

Risk Prediction Model for Non-Suicidal Self-Injury in Chinese Adolescents with Major Depressive Disorder Based on Machine Learning

Ting Sun^{1,2,*}, Jingfang Liu^{3,*}, Hui Wang^{3,*}, Bing Xiang Yang³⁻⁵, Zhongchun Liu³, Jie Liu⁶, Zhiying Wan³, Yinglin Li¹, Xiangying Xie¹, Xiaofen Li³, Xuan Gong³, Zhongxiang Cai¹

¹Department of Nursing, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China; ²Health Science Center, Yangtze University, Jingzhou, People's Republic of China; ³Department of Psychiatry, Renmin Hospital of Wuhan University, Wuhan, People's Republic of China; ⁴School of Nursing, Wuhan University, Wuhan, People's Republic of China; ⁵Population and Health Research Center, Wuhan University, Wuhan, People's Republic of China; ⁶Anesthesiology, Virginia Commonwealth University Health System, Richmond, VA, USA

*These authors contributed equally to this work

Correspondence: Zhongxiang Cai, Department of Nursing, Renmin Hospital of Wuhan University, No. 238 Jiefang Road, Wuhan, Hubei Province, 430060, People's Republic of China, Email tg20201228@163.com; Xuan Gong, Department of Psychiatry, Renmin Hospital of Wuhan University, No. 238 Jiefang Road, Wuhan, Hubei Province, 430060, People's Republic of China, Email 12048387@qq.com

Background: Non-suicidal self-injury (NSSI) is a significant social issue, especially among adolescents with major depressive disorder (MDD). This study aimed to construct a risk prediction model using machine learning (ML) algorithms, such as XGBoost and random forest, to identify interventions for healthcare professionals working with adolescents with MDD.

Methods: This study investigated 488 adolescents with MDD. Adolescents were randomly divided into 75% training set and 25% test set to testify the predictive value of risk prediction model. The prediction model was constructed using XGBoost and random forest algorithms. We evaluated the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, recall, F Score of the two models for comparing the performance of the two models.

Results: There were 161 (33.00%) participants having NSSI. Compared without NSSI, there were statistically significant differences in gender ($P=0.035$), age ($P=0.036$), depressive symptoms ($P=0.042$), sleep quality ($P=0.030$), dysfunctional attitudes ($P=0.048$), childhood trauma ($P=0.046$), interpersonal problems ($P=0.047$), psychoticism ($P=0.049$), neuroticism ($P=0.044$), punishing and Severe ($F2$) ($P=0.045$) and Overly-intervening and Protecting ($M2$) ($P=0.047$) with NSSI. The AUC values for random forest and XGBoost were 0.780 and 0.807, respectively. The top five most important risk predictors identified by both machine learning methods were dysfunctional attitude, childhood trauma, depressive symptoms, $F2$ and $M2$.

Conclusion: The study demonstrates the suitability of prediction models for predicting NSSI behavior in Chinese adolescents with MDD based on ML. This model improves the assessment of NSSI in adolescents with MDD by health care professionals working. This provides a foundation for focused prevention and interventions by health care professionals working with these adolescents.

Keywords: non-suicidal self-injury, adolescents, major depressive disorder, risk prediction model

Introduction

Non-suicidal self-injury is the deliberate destruction of one's own body without suicidal intention, such as cutting, burning, biting, or scratching skin and other self-inflicted injuries.¹ In Chinese culture, the value of collectivism and the emphasis on "face" may lead to conflicts between adolescents' sense of social expectations and personal identity, which could be a reason for their non-suicidal self-injury (NSSI) behavior. At the same time, the immense pressure of the Gaokao and academic competition can lead to mental health issues such as depression and anxiety in adolescents, thereby increasing their risk of engaging in NSSI. Adolescents have a high incidence of NSSI, with a global incidence of 14% ~ 39% and an increasing trend year by year.² NSSI rarely manifests in isolation, NSSI and major depressive disorder

(MDD) often co-occur currently.³ The prevalence of MDD in adolescents is 11% - 20%.⁴ The lifetime prevalence rate of NSSI is 17% in the general population and 74% in adolescents with mental disorders.⁵ Studies have shown that the incidence of NSSI behavior in patients with depression is 37%.⁶ The incidence of adolescent depression has been increasing annually along with the incidence of NSSI. Non-suicidal self-injury seriously endangers teenagers' mental health and affects their social functions.⁷ More than 70% of adolescents with NSSI had suicidal ideation.⁸ It is an important predictor of suicidal behavior in adolescents.⁹ Therefore, early detection of the risk of NSSI in adolescent patients with depression is very important in initiating early monitoring and prevention interventions.

This study adopts the Biopsychosocial Model as its theoretical framework, which was initially proposed by George Engel in 1977 to explain psychological conditions such as emotional dysregulation and borderline personality disorder. This model integrates biological, psychological, and social environmental factors to assess the impact of these factors on an individual's mental health.¹⁰ On a biological level, the model focuses on individual differences and variations in brain structure and function. Hormonal differences due to gender may lead to changes in emotional regulation and impulse control. For instance, fluctuations in estrogen levels during a woman's menstrual cycle can exacerbate emotional volatility and are associated with an increase in non-suicidal self-injury (NSSI) behaviors (Bresin & Schoenleber, 2015). The neurodevelopmental stages, which are related to age, significantly influence the structure and function of the brain, particularly the maturity of the prefrontal cortex, which is involved in higher cognitive functions such as decision-making and impulse control. The incomplete neurodevelopment during adolescence can lead to vulnerabilities in emotional regulation and behavioral control, potentially increasing the risk of NSSI behaviors.¹¹

From a psychological perspective, cognitive biases formed during an individual's growth process can lead to dysfunctional attitudes, such as low self-esteem, pessimism about the future, or a lack of control. These cognitive biases can result in difficulties with emotional regulation, increasing the risk of mood disorders like anxiety and depression, which may then trigger NSSI behaviors. Personality traits, as psychological factors, influence an individual's cognition, emotions, and behaviors. Certain personality traits, such as those associated with borderline personality disorder, are linked to difficulties in emotional regulation and impulse control disorders, which may lead individuals to adopt NSSI behaviors as a coping strategy under stress.¹²

On the social environmental level, parenting styles significantly affect the mental health and behavioral development of adolescents. A family environment characterized by positive communication and strong support can help reduce the risk of NSSI behaviors in adolescents.¹³ Conversely, family conflicts, neglect, or abuse increase the likelihood of adolescents engaging in NSSI behaviors. Active parental involvement and emotional support not only provide adolescents with a necessary social support network but also help alleviate their stress and emotional distress, thereby reducing the occurrence of NSSI behaviors. Childhood trauma and negative life events as social environmental factors can severely impact the establishment of stable and supportive relationships during an individual's growth process, thereby affecting their emotional regulation abilities.¹⁴ Childhood trauma can also lead to interpersonal difficulties and increased emotional stress, thus increasing the likelihood of NSSI behaviors. Insomnia, resulting from the interplay of biopsychosocial factors, may exacerbate depressive symptoms and impulsive behaviors, potentially leading to NSSI.¹⁵

Furthermore, the Biopsychosocial Model emphasizes that individuals develop a range of coping strategies to deal with emotional stress. In the absence of healthy coping strategies, these may become maladaptive, such as engaging in self-harming behaviors. The model also provides guidance for the treatment of NSSI, such as Dialectical Behavior Therapy (DBT). DBT aims to help individuals learn more effective cognitive-behavioral skills, improve emotional regulation, enhance interpersonal effectiveness, and reduce NSSI behaviors. Please note that I have corrected the attribution of the model to George Engel and have attempted to simplify some sentences to improve clarity and fluency. Additionally, I have made sure the citation format adheres to academic standards, pending the inclusion of specific publication years and page numbers according to the actual literature.

Machine learning, as one of the primary tools of data mining, is integral to artificial intelligence. It is extensively employed across various fields, including healthcare, particularly in predictive modeling.¹⁶ Machine learning involves using algorithms to make predictions from data¹⁷ and is already used in identifying predictive tasks.¹⁸ These learning methods rely on patients' clinical test data or monitoring data, offering valuable guidance for early intervention and treatment, thus playing a crucial role in disease prevention and control. The Random Forest algorithm is highly regarded

for its robustness and its ability to capture nonlinear relationships between the dependent and multiple independent variables. This algorithm can provide accurate predictions for large-scale complex datasets and is relatively easy to implement and efficient. The XGBoost algorithm has also shown its strength in handling missing data. It constructs a high-performance predictive model by integrating multiple weak predictive models.¹⁹ XGBoost can automatically analyze complex medical data, revealing nonlinear relationships within the data, offering a clear advantage over traditional regression analysis methods. In terms of the area under the Receiver Operating Characteristic (ROC) curve and decision curve analysis, XGBoost has shown superior predictive performance, indicating its potential as a powerful tool for clinical predictive modeling.²⁰ However, to date, there has been no research report comparing the predictive accuracy of these two algorithms in forecasting NSSI behavior in adolescents with depression.

Early and effective prevention of NSSI in adolescents is a prerequisite for preventing adolescent suicide. The risk of NSSI varies based on demographics, personality characteristics, physiology, psychological factors and the environment, involving the interaction of multiple factors.²¹ Current NSSI risk assessment tools in clinical practice have limited accuracy in identifying. Accurate risk detection is the most critical step in identifying NSSI behavior. In this study, data from adolescents with MDD were extracted to explore the contribution rate of independent variables to self-injury behavior. Random forest and XGBoost algorithms were used to create a prediction model.

Methods

Participants and Setting

This cross-sectional study is part of the Early-Warning System and Comprehensive Intervention for Depression (ESCID) study. The purpose of the ESCID study was to explore the mechanisms of the onset and recurrence of depression and develop interventions.²² A total of 488 adolescents aged 13–25 years (Chinese public security limit the age of adolescents to 13 to 25)²³ were recruited from the psychiatric outpatient clinic of Renmin Hospital of Wuhan University between April 2019 and April 2021. These adolescents were diagnosed with major depressive disorder (MDD) by two experienced psychiatrists in the psychiatric department in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).²⁴ The following exclusion criteria were applied: (1) inability to continue participation in the study due to suicidal behavior, refusal to eat or stupor; (2) history of serious adverse drug reaction; (3) presence of serious physical disease; (4) pregnancy; (5) symptoms of mania; (6) history of cerebral organic disease; (7) dependence on alcohol and other addictive substances.

Survey Process

Participants and their parents were provided with detailed information about the purpose of the study and their informed consent was obtained prior to their participation and their parents (If the participant is a minor, the consent of their parents is required). Participants were assured that they could withdraw from the study at any time without facing any negative consequences.

To ensure a conducive environment for completing the questionnaire, participants were escorted to a quiet room where they were provided with an iPad to complete the questionnaire. The socio-demographic questionnaire, which captured information such as age, gender, and location of residence, was designed by the researchers. Additionally, participants were asked to complete self-report scales, and trained researchers completed an observer rating scale.

Ethical Considerations

Ethical considerations were strictly followed throughout the study, adhering to the latest version of the Declaration of Helsinki. This study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (NO: WCRY2020-K004).

Measurements

Socio-Demographic Characteristics

Gender (male or female), age (in years).

Depressive Symptoms

The Hamilton Depression Scale-17 (HAMD-17) was used to measure the severity of depression symptoms.²⁵ Most of the items are scored on a scale of 0 to 4: 0 = “normal”, 1 = “mild”, 2 = “moderate”, 3 = “severe”, 4 = “extremely severe”. Some items are rated on a scale of 0 to 2: 0 = “normal”, 1 = “mild to moderate”, and 2 = “severe”. The overall level of depression is determined by the total score on the HAMD-17: no depression (0–7 points), mild depression (8–17 points), moderate depression (18–24 points), and severe depression (> 24). The HAMD-17 has been validated and applied to Chinese individuals, demonstrating good reliability in measuring the severity of depression symptoms.

Sleep Quality

The Pittsburgh Sleep Quality Index (PSQI) was used to evaluate the sleep quality of the participants over the past month. The PSQI consists of 18 items that cover seven factors related to sleep quality: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. Each factor is rated on a scale of 0 to 3, with higher scores indicating poorer sleep quality. The scores for all factors are then totaled. A PSQI score >7 indicates the presence of a sleep disorder. The PSQI has been validated and applied to Chinese individuals, demonstrating good reliability in assessing sleep quality.

Personality

The Eysenck Personality Questionnaire (EPQ) was used to evaluate the personality of the participants.²⁶ The questionnaire consisted of 88 items, which were divided into several subscales: internal and psychoticism (P), the extraversion (E), the neuroticism (N), and the lie (L). The EPQ is a widely used measure in China and has demonstrated high reliability and validity for assessing personality traits.

Dysfunctional Attitude

A person's negative attitudes and beliefs were assessed using the Dysfunctional Attitude Scale (DAS).²⁷ This 40-item scale consists of eight sub-scales including vulnerability, attraction and repulsion, perfectionism, compulsion, seeking applause, dependence, self-determination attitude and cognition philosophy. Higher scores in each domain indicate more severe cognitive distortions. The Cronbach's alpha of this scale was 0.87 and split-half reliability of 0.835.

Childhood Trauma

The Childhood Trauma Questionnaire (CTQ) is a 28-item scale designed to assess childhood trauma. It consists of five sub-scales: sexual abuse, physical abuse, emotional abuse, emotional neglect and physical neglect. The total score on the CTQ reflects the extent of childhood trauma experienced, with higher scores indicating greater trauma. For the specific sub-scales, a score of 8 or higher for sexual abuse, 10 or higher for physical abuse, 13 or higher for emotional abuse, 15 or higher for emotional neglect, and 10 or higher for physical neglect is considered indicative of childhood trauma. The Chinese version of the CTQ has shown good reliability and validity with a Cronbach's alpha of 0.77 and the Cronbach's alpha of this sub-scale was 0.41–0.68.²⁸

Interpersonal Problems

The Interpersonal Relationship Comprehensive Diagnostic Scale (ICDS) is a 28-item scale developed by Richang Zheng to measure the degree of interpersonal distress.²⁹ Each item is scored based on whether the participant feels it accurately represents their situation, with 1 point given for a fitting item and no point for an unfitting item. A total score of 0 to 8 indicates low level of relationship distress; a score of 9 to 14 indicates some level of relationship distress; and a score of 15 to 28 indicates severe relationship distress.

Parenting Style

The Chinese version of the Egna Minnen av Barndoms Uppfostran (EMBU) was used to measure participants' self-perceived parenting style.³⁰ This scale consists of 66 items and six subscales: warm and affectionate (F1), punishing and severe (F2), overly-intervening (F3), favoring (F4), rejecting (F5), and overly-protecting (F6). The mothering parenting style consists of five subscales: warm and affectionate (M1), overly-intervening and protecting (M2), rejecting (M3),

punishing and severe (M4), and favoring (M5). The EMBU questionnaire provides insights into participants' perceptions of their parents' parenting behaviors across these various dimensions.

NSSI Behavior

Experienced physicians utilized the Kiddie-SADS-Lifetime Version (K-SADS-PL) screening interview to assess NSSI behavior in adolescents with MDD.³¹ The interview included specific questions such as: "Did you try to hurt yourself and how?", "Why did you do this?", "Were you intending to kill yourself or not?", "How many times have you done this kind of behavior?" Based on the interview results, clinicians would determine whether the patient has a history of NSSI behavior, the methods employed, and the frequency of engagement in such behaviors. The K-SADS-PL interview serves as a valuable tool for evaluating and documenting NSSI behavior in this population.

Statistical Analysis

The sample size of the model was calculated based on Events per variable (EPV). The independent variables included in the model were 10, the prevalence of NSSI behaviors among Chinese adolescents with MDD was 44%,³² and the EPV was taken as 10. The sample size was $10 \times 10 / 0.44 = 228$. Considering 20% invalid samples, the final sample size was 274. Therefore, 488 participants recruited for this study met the requirements.

The dependent variables were defined as NSSI behavior (no=0, yes=1), and the independent variables were gender (0=Male, 1=Female), age, depressive symptoms, sleep quality, personality, dysfunctional attitudes, childhood trauma, interpersonal relationship, fathering parenting style, and mothering parenting style.

The data set was randomly divided into 75% training set and 25% test set. The training set and test set are analyzed descriptively. Continuous variables were described by means (M) and standard deviation (SD), and categorical variables by frequency and percentage. Single factor analysis was performed on the training set. The *t*-test and Chi-square test were used to test whether the independent variables were different between the NSSI group and the non-NSSI group, excluding variables with a P value greater than 0.05.

After the screening process, two well-known machine learning algorithms, namely random forest and XGBoost were selected to build the predictive model for outcome variable. These algorithms have been widely used in the literature and known for their effectiveness in predictive modeling.

Random Forest

Random forest is a machine learning algorithm that utilizes the concept of ensemble learning by combining multiple decision trees. Each decision tree in the random forest is a classifier that independently makes predictions on the input data. When making predictions, the random forest algorithm aggregates the classification results from all the decision trees and selects the class with the highest number of votes as the final output. The use of random forest provides several benefits, including improved accuracy, robustness against overfitting, and the ability to handle high-dimensional data. It is a popular algorithm for classification and regression tasks, known for its versatility and good performance in a variety of domains.

eXtreme Gradient Boosting

XGBoost belongs to the Gradient Boosting Decision Tree (GBDT) model family. The fundamental principle of GBDT is to iteratively build new models that fit the residual errors of the previous models, aiming to progressively reduce the overall error of the additive model. XGBoost improves upon the classic GBDT algorithm by introducing several enhancements that contribute to its effectiveness and performance. Firstly, XGBoost retains more information about the target function during the model building process, allowing for better model representation and learning. Secondly, XGBoost incorporates regularization techniques by adding penalty terms to the model's objective function, resulting in lower variances and better generalization performance. Thirdly, XGBoost includes a built-in mechanism for handling missing values in the data. Lastly, XGBoost introduces several engineering features, such as candidate loci cutting and feature parallelism, which improve its efficiency and scalability. Overall, XGBoost has gained popularity in the machine learning community due to its powerful modeling capabilities, excellent performance, and the ability to handle various types of data.

Performance Measurement

In this study, the importance of variables in predicting the outcome was assessed. Importance refers to the degree of influence that a variable has on the outcome. A higher importance value indicates a stronger impact on the outcome. To evaluate the predictive performance of the models, a Receiver Operating Characteristic (ROC) analysis was conducted. The ROC curve is a graphical representation of the trade-off between the true positive rate and the false positive rate for different classification thresholds. The area under the receiver operating characteristic curve (AUC) is a commonly used measure of the model's predictive accuracy. A higher AUC value (>0.5) indicates better predictive performance. In addition, sensitivity, specificity, accuracy, recall, F Score of these two models were calculated and compared.

To compare the AUC between different prediction models, the Delong test was employed. A p-value <0.05 was considered as statistically significant.

The above analysis was completed using the R (version 4.1.3). The R package “caret” is used to divide the training set and test set. The R package “randomForest” is used to perform a random forest. The R package XGBoost is used to perform XGboost. The R package “pROC” is used to perform ROC analysis.

Results

Participants' Characteristics

A total of 488 adolescents with MDD participated in the current study. There were 161 participants (33.00%) who reported NSSI. The average age of the participants was 21.74 ± 1.71 years. Females comprised 76.2% of the participant. Participants were randomly divided into a training data set (N = 367) and a test data set (N = 121). A total of 121 (32.97%) participants in the training data set and 40 (33.06%) in the test data set reported NSSI. Detailed information is shown in Table 1.

Table 1 Characteristics of Variables

Variable		Training Data Set		Test Data Set	
		NSSI (N=121)	Not NSSI (N=246)	NSSI (N=40)	Not NSSI (N=81)
		$\bar{x} \pm s, N (\%)$	$\bar{x} \pm s, N (\%)$	$\bar{x} \pm s, N (\%)$	$\bar{x} \pm s, N (\%)$
Gender	Male	23 (19.01%)	72 (29.26%)	3 (7.5%)	18 (22.22%)
	Female	98 (80.99%)	174 (70.73%)	37 (92.5%)	63 (77.78%)
Age		21.43 ± 1.72	21.83 ± 1.72	21.52 ± 1.75	22.04 ± 1.55
Depressive symptoms		18.72 ± 7.68	17.07 ± 7.10	19.50 ± 7.45	15.04 ± 8.33
Sleep quality		10.61 ± 4.29	9.61 ± 4.07	11.68 ± 3.74	8.44 ± 3.67
Personality	Psychoticism (P)	53.68 ± 10.18	51.58 ± 9.26	56.38 ± 11.71	50.40 ± 10.01
	Extroversion (E)	44.59 ± 10.05	46.10 ± 11.08	46.25 ± 9.39	49.75 ± 11.40
	Neuroticism (N)	65.41 ± 10.30	63.01 ± 10.92	67.25 ± 8.16	60.80 ± 12.23
	Lying (L)	40.91 ± 9.13	40.87 ± 8.79	39.25 ± 9.37	42.28 ± 8.59
Dysfunctional attitude		171.41 ± 27.37	165.29 ± 28.08	178.22 ± 28.06	158.99 ± 29.03
Childhood trauma		46.26 ± 15.28	43.11 ± 13.61	52.25 ± 18.97	41.83 ± 15.11
Interpersonal problem		15.31 ± 5.46	14.05 ± 5.76	15.62 ± 5.80	12.86 ± 6.39
Fathering parenting style	Warm and affectionate (F1)	42.00 ± 9.33	42.51 ± 10.61	42.92 ± 11.53	45.10 ± 10.20
	Punishing and Severe (F2)	20.26 ± 7.64	18.64 ± 6.99	21.55 ± 8.51	16.74 ± 6.40
	Overly-intervening (F3)	20.40 ± 4.65	19.69 ± 4.24	21.90 ± 4.58	19.37 ± 4.25
	Favoring (F4)	9.90 ± 3.64	9.80 ± 4.08	9.18 ± 4.25	10.21 ± 4.21
	Rejecting (F5)	9.97 ± 3.67	10.35 ± 3.33	11.18 ± 4.77	9.09 ± 3.57
Mothering parenting style	Overly-protecting (F6)	10.69 ± 2.97	10.33 ± 2.87	11.38 ± 3.49	9.80 ± 2.69
	Warm and affectionate (M1)	47.10 ± 11.41	46.95 ± 11.84	46.08 ± 13.19	49.98 ± 10.77
	Overly-intervening and Protecting (M2)	37.00 ± 7.05	35.31 ± 7.91	38.45 ± 8.11	34.10 ± 7.39
	Rejecting (M3)	15.02 ± 5.00	14.32 ± 4.70	15.97 ± 6.13	13.04 ± 4.77
	Punishing and severe (M4)	14.81 ± 5.08	13.95 ± 4.78	16.12 ± 5.72	12.75 ± 4.76
	Favoring (M5)	10.47 ± 3.07	10.31 ± 3.51	9.80 ± 3.57	10.81 ± 3.34

Differences in Risk Factors for NSSI Between Participants with and without NSSI

Compared without NSSI, there were statistically significant differences in gender ($P=0.035$), age ($P=0.036$), depressive symptoms ($P=0.042$), sleep quality ($P=0.030$), dysfunctional attitudes ($P=0.048$), childhood trauma ($P=0.046$), interpersonal problems ($P=0.047$), P ($P=0.049$), N ($P=0.044$), F2 ($P=0.045$) and M2 ($P=0.047$) with NSSI. See Table 2.

Comparison of Predictive Power of Models

Random forest model had an AUC of 0.780, accuracy of 0.760, recall of 0.650, sensitivity of 0.650, specificity of 0.815, and an F1 of 0.642. The XGBoost model had an AUC of 0.807, accuracy of 0.719, recall of 0.850, sensitivity of 0.850, specificity of 0.654, and an F1 of 0.667. Delong test showed that there was no significant difference in the AUC between random forest and XGBoost ($Z=0.848$, $P=0.396$), as shown in Table 3 and Figure 1.

Model Influencing Factors Analysis

In the random forest model, the top five variables of importance were dysfunctional attitude (importance score = 13.802), childhood trauma (importance score = 13.380), depressive symptoms (importance score = 13.141), F2 (importance score = 13.104) and M2 (importance score = 12.627). Figure 2 present the results of variable importance ranking in the random forest model.

In the XGBoost model, the top five variables of importance were dysfunctional attitude (importance score = 0.148), depressive symptoms (importance score = 0.131), F2 (importance score = 0.117), childhood trauma (importance score = 0.113), and M2 (importance score = 0.112). Figure 3 present the results of variable importance ranking in the XGBoost model.

Table 2 Differences in Risk Factors for NSSI Between Participants with and without NSSI

Variable		t, X2	P
Gender	Male	4.450	0.035*
	Female		
Age		2.111	0.036*
Depressive symptoms		-2.036	0.042*
Sleep quality		-2.176	0.030*
Personality	Psychoticism (P)	-1.972	0.049*
	Extroversion (E)	1.266	0.206
	Neuroticism (N)	-2.020	0.044*
	Lying (L)	-0.036	0.972
Dysfunctional attitude		-1.981	0.048*
Childhood trauma		-2.006	0.046*
Interpersonal problem		-1.993	0.047*
Fathering parenting style	Warm and affectionate (F1)	0.448	0.654
	Punishing and Severe (F2)	-2.017	0.045*
	Overly-intervening (F3)	-1.452	0.147
	Favoring (F4)	-0.219	0.827
	Rejecting (F5)	1.001	0.317
	Overly-protecting (F6)	-1.119	0.264
Mothering parenting style	Warm and affectionate (M1)	-0.114	0.909
	Overly-intervening and Protecting (M2)	-1.995	0.047*
	Rejecting (M3)	-1.320	0.188
	Punishing and severe (M4)	-1.592	0.112
	Favoring (M5)	-0.422	0.673

Note: * $P<0.05$.

Table 3 Predictive Power of Models

	AUC	Accuracy	Recall	Sensitivity	Specificity	F1 Score	
Random Forest	0.780	0.760	0.650	0.650	0.815	0.642	P=0.397
XGBoost	0.807	0.719	0.850	0.850	0.654	0.667	

Abbreviation: AUC, area under the receiver operating characteristic curve.

Discussion

The present study found that using an XGBoost model based on a machine learning (ML) method has better predictive performance in predicting the risk factors of NSSI in adolescents with moderate or major depression disorder. Based on the established ML models, 11 risk factors were predicted for NSSI in adolescents, including childhood trauma, dysfunctional attitude, M2, depressive symptom, F2, and personality.

Non-suicidal self-injury in adolescents poses an increasing social challenge and effective strategies to prevent these behaviors have not been identified. Therefore, there is an urgent need to develop an effective and predictive screening tool for NSSI. This study was the first to use the ML method to investigate the predicted risk factors for NSSI in adolescents with MDD. Previous evidence has indicated that ML might be well-suited to prediction in suicide research,^{33–35} but its applicability in predicting risk factors of NSSI in adolescents was unknown. A previous study suggested that compared to traditional model-based statistical methods, an ML method has been shown to be more sensitive and precise in predicting suicidal behavior.³⁶ Consistent with a previous study, the prediction models based on the ML method (random forest model and XGBoost model) achieved the best AUC (0.780 and 0.807), accuracy (0.760 and 0.719), sensitivity (0.650 and 0.850), and specificity (0.815 and 0.654), which may help in reducing the occurrence of NSSI in adolescents.

The demographic results showed that the occurrence rate of NSSI in adolescents with MDD was 32.99%, which differed from other studies. This difference may be attributed to the adopted NSSI definition, sample heterogeneity, and the assessment tools used.³⁷ The current study found that the majority of with NSSI were females, which was partially consistent with a previous study, considering the high rate of female in participants.³⁸ However, this gender difference was more significant in clinical populations compared to studies conducted in the general population. According to the

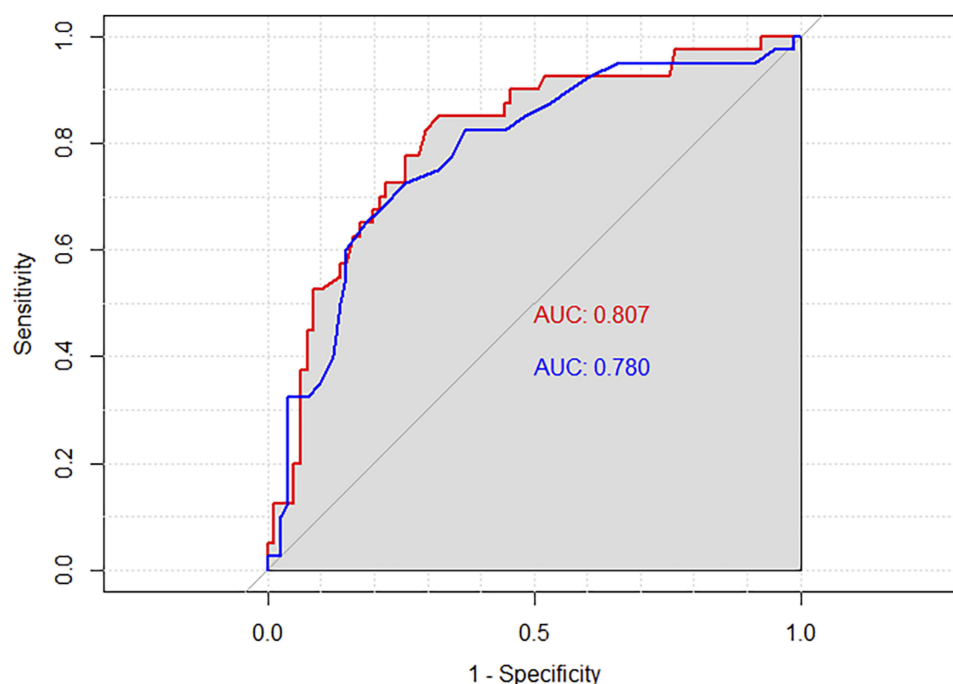


Figure 1 ROC curves of XGBoost (red) and Random Forest (blue).

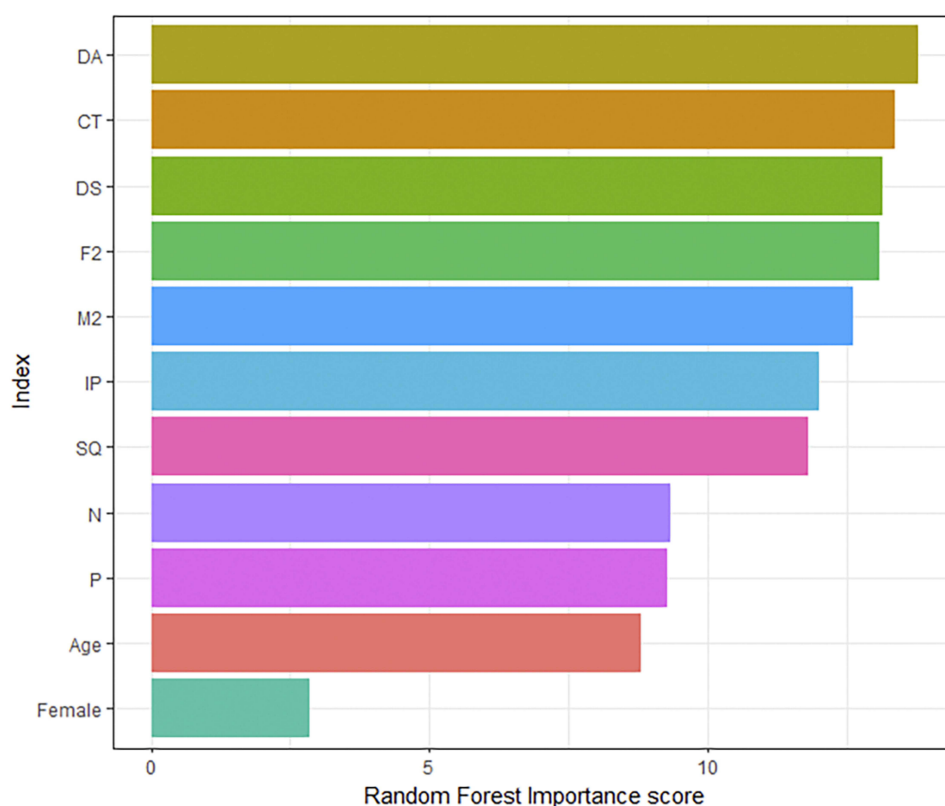


Figure 2 Random Forest variable importance ranking diagram.

Abbreviations: DA, Dysfunctional attitude; CT, Childhood trauma; DS, Depressive symptoms; IP, Interpersonal problem; SQ, Sleep quality.

study, cutting was the most common method among girls, while hitting against a wall was the most common method among boys.³⁹

In addition to the demographic factors, personal factors including concurrent psychiatric disease, sleep quality, personality, dysfunctional attitude, childhood trauma, and interpersonal problems, may influence the occurrence of NSSI in adolescents with MDD. Adolescents with psychiatric disease risk factors may have higher occurrences rates of NSSI.⁴⁰ Previous studies have indicated that adolescents with depression have rate of NSSI as high as a 50–60%,⁴¹ which was confirmed in the present study. This may be related to specific biological and sociological factors.⁴² The current results indicated that poor sleep quality was significantly associated with NSSI.⁴³ Another previous study showed poor sleep quality and nightmares were independent risk factors for NSSI, and depression mediated the relation between nightmares and NSSI,⁴⁴ which could help to understand the results in the current study.

A previous study indicated that personality was important for developing NSSI.⁴⁵ The current results showed that adolescents with psychoticism or neuroticism were more likely to engage in NSSI.⁴⁶ Supporting evidence suggests that individuals with high psychoticism may exhibit increased impulsivity, aggression, which could promote NSSI in individuals with depression.⁴⁷ The interpersonal function was significantly impaired in adolescents with NSSI, and another study has indicated that interpersonal conflict may be one of the factors promoting NSSI,⁴⁸ thus supporting the present results. A recent study indicated that dysfunctional attitude increased susceptibility to suicide ideation.⁴⁹ The current study showed that most individuals with NSSI had suicidal thoughts, which could indirectly explain the dysfunctional attitude in adolescents with NSSI. The direct mechanism by which dysfunctional attitude increased the NSSI susceptibility needs to be studied in future research.

Multiple studies have shown that the experiences of childhood trauma, such as abuse, parental neglect, or deprivation, seems to increase the vulnerability to NSSI.^{50,51} Recent evidence on the effect of childhood trauma has shown some differences. A meta-analysis review indicated that sexual abuse was moderately related to the occurrence of NSSI, which aligned with another study.^{13,52,53} Other evidence suggested that childhood maltreatment (witnessing domestic violence)

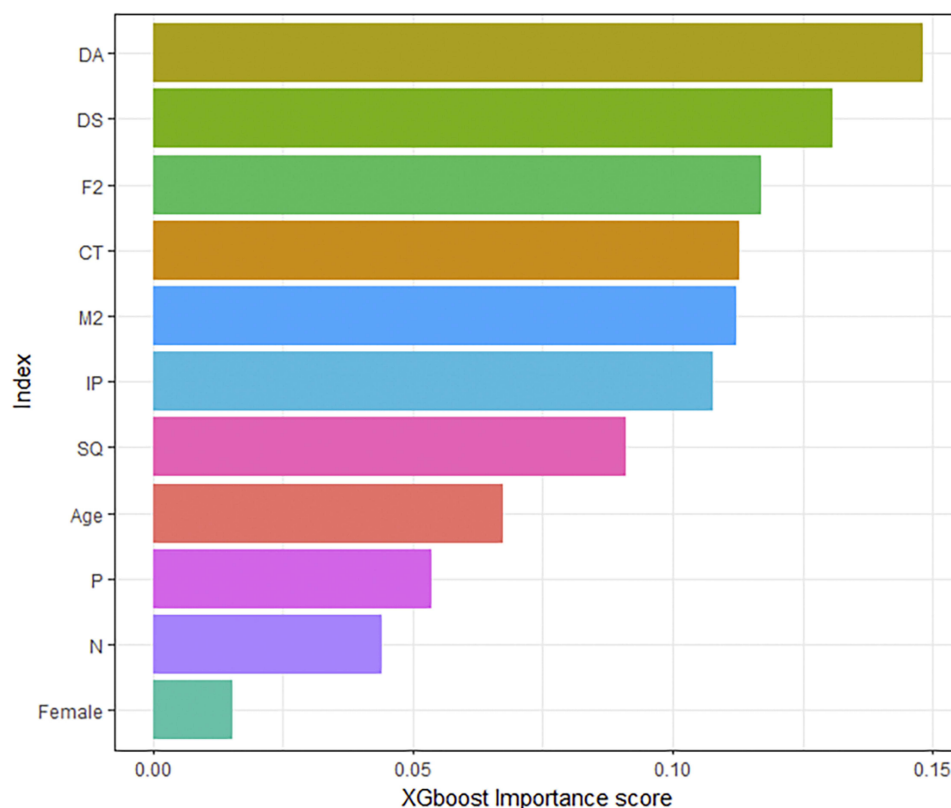


Figure 3 XGBoost variable importance ranking diagram.

Abbreviations: DA, Dysfunctional attitude; CT, Childhood trauma; DS, Depressive symptoms; IP, Interpersonal problem; SQ, Sleep quality.

was associated with NSSI, but emotional abuse was not assessed in this study.⁴⁰ However, the present study did not analyze the specific experience of childhood trauma, which should be included in future research.

Additionally, parenting styles were also important for adolescents to develop NSSI. The present study found that punishing and severe fathering parenting styles and overly-intervening and protecting mothering parenting styles were significantly associated with NSSI behaviors in adolescents with MDD. The results from a recent study showed NSSI was positively associated with a high level of mothers' over-protection, but the study did not assess punishing and severe fathering parenting style.⁵⁴ Attachment theory suggests that mothers and fathers play different roles in shaping the social, cognitive, and emotional development in adolescents.³⁸ The different associations observed with fathers and mothers may result from the primary companion during children's growth stages and the greater involvement of mothers in the daily lives of adolescents compared to fathers.⁵⁴ Mothers' over-protection is related to childrens' unhealthy self-conscious emotions, which are associated with NSSI.⁵⁵ This highlights the need for specific interventions in the clinical setting to assess the parenting style and determine what is needed for future adolescents with MDD.

The clinical implication of practice is that the construction of risk predicted model based on ML method and the top five most important risk predictors identified were dysfunctional attitude, childhood trauma, depressive symptoms, punishing and Severe (F2) and Overly-intervening and Protecting (M2) in adolescents with MDD. This may explain that the high rate of female in population with MDD, which suggests we should pay more attention to female patients in MDD to some extent. We could treat the adolescents with MDD with psychotherapy treatment from dysfunctional attitude, childhood trauma, punishing and Severe (F2) and Overly-intervening and Protecting (M2) to reduce the rate of NSSI.

Limitations

Several limitations in this study should be considered. Firstly, this is a cross-sectional study, which may have recall bias and prevalence-incidence bias, especially in the case of longer-lasting diseases, any risk factors that results in death will be under-represented among those with the disease. Additionally, it is difficult to make causal inference. Therefore, longitudinal studies are

needed to determine the cause-consequence relationship between risk factors identified by the ML method and NSSI. Techniques to reduce the level of nonresponse include telephone and mail prompting, second and third mailing of surveys, letters outlining the importance of replying and a range of incentives would be carried out in longitudinal study. Reducing the amount of information collected and maintain optimum response levels may decrease the loss to follow-up. Cross-sectional study may not reflect the fluctuation of the NSSI over time and the temporal relationship between predictors and the NSSI. Therefore, longitudinal studies are needed to determine the fluctuation of the NSSI over time and how changes in risk factors over time affect the NSSI. In future studies, risk prediction models using NSSI future trends as an outcome variable will be constructed in longitudinal data. Secondly, the frequency and intensity of NSSI were not considered in the present study, and it was not clear whether the independent predictors were related to the severity of NSSI. Thirdly, the female rate of the participants was significantly higher than the male rate of that and the results may be affected by the sample imbalance, therefore, the sexual rate of participants will be seriously considered. Fourthly, this study was conducted in Chinese adolescents with MDD which is not representative of the entire population. In addition, the validation approach of the current study is internal validation and not involving external validation. The findings of the current study will be validated in a larger sample in the future, which can increase the applicability of the findings to other cultures or population groups and the generalization of the model. Finally, no interventions on MDD were conducted in this study, and they will be included in future studies.

Conclusion

This study establishes a risk prediction model for NSSI in adolescents with MDD using the ML method. This prediction model appears to have better AUC, accuracy, sensitivity and specificity and provides relatively accurate risk factors for NSSI. This may help to develop individualized prevention and intervention measures for NSSI in adolescents with MDD. The application of a risk prediction model for NSSI in adolescents with MDD could support Healthcare professionals, mental health center personnel, and hospitals' decision-making on early psychological interventions and proper mental health management. Future research focuses on the external validation of predictive models. The model needs to be validated repeatedly in larger samples and continuous revisions to the model before it can be translated into clinical practice.

Acknowledgments

We sincerely thank Dr. Sharon R. Redding (EdD, RN, CNE) for her help in editing.

Funding

This work was supported by grants from the National Key R&D Program of China (grant numbers: 2018YFC1314600), the Key opening projects of Hubei Province (grant numbers: 2021KFH012, 2022KFH001), the Research project of Wuhan Nursing Association (grant numbers: WHHL202202), the Clinical Nursing Special Research and Cultivation Fund Project of Wuhan University (grant numbers: LCHL202301), and the Interdisciplinary Innovative Talents Foundation from Renmin Hospital of Wuhan University (grant numbers: JCRCGW-2022-010).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Buelens T, Luyckx K, Kiekens G, et al. Investigating the dsm-5 criteria for non-suicidal self-injury disorder in a community sample of adolescents. *J Affect Disord.* 2020;260:314–322. doi:10.1016/j.jad.2019.09.009
2. Bresin K, Gordon KH. Endogenous opioids and nonsuicidal self-injury: a mechanism of affect regulation. *Neurosci Biobehav Rev.* 2013;37(3):374–383. doi:10.1016/j.neubiorev.2013.01.020
3. Auerbach RP, Kim JC, Chango JM, et al. Adolescent nonsuicidal self-injury: examining the role of child abuse, comorbidity, and disinhibition. *Psychiatry Res.* 2014;220(1–2):579–584. doi:10.1016/j.psychres.2014.07.027
4. Avenevoli S, Swendsen J, He JP, Burstein M, Merikangas KR. Major depression in the national comorbidity survey–adolescent supplement: prevalence, correlates, and treatment. *J Am Acad Child Adolesc Psychiatry.* 2015;54(1):37–44. doi:10.1016/j.jaac.2014.10.010
5. Swannell SV, Martin GE, Page A, Hasking P, St John NJ. Prevalence of nonsuicidal self-injury in nonclinical samples: systematic review, meta-analysis and meta-regression. *Suicide Life Threaten Behav.* 2014;44(3):273–303. doi:10.1111/sltb.12070

6. Weintraub MJ, Van de Loo MM, Gitlin MJ, Miklowitz DJ. Self-harm, affective traits, and psychosocial functioning in adults with depressive and bipolar disorders. *J Nerv Ment Dis.* 2017;205(11):896–899. doi:10.1097/NMD.0000000000000744
7. Lüdtke J, In-Albon T, Schmeck K, et al. Nonsuicidal self-injury in adolescents placed in youth welfare and juvenile justice group homes: associations with mental disorders and suicidality. *J Abnorm Child Psychol.* 2018;46(2):343–354. doi:10.1007/s10802-017-0291-8
8. Wolff J, Frazier EA, Esposito-Smythers C, et al. Cognitive and social factors associated with nssi and suicide attempts in psychiatrically hospitalized adolescents. *J Abnorm Child Psychol.* 2013;41(6):1005–1013. doi:10.1007/s10802-013-9743-y
9. Castellví P, Lucas-Romero E, Miranda-Mendizábal A, et al. Longitudinal association between self-injurious thoughts and behaviors and suicidal behavior in adolescents and young adults: a systematic review with meta-analysis. *J Affect Disord.* 2017;215:37–48. doi:10.1016/j.jad.2017.03.035
10. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science.* 1977;196(4286):129–136. doi:10.1126/science.847460
11. Brunner R, Kaess M, Parzer P, et al. Life-time prevalence and psychosocial correlates of adolescent direct self-injurious behavior: a comparative study of findings in 11 European countries. *J Child Psychol Psychiatry.* 2014;55(4):337–348. doi:10.1111/jcpp.12166
12. Vega D, Torrubia R, Soto Á, et al. Exploring the relationship between non suicidal self-injury and borderline personality traits in young adults. *Psychiatry Res.* 2017;256:403–411. doi:10.1016/j.psychres.2017.07.008
13. Thomassin K, Shaffer A, Madden A, Londino DL. Specificity of childhood maltreatment and emotion deficit in nonsuicidal self-injury in an inpatient sample of youth. *Psychiatry Res.* 2016;244:103–108. doi:10.1016/j.psychres.2016.07.050
14. Shao C, Wang X, Ma Q, Zhao Y, Yun X. Analysis of risk factors of non-suicidal self-harm behavior in adolescents with depression. *Ann Palliat Med.* 2021;10(9):9607–9613. doi:10.21037/apm-21-1951
15. Khazaie H, Zakiei A, McCall WV, et al. Relationship between sleep problems and self-injury: a systematic review. *Behav. Sleep Med.* 2021;19(5):689–704. doi:10.1080/15402002.2020.1822360
16. Yu HJ, Liu X, Yang HG, Chen R, He QQ. The association of adverse childhood experiences and its subtypes with adulthood sleep problems: a systematic review and meta-analysis of cohort studies. *Sleep Med.* 2022;98:26–33. doi:10.1016/j.sleep.2022.06.006
17. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins.* 2006;62(4):1125–1132. doi:10.1002/prot.20810
18. Han B, Chen H, Yao Y, et al. Genetic and non-genetic factors associated with the phenotype of exceptional longevity & normal cognition. *Sci Rep.* 2020;10(1). doi:10.1038/s41598-020-75446-2
19. Little MP, Rosenberg PS, Arsham A. Alternative stopping rules to limit tree expansion for random forest models. *Sci Rep.* 2022;12(1). doi:10.1038/s41598-022-19281-7
20. Hou N, Li M, He L, et al. Predicting 30-days mortality for mimic-iii patients with sepsis-3: a machine learning approach using xgboost. *J Transl Med.* 2020;18(1):462. doi:10.1186/s12967-020-02620-5
21. Huang X, Ribeiro JD, Franklin JC. The differences between individuals engaging in nonsuicidal self-injury and suicide attempt are complex (vs. Complicated or simple). *Front Psychiatry.* 2020;11. doi:10.3389/fpsy.2020.00239
22. Kang L, Li R, Liu H, et al. Nonsuicidal self-injury in undergraduate students with major depressive disorder: the role of psychosocial factors. *J Affect Disord.* 2021;290:102–108. doi:10.1016/j.jad.2021.04.083
23. Xie X, Li Y, Liu J, et al. The relationship between childhood maltreatment and non-suicidal self-injury in adolescents with depressive disorders. *Psychiatry Res.* 2024;331:115638. doi:10.1016/j.psychres.2023.115638
24. Arbanas G. Diagnosing the diagnostic and statistical manual of mental disorders. *J Am Acad Child Adolesc Psychiatry.* 2015;54(1):37–44.
25. Zheng Y, Zhao J, Phillips M, et al. Validity and reliability of the Chinese Hamilton depression rating scale. *Br J Psychiatry.* 1988;152(5):660–664. doi:10.1192/bjp.152.5.660
26. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Personal Individ Differ.* 1985;1(6):21–29.
27. Chen L, Xu J, Yan S, et al. Preliminary study on reliability and validity of dysfunctional attitude scale. *Chin Mental Health J.* 1998;05:10–12.
28. Zhao X, Zhang Y, Li L, Zhou Y, Li H. Reliability and validity of the Chinese version of childhood trauma questionnaire. *Chin J Clin Rehabil.* 2005;20:105–107.
29. Zheng R. *Psychological Diagnosis of College Students*. Shandong Education Press; 1999.
30. Perris C, Jacobsson L, Linnström H, von Knorring L, Perris H. Development of a new inventory for assessing memories of parental rearing behaviour. *Acta Psychiatrica Scandinavica.* 1980;61(4):265–274. doi:10.1111/j.1600-0447.1980.tb00581.x
31. Kaufman J, Birmaher B, Brent D, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (k-sads-pl): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry.* 1997;36(7):980–988. doi:10.1097/00004583-199707000-00021
32. Xu Z, Huang Y, Zhu G, Wang X, Wang F. Prevalence of non-suicidal self-injury behavior among Chinese adolescents with depression: a meta-analysis. *Sichuan Mental Health.* 2022;35(4):331–336.
33. Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry.* 2018;59(12):1261–1270. doi:10.1111/jcpp.12916
34. McHugh CM, Large MM. Can machine-learning methods really help predict suicide? *Curr Opin Psychiatry.* 2020;33(4):369–374. doi:10.1097/YCO.0000000000000609
35. Jiang T, Rosellini AJ, Horváth-Puhó E, et al. Using machine learning to predict suicide in the 30 days after discharge from psychiatric hospital in Denmark. *Br J Psychiatry.* 2021;219(2):440–447. doi:10.1192/bjp.2021.19
36. Grendas LN, Chiapella L, Rodante DE, Daray FM. Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour. *J Psychiatr Res.* 2022;145:85–91. doi:10.1016/j.jpsychires.2021.11.029
37. Costa RPDO, Peixoto ALRP, Lucas CCA, et al. Profile of non-suicidal self-injury in adolescents: interface with impulsiveness and loneliness. *Jornal De Pediatria.* 2021;97(2):184–190. doi:10.1016/j.jped.2020.01.006
38. Bresin K, Schoenleber M. Gender differences in the prevalence of nonsuicidal self-injury: a meta-analysis. *Clinic Psychol Rev.* 2015;38:55–64. doi:10.1016/j.cpr.2015.02.009
39. Barrocas AL, Hankin BL, Young JF, Abela JRZ. Rates of nonsuicidal self-injury in youth: age, sex, and behavioral methods in a community sample. *Pediatrics.* 2012;130(1):39.
40. Brown RC, Plener PL. Non-suicidal self-injury in adolescence. *Curr Psychiatry Rep.* 2017;19(3). doi:10.1007/s11920-017-0767-9
41. Kaess M, Parzer P, Mattern M, et al. Adverse childhood experiences and their impact on frequency, severity, and the individual function of nonsuicidal self-injury in youth. *Psychiatry Res.* 2013;206(2–3):265–272. doi:10.1016/j.psychres.2012.10.012

42. Cyranowski JM, Frank E, Young E, Shear MK. Adolescent onset of the gender difference in lifetime rates of major depression: a theoretical model. *Arch Gen Psychiatry*. 2000;57(1):21–27. doi:10.1001/archpsyc.57.1.21
43. Liu X, Chen H, Bo Q, Fan F, Jia C. Poor sleep quality and nightmares are associated with non-suicidal self-injury in adolescents. *Eur Child Adolesc Psychiatry*. 2017;26(3):271–279. doi:10.1007/s00787-016-0885-7
44. Liu Z, Tein J, Jia C, Liu X. Depression as a mediator between frequent nightmares and non-suicidal self-injury among adolescents: a 3-wave longitudinal model. *Sleep Med*. 2021;77:29–34. doi:10.1016/j.sleep.2020.11.015
45. Gromatsky MA, He S, Perlman G, et al. Prospective prediction of first onset of nonsuicidal self-injury in adolescent girls. *J Am Acad Child Adolesc Psychiatry*. 2020;59(9):1049–1057. doi:10.1016/j.jaac.2019.08.006
46. Burke TA, Anne McArthur B, Daryanani I, Abramson LY, Alloy LB. Latent classes of trait affect and cognitive affective regulation strategies are associated with depression, non-suicidal self-injury, and well-being. *J Affect Disord*. 2018;225:180–187. doi:10.1016/j.jad.2017.08.015
47. Tobore TO. On the neurobiological role of oxidative stress in alcohol-induced impulsive, aggressive and suicidal behavior. *Subst Use Misuse*. 2019;54(14):2290–2303. doi:10.1080/10826084.2019.1645179
48. Guerry JD, Prinstein MJ. Longitudinal prediction of adolescent nonsuicidal self-injury: examination of a cognitive vulnerability-stress model. *J Clin Child Adolesc Psychol*. 2009;39(1):77–89. doi:10.1080/15374410903401195
49. Rohani FF, Esmaeili MM. Psychological factors of vulnerability to suicide ideation: attachment styles, coping strategies, and dysfunctional attitudes. *J Educ Health Promot*. 2020;9(1):50. doi:10.4103/jehp.jehp_260_19
50. Yates TM, Carlson EA, Egeland B. A prospective study of child maltreatment and self-injurious behavior in a community sample. *Dev psychopathol*. 2008;20(2):651–671. doi:10.1017/S0954579408000321
51. Muehlenkamp JJ, Kerr PL, Bradley AR, Adams Larsen M. Abuse subtypes and nonsuicidal self-injury: preliminary evidence of complex emotion regulation patterns. *J Nerv Ment Dis*. 2010;198(4):258–263. doi:10.1097/NMD.0b013e3181d612ab
52. Klonsky ED, Moyer A. Childhood sexual abuse and non-suicidal self-injury: meta-analysis. *Br J Psychiatry*. 2008;192(3):166–170. doi:10.1192/bjp.bp.106.030650
53. Maniglio R. The role of child sexual abuse in the etiology of suicide and non-suicidal self-injury. *Acta Psychiatrica Scandinavica*. 2011;124(1):30–41. doi:10.1111/j.1600-0447.2010.01612.x
54. Liu Y, Xiao Y, Ran H, et al. Association between parenting and non-suicidal self-injury among adolescents in Yunnan, China: a cross-sectional survey. *PeerJ*. 2020;8:e10493. doi:10.7717/peerj.10493
55. Spitz TL, Tull MT, Baer MM, et al. Predicting engagement in nonsuicidal self-injury (nssi) over the course of 12 months: the roles of borderline personality disorder pathology and emotional consequences of nssi. *J Affect Disord*. 2020;277:631–639. doi:10.1016/j.jad.2020.08.049

Neuropsychiatric Disease and Treatment

Dovepress

Publish your work in this journal

Neuropsychiatric Disease and Treatment is an international, peer-reviewed journal of clinical therapeutics and pharmacology focusing on concise rapid reporting of clinical or pre-clinical studies on a range of neuropsychiatric and neurological disorders. This journal is indexed on PubMed Central, the 'PsycINFO' database and CAS, and is the official journal of The International Neuropsychiatric Association (INA). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/neuropsychiatric-disease-and-treatment-journal>