Second, although ML captures nonlinearities and high-order interactions without prior specification, it remains inherently data-driven. SHAP interpretations are model-specific and descriptive rather than causal, and feature importance

may vary with feature selection, hyperparameter tuning, and class-imbalance handling. Future work should examine the robustness of SHAP results across different modelling pipelines. Third, LSA and NSSI were measured using single-item, retrospective self-reports, which are prone to recall bias and misclassification (Su *et al.*, 2023). Stigma, especially among older and rural populations, may further contribute to underreporting and underestimation of prevalence (Wu *et al.*, 2024a, 2024b). Fourth, the absence of psychiatric diagnoses may have reduced model performance and limited insights into mental health–related risk factors, despite their known associations with suicidal behaviour (Mullins *et al.*, 2022). Finally, model development and validation were conducted within a single dataset, which limits generalizability and highlights the need for external validation in future research.

Conclusion

The prevalence of LSA among Chinese adults remains relatively high, with a clear age gradient—peaking in young adults and declining with age. Risk profiles revealed both shared and age-specific predictors, reflecting distinct life-stage vulnerabilities. These findings highlight the need for age-tailored suicide prevention strategies in China.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S2045796025100231>.

Availability of data and materials. Data are available on reasonable request. The data that support the findings of this study are available from the PBICR team (Yibo Wu) upon reasonable request.

Acknowledgements. We gratefully acknowledge the PBICR team for their significant efforts in conducting the survey.

Author contributions. Yu Wu: conceptualization, methodology, formal analysis, writing-original draft, and writing-review & editing; Yihao Zhao, Chen Chen, Panliang Zhong: methodology, visualization and validation, and writing-review & editing; Yibo Wu: project administration, validation, and writing-review & editing; Xiaoying Zheng: conceptualization, supervision, project administration, writing-original draft, and writing-review & editing. Yibo Wu can also be contacted for correspondence, email bjmuwuyibo@outlook.com.

Financial support. This work was supported by the National Key Research and Development Program of China (2022YFC3600800), the Population and Aging Health Science Program (WH10022023035).

Competing interests. There are no actual or potential conflicts of interest, including any financial, personal, or other relationships with other people or organizations.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2000. Ethical approval was granted by the Key Laboratory of Health Economics and Policy Research of the National Health Commission (NHC-HEPR202401) and the Ethics Committee of Shanghai Jiao Tong University (H202402371). The study was registered in the Chinese Clinical Trial Registry (ChiCTR). All participants provided informed consent.

References

- Anestis MD (2016) Prior suicide attempts are less common in suicide decedents who died by firearms relative to those who died by other means. *Journal of Affective Disorders* **189**, 106–109. <https://doi.org/10.1016/j.jad.2015.09.007>

- Angelakis I, Gillespie EL and Panagioti M (2019) Childhood maltreatment and adult suicidality: A comprehensive systematic review with meta-analysis. *Psychological Medicine* 49(7), 1057–1078. <https://doi.org/10.1017/s0033291718003823>
- Borritz M, Bültmann U, Rugulies R, Christensen KB, Villadsen E and Kristensen TS (2005) Psychosocial work characteristics as predictors for burnout: Findings from 3-year follow up of the PUMA Study. *J Occup Environ Medicine* 47(10), 1015–1025. <https://doi.org/10.1097/01.jom.0000175155.50789.98>
- Bostwick JM, Pabbati C, Geske JR and McKean AJ (2016) Suicide attempt as a risk factor for completed suicide: Even more lethal than we knew. *American Journal of Psychiatry* 173(11), 1094–1100. <https://doi.org/10.1176/appi.ajp.2016.15070854>
- Breiman L (2001) Random forests. *Machine Learning* 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chen F, Chi J, Niu F, Gao Q, Mei F, Zhao L, Hu K, Zhao B and Ma B (2022) Prevalence of suicidal ideation and suicide attempt among patients with traumatic brain injury: A meta-analysis. *Journal of Affective Disorders* 300, 349–357. <https://doi.org/10.1016/j.jad.2022.01.024>
- Chen R, Zhu X, Wright L, Drescher J, Gao Y, Wu L, Ying X, Qi J, Chen C, Xi Y, Ji L, Zhao H, Ou J and Broome MR (2019) Suicidal ideation and attempted suicide amongst Chinese transgender persons: National population study. *Journal of Affective Disorders* 245, 1126–1134. <https://doi.org/10.1016/j.jad.2018.12.011>
- Chen T and Guestrin C (2016) *XGBoost: a Scalable Tree Boosting System* Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. <https://doi.org/10.1145/2939672.2939785>
- Conner A, Azrael D and Miller M (2019) Suicide case-fatality rates in the United States, 2007 to 2014: A nationwide population-based study. *Annals of Internal Medicine* 171(12), 885–895. <https://doi.org/10.7326/m19-1324>
- Domingos P and Pazzani M (1997) On the optimality of the simple Bayesian Classifier under Zero-One Loss. *Machine Learning* 29(2), 103–130. <https://doi.org/10.1023/A:1007413511361>
- Fan M, Sun D, Zhou T, Heianza Y, Lv J, Li L and Qi L (2020) Sleep patterns, genetic susceptibility, and incident cardiovascular disease: A prospective study of 385 292 UK biobank participants. *European Heart Journal* 41(11), 1182–1189. <https://doi.org/10.1093/eurheartj/ehz849>
- Feng X and Chen XY (2012) Reliability and validity of the New General Self-Efficacy Scale (NGSES). *Journal of Mudanjiang Normal University (Philosophy and Social Sciences Edition)* 21(4), 127–129. <https://doi.org/10.13815/j.cnki.jmtc.2012.04.042>
- Foster HME, Gill JMR, Mair FS, Celis-Morales CA, Jani BD, Nicholl BJ, Lee D and O'Donnell CA (2023) Social connection and mortality in UK Biobank: A prospective cohort analysis. *Bmc Medicine* 21(1), 384. <https://doi.org/10.1186/s12916-023-03055-7>
- Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, Musacchio KM, Jaroszewski AC, Chang BP and Nock MK (2017) Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin* 143(2), 187–232. <https://doi.org/10.1037/bul0000084>
- García de la Garza Á, Blanco C, Olfson M and Wall MM (2021) Identification of suicide attempt risk factors in a national US survey using machine learning. *Jama Psychiatry* 78(4), 398–406. <https://doi.org/10.1001/jamapsychiatry.2020.4165>
- Gordon JA, Avenevoli S and Pearson JL (2020) Suicide prevention research priorities in health care. *Jama Psychiatry* 77(9), 885–886. <https://doi.org/10.1001/jamapsychiatry.2020.1042>
- Haibo H, Yang B, Garcia EA and Shutao L (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence),
- Huang F, Wang H, Wang Z, Zhang J, Du W, Su C, Jia X, Ouyang Y, Wang Y, Li L, Jiang H and Zhang B (2020) Psychometric properties of the perceived stress scale in a community sample of Chinese. *BMC Psychiatry* 20(1), 130. <https://doi.org/10.1186/s12888-020-02520-4>
- Jacobucci R, Littlefield AK, Millner AJ, Kleiman EM and Steinley D (2021) Evidence of inflated prediction performance: a commentary on machine learning and suicide research. *Clinical Psychological Science* 9(1), 129–134. <https://doi.org/10.1177/2167702620954216>
- Kearns JC, Coppersmith DDL, Santee AC, Insel C, Pigeon WR and Glenn CR (2020) Sleep problems and suicide risk in youth: A systematic review, developmental framework, and implications for hospital treatment. *General Hospital Psychiatry* 63, 141–151. <https://doi.org/10.1016/j.genhosppsych.2018.09.011>
- Klonsky ED, May AM and Saffer BY (2016) Suicide, suicide attempts, and suicidal ideation. *Annual Review of Clinical Psychology* 12, 307–330. <https://doi.org/10.1146/annurev-clinpsy-021815-093204>
- Lee S, Fung SC, Tsang A, Liu ZR, Huang YQ, He YL, Zhang MY, Shen YC, Nock MK and Kessler RC (2007) Lifetime prevalence of suicide ideation, plan, and attempt in metropolitan China. *Acta Psychiatrica Scandinavica* 116(6), 429–437. <https://doi.org/10.1111/j.1600-0447.2007.01064.x>
- Lee S, Gornitz N, Xing EP, Heckerman D and Lippert C (2018) Ensembles of Lasso Screening Rules. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(12), 2841–2852. <https://doi.org/10.1109/tpami.2017.2765321>
- Leung H, Pakpour AH, Strong C, Lin YC, Tsai MC, Griffiths MD, Lin CY and Chen IH (2020) Measurement invariance across young adults from Hong Kong and Taiwan among three internet-related addiction scales: Bergen Social Media Addiction Scale (BSMAS), Smartphone Application-Based Addiction Scale (SABAS), and Internet Gaming Disorder Scale-Short Form (IGDS-SF9) (Study Part A). *Addictive Behaviors* 101, 105969. <https://doi.org/10.1016/j.addbeh.2019.04.027>
- Levis B, Benedetti A and Thombs BD (2019) Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 365, 11476. <https://doi.org/10.1136/bmj.11476>
- Lin L, Wang HH, Lu C, Chen W and Guo VY (2021) Adverse childhood experiences and subsequent chronic diseases among middle-aged or older adults in China and Associations With Demographic and Socioeconomic Characteristics. *JAMA Network Open* 4(10), e2130143. <https://doi.org/10.1001/jamanetworkopen.2021.30143>
- Lyu J, Wang Y, Shi H and Zhang J (2018) Early warnings for suicide attempt among Chinese rural population. *Journal of Affective Disorders* 238, 353–358. <https://doi.org/10.1016/j.jad.2018.06.009>
- Miller M, Azrael D and Barber C (2012) Suicide mortality in the United States: The importance of attending to method in understanding population-level disparities in the burden of suicide. *Annual Review of Public Health* 33, 393–408. <https://doi.org/10.1146/annurev-publhealth-031811-124636>
- Miller M, Azrael D and Hemenway D (2004) The epidemiology of case fatality rates for suicide in the northeast. *Annals of Emergency Medicine* 43(6), 723–730. <https://doi.org/10.1016/j.annemergmed.2004.01.018>
- Mullins N, Kang J, Campos AI, Coleman JRI, Edwards AC, Galfalvy H, Levey DE, Lori A, Shabalin A, Starnawska A, Su MH, Watson HJ, Adams M, Awasthi S, Gandal M, Hafferty JD, Hishimoto A, Kim M, Okazaki S, Otsuka I, Ripke S, Ware EB, Bergen AW, Berrettini WH, Bohus M, Brandt H, Chang X, Chen WJ, Chen HC, Crawford S, Crow S, DiBlasi E, Duriez P, Fernández-Aranda F, Fichter MM, Gallinger S, Glatt SJ, Gorwood P, Guo Y, Hakonarson H, Halmi KA, Hwu HG, Jain S, Jamain S, Jiménez-Murcia S, Johnson C, Kaplan AS, Kaye WH, Keel PK, Kennedy JL, Klump KL, Li D, Liao SC, Lieb K, Lilienfeld L, Liu CM, Magistretti PJ, Marshall CR, Mitchell JE, Monson ET, Myers RM, Pinto D, Powers A, Ramoz N, Roepke S, Rozanov V, Scherer SW, Schmahl C, Sokolowski M, Strober M, Thornton LM, Treasure J, Tsuang MT, Witt SH, Woodside DB, Yilmaz Z, Zillich L, Adolfsson R, Agartz I, Air TM, Alda M, Alfredsson L, Andreassen OA, Anjorin A, Appadurai V, Soler Artigas M, Van der Auwera S, Azevedo MH, Bass N, Bau CHD, Baune BT, Bellivier F, Berger K, Biernacka JM, Bigdeli TB, Binder EB, Boehnke M, Boks MP, Bosch R, Braff DL, Bryant R, Budde M, Byrne EM, Cahn W, Casas M, Castela E, Cervilla JA, Chaumette B, Cichon S, Corvin A, Craddock N, Craig D, Degenhardt F, Djurovic S, Edenberg HJ, Fanous AH, Foo JC, Forstner AJ, Frye M, Fullerton JM, Gatt JM, Gejman PV, Giegling I, Grabe HJ, Green MJ, Grevet EH, Grigoriou-Serbanescu M, Gutierrez B, Guzman-Parra J, Hamilton SP, Hamshere ML, Hartmann A, Hauser J, Heilmann-Heimbach S, Hoffmann P, Ising M, Jones I, Jones LA, Jonsson L, Kahn RS, Kelsoe JR, Kendler KS, Kloiber S, Koenen KC, Kogevinas M, Konte B, Krebs MO, Landén M, Lawrence J, Leboyer M,

- Lee PH, Levinson DE, Liao C, Lissowska J, Lucae S, Mayoral F, McElroy SL, McGrath P, McGuffin P, McQuillin A, Medland SE, Mehta D, Melle I, Milaneschi Y, Mitchell PB, Molina E, Morken G, Mortensen PB, Müller-Myhsok B, Nievergelt C, Nimgaonkar V, Nöthen MM, O'Donovan MC, Ophoff RA, Owen MJ, Pato C, Pato MT, Penninx B, Pimm J, Pistis G, Potash JB, Power RA, Preisig M, Quesed D, Ramos-Quiroga JA, Reif A, Ribasés M, Richarte V, Rietschel M, Rivera M, Roberts A, Roberts G, Rouleau GA, Rovaris DL, Rujescu D, Sánchez-Mora C, Sanders AR, Schofield PR, Schulze TG, Scott LJ, Serretti A, Shi J, Shyn SI, Sirignano L, Sklar P, Smeland OB, Smoller JW, Sonuga-Barke EJS, Spalletta G, Strauss JS, Świątkowska B, Trzaskowski M, Turecki G, Vilar-Ribó L, Vincent JB, Völzke H, Walters JTR, Shannon Weickert C, Weickert TW, Weissman MM, Williams LM, Wray NR, Zai CC, Ashley-Koch A, Beckham JC, Hauser ER, Hauser MA, Kimbrel NA, Lindquist JH, McMahon B, Oslin DW, Qin X, Agerbo E, Borglum AD, Breen G, Erlangen A, Esko T, Gelernter J, Hougaard D, Kessler RC, Kranzler HR, Li QS, Martin NG, McIntosh AM, Mors O, Nordentoft M, Olsen CM, Porteous D, Ursano RJ, Wasserman D, Werge T, Whiteman DC, Bulik CM, Coon H, Demontis D, Docherty AR, Kuo PH, Lewis CM, Mann JJ, Rentería ME, Smith DJ, Stahl EA, Stein MB, Streit F, Willour V and Ruderfer DM (2022) Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. *Biological Psychiatry* 91(3), 313–327. <https://doi.org/10.1016/j.biopsych.2021.05.029>
- National Bureau of Statistics of China (2023) China Statistical Yearbook 2023. Beijing: China Statistics Press.
- Nock MK, Borges G, Bromet EJ, Alonso J, Angermeyer M, Beautrais A, Bruffaerts R, Chiu WT, de Girolamo G, Gluzman S, de Graaf R, Gureje O, Haro JM, Huang Y, Karam E, Kessler RC, Lepine JB, Levinson D, Medina-Mora ME, Ono Y, Posada-Villa J and Williams D (2008a) Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *British Journal of Psychiatry* 192(2), 98–105. <https://doi.org/10.1192/bjp.bp.107.040113>
- Nock MK, Borges G, Bromet EJ, Cha CB, Kessler RC and Lee S (2008b) Suicide and suicidal behavior. *Epidemiologic Reviews* 30(1), 133–154. <https://doi.org/10.1093/epirev/mxn002>
- Norman RE, Byambaa M, De R, Butchart A, Scott J and Vos T (2012) The long-term health consequences of child physical abuse, emotional abuse, and neglect: A systematic review and meta-analysis. *PLOS Medicine* 9(11), e1001349. <https://doi.org/10.1371/journal.pmed.1001349>
- Oquendo MA, Wall M, Wang S, Olfson M and Blanco C (2024) Lifetime suicide attempts in otherwise psychiatrically healthy individuals. *Jama Psychiatry* 81(6), 572–578. <https://doi.org/10.1001/jamapsychiatry.2023.5672>
- Pan X and Xu Y (2022) A Safe Feature Elimination Rule for L(1)-Regularized Logistic Regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(9), 4544–4554. <https://doi.org/10.1109/tpami.2021.3071138>
- Rezvani S and Wu J (2023) Handling multi-class problem by intuitionistic fuzzy twin support vector machines based on relative density information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(12), 14653–14664. <https://doi.org/10.1109/tpami.2023.3310908>
- Rudd MD (2000) The suicidal mode: A cognitive-behavioral model of suicidality. *Suicide and Life-Threatening Behavior* 30(1), 18–33.
- Song X, Liu X, Zhou Y and Zhang X (2023) Prevalence and correlates of suicide attempts in young patients with first-episode and drug-naïve major depressive disorder: A large cross-sectional study. *Journal of Affective Disorders* 340, 340–346. <https://doi.org/10.1016/j.jad.2023.08.006>
- Stekhoven DJ and Buhlmann P (2012) MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28(1), 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Su R, John JR and Lin PI (2023) Machine learning-based prediction for self-harm and suicide attempts in adolescents. *Psychiatry Research* 328, 115446. <https://doi.org/10.1016/j.psychres.2023.115446>
- Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA and Joiner Jr TE (2010) The interpersonal theory of suicide. *Psychological Review* 117(2), 575–600. <https://doi.org/10.1037/a0018697>
- Vidal C, Reinert M, Nguyen T and Jun HJ (2024) Chronic stress and lack of social support: Role in adolescent depression and suicide-related behaviors in the context of the COVID-19 pandemic. *Journal of Affective Disorders* 365, 437–442. <https://doi.org/10.1016/j.jad.2024.08.090>
- Wang F, Wu Y, Wang S, Du Z and Wu Y (2024) Development of an optimal short form of the GAD-7 scale with cross-cultural generalizability based on Riskslim. *General Hospital Psychiatry* 87, 33–40. <https://doi.org/10.1016/j.genhosppsych.2024.01.010>
- WHO. (2021) *Suicide worldwide in 2019: Global Health Estimates*. <https://www.who.int/publications/i/item/9789240026643> (accessed 5 November 2023)
- Wu R, Zhu H, Wang ZJ and Jiang CL (2021) A Large Sample Survey of Suicide Risk among University Students in China. *BMC Psychiatry* 21(1), 474. <https://doi.org/10.1186/s12888-021-03480-z>
- Wu Y, Su B, Chen C, Zhao Y, Zhong P and Zheng X (2023) Urban-rural disparities in the prevalence and trends of depressive symptoms among Chinese elderly and their associated factors. *Journal of Affective Disorders* 340, 258–268. <https://doi.org/10.1016/j.jad.2023.07.117>
- Wu Y, Su B, Zhao Y, Chen C, Zhong P and Zheng X (2024a) Epidemiological features of suicidal ideation among the elderly in China based meta-analysis. *BMC Psychiatry* 24(1), 562. <https://doi.org/10.1186/s12888-024-06010-9>
- Wu Y, Su B, Zhong P, Wang Y, Huang Y and Zheng X (2024b) The long-term changing patterns of suicide mortality in China from 1987 to 2020: Continuing urban-rural disparity. *BMC Public Health* 24(1), 1269. <https://doi.org/10.1186/s12889-024-18743-z>
- Wu Y, Tang J, Du Z, Chen K, Wang F, Sun X, Zhang G and Wu Y (2025) Development of a short version of the perceived social support scale: Based on classical test theory and ant colony optimization. *BMC Public Health* 25(1), 232. <https://doi.org/10.1186/s12889-025-21399-y>
- Yang LS, Zhang ZH, Sun L, Sun YH and Ye DQ (2015) Prevalence of suicide attempts among college students in China: A meta-analysis. *PLoS One* 10(2), e0116303. <https://doi.org/10.1371/journal.pone.0116303>
- Yazdi-Ravandi S, Khazaei S, Davari H, Matinnia N, Karami M, Taslimi Z, Afkhami MR and Ghaleiha A (2023) Gender and age differences in suicide attempt: A large population study in the West of Iran. *Asian Journal of Psychiatry* 81, 103470. <https://doi.org/10.1016/j.ajp.2023.103470>
- Yount KM, Cheong YE, Khan Z, Bergenfeld I, Kaslow N and Clark CJ (2022) Global measurement of intimate partner violence to monitor Sustainable Development Goal 5. *BMC Public Health* 22(1), 465. <https://doi.org/10.1186/s12889-022-12822-9>