

Predicting suicidal behavior in individuals with depression over 50 years of age: Evidence from the UK biobank

DIGITAL HEALTH
Volume 10: 1–13
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241287450
journals.sagepub.com/home/dhj



Jian Zhang^{1,2}, Yujun Liu^{2,3} , Chao Zhang^{2,3}, Yilong Chen^{2,3}, Yao Hu^{2,3},
Xiujia Yang⁴, Wentao Liu^{2,3}, Wei Zhang^{1,2,3}, Di Liu^{2,3,5} and Huan Song^{2,3,6} 

Abstract

Objective: To construct applicable models suitable for predicting the risk of suicidal behavior among individuals with depression, particularly on the progression from no history of suicidal behavior to suicide attempts, as well as from suicidal ideation to suicide attempts.

Methods: Based on a prospective cohort from the UK Biobank, a total of 55,139 individuals aged 50 and above with depression were enrolled in the study, among whom 29,528 exhibited suicidal behavior. Specifically, they were divided into control (25,611), suicidal ideation (24,361), and suicide attempt (5167) groups. Least absolute shrinkage and selection operator (LASSO) regression was used to identify a subset of important features for distinguishing suicidal ideation and suicide attempts. We used the Gradient Boosting Decision Tree (GBDT) algorithm with stratified 10-fold cross-validation and grid-search to construct the prediction models for suicidal ideation or suicide attempts. To address the dataset imbalance in classifying suicide attempts, we used random under-sampling. The SHapley Additive exPlanations (SHAP) were used to estimate the important variables in the GBDT model.

Results: Significant differences in sociodemographic, economic, lifestyle, and psychological factors were observed across the three groups. Each classifier optimally utilized 8–11 features. Overall, the algorithms predicting suicide attempts demonstrated slightly higher performance than those predicting suicidal ideation. The GBDT classifier achieved the highest accuracy, with AUROC scores of 0.914 for suicide attempts and 0.803 for suicidal ideation. Distinctive predictive factors were identified for each group: while depression's inherent characteristics crucially distinguished the suicidal ideation group from controls, some key predictors, including the age of depression onset and childhood trauma events, were identified for suicide attempts.

Conclusions: We established applicable machine learning-based models for predicting suicidal behavior, particularly suicide attempts, in individuals with depression, and clarified the differences in predictors between suicidal ideation and suicide attempts.

Keywords

Depression, suicide, prevention, risk factors, machine learning

Submission date: 4 March 2024; Acceptance date: 10 September 2024

¹Mental Health Center, West China Hospital, Sichuan University, Chengdu, China

²West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China

³Med-X Center for Informatics, Sichuan University, Chengdu, China

⁴University of Illinois at Urbana and Champaign, Urbana, IL, USA

⁵Industrial Engineering, Pittsburgh Institute, Sichuan University, Chengdu, China

⁶Center of Public Health Sciences, Faculty of Medicine, University of Iceland, Reykjavík, Iceland

Corresponding author:

Huan Song, West China Biomedical Big Data Center, West China Hospital, Sichuan University, Guo Xue Lane 37, Chengdu, China.
Email: songhuan@wchscu.cn

Di Liu, Industrial Engineering, Pittsburgh Institute, Sichuan University, Chengdu, China.
Email: di.liu@scupi.cn



Introduction

Suicide is a critical public health issue all over the world because one person dies by suicide every 40 s, leading to substantial personal, family, and economic burdens.¹ Suicides can occur at almost any age across the world, with 79% occurring in low- and middle-income countries, while the highest suicide rates are in high-income countries.¹ Despite various public-health efforts since the World Health Organization's (WHO) first global report on suicide in 2014, limited progress has been made in suicide prevention due to the lack of reliable methods for predicting suicidality and a poor understanding of suicide's biological etiology.² Current methods of predicting suicide still mainly rely on self-report measures, such as questionnaires and interviews, which can be too subjective. Therefore, developing effective models for predicting suicidality is crucial to preventing suicide, saving lives, and mitigating the societal risks and negative impacts of suicide.

Suicide is thought to be a complicated dimensional trait that encompasses a broad range of experiences and behaviors, ranging from suicidal ideation to acts of deliberate self-harm and attempted suicide, which occur along a spectrum toward committing suicide.³ Though there are various risk factors for suicide (e.g., sociodemographic, biological, and psychological factors), studies have found that approximately 90% of individuals who die by suicide were diagnosed with a mental disorder prior to their death,^{2,4} and more than half of the persons who attempt suicide meet the criteria of Major Depressive Disorder (MDD).⁵ Thus, depression and suicide, as independent mental health problems, have received increasing attention among researchers in recent years.

Given the complexity of suicide, recent research has focused on the risk for different components of suicide (e.g., suicidal ideation, non-suicidal-injury, and attempted suicide) and different population groups,^{6–9} using traditional statistical approaches. However, traditional approaches greatly reduce the number of predictors and interactions that can be examined simultaneously, and may mistake some more complex associations for linearity, thereby weakening the predictive power of findings. Machine Learning (ML) has been introduced into suicide prediction research to compensate for the deficiency of traditional statistical methods, by identifying novel risk factors and interactions, and improving prediction accuracy, with AUCs (Area Under the Curve) ranging from 0.71 to 0.89.^{10–13} However, research on suicide prediction, including limited research on depression and suicide, has yielded inconsistent and disparate findings, possibly due to relatively small sample sizes and the lack of external validation of models.

The present study was designed to examine suicide behavior in the general population using the UK Biobank, a very large and population-based prospective study, to confirm that depressive disorder is an important predictor of suicide. Moreover, we explored the

associations among different stages of suicide in a depressive population and developed prognostic prediction models for suicidal behavior, with a particular focus on suicide attempts.

Methods

Study population

The UK Biobank is a well-characterized cohort study of over 500,000 adults, 40–69 years of age, at baseline (2006–2010), which includes their extensive phenotypic and genotypic details, based on questionnaires and physical measures. Importantly, it also contains longitudinal follow-up measures of health-related outcomes.^{14,15} Further information about the UK Biobank is available elsewhere (<https://www.ukbiobank.ac.uk/>). We derived the mental health outcomes from the UK Biobank Mental Health Questionnaire (MHQ), using established criteria to identify likely disorders and risk profiles, including lifetime depression, manic or hypomania, generalized anxiety disorder, unusual experiences, self-harm, current post-traumatic stress, and hazardous or harmful alcohol consumption, as described elsewhere.^{15–17} More information on the mental health variables derived in the Supplementary File 1. The present study included a subset of 157,366 individuals (46.4% of the original UK Biobank sample) who completed the online MHQ by August 2017.

Individuals with MDD were identified using the following criteria: (i) lifetime MDD based on the Composite International Diagnostic Interview (CIDI) Short Form; (ii) ICD-coded MDD based on linked hospital admission records; and (iii) self-reported MDD as part of past and current medical conditions. Among the 56,820 individuals who met the criteria for at least one of the three definitions, we excluded 18 individuals who withdrew their informed consent, as well as 1662 individuals who preferred not to answer questions about self-harm, and 249 individuals who had more than 30% missing answers on the MHQ. The UK Biobank has received full ethical approval from the NHS National Research Ethics Service (16/NW/0274), and all the participants provided written informed consent before data collection. The current study was approved by the biomedical research ethics committee of West China Hospital (2020.661). The flow chart of participant selection for the present study is illustrated in Figure 1. We have introduced and followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guideline,^{18,19} with a detailed checklist provided in Supplementary File 2-Table 1.

Measures of suicide phenotypes

The suicide phenotypes in this study were based on the online MHQ administered in 2016/2017.²⁰ Suicidal ideation was measured by two questions from the MHQ-Field 20479

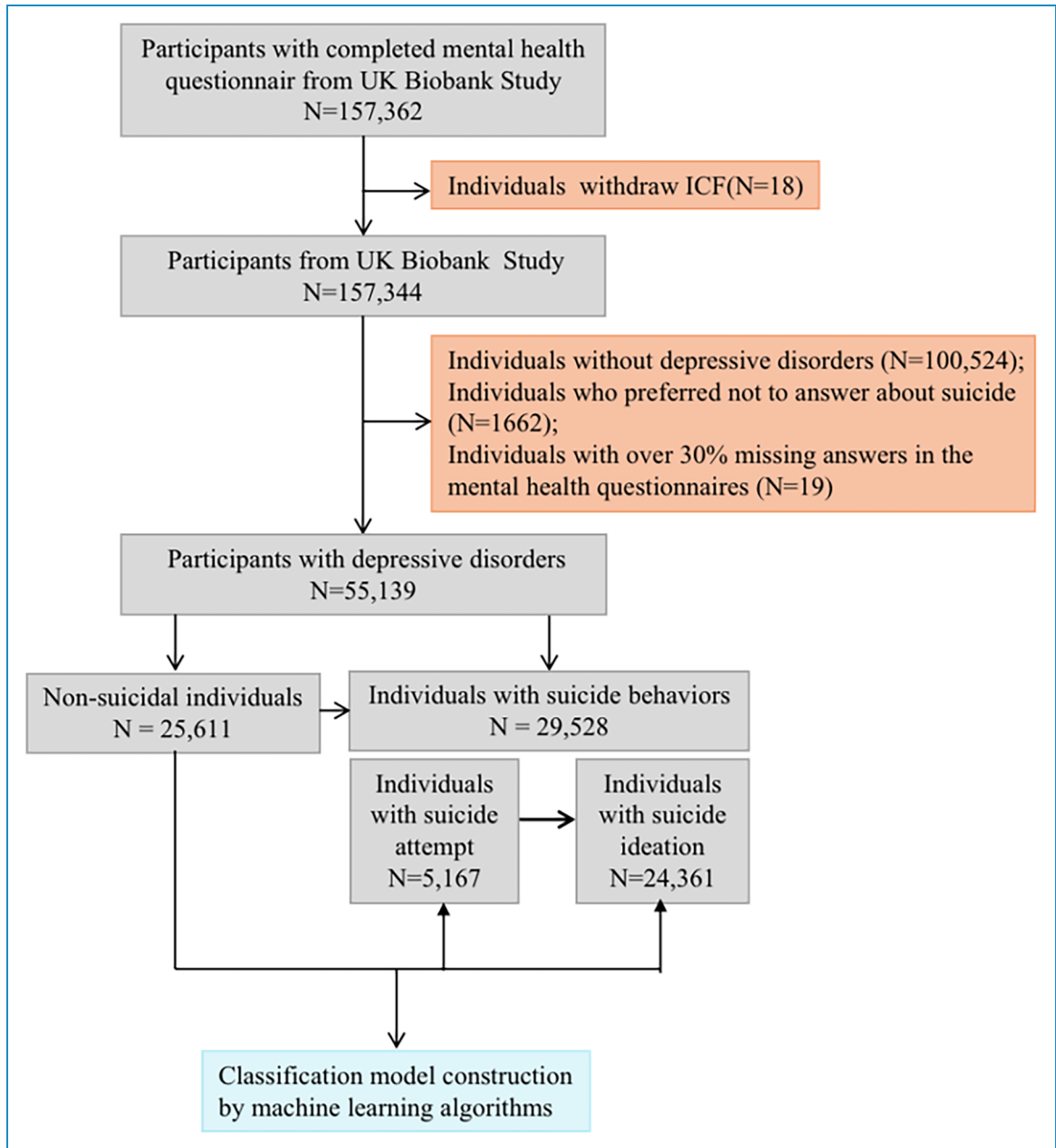


Figure 1. Flow chart of the study population selection.

(Ever thought that life is not worth living) and Field 20513 (Thoughts that you would be better off dead or of hurting yourself in some way). We re-coded these into a binary variable in which those who responded 'No' or 'Not at all' to either question were coded as 0 and everyone else was coded as 1. Similarly, suicide attempts were measured using three questions: Field 20480 (Ever deliberately harmed yourself), Field 20481 (Ever harmed yourself in the last 12 months), and Field 20483 (Ever attempted suicide). We used these measures to derive a binary measure of suicide attempts in which individuals were given a score of 0 if they responded negatively to any of

the three questions, and a score of 1 if they responded affirmatively to all three questions. To be specific, non-fatal suicide attempts and intentional self-harm were also defined as suicide attempts in this study according to any of the following ICD-10 codes X71–X83, T36–T71, T14. Accordingly, individuals with MDD ($n = 55,139$) were then classified into three subgroups based on a composite variable, with categories ordered by increasing suicide risk (Figure 1): non-suicidal depressed control participants ($n = 25,611$), depressed participants with suicidal ideation ($n = 24,361$), and depressed participants with a history of suicide attempts ($n = 5167$).

Feature engineering

Data pre-processing

Based on suggestions from psychiatrists and valuable current documents,^{21–23} we manually inspected and classified 403 candidate variables that were assessed at the time of study enrollment across the domains of social-demographics, lifestyle, family history, physical measures, and longitudinal follow-up of health-related outcomes. The term feature engineering encompasses feature filtering and transformations. The term feature filter refers to eliminating irrelevant features that act as noise to improve prediction accuracy. First, we eliminated 123 variables with more than 20% missing values. For the rest 280 variables for further processing (Supplementary File 2-Table 2), we added an “unknown” category to multi-category variables if some observed values of them were missing. Subsequently, the missing values of categorical variables were assigned the mode, whereas continuous variables were assigned the mean. Categorical variables with n categories were split into n binary variables, namely “one-hot coding.” All continuous variables were normalized; that is, the mean of the scale was set to 0 and the variance was set to 1.

Feature selection

Feature selection has been used to reduce the complexity of predictive models to facilitate better visualization of data and improve the performance of prediction.²⁰ Based on the above 280 variables, we performed Least Absolute Shrinkage and Selection Operator (LASSO) regression to select a minimum subset of features that accurately classified suicidal behaviors. LASSO is a penalized logistic regression model that selects important predictors by shrinking the coefficients of less important predictors to zero.²⁴ LASSO uses an embedded feature selection framework through L1 regularization.²⁵ L1 was set to 0.15 in this study, and four classification models (see Supplementary File 2-Figure 1) were tested to obtain better performance. We calculated variable importance by the principle of “mean decreases in accuracy”¹² to see which variables contributed significantly to optimization of the predictive model. Therefore, the important features that were selected using LASSO were used to build the predictive models (Supplementary File 2-Figure 1).

Statistical analyses

Model building

Imbalance handling and ML algorithms

Four binary classification tasks were delineated, encompassing the following distinctions: individuals without a history of suicidal behavior versus those with suicidal ideation (25,611 vs. 24,361), individuals without suicidal behavior versus those who have made suicide attempts

(25,611 vs. 5167), individuals with suicidal ideation versus those who have made suicide attempts (24,361 vs. 5167), and individuals without suicidal behavior versus those exhibiting any suicidal behavior (25,611 vs. 29,449). Given the extreme imbalance in the datasets for the classifications involving suicide attempts (25,611 vs. 5167 and 24,361 vs. 5167), we adopted the random under-sampling method to address this issue.²⁶

We used the Gradient Boosting Decision Tree (GBDT) algorithm in this study, which is one of the best algorithms for fitting real distributions in traditional ML algorithms. GBDT models inherently possess greater complexity compared to simpler models, primarily due to their iterative nature. Specifically, GBDT is an iterative decision tree algorithm that consists of multiple decision trees.²⁷ This iterative process involves sequentially adding trees to correct the errors of previous iterations, thereby enhancing the model’s predictive performance. GBDT generates a weak classifier through multiple rounds of iteration, in which each classifier is trained on the basis of the residual of the previous round of classifiers, and the final classifier is improved by continuously reducing the residual generated during the training process. Despite this complexity, GBDT models offer significant interpretability through feature importance scores.²⁸ These scores elucidate the relative contribution of each feature to the model’s predictions, facilitating a deeper understanding of which features are most influential in driving the predictive outcomes. In addition, five other ML algorithms were used to compare model performance: K-Nearest Neighbors, Gaussian Naive Bayes (Gaussian NB), logistic regression, Random Forest (RF), and Support Vector Machine (SVM). These six algorithms have been broadly accepted to solve binary classification problems between suicide attempts and non-suicidal attempts due to their performance and ease of interpretation, implementation, accuracy, and robustness.^{29,30}

Repeated cross-validation and parameter tuning

To avoid reporting biased results and limit over-fitting, ML algorithms are usually validated through cross-validation.³¹ K-fold cross-validation is a widely used method to evaluate ML models. After dividing the original data into K subsets (K-fold), the process involves training and evaluating the model K times, with a different fold of the data serving as the validation set each time, and the remaining K-1 folds being used for training the model.³² To ensure the stability and generalization ability of our models, we adopted a 10-fold cross-validation ($K = 10$) approach. Specifically, we divided the dataset into training and validation sets in an 8:2 ratio and performed 10-fold cross-validation on the training set. This approach effectively reduces the model’s dependence on a single dataset division and provides a more reliable assessment of model performance.

We optimized the model's hyperparameters using a grid search strategy. During the grid search process, we searched for optimal parameter combinations across a wide range of hyperparameters, including the number of trees, learning rate, and maximum depth. Our experimental results indicate that GBDT models achieved optimal performance with default parameter settings. Consequently, we used these simplified default settings for all GBDT classification models, which not only streamlined the training process but also ensured high efficiency. To enhance the robustness of GBDT models against overfitting, we implemented several practical and effective measures. We used regularization techniques (both L1 and L2) to reduce model complexity and employed early stopping to halt training when the validation error ceased to improve, thereby preventing overfitting. Additionally, we limited the maximum depth of each tree to prevent them from becoming overly complex. Together, these measures ensure the stability and generalization ability of the GBDT model during training and validation.

Explanation by SHAP

The correct interpretation of a prediction model for ML is also a challenge. We used SHapley Additive exPlanation (SHAP) values to provide consistent and locally accurate attribution values for each feature within each prediction model.³³ This is a unified approach for explaining the outcome of any ML model. SHAP values evaluate the importance of the output resulting from the inclusion of feature A for all combinations of features other than A. According to the prediction model, the higher the SHAP value of a feature, the more likely suicidal behaviors will occur. We visualized how a feature's attributed importance changed as its values varied in the plot.

Performance measurement and comparison

The classification fitted models were used to predict the classes (non-suicidal depressed control individuals, and individuals with suicidal behaviors) in the dataset, and the predicted classes were compared with the actual classes. Accuracy, specificity, sensitivity, and F1 scores were calculated along with the Area Under the Curve (AUC) derived from Receiver Operating Characteristics (ROC) curves as performance evaluation indicators. Sensitivity and specificity were measured at different binary-classification thresholds that the authors deemed relevant for clinical practice.

Statistical divergences in AUC between groups were assessed using the Hanley and McNeil method. The cut-off point was determined by maximizing the Youden function³⁴ ($J[c] = \text{Sensitivity}[c] + \text{Specificity}[c] - 1$), where c represents any given cut-off to differentiate individuals with suicidal behaviors from individuals with depressive disorders. All tests of significance were two-sided, with

$p < 0.05$ indicating statistical significance. Data analyses were performed using R (version 3.6.2, R Foundation for Statistical Computing, Vienna, Austria) and Python 3.7 (Python Software Foundation Delaware, USA). Relevant Python libraries included the SciPy ecosystem, NumPy, Pandas, and Scikit-learn.

Results

Characteristics of the participants

The final study population consisted of 55,139 participants with MDD, age > 50 years, among whom 29,528 (52.06%) individuals had ever exhibited suicidal behaviors. The class distribution is shown in Figure 1, with 25,611 individuals without suicidal behaviors, 24,361 with suicidal ideation, and 5167 with suicide attempts.

Table 1 illustrates the differences in the variables, such as sociodemographic and economic characteristics, and lifestyle in participants with and without suicidal behaviors. The mean age was 54.6 and 52.3 years for individuals with suicidal ideation and suicide attempts, respectively. Compared with the control group, individuals with suicidal ideation were more likely to be female (72.21% vs. 65.59%), obese (BMI ≥ 29.9 , 26.42% vs. 21.24%), cigarette smokers (current or former, 56.82% vs. 43.63%), former drinkers (7.32% vs. 3.00%), and have higher neuroticism scores (mean score, 6.34 vs. 4.79) and had an earlier age of depression onset (24.58 vs. 34.52). However, they were less likely to have a higher household income ($\geq 52,000\text{£}$, 25.02% vs. 29.21%). More negative life events, including a long-standing illness (46.41% vs. 30.94%) and the non-accidental death of a close genetic family member (10.51% vs. 9.20%), were observed among individuals with suicide attempts compared to the control group, with the exception of depression possibly related to stressful or traumatic events. Similar results were observed when comparing individuals with suicidal ideation to the control group, and comparing individuals with suicide attempts to those with suicidal ideation. Strikingly, individuals with suicidal ideation had the highest scores on both the PHQ-9 and GAD-7 among the study population, whose causes of depression were possibly related to stressful or traumatic events.

Feature importance

Of the 280 predictors, the optimal number of input features was 8–11, as selected by LASSO regression for the four different classifiers that achieved the highest accuracy, varying according to the context and distribution of the data. The importance matrix plot for the LASSO method is shown in Supplementary File 2-Figure 1. These predictors include mental health assessments (i.e., PHQ-9, GAD-7 scores, neuroticism score), depression history (i.e., age at first and last episodes, lifetime number of depressive periods, frequency of

Table 1. Baseline characteristics of the study participants with MDD

Characteristics	Non-suicidal (N = 25,611)	Suicidal ideation (N = 24,361)	Suicide attempt (N = 5167)
Age at recruitment, years	55.52 (7.58)	54.65 (7.54)	52.35 (7.50)
Male (n, %)	8813 (34.41%)	7111 (29.19%)	1436 (27.79%)
Townsend deprivation index			
<-3.64	6930 (27.06%)	6255 (25.68%)	988 (19.12%)
-3.64 to -2.14	6540 (25.54%)	6052 (24.84%)	1031 (19.95%)
-2.14 to 0.55	6717 (26.23%)	6302 (25.87%)	1428 (27.64%)
≥0.55	5424 (21.18%)	5752 (23.61%)	1720 (33.29%)
Household income, £			
<18,000	3447 (13.46%)	3678 (15.10%)	1034 (20.01%)
18,000 to 52,000	12,440 (48.57%)	11,731 (48.15%)	2465 (47.71%)
≥52,000	7480 (29.21%)	6668 (27.37%)	1293 (25.02%)
Unknown	2224 (8.68%)	2284 (9.38%)	375 (7.26%)
Body mass index, kg/m ²			
<24.1	7584 (29.61%)	7110 (29.19%)	1557 (30.13%)
24.1 to 29.9	12,535 (48.94%)	11,461 (47.05%)	2229 (43.14%)
≥29.9	5440 (21.24%)	5736 (23.55%)	1365 (26.42%)
Unknown	52 (0.20%)	54 (0.22%)	16 (0.31%)
Smoking status			
Current smoker	1833 (7.16%)	2161 (8.87%)	793 (15.35%)
Former smoker	9341 (36.47%)	9124 (37.45%)	2143 (41.47%)
Never smoker	14,375 (56.13%)	13,035 (53.51%)	2219 (42.95%)
Unknown	62 (0.24%)	41 (0.17%)	12 (0.23%)
Alcohol drinking status ^a			
Low risk drinking	14,789 (57.74%)	14,350 (58.91%)	2896 (56.05%)
Hazardous drinking	7854 (30.67%)	6926 (28.43%)	1334 (25.82%)
Harmful drinking	1250 (4.88%)	1290 (5.30%)	375 (7.26%)
Former drinker	766 (2.99%)	917 (3.76%)	378 (7.32%)

(continued)

Table 1. Continued.

Characteristics	Non-suicidal (N = 25,611)	Suicidal ideation (N = 24,361)	Suicide attempt (N = 5167)
Never drinker	669 (2.61%)	641 (2.63%)	128 (2.48%)
Unknown	283 (1.10%)	237 (0.97%)	61 (1.18%)
Age at first episode of depression, years	34.52 (16.52)	33.58 (17.43)	24.58 (14.54)
Neuroticism score	4.79 (2.85)	5.60 (3.03)	6.34 (3.13)
Life events			
Depression possibly related to stressful or traumatic event	16,682 (65.14%)	18,885 (77.52%)	3278 (63.44%)
Longstanding illness disability or infirmity	7924 (30.94%)	8470 (34.77%)	2398 (46.41%)
Nonaccidental death in close genetic family	2537 (9.91%)	2517 (10.33%)	543 (10.51%)
PHQ-9 score	6.35 (4.69)	8.94 (6.27)	7.90 (5.51)
GAD-7 score	2.47 (3.36)	4.26 (4.62)	3.96 (4.51)

The values are reported as median (lower quantile-upper quantile) for continuous variables and number (%) for categorical variables.

aFor current drinker, alcohol consumption level was calculated by converting the reported number of glasses to UK standard unit for each type of alcohol and summing up different types of alcohol. Low risk drinking was defined as alcohol consumption level ≤ 14 units/week, hazardous drinking was defined as 14–35 units/week (for females) or 14–50 units/week (for males), and harmful drinking was defined as ≥ 35 units/week (for females) or ≥ 50 units/week (for males).

worst episodes), sleep and trauma (i.e., trouble falling asleep, depression related to stressful or traumatic events, adverse events in adult life and childhood), demographics (i.e., year of birth, Townsend deprivation index), and health and diet (i.e., number of self-reported non-cancer illnesses, beef and poultry intake). There was substantial overlap in the top-ranked predictors across different models.

Model performance evaluation and contribution of variables

We used the final models with the important features selected by LASSO as input variables. The performances of four classification prediction models for differentiating suicidal behaviors are summarized in Figure 2 and Supplementary File 2-Table 3. A preliminary investigation was carried out to evaluate the performance of various machine learning models. Among these, GBDT consistently outperformed others, including SVM and RF. Figure 2A depicts the ROC curves for different classifications in detail. It shows that the GBDT classifier for differentiating between suicide attempters and controls had the largest value (0.914), followed by distinguishing between suicide attempters and controls (0.803). The confusion matrices and the comparisons of other performance evaluation indicators are shown in Figure 2B and Supplementary File 2-Table 3, respectively.

The SHAP model to estimate important variables with the GBDT model used the variables selected by LASSO. The

features were sorted in descending order by Shapley values (Figure 3). The advantage of SHAP values is that the SHAP mean value reflects the impact of the feature on the sample identification or the contribution of the feature to sample identification. The larger the SHAP value is, the larger the contribution of the feature is to sample identification.³³ Figure 3A presents the significance of depression history, including related trauma events, severity, duration, and frequency, in the classification model that distinguishes the suicidal ideation group from the control group. In distinguishing the suicide attempt group from the control group or the suicidal ideation group, the important predictors, as shown in Figure 3B and C, are the age of onset of depression and adverse events in adult life and childhood. There were only five common features that differentiated persons with suicide attempts from controls and persons with suicidal ideation, including adverse events in adult life and childhood, age, the Townsend deprivation index, and age at first-onset of depression.

Categorization of ML scores

The population was separated into two groups (low and high ML scores), using the maximal Youden's index as the optimal cut-off value for distinguishing controls from the and suicidal ideation group. Figure 4 shows the suicide attempt risk scores relative to the control group (0.235) and the suicidal ideation group (0.21).

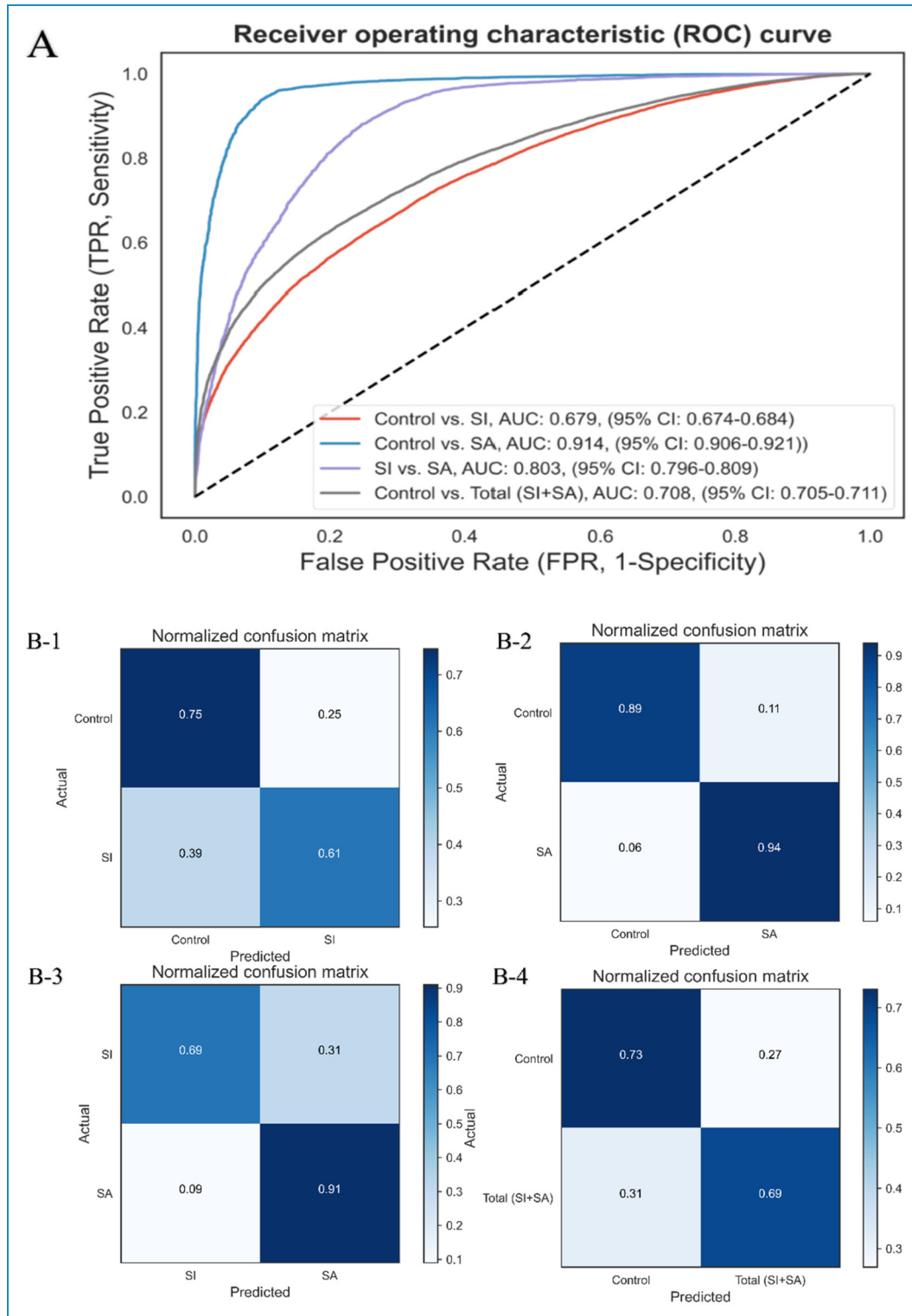


Figure 2. Receiver operating characteristics (ROC) curves and normalized confusion matrices of classifications of GBDT in identifying suicidal behaviors. Abbreviation: CI, confidence interval; ROC, receive operating characteristics. A, The ROC curves for four binary classifications of suicide prediction in individuals with depressive disorders. B, Corresponding normalized confusion matrixes. B-1, individuals without a history of suicidal behavior versus those with suicidal ideation; B-2, individuals without suicidal behavior versus those who have made suicide attempts; B-3, individuals with suicidal ideation versus those who have made suicide attempts; B-4, individuals without suicidal behavior versus those exhibiting any suicidal behavior.

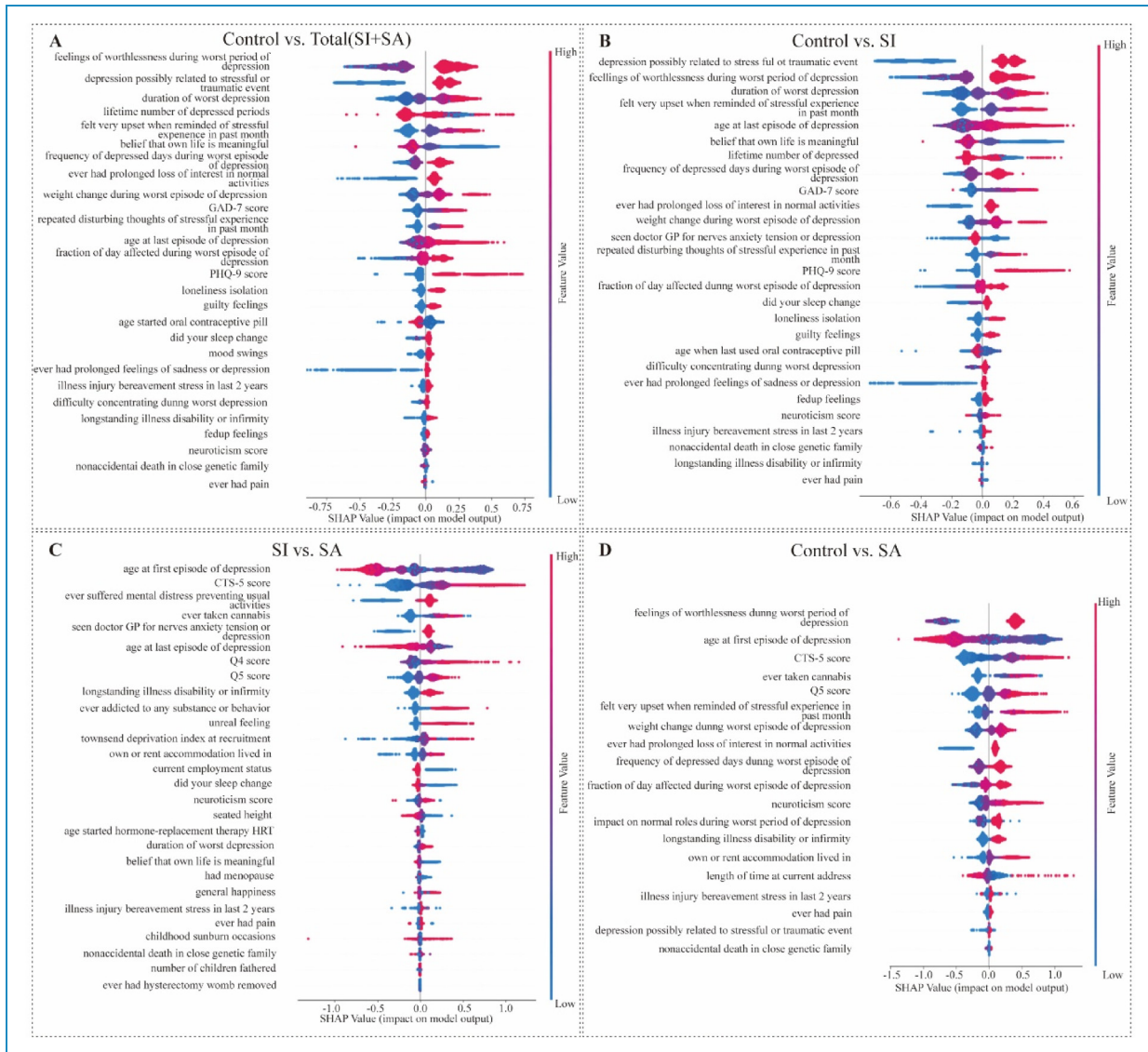


Figure 3. SHAP framework for the top features in the suicide prediction model. Abbreviation: CTS-5, Childhood Trauma Screener-5 item; GAD-7, Generalized Anxiety Disorder-7 questions; Q5, Questions-5 item survey for adult domestic violence; SHAP, SHapley Additive exPlanations. Features on the right of the risk explanation bar pushed the risk higher, and features on the left pushed the risk lower. A, individuals without a history of suicidal behavior versus those with suicidal ideation; B, individuals without suicidal behavior versus those who have made suicide attempts; C, individuals with suicidal ideation versus those who have made suicide attempts; D, individuals without suicidal behavior versus those exhibiting any suicidal behavior.

Discussion

In this study, leveraging a large community-based UK Biobank cohort, focusing on individuals over 50 years of age with depression, recognized as a high-risk group for suicidal behavior,^{2,35,36} we developed machine learning models to accurately predict the risk of suicidal behavior, particularly focusing on the risk of suicide attempts among individuals with depression and those with depressive disorders who have suicidal ideation, achieving AUROC scores of 0.914 and 0.803, respectively.

We found that there was a significant overlap among the top-ranked predictors across different models. When predicting overall suicidal behavior, including both ideation and attempts in individuals with depression, we identified key factors such as neuroticism, GAD-7 scores, and trauma experienced during childhood and adulthood, which align with findings from a substantial body of previous research on suicide prediction.^{4,37–39} Notably, basic sociodemographic variables had relatively limited predictive value, suggesting that a combination of clinical and psychosocial factors may provide more promising prediction accuracy.^{40,41} Furthermore,

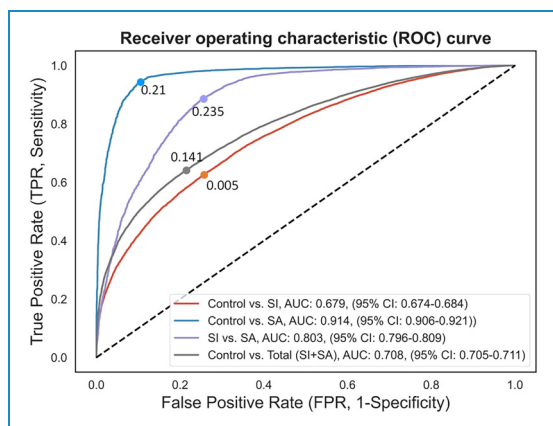


Figure 4. Receiver-operating characteristics (ROC) curves for suicidal behavior prediction using the maximum Youden's index. Abbreviation: CI, confidence interval; ROC, receive operating characteristics.

the algorithms in this study predicting suicide attempts demonstrated slightly higher classification performance than those predicting suicidal ideation, consistent with findings from a population-based study of Korean adults.⁴² Interestingly, the inherent characteristics of depression, including related trauma events, severity, duration, and frequency, play a crucial role in distinguishing the suicidal ideation group from the control group in the classification model. In contrast, when distinguishing the suicide attempt group from the control group or the suicidal ideation group, key predictors include the age of onset of depression and childhood trauma events, among other intriguing findings. We also noted the contribution of other factors, such as beef or poultry intake and the number of non-cancerous diseases. Consistent with earlier research,⁴³ the incidence of non-cancerous diseases underscores the strong relationship between lifestyle factors and mental health. While previous studies on suicide did not investigate the intake of beef and poultry, our analysis suggests a potential link between protein intake and depressive episodes. For instance, a study in a Brazilian population revealed a 165% increase in the risk of depression among individuals who abstain from meat, highlighting a possible association.⁴⁴

Extreme imbalance is a characteristic feature of medical big data, where individuals with a specific disease represent only a small portion of the total population. The low prevalence rate of suicidal behavior significantly complicates the task of precise predictive modeling. Given the nearly 1:5 ratio of individuals who have attempted suicide to those experiencing suicidal ideation or exhibiting no suicidal actions in our study, we identified an imbalance that required a strategic solution. Consequently, we applied the under-sampling technique, which has a proven track record for enhancing model efficacy in the medical research field.^{45,46} This imbalance issue also explains why some AI models do not meet clinical expectations in real-world implementation. Typically, supervised machine learning

models are evaluated on balanced datasets, leading to potential deviations from real-world scenarios. Our GBDT classifier demonstrated high accuracy and F1 score on the test set when distinguishing between suicide attempters and controls, as well as between suicide attempters and those with suicidal ideation. However, the specificity was relatively low. In the real world, most AI models struggle to achieve both a low misdiagnosis rate and a low missed diagnosis rate simultaneously. Therefore, while not all individuals predicted by the model to be at risk will attempt suicide—only a small portion may—the model effectively helps identify high-risk individuals who need early intervention and close observation. This makes the model suitable for clinical application and provides it with high practical value.

The suicide prediction model we have developed targets a high-risk group widely recognized as being at risk—those with depression. These individuals regularly visit relevant health institutions for treatment needs, so early identification can lead to effective intervention. For instance, this system helps care providers identify individuals with depression who may progress to suicidal ideation, allowing for preventive measures to be taken before any suicide attempts occur. If an individual demonstrates a notable increase in the severity, duration, and frequency of their condition, they may be identified as being at high risk of suicidal ideation. Subsequently, they would be recommended for appropriate follow-up, such as psychological counseling services and professional mental health care. More importantly, in clinical practice, it is well understood that a series of questions related to suicidal behaviors are an essential part of the psychiatric assessment for patients with depression. Based on our suicide attempt prediction model, physicians can now identify individuals at risk of suicide attempts with a reduced set of questions, and pay particular attention to those who have experienced adverse events in childhood or adulthood, thereby enhancing the therapeutic value of existing suicide prevention treatments and care.

Despite a widespread acknowledgment of the ethical challenges associated with using machine learning in suicide research and prevention,^{47,48} there is a consensus that bias and the lack of consistency and transparency are among the primary concerns. Therefore, implementing such algorithms could lead to a high rate of false positives and unnecessary interventions. As discussed earlier in our review, false positives may not be problematic when the cost of intervention is low. Moreover, adhering to current reporting guidelines for clinical prediction and diagnostic models using machine learning (TRIPOD-ML and STARD-AI) promotes consistency and transparency in reporting.

The study aims to identify the characteristics and important predictors of suicidal behaviors among individuals over 50 years of age, as well as to develop and evaluate ML models for accurately differentiating between individuals with and without suicidal ideation and suicide attempts. Characteristics of depression significantly influence predicting the risk of

suicidal ideation. In contrast, key predictors for suicide attempts include the age of onset of depression and childhood trauma events, highlighting distinct factors for each outcome. Our focus was particularly on comparing individuals without a history of suicidal behavior to those who have made suicide attempts, as well as comparing individuals with suicidal ideation to those who have made suicide attempts. The findings of this study have significant implications for suicide prevention and mental health interventions for middle-aged and older adults.

A notable limitation of this study includes its cross-sectional design, which does not account for the frequency or sequence of suicidal behaviors, thus hindering causal inferences and longitudinal assessment of risk factors. Causal inferences and longitudinal assessment of risk factors. When defining outcomes, we relied on self-reports and inpatient ICD coding, missing data from emergency departments that have traditionally been the primary source for identifying suicide cases in previous research.^{49,50} Therefore, cases with less severe outcomes require further investigation. Finally, to ensure high-quality data in mental health research, the sample was limited to individuals with depression who completed the MHQ, with the youngest participant being 50 years old. The candidate predictors were restricted by the available data sources, and the resulting models were not clinically derived, leading to limited validation. Consequently, our findings may not be generalizable to the entire UK population or other populations.

Conclusions

In conclusion, using the UK Biobank cohort, we have established applicable machine learning-based models that accurately predict the risk of suicidal behavior in individuals with depression.

These models identify distinct predictive factors at various stages of suicidal behavior and have demonstrated strong performance in assessing the risk of suicide attempts. This underscores the potential of these models to identify individuals at medium to high risk within the general population of those over 50 with depression, thereby aiding in the development of cost-effective suicide prevention strategies.

Abbreviations:

AUC	Area Under the Curve
CI	Confidence Interval
CIDI	Composite International Diagnostic Interview
CTS-5	Childhood Trauma Screener-5 item
GAD-7	Generalized Anxiety Disorder-7 310 questions
Gaussian NB	Gaussian Naive Bayes
GBDT	Gradient Boosting Decision Tree
LASSO	Least Absolute Shrinkage and Selection Operator
MDD	Major Depressive Disorder
MHQ	Mental Health Questionnaire

ML	Machine Learning
PHQ-9	Patient Health Questionnaire
Q5	Questions-5 item survey for adult domestic violence
RF	Random Forest
ROC	Receive Operating Characteristics
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine

Acknowledgements: This research has been conducted using the UK Biobank Resource under Application 54803. This work uses data provided by patients and collected by the NHS as part of their care and support. This research used data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (grant ref: MC_PC_20029 and MC_PC_20058). We thank all the sponsors and team members involved in West China Biomedical Big Data Center and Med-X Center for Informatics, Sichuan University.

Author contributions: HS and WZ were responsible for the study concept and design. JZ, YC, and YH did the data and project management. JZ, and YC did the data cleaning and analysis. YL, CZ, WL, and XY interpreted the data. JZ and DL drafted the manuscript. All the authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Availability of data and materials: Data from the UK Biobank (<http://www.ukbiobank.ac.uk/>) are available to all researchers upon making an application. This research was done using the UK Biobank Resource under Application 54803 (approved on October 29th 2019).



Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Declaration of generative Ai in scientific writing: We didn't use generative artificial intelligence (AI) and AI-assisted technologies in the writing process.

Ethical approval: The UK Biobank study has received full ethical approval from the NHS National Research Ethics Service (16/NW/0274), and all the participants provided written informed consent before data collection. The current study was approved by the biomedical research ethics committee of West China Hospital (2020.661).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by The 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University, (grant number ZYYC21005 to HS).

Guarantor: Huan Song

ORCID iDs: Yujun Liu  <https://orcid.org/0000-0002-5818-2543>
Huan Song  <https://orcid.org/0000-0003-3845-8079>

Supplemental material: Supplemental material for this article is available online.

References

- World Health Organization. Suicide: one person dies every 40 seconds. 2019. <https://www.who.int/news/item/09-09-2019-suicide-one-person-dies-every-40-seconds>.
- Chen Q, Zhang-James YL, Barnett EJ, et al. Predicting suicide attempt or suicide death following a visit to psychiatric specialty care: a machine learning study using Swedish national registry data. *PLoS Med* 2020; 17: e1003416.
- O'Connor RC and Nock MK. The psychology of suicidal behaviour. *Lancet Psychiatry* 2014; 1: 73–85.
- Plana-Ripoll O, Pedersen CB, Agerbo E, et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. *Lancet* 2019; 394: 1827–1835.
- National Institute of Mental Health. Major Depression. 2021. <https://www.nimh.nih.gov/health/topics/depression/index.shtml>.
- Lee HS, Park KJ, Kwon Y, et al. Clinical characteristics associated with suicidal attempt and non-suicidal self-injury in Korean adolescents. *Psychiatry Investig* 2021; 18: 561–569.
- Robinson K, Garisch JA and Wilson MS. Nonsuicidal self-injury thoughts and behavioural characteristics: associations with suicidal thoughts and behaviours among community adolescents. *J Affect Disord* 2021; 282: 1247–1254.
- Voss C, Hoyer J, Venz J, et al. Non-suicidal self-injury and its co-occurrence with suicidal behavior: an epidemiological-study among adolescents and young adults. *Acta Psychiatr Scand* 2020; 142: 496–508.
- Hafferty JD, Navrady LB, Adams MJ, et al. The role of neuroticism in self-harm and suicidal ideation: results from two UK population-based cohorts. *Soc Psychiatry Psychiatr Epidemiol* 2019; 54: 1505–1518.
- Alexopoulos GS, Raue PJ, Banerjee S, et al. Modifiable predictors of suicidal ideation during psychotherapy for late-life major depression. A machine learning approach. *Transl Psychiatry* 2021; 11: 536.
- Kim S, Lee HK and Lee K. Which PHQ-9 items can effectively screen for suicide? Machine learning approaches. *Int J Env Res Pub He* 2021; 18: 3339.
- Nordin N, Zainol Z, Mohd Noor MH, et al. A comparative study of machine learning techniques for suicide attempts predictive model. *Health Informatics J* 2021; 27: 1460458221989395.
- Na KS, Geem ZW and Cho SE. The development of a suicidal ideation predictive model for community-dwelling elderly aged >55 years. *Neuropsychiatr Dis Treat* 2022; 18: 163–172.
- Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; 12: e1001779.
- Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open* 2018; 4: 83–90.
- Davis KAS, Cullen B, Adams M, et al. Indicators of mental disorders in UK Biobank—a comparison of approaches. *Int J Meth Psych Res* 2019; 28: e1796.
- Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank - development, implementation and results from an online questionnaire completed by 157 366 participants: a reanalysis. *BJpsych Open* 2020; 6: e18.
- Collins GS, Dhiman P, Navarro CLA, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *Bmj Open* 2021; 11: e048008.
- Collins GS, Moons KGM, Dhiman P, et al. TRIPOD Plus AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ-Brit Med J* 2024; 385: e078378.
- Peters EM, John A, Bowen R, et al. Neuroticism and suicide in a general population cohort: results from the UK Biobank Project. *BJPsych Open* 2018; 4: 62–68.
- Chai Y, Luo H, Wong GHY, et al. Risk of self-harm after the diagnosis of psychiatric disorders in Hong Kong, 2000–10: a nested case-control study. *Lancet Psychiatry* 2020; 7: 135–147.
- Zheng L, Wang O, Hao S, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl Psychiatry* 2020; 10: 72.
- Kyron MJ, Hooke GR and Page AC. Assessing interpersonal and mood factors to predict trajectories of suicidal ideation within an inpatient setting. *J Affect Disorders* 2019; 252: 315–324.
- Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* 1996; 58: 267–288.
- Hastie TTR and Friedman JH. *The Elements of Statistical Learning*. 2nd ed. New York: Springer, 2009.
- Drummond C and Holte RC C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In Proceedings of the international conference on machine learning (ICML), 2003, 2003, p. 123–130.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist* 2001; 29: 1189–1232.
- Natekin A and Knoll A. Gradient boosting machines, a tutorial. *Front Neuroinformatics* 2013; 7: 21.
- Richter AN and Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. *Artif Intell Med* 2018; 90: 1–14.
- Cho SE, Geem ZW and Na KS. Development of a suicide prediction model for the elderly using health screening data. *Int J Environ Res Public Health* 2021; 18: 10150.
- Cawley GC and Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 2010; 11: 2079–2107.
- Varma S and Simon R. Bias in error estimation when using cross-validation for model selection. *Bmc Bioinformatics* 2006; 7: 91.
- Lundberg S and Lee SI. A Unified approach to interpreting model predictions. 2017.
- Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32–35.
- Wang JR, Qiu JJ, Zhu T, et al. Prediction of suicidal behaviors in the middle-aged population: machine learning analyses of UK Biobank. *Jmir Public Hlth Sur* 2023; 9: e43419.

36. Zheng L, Wang O, Hao SY, et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl Psychiat* 2020; 10: 72.
 37. Tyrer P, Tyrer H and Yang M. The influence of personality disorder in predicting suicidal behaviour in common mental disorders: a 30-year study. *Personal Ment Health* 2022; 16: 111–119.
 38. Wiebenga JXM, Heering HD, Eikelenboom M, et al. Associations of three major physiological stress systems with suicidal ideation and suicide attempts in patients with a depressive and/or anxiety disorder. *Brain Behav Immun* 2022; 102: 195–205.
 39. Gaweda L, Pionke R, Krezolek M, et al. The interplay between childhood trauma, cognitive biases, psychotic-like experiences and depression and their additive impact on predicting lifetime suicidal behavior in young adults. *Psychol Med* 2020; 50: 116–124.
 40. Gale CR, Batty GD, Osborn DPJ, et al. Association of mental disorders in early adulthood and later psychiatric hospital admissions and mortality in a cohort study of more than 1 million men. *Arch Gen Psychiat* 2012; 69: 823–831.
 41. Klonsky ED, May AM and Glenn CR. The relationship between nonsuicidal self-injury and attempted suicide: converging evidence from four samples. *J Abnorm Psychol* 2013; 122: 231–237.
 42. Lee J and Pak TY. Machine learning prediction of suicidal ideation, planning, and attempt among Korean adults: a population-based study. *Ssm-Popul Hlth* 2022; 19: 101231.
 43. Nafilyan V, Morgan J, Mais D, et al. Risk of suicide after diagnosis of severe physical health conditions: a retrospective cohort study of 47 million people. *Lancet Reg Health-Eu* 2023; 25: 100562.
 44. Kohl IS, Luft VC, Patra AL, et al. Association between meatless diet and depressive episodes: a cross-sectional analysis of baseline data from the longitudinal study of adult health (ELSA-Brasil). *J Affect Disorders* 2023; 320: 48–56.
 45. Liu R, Wang MY, Zheng T, et al. An artificial intelligence-based risk prediction model of myocardial infarction. *BMC Bioinformatics* 2022; 23: 217.
 46. Lu HX, Barrett A, Pierce A, et al. Predicting suicidal and self-injurious events in a correctional setting using AI algorithms on unstructured medical notes and structured data. *J Psychiatr Res* 2023; 160: 19–27.
 47. Jacobson NC, Bentley KH, Walton A, et al. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. *B World Health Organ* 2020; 98: 270–276.
 48. Kirtley OJ, van Mens K, Hoogendoorn M, et al. Translating promise into practice: a review of machine learning in suicide research and prevention. *Lancet Psychiatry* 2022; 9: 243–252.
 49. Gradus JL, Rosellini AJ, Horváth-Puhó E, et al. Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark. *Jama Psychiat* 2020; 77: 25–34.
 50. Sanderson M, Bulloch AGM, Wang JL, et al. Predicting death by suicide following an emergency department visit for parasuicide with administrative health care system data and machine learning. *Eclinicalmedicine* 2020; 20: 100281.
-