Observational Study

Medicine

OPEN

Machine learning approaches for predicting suicidal behaviors among university students in Bangladesh during the COVID-19 pandemic

A cross-sectional study

Sultan Mahmud, MS[a,*], Md Mohsin, MS[b], Abdul Muyeed, MS[c], Shaila Nazneen, MS[d], Md. Abu Sayed, MS[e], Nabil Murshed, MS[e], Tajrin Tahrin Tonmon, MS[d], Ariful Islam, MS[e]

Abstract

Psychological and behavioral stress has increased enormously during Coronavirus Disease 2019 (COVID-19) pandemic. However, early prediction and intervention to address psychological distress and suicidal behaviors are crucial to prevent suicide-related deaths. This study aimed to develop a machine algorithm to predict suicidal behaviors and identify essential predictors of suicidal behaviors among university students in Bangladesh during the COVID-19 pandemic. An anonymous online survey was conducted among university students in Bangladesh from June 1 to June 30, 2022. A total of 2391 university students completed and submitted the questionnaires. Five different Machine Learning models (MLMs) were applied to develop a suitable algorithm for predicting suicidal behaviors among university students. In predicting suicidal behaviors, the most crucial background and demographic features were relationship status, friendly environment in the family, family income, family type, and sex. In addition, features related to the impact of the COVID-19 pandemic were identified as job loss, economic loss, and loss of family/relatives due to COVID-19. Moreover, factors related to mental health include depression, anxiety, stress, and insomnia. The performance evaluation and comparison of the MLM showed that all models behaved consistently and were comparable in predicting suicidal risk. However, the Support Vector Machine was the best and most consistent performing model among all MLMs in terms of accuracy (79%), Kappa (0.59), receiver operating characteristic (0.89), sensitivity (0.81), and specificity (0.81). Support Vector Machine is the best-performing model for predicting suicidal risks among university students in Bangladesh and can help in designing appropriate and timely suicide prevention interventions.

Abbreviations: COVID-19 = coronavirus disease 2019, CT = classification tree, CV = cross-validation, ISI = Insomnia Severity Index, KNN = K-nearest neighbors algorithm, LR = logistic regression, NB = Naïve Bayes, RF = Random Forest, RFE = Recursive Feature Elimination, ROC = receiver operating characteristic curve, SBQ-R = Suicidal Behaviors Questionnaire-Revised, SVM = Support Vector Machine.

Keywords: COVID-19, machine learning, suicidal ideation/risk, support vector machine (SVM)

1. Introduction

Fear, concern, and anxiety around the globe are direct aftereffects of the global emergence of the Coronavirus Disease 2019 (COVID-19) pandemic. Besides the physical symptoms, COVID-19 can result in several psychological symptoms such as nervousness, decreased consciousness (with seizures), traumatic stress, depressed mood, confusion, insomnia, anxiety disorders, brain inflammation, stroke, delirium, and nerve damage.[1,2] Moreover, the lockdown measures have intensified the emotional and behavioral difficulties that led to a lack of concentration and poor job performance.[3–5] Furthermore, the World Health Organization has reported an increasing global rate of mental health issues as a direct consequence of COVID-19.[1] This higher rate reflects people's vulnerability regarding their mental well-being and physical predicaments. The psychosomatic stress imposed by COVID-19 affects the quality of life of people suffering from the disease and

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

This manuscript was previously posted to Research Square: doi: https://doi.org/10.21203/rs.3.rs-2069873/v1

Supplemental Digital Content is available for this article.

[a] International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh, [b] Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh, [c] Department of Statistics, Jatiya Kabi Kazi Nazrul Islam University, Mymensingh, Bangladesh, [d] North South University, Dhaka, Bangladesh, [e] University of Dhaka, Dhaka, Bangladesh.

*Correspondence: Sultan Mahmud, International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), 68, Shaheed Tajuddin Ahmed Sarani Mohakhali, Dhaka 1212, Bangladesh (e-mail: smahmud@isrt.ac.bd).

How to cite this article: Mahmud S, Mohsin M, Muyeed A, Nazneen S, Abu Sayed M, Murshed N, Tonmon TT, Islam A. Machine learning approaches for predicting suicidal behaviors among university students in Bangladesh during the COVID-19 pandemic: A cross-sectional study. Medicine 2023;102:28(e34285).

Received: 13 April 2023 / Received in final form: 12 June 2023 / Accepted: 21 June 2023

http://dx.doi.org/10.1097/MD.0000000000034285

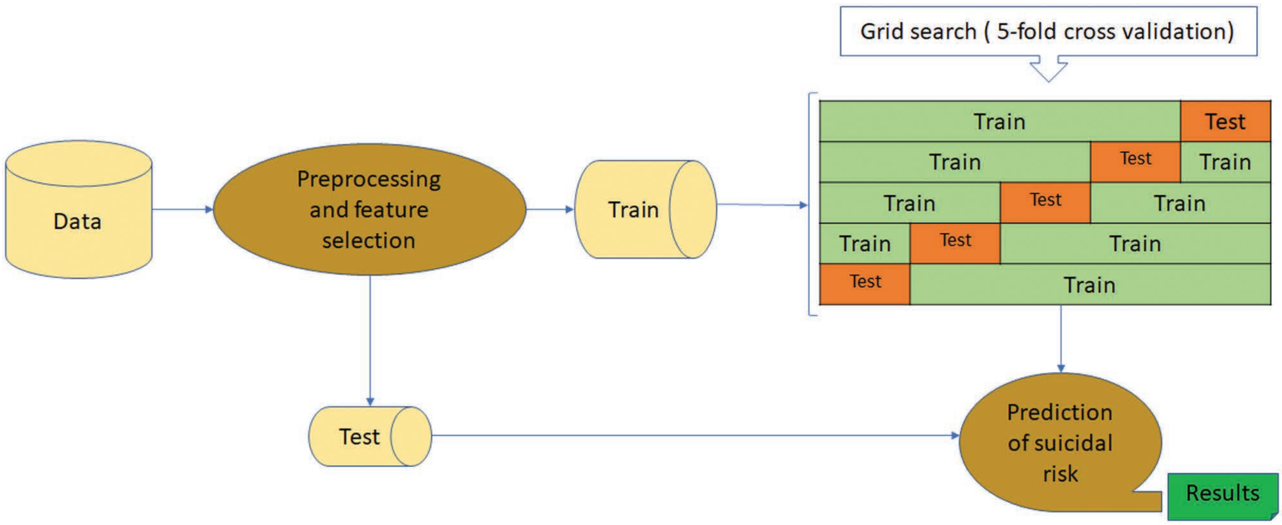

Figure 1. Diagram of the workflow of suicidal risk prediction.

Table 1

Confusion matrix.

	Prediction	
True value	**P**	**N**
P	TP	FN
N	FP	TN

is a critical determinant of later psychological disorders and suicide incidents.[6] According to the World Health Organization, it has become the second leading cause of mortality among people between 15 and 29 years of age.[7] Like many developing countries, the covid-19 pandemic had a devastating impact on Bangladesh as well. The Asian Development Bank projections estimated nearly 9 million job losses and approximately US$3 billion loss of GDP due to COVID-19.[8] Therefore, studies are examining the long-term mental health impact on the Bangladeshi population who have faced economic hardships,[9,10] which could potentially contribute to an increase in suicidal incidences. Within the initial year of the pandemic (from March 8, 2020, to March 8, 2021), Bangladesh witnessed approximately 14,436 suicide deaths, which is 70% higher than the fatalities attributed to the pandemic.[11]

Psychologically vulnerable people commit suicide, an extreme action that requires several preparatory steps, such as suicidal ideation or thoughts, suicide plans, and suicide attempts before finally committing suicide to end their lives.[12,13] Reducing the risk factors and boosting resilience are the 2 main objectives of suicide prevention. Therefore, early detection and intervention to address psychological distress and suicidal behaviors are crucial to prevent suicide-related deaths.[14] Limited resources on campuses and students' unwillingness to share information regarding their mental health[15,16] makes detecting university students' suicidal behaviors challenging.

Several recent studies have observed the prevalence and provided association measures of suicidal behaviors among Bangladeshi inhabitants using logistic regression,[17–19] some of which targeted students.[20,21] Nevertheless, identifying factors associated with suicidal behaviors is not necessarily helpful in predicting future suicidal behaviors. Therefore, a model designed explicitly for prediction is needed.[16,22] Additionally, psychiatric assessment tools used in previous studies require the expertise of trained clinicians and are primarily impractical for assessing suicidal ideation in large populations.[23] Therefore, it is now time to focus on developing predictive algorithms using machine-learning models. To our knowledge, predictive tools using machine-learning models have been developed in 2 previous studies. Both used only random forest models with 3 or 4 performance measures.[16,24] To the best of our knowledge, this is the first study to develop a suitable predictive model using machine learning algorithms with 5 performance measures to identify or predict the risk of suicidal behaviors among university students in Bangladesh. In addition, this study also assessed the potential socio-demographic, behavioral, health, and psychopathological determinants to predict the risk of suicidal behaviors.

2. Related works

In recent years, there has been a growing interest in utilizing machine learning techniques to predict suicidal behaviors. Several studies have focused on applying machine learning algorithms to identify early warning signs and risk factors associated with suicidal behaviors in vulnerable populations. For instance, Macalli et al[16] conducted a study using the random forests model to predict suicidal behaviors among college students. Although they considered a large number of predictors (70), their evaluation included only 3 performance measures: the receiver operating curve (AUC), sensitivity, and positive predictive value. Another study by Rezapour and Hansen[25] explored various machine learning algorithms, such as Logistic Regression, Naive Bayes, k-Nearest Neighbors, Support Vector Machines, Neural Networks, and Random Forests. While they emphasized the importance of identifying risk factors contributing to suicidal behaviors, they did not specify the most suitable predictive model. Furthermore, Chen et al[26] developed a model based on ensemble learning by combining several machine learning models for predicting suicide attempts or suicide death in the Swedish population. A narrative review of 56 studies and an analysis of 54 machine learning models provided an overview of the strong overall performance of these models in predicting suicidal behaviors. Although there was variability in model performance, no significant differences was observed based on outcome, data type, or model type.[27] However, it is crucial to note that further research is necessary to validate and refine these models, incorporating a considerable number of performance evaluation metrics. Additionally, considering the specific factors related to the COVID-19 pandemic and the cultural context of Bangladesh would enhance the applicability and effectiveness of these models.

Table 2

Demographic and background characteristics of the participants.

Characteristics	Frequency (n)	Percentage or Mean (SD, Rage)
Age (yr)		22.68 (1.93, 19–27)
Respondents' age (Category)		
19–22 yr	329	13.76
23–27 yr	2062	86.24
Sex		
Male	1483	62.02
Female	908	37.98
Residence		
Urban	1053	44.04
Rural	1338	55.96
Religious status		
Muslim	2273	95.06
Hindu	92	3.85
Others	26	1.09
Studentship status		
Bachelor's 1st year	764	31.95
Bachelor's 2nd year	587	24.55
Bachelor's 3rd year	397	16.6
Bachelor's 4th year	252	10.54
Master's degree	391	16.35
Disability status (physical or mental)		
Not disable	1986	83.06
Disable	405	16.94
Relationship status		
Single	2041	85.37
Engaged	198	8.28
Married	125	5.23
Separated	27	1.12
Respondents' body mass index		
Underweight	318	13.30
Normal weight	1591	66.54
Overweight	365	15.27
Obese	117	4.89
Family environment		
Friendly	2099	87.79
Unfriendly	292	12.21
Studied faculty		
Fundamental science	539	22.54
Engineering	493	20.62
Social science	848	35.47
Humanities	269	11.25
Business administration	242	10.12
Current residency type		
Hall	551	23.04
Rented house or mess	1257	52.57
Own house	583	24.38
Family income (category)		
≤10,000 BDT	180	7.53
>10,000–20,000 BDT	604	25.26
>20,000–30,000 BDT	726	30.36
>30,000 BDT	881	36.85
Genre of family		
Nuclear	1970	82.39
Joint	421	17.61
No. of Siblings		
≤3 Siblings	985	41.20
>3 Siblings	1406	58.80
Fathers' highest level of education		
No education	293	12.25
Primary education	435	18.19
Secondary education	1033	43.20
Higher education	630	26.35
Fathers' occupation		
Agricultural worker (farmer)	727	30.41
Businessman	742	31.03
Service holder	922	38.56
Mothers' highest level of education		
No education	408	17.06
Primary education	660	27.60
Secondary education	1115	46.63
Higher education	208	8.70
Mothers' occupation		
Homemaker	2084	87.16
Service holder	229	9.58
Others	78	3.26

3. Methods

3.1. Data collection

We collected data through an anonymous online survey using convenience sampling from June 1, 2022, to June 30, 2022, from university students across Bangladesh. We used a structured questionnaire (Supplemental Digital Content, http://links.lww.com/MD/J342) to collect data regarding suicidal behavioral patterns, sociodemographic information, academic information, psychological morbidities, behaviors, and the impact of the COVID-19 pandemic.

Participants in the study were interviewed using an electronic questionnaire. The researchers distributed an online survey link through platforms such as Facebook, Messenger, and WhatsApp using authors' connections. Potential respondents were requested to participate in the survey and were presented with a detailed consent section. This consent section outlined the study's purpose, the types of questions that would be asked, the assurance of anonymity, and the voluntary nature of participation. The survey continued only if the participants agreed to proceed with the survey after providing their consent. They were explicitly informed of their right to skip questions, refuse to answer, or withdraw from the study at any point.

The survey questionnaire was comprised of 3 primary sections: Demographic and background information; Assessment of suicidal behavior; Evaluation of insomnia severity; and Measurement of depression, anxiety, and stress levels.

3.2. Participants and sampling

A previous study showed that 14% of university students in Bangladesh had suicidal ideation,[28] and the total number of university students was 853,267.[29] The minimum required sample size was 291, considering the expected response rate[28] of 71%, the expected proportion with 5% absolute precision, and a 95% confidence interval. However, 291 is a small sample size for training machine learning models which can lead to biased model performance estimates, underfitting, and poor performance.[30,31] Therefore, we targeted to recruit at least 2000 participants using convenience sampling. Participants were included if they were a university student, aged 18 years or older, able to read Bangla, and living in Bangladesh.

3.3. Assessment of anxiety, depression, stress, and insomnia

We utilized the Depression Anxiety Stress Scale tool, which consists of 42 items. The scale employs a four-point Likert scale and is divided into 3 subsections, each comprising 7 questions. The purpose of the scale was to evaluate the levels of depression, anxiety, and stress among the participants. The participants were asked to rate the severity of their experiences over the past week using the 4-point Likert scale, emphasizing temporary states rather than enduring traits. The scale ranged from 0, indicating that the item did not apply to the participant at all, to 3, indicating that the item applied to the participant very much or most of the time. Items 1 and 2 fell within an intermediate level of rating. The instructions explicitly clarified that there were no correct or wrong answers.

The Depression Anxiety Stress Scale, originally developed in English, is designed to measure negative emotional states such as depression, anxiety, and stress.[32,33] The Bangla version of this scale has been validated among students in Bangladesh.[34,35] Additionally, the Insomnia Severity Index (ISI) scale, another widely used self-report questionnaire, was employed to assess the severity of insomnia symptoms in the patients. The ISI scale consists of 7 items that measure various dimensions of insomnia, including difficulties with sleep initiation, sleep maintenance, early morning awakening, sleep satisfaction, interference with daily functioning, noticeability of sleep problems by others, and distress caused by insomnia symptoms.[36] Each item is rated on a 5-point Likert scale, ranging from 0 to 4, with higher scores indicating greater severity of insomnia symptoms. The Bangla version of the ISI scale was validated by Mamun et al.[37]

3.4. Assessment of suicidal behaviors

Previous instances of suicidal behaviors, including thoughts and attempts, have been acknowledged as important risk factors for future suicidal behaviors. To identify such behaviors and assess the associated risk, the widely used Suicidal Behaviors Questionnaire-Revised (SBQ-R)[38] is employed. In this study, we utilized a validated SBQ-R questionnaire[39,40] consisting of 4 questions. The first question aimed to determine if the participant had ever experienced suicidal thoughts or attempted suicide. The second question assessed the frequency of suicidal thoughts in the past 12 months. The third question inquired whether the individual had disclosed their suicidal thoughts or intentions to anyone else.

Furthermore, the final question aimed to gauge the self-reported likelihood of future suicidal behaviors. It is important to note that the questionnaire was modified specifically for capturing suicidal behaviors during the COVID-19 pandemic. The SBQ-R score ranges from 3 to 18, with a cutoff point of 7 or higher indicating a significant risk of suicidal behaviors, as determined by previous research.[38]

3.5. Consent and ethical considerations

Participants in the survey willingly provided their consent online by accessing the shared link and agreeing to complete the survey form shared on Facebook, Messenger inbox, or WhatsApp inbox. It is crucial to emphasize that the survey guaranteed participant anonymity, which implies that their identities remained undisclosed and unlinked to their responses throughout the entire data collection and analysis process. Ethical approval for the study was granted by the Ethics Committee of Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh-2224, Bangladesh, with Ref: JKKNIU/PS/Ethical/2022/60.

3.6. Statistical analysis

We performed all statistical analyses using the packages "mlbench" and "Caret" in the statistical programming language R to predict suicidal ideation based on the most important psychological and demographic features. The workflow of suicide prediction is presented in Figure 1.

3.7. Pre-processing and feature selection

Data preprocessing is the first step in predictive model fitting and statistical analysis. This step included removing outliers, replacing missing values, and data standardization. Outliers are observations that differ significantly from other observations and significantly affect classifiers. Standardization or normalization is an approach to convert data to achieve a distribution with a mean of zero and variance of one.[41,42]

Feature selection is a crucial part of machine learning model building to improve the predictive performance of machine learning models. In this study, we used the Recursive Feature Elimination (RFE) method, which is the most commonly used method because of its configuration and effectiveness, to identify the most important features for predicting diseases.[43] These crucial features are strongly correlated with the disease of concern. The RFE algorithm identifies the relevant and most significant predictors by eliminating useless, less critical, or less correlated features.[44] The capacity to distinguish the closest distance between 2 classes (yes/no) was used to determine the significance of a feature.[45] Several metrics can be used to measure the capacity of the features, such as accuracy, Kappa, and receiver operating characteristic curve (ROC). In this study, we used 2 metrics to select the importance of a relevant subset of features. The following section discusses details regarding the evaluation metrics of the machine learning models.

3.8. Cross-validation

Cross-validation (CV) is a method to improve the generalizability of machine learning models and prevent overfitting.[46] The CV has also been widely used for model selection and classifier error estimation. This approach repeatedly and randomly divides data into K subsets. First, the model was trained, the hyperparameters were fine-tuned in the inner loop where the grid search algorithm was applied using (K-1) datasets, and the model was evaluated on the remaining data (test) set. For imbalance in the dataset, CV stratified it so that each of the K folds contained the same proportion of negative and positive observations as in the original dataset.[47] This approach was repeated K times to determine the best-performing model[41] based on the final metric. The final matrix is estimated[41] as

$$\bar{M} = \frac{1}{K}\sum_{i=1}^{K} m_i \mp \sqrt{\frac{\sum_{i=1}^{K}(m_i - \overline{m})}{K-1}}, \tag{1}$$

where $\bar{M}$ is the final performance metric for the classifiers and $m_i \in r$, $i = 1,2,\ldots, K$ is the performance metric for each fold.

3.9. Machine learning model

In our study, we aimed to compare the performance of 6 popular machine learning models in predicting suicidal ideation. The models we examined were logistic regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), K-nearest neighbors (KNN), Classification Tree (CT), and Random Forest (RF). Our objective was to determine which model exhibited the highest efficiency in identifying individuals at risk of suicidal ideation.

Firstly, LR is a widely used model for binary classification problems.[48] It calculates the probability of an event occurring based on the input features and estimates coefficients for each feature to make predictions. LR is known for its simplicity and interpretability, as it provides insight into the importance of each feature in predicting the outcome. For a target variable Y and a set of features $x_1.x_2.x_3 \ldots x_n$ the logistic regression classifier can be defined as

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_n * X_n \tag{2}$$

Where $\beta_1, \beta_2, \beta_3 \ldots \beta_n$ is a set of coefficients $\frac{1}{1+e^{-z}}$ is the probability of occurring the event.

SVM, on the other hand, is a versatile model that can handle both classification and regression tasks.[48] It seeks to find an optimal hyperplane that separates different classes or predicts target variable values. By maximizing the margin between

Table 3

The features used for machine learning.

Feature name	Feature label	Feature type	Value label
Academic_year	Academic year	Categorical	Values indicate the year of education in university (1st, 2nd, 3rd, 4th, 5th year or Masters)
gender	Gender	Binomial	F: Female, M: Male
religion	Religion	Categorical	Values in Islam, Hindu, and Others
friendly_environment	Do you think your family environment is friendly?	Categorical	Values as Strongly-disagree, Dis-agree, Neutral, Agree, Strongly-agree
bmi_category	RECODE from weight (kg) and height (m) as kg/m^2	Categorical	Values as Underweight, Normal weight, Overweight, Obese
siblings_new	Number of siblings	Numeric	Number of siblings in the household
age_new	Age	Numeric	Age in year
faculty	Name of the subject/group you are studying	Categorical	Values indicate the name of the faculty he/she is studying (Science, Engineering, Medical Science, Arts, Social Science)
univer_type	The type of university you are studying	Categorical	Values indicate the type of the university (Public, Private, National, Medical College)
permanent_residence	Origin/ Permanent residence	Binomial	values as Rural, Urban
current_residence	Current residence	Categorical	Values as Hall, Rented house or Mess, Own house
family_type	Family type	Categorical	Values as Nuclear, Joint or Extended
fathers_education	Father's Education	Categorical	Values as Illiterate, Primary, Secondary, Higher-secondary, Above
fathers_occupation	Father's Occupation	Categorical	Values as Service holder, Businessman, Farmer, Others
mothers_education	Mother's Education	Categorical	Values as Illiterate, Primary, Secondary, Higher-secondary, Above
mothers_occupation	Mother's Occupation	Categorical	Values as Housewife, Service holder, Others
family_income	Family income (monthly)	Numeric	In Taka
daily_study_hour	Daily average study hour	Categorical	Values as 1–3 h, 4–6 h, 7–9 h, ≥9 h
relationship_status	Relationship status	Categorical	Values indicate marital Status as Single, Married, Engaged
smoking_status	Smoking status	Binomial	Values as Yes, No
physical_or_mental_disability	Physical or mental disability	Binomial	Values as Yes, No
religious_practice	Do you perform religious practice regularly?	Binomial	Values as Yes, No
academic_pressure	Are you satisfied with your academic workload (i.e., presentations, assignments, tutorials)?	Binomial	Values as Yes, No
extracurricular_activities	Do you usually do extracurricular activities (sports, student government, community service, employment, arts, hobbies, and educational clubs)?	Binomial	Values as Yes, No
have_session_jam_covid	Do you face session jam in your department due to COVID-19?	Binomial	Values as Yes, No
infected_covid	Did you get infected by the novel coronavirus?	Binomial	Values as Yes, No
economic_loss_covid	Did you/any of your family members face economic loss due to COVID-19?	Binomial	Values as Yes, No
job_loss_covid	Did you/any of your family members lose their job due to COVID-19?	Binomial	Values as Yes, No
relatives_infected_covid19	Did your family member (s) or relatives get infected by the novel coronavirus?	Binomial	Values as Yes, No
relatives_died_of_covid19	Did your family member (s) or relatives die of the novel coronavirus?	Binomial	Values as Yes, No
subject_opportunity	Do you feel that in the aspect of career building your subject has an adequate opportunity in our country?	Categorical	Values as Strongly-disagree, Disagree, Neutral, Agree, Strongly-agree
subject_related_job	Do you think your subject-related job gets enough social value in our country?	Categorical	Values as Strongly-disagree, Dis-agree, Neutral, Agree, Strongly-agree
professional_environment	Do you think you have a good professional environment in our country?	Categorical	Values as Strongly-disagree, Dis-agree, Neutral, Agree, Strongly-agree
insomnia_new	The score was calculated using Insomnia Severity Index (ISI) and categorized as No clinically significant insomnia and Clinically significant insomnia	Binomial	Values as Yes: Clinically significant insomnia, No: No clinically significant insomnia
depression_new	Generated as no depression and depression from DASS-21 Scoring	Binomial	Values as Yes, No
anxiety_new	Generated as no anxiety and anxiety from DASS-21 Scoring	Binomial	Values as Yes, No
stress_new	Generated as no stress and stress from DASS-21 Scoring	Binomial	Values as Yes, No

classes, SVM aims for better generalization and robustness. It can handle linearly separable data as well as nonlinear relationships through the use of kernel functions.

In the case of linear SVM, the equation for the decision function is

$$f(x) = W^T * X + b \tag{3}$$

Where $f(x)$ represents the predicted class label for a given input sample X, W is the weight vector perpendicular to the hyperplane and b is the bias term. W^T denotes the transpose of w.

NB is a probabilistic model based on Bayes' theorem.[48] It assumes independence among features and calculates the probability of an event based on the prior probabilities and conditional probabilities of the features. NB is known for its simplicity, scalability, and fast training speed. Despite its naive assumption of feature independence, it often performs well in practice, especially when dealing with text classification tasks. The equation for the Naïve Bayes classifier is as follows:

$$P(k \mid X) = \frac{P(X \mid k) * P(K)}{P(X)} \tag{4}$$

where $P(k \mid X)$ is the probability of class k given the input features X.

The KNN algorithm is a classification technique that relies on the proximity of training examples in the problem space. It belongs to the category of instance-based learning or lazy learning, where the function approximations are made locally, and computation is deferred until the classification stage.[49] In KNN, when classifying an object, the algorithm looks at the k nearest neighbors of that object. The value of k is a positive integer, usually small. By default, the algorithm employs a majority voting approach to determine the class label for the object. This means that the object is assigned to the class that is most common among its k nearest neighbors.

CT is a decision tree-based model that recursively splits the data based on feature thresholds to create a hierarchical structure of decisions.[48,50] It partitions the data into subsets, making predictions based on majority voting within each subset. CT is easy to understand and interpret, and it can handle both categorical and numerical features.

Lastly, RF is an ensemble model that combines multiple decision trees to make predictions.[48] It creates a diverse set of decision trees by training each tree on a random subset of the data and random subsets of features. RF then aggregates the predictions of individual trees to make a final prediction. This ensemble approach improves generalization and reduces overfitting. Random Forest is known for its robustness and ability to handle high-dimensional data.

3.10. Evaluation metrics

We evaluated the prediction accuracy of the machine learning models using accuracy, Kappa, ROC, specificity, and sensitivity. Accuracy is the average proportion of correct predictions. It can be calculated from the confusion matrix (Table 1) as $\frac{TP+TN}{T+N}$. The ROC curve is a graphical tool widely used to evaluate the predictive ability of binary classifiers. It shows the relationship between the specificity (true positive rate) and 1-sensitivity (false positive rate) of the classifiers.[50] Specificity is the ability of a model to correctly identify a patient who does not have the disease of interest and can be computed from a confusion matrix as $\frac{TN}{TP+TN}$. Similarly, the ability of a model to correctly identify a patient who actually has the disease is called sensitivity, and can be determined from the confusion matrix as $\frac{TP}{TP+FN}$.

For many years, the accuracy of predictive models has been assessed using Cohen's kappa in fields including statistics, psychology, biology, and medicine.[51] It can be defined as

$$\mathrm{k} = \frac{P_{a-P_{cb}}}{1 - P_{cb}},$$

where P_a is the probability of true classification or accuracy and P_{cb} is the probability of true classification due to chance. The value of Cohen's kappa varies from –1 to 1.

4. Results

In total, 2391 students completed the survey. Table 2 presents the demographic and background characteristics of the participants. The average age of all respondents was 22.68 (SD = 1.93 and range = 19–27), and more than 86% were 23 to 27 years old. Most respondents were males (62%) and Muslims (95%). All students had bachelor's (83.65%) or master's (16.35%) degrees. Only 26.35% of fathers and 8.70% of mothers of the participating students had above secondary-level education. Among the participants, 12.21% described their family environment as being unfriendly. An overwhelming majority (82.39%) of participants belonged to nuclear families.

As a part of the process, we first checked for outliers and missing values and found no outliers or missing values. Then, we perform standardization or normalization to convert the data to achieve a mean zero and variance one distribution.

We considered 80% of the data points as training data and the rest as test data. During the model training, we used a training dataset and 5-fold cross-validation in which the datasets were divided again into 5 subsets. Four were used for model training, and one was used as a test set. This procedure was repeated 5 times to obtain the best-performing model. The model with the highest average performance metrics was considered the best predictive model for this study. The final setup of the models was also validated using the test dataset (20% of the primary dataset).

This study used the features related to socio-demographic characteristics, behavior, health, mental health, and the impact of COVID-19 was used to train the machine learning model for predicting suicidal risk. Table 3 highlights all 37 features available in the suicidal ideation dataset. The importance of the most critical 24 features is presented in Figure 2, which shows that depression has the highest association (18.71) and subject-related opportunities have the lowest score (1.2). Figure 3 shows that the optimal number of features is 15, for which RFE shows the highest cross-validated accuracy and Kappa score. Therefore, Figures 2 and 3 indicate that the most important psychological factors or features for predicting suicidal behaviors were depression (18.71), insomnia (7.25), anxiety (4.05), and stress (4.01). Among the background and demographic features, the most important feature for predicting suicidal risk was relationship status (9.97), followed by the friendly environment in the family (5.50), family income (4.70), family type (3.51), and sex (3.42). Among the features related to the impact of the COVID-19 pandemic, job loss due to COVID-19 (4.42), economic loss due to COVID-19 (4.11), and loss of family/relatives due to COVID-19 (3.60) were significant predictive factors for risk of suicidal behaviors among Bangladeshi university students. Among the behaviors and health-related features, smoking status (4.90) and extracurricular activities (4.51) were the most important predictors of suicidal behaviors among university students.

The correlation matrix, represented in the Heatmap (Fig. 4), illustrates the relationships between features, both among themselves and with the target feature. By visually presenting the data, the Heatmap facilitates the identification of the features that exhibit the strongest associations with the target feature. Remarkably, the set of features identified through the RFE method also demonstrates a notable correlation with suicidal behavior.

Models with the relevant subset of features (optimal 15) were trained and cross-validated using 5 folds of the training dataset. The comparative performance of the models from the cross-validation is presented in Figures 5 and 6. Figure 5 shows each model's average accuracy and kappa coefficient with a 95% confidence interval. All models showed more than 70% accuracy and more than 0.5 value of Kappa. Moreover, the Support Vector Machine (SVM) had the highest average accuracy (approximately 79%) and kappa (0.60). The average values of ROC, sensitivity, and specificity with 95% CI for each model are presented in Figure 6, which indicates that SVM has the highest average value of ROC and sensitivity.

Regarding specificity, the 2 models, SVM, and Naïve Bayes, performed best.

These models, with the same setup and the same subset of features, were also fitted for the test dataset. The findings are presented in Table 4. The Logistic Regression model achieved an accuracy of 0.78, indicating that it correctly classified 78% of the instances. It also had a Kappa score of 0.57, which measures the agreement between the model's predictions and the actual outcomes. The ROC value of 0.87 suggests that the model performed well in distinguishing between the positive and negative classes. Additionally, it had a sensitivity of 0.80 and a specificity of 0.76, showing its ability to correctly identify both true positives and true negatives.

The Support Vector Machine (SVM) model performed slightly better than Logistic Regression, with an accuracy of 0.79 and a Kappa score of 0.59. It had an ROC value of 0.89, indicating a strong ability to separate the classes. The sensitivity and specificity values of 0.81 and 0.81, respectively, demonstrate its balanced performance in correctly identifying positive and negative instances.

The K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 0.77 and a Kappa score of 0.54. While it had a lower ROC value of 0.84 compared to the previous models, it still showed a satisfactory ability to distinguish between classes. The sensitivity and specificity values of 0.77 and 0.75 indicate moderate performance in correctly classifying positive and negative instances.

The Naïve Bayes classifier achieved an accuracy of 0.78 and a Kappa score of 0.57, similar to the Logistic Regression model. It had an ROC value of 0.87, suggesting a good discriminatory capability. Notably, it had a sensitivity of 0.76 and a specificity of 0.82, indicating a higher ability to correctly identify true negatives.

The Random Forest model demonstrated an accuracy of 0.79 and a Kappa score of 0.58. With an ROC value of 0.87, it performed well in distinguishing between the classes. It had a sensitivity of 0.81 and a specificity of 0.77, indicating balanced performance in identifying positive and negative instances.

Lastly, the Classification Tree model had an accuracy of 0.76 and a Kappa score of 0.52. It achieved an ROC value of 0.76, suggesting a moderate discriminatory capability. The sensitivity and specificity values of 0.76 and 0.72, respectively, indicate its ability to correctly classify true positives and true negatives.

Overall, the Support Vector Machine model performed consistently well across multiple metrics, indicating their suitability for the classification task.

5. Discussion

University students in Bangladesh are at a higher risk of suicide than the general population. This study investigated 2391 university students during the COVID-19 pandemic to identify people at an increased risk of suicidal behavior using suitable machine-learning models. The study aimed to identify the most crucial socio-demographic, behavioral, and psychological determinants in order to build an efficient predictive algorithm for forecasting suicidal behavior risks. Out of the 35 predictors used, 15 were found to be significant in building a predictive algorithm that can accurately identify students with the potential risk of suicidal behaviors. This algorithm has the potential to identify suicidal ideation among university students in Bangladesh, which can aid university authorities and policymakers in implementing targeted interventions and support systems.

In this study, features related to socio-demographic characteristics, behaviors, physical and psychological health, and the impact of COVID-19 were included in the machine learning model to predict suicidal risk. Based on these findings, the essential psychological determinants of suicidal behaviors among university students include depression, insomnia, anxiety, and stress. A similar set of features was identified as a significant predictor of suicide risk among the general Korean population.[52] Moreover, mood, anxiety, psychotic, and trauma-related disorders were also significantly associated with suicidal behaviors among Bangladeshi students.[53] During the COVID-19 pandemic, students with psychopathological conditions, including depression, anxiety, and stress, were found to be more likely to have suicidal risk factors,.[20,54]

Among demographic features, sex, relationship status, family type, and income were crucial predictors of suicidal risk. While several cross-sectional studies have found that relationship problems are the leading cause of suicide among women, financial concerns and illness are the leading causes of suicide among men.[53,55,56] A Study done in India also highlighted that the demanding nature and stress of nuclear families might influence the attempts to suicide.[57] During the COVID-19 pandemic, studies have reported that sex and marital status were significantly associated with suicidal behaviors among Bangladeshi citizens.[18–20] Several studies have identified academic pressure as an essential feature and significant predictor in envisaging the

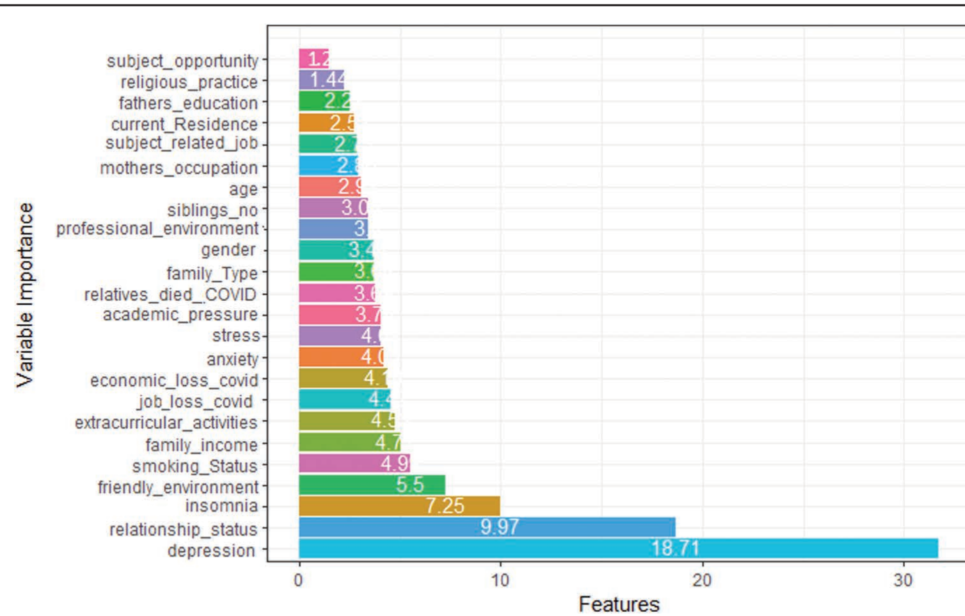

Figure 2. Ranking of the importance of the features according to the random forest model.

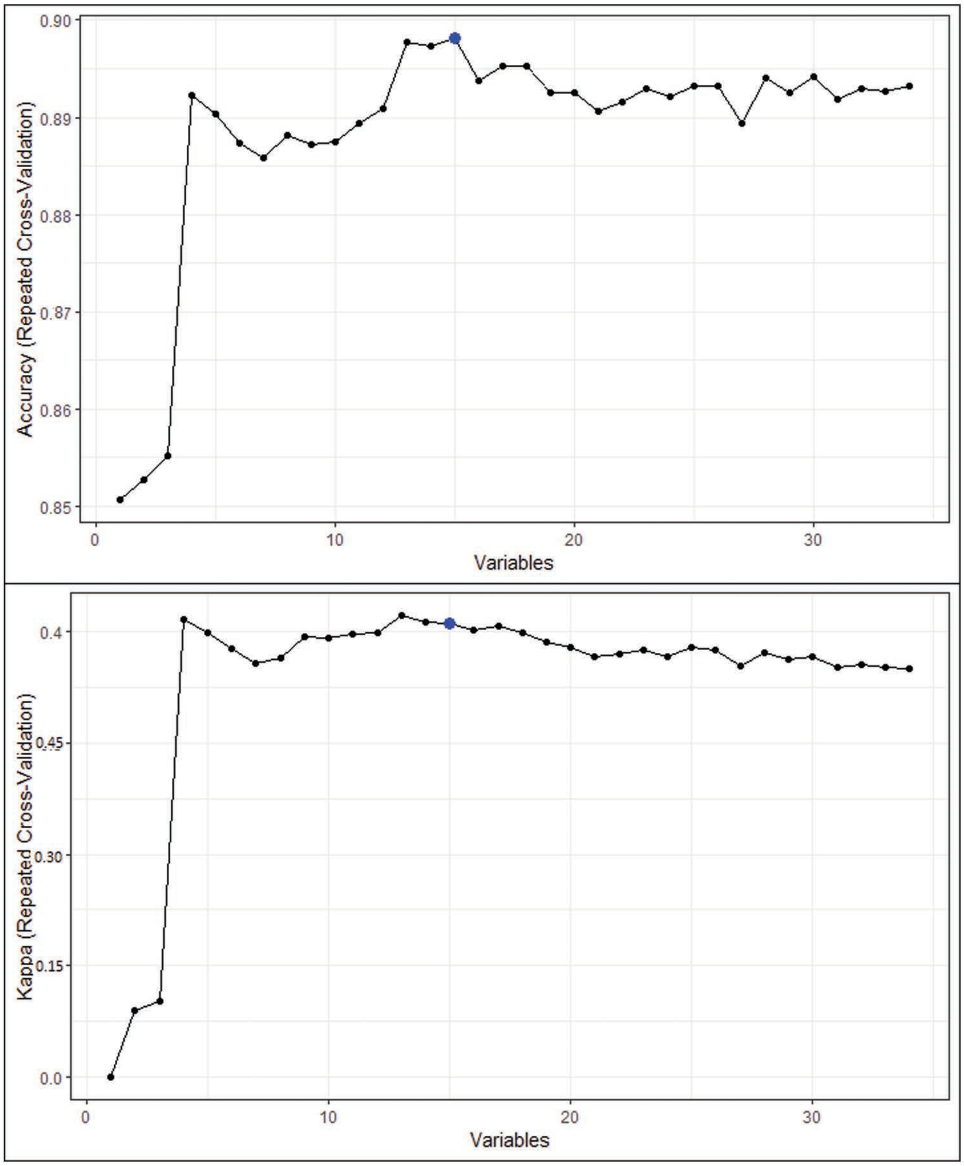

Figure 3. The selection of the optimal number of features in recursive feature elimination.

risks of suicide[58–60]; however, other studies failed to show any substantial contribution of academic pressure in suicidal risk prediction.[61]

The impact of COVID-19, including financial damage, loss of jobs of family members due to COVID-19, and the death of relatives or acquaintances from COVID-19, were critical predictive determinants of the risk of suicidal behaviors among university students in Bangladesh. These findings are consistent with those of studies conducted in Bangladesh among the general population during the COVID-19 pandemic[19] and in other countries among children and adolescents.[62,63] Among the behavioral and health-related features, extracurricular activities (sports, student government, community service, employment, arts, hobbies, and educational clubs) and smoking cigarettes were significant predictors. Students with unhealthy lifestyles with little or no physical exercise and cigarette smoking were identified as having high-risk behaviors in suicide cases in Bangladesh during the pandemic.[20]

We tested the features using the 6 machine-learning models mentioned earlier to assess their capability and performance in appropriately identifying the predictors of the potential risk of suicidal behavior/events. The performance evaluation and comparison of machine learning models showed that all 6 models behaved consistently and were comparable in predicting suicidal risk for the test and training datasets. The accuracy range was 0.76 to 0.79; ROC was 0.76 to 0.89; Kappa was 0.52 to 0.59; sensitivity was 0.76 to 0.81; specificity was 0.72 to 0.82; and the test data set, respectively (Table 4). However, SVM showed the highest and most consistent performance as opposed to the other 5 models in terms of all metrics, namely, accuracy (79%), kappa (0.59), ROC (0.89), sensitivity (0.81), and specificity (0.81). The predictive ability of machine learning algorithms highly depends on data quality and dataset strategy.[50,64–66] Several studies have shown that the prediction performance of suicide risk prediction models varies according to different populations.[67–69] Therefore, SVM can be used to develop the best predictive algorithm to identify at-risk university students with suicidal ideation/behaviors. Health professionals and university authorities can utilize this computerized system to identify people, specifically university students, at risk for suicidal events. Likewise, this predictive model can guide policymakers and university authorities in designing and implementing appropriate, timely, early intervention and suicide prevention programs.

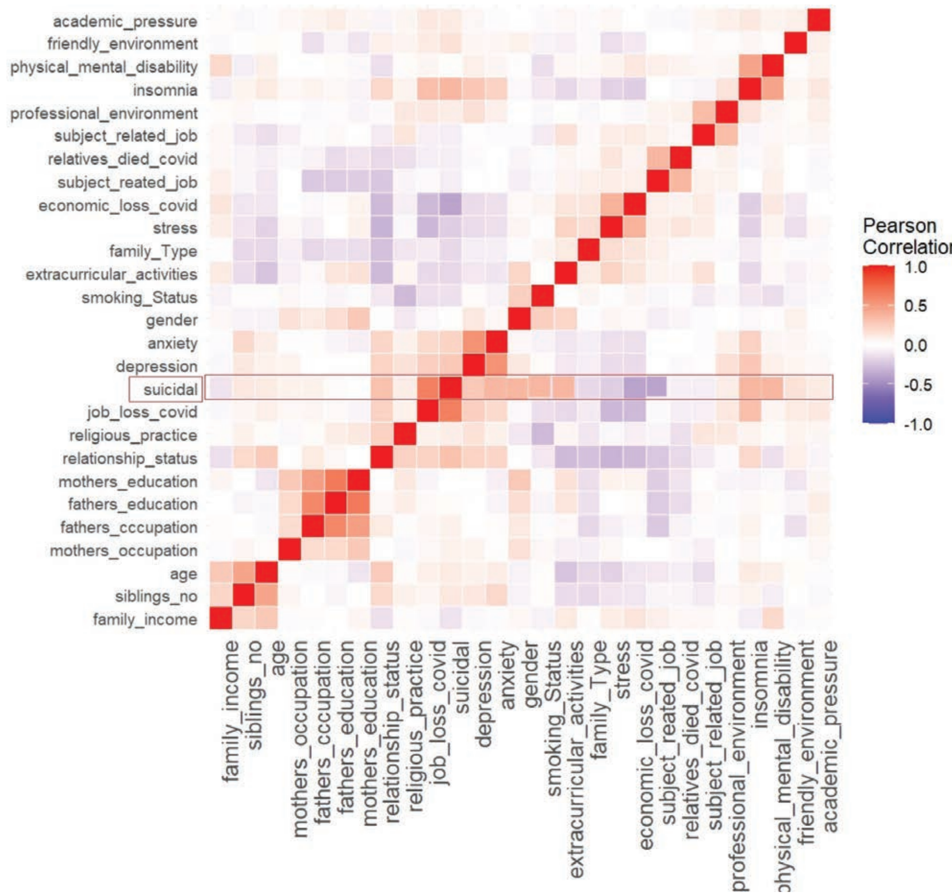

Figure 4. Heatmap of the Correlation Matrix for the suicidal ideation data set.

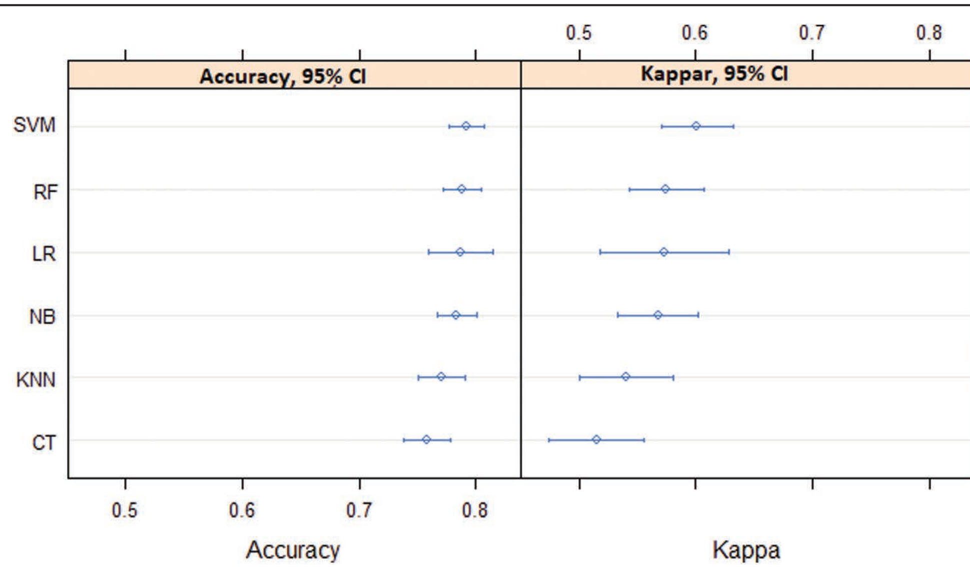

Figure 5. Model's accuracy and Kappa with 95% confidence interval from cross-validation. CI = confidence intervals, CT = classification tree, KNN = k-nearest neighbors algorithm, LR = logistic regression, NB = Naïve Bayes, RF = Random Forest, SVM = Support Vector Machine.

6. Strengths and limitations

The strength of our study lies in the comprehensive comparison of multiple machine learning models using 5 performance measures. Additionally, the use of online data collection through a self-administered questionnaire increased the likelihood of students sharing their mental health experiences.[16] Furthermore, this study represents the first attempt to predict suicidal behaviors and develop a diagnostic system for university students in Bangladesh using machine learning approaches. However, there are certain limitations to consider:

1. The study participants primarily consisted of Muslim and male university students. Therefore, caution should be exercised in generalizing the study findings to the broader population.

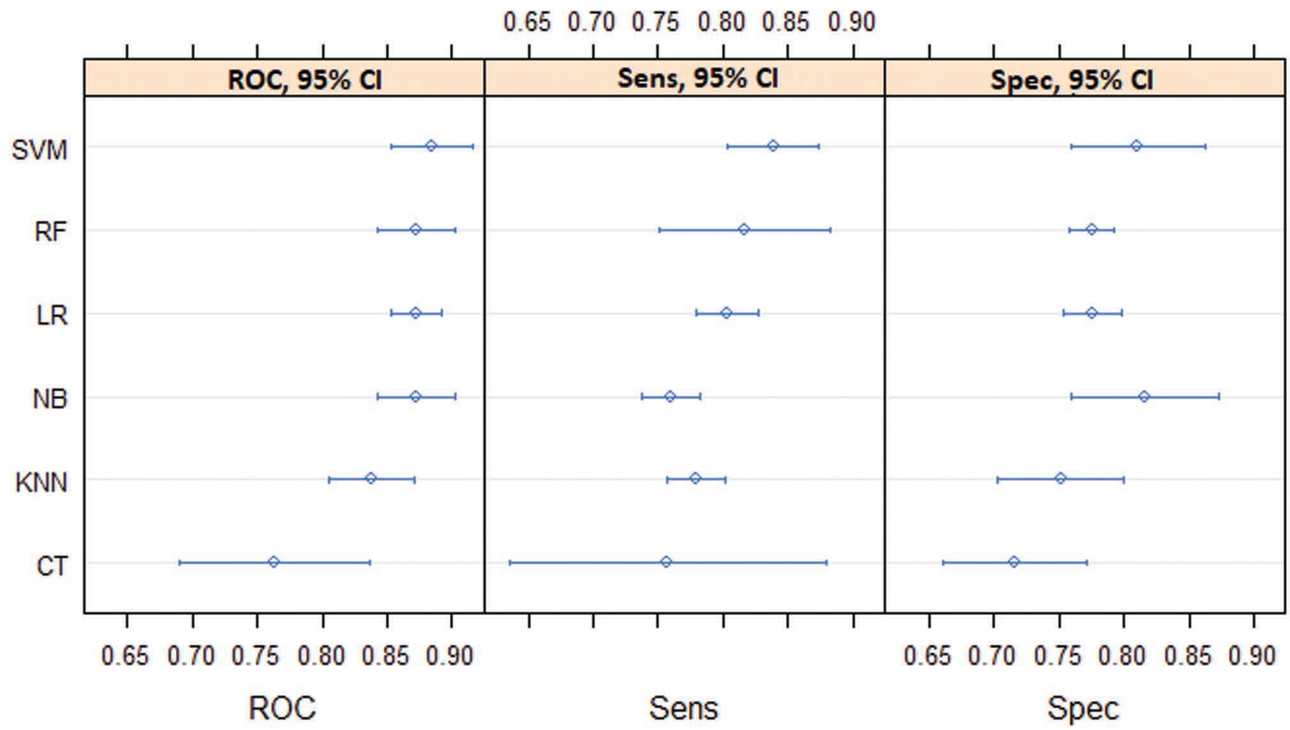

Figure 6. Model's ROC, Sensitivity, and Specificity with 95% confidence interval from cross-validation. CI = confidence intervals, CT = classification tree, KNN = k-nearest neighbors algorithm, LR = logistic regression, NB = Naïve Bayes, RF = Random Forest, ROC = receiver operating characteristic curve, SVM = Support Vector Machine.

Table 4

Comparative predictive performance of the machine learning model on the test data set.

	Matrics				
Model/classifier	**Accuracy**	**Kappa**	**ROC**	**Sensitivity**	**Specificity**
Logistic regression	0.78	0.57	0.87	0.80	0.76
Support Vector Machine	0.79	0.59	0.89	0.81	0.81
K-nearest neighbors algorithm	0.77	0.54	0.84	0.77	0.75
Naïve Bayes	0.78	0.57	0.87	0.76	0.82
Random Forest	0.79	0.58	0.87	0.81	0.77
Classification Tree	0.76	0.52	0.76	0.76	0.72

ROC = receiver operating characteristic curve.

2. Convenience sampling was employed in this study, which may introduce selection biases.
3. The data collection method relied on self-reported online surveys, which could potentially lead to information biases.

7. Conclusion

One significant intangible public health challenge is detecting the risk of suicidal ideation/behavior at the initial stage. In this study, we systematically designed a system that predicts suicidal behaviors among university students in Bangladesh. Six popular machine-learning approaches were studied and evaluated based on 5 metrics using cross-sectional data collected during the COVID-19 pandemic from university students in Bangladesh. SVM outperformed other Machine Learning Model approaches, making it the best machine learning model to predict suicidal risks more accurately among university students in Bangladesh. The authorities in universities and policymakers in charge of suicide prevention and treatment can utilize this algorithm to identify suicidal ideation among university students in Bangladesh, which can assist in preventing suicide incidents among university students.

Acknowledgments

We appreciate everyone who willingly participated in this survey.

Author contributions

Conceptualization: Sultan Mahmud, Md Mohsin.
Data curation: Sultan Mahmud, Abdul Muyeed.
Formal analysis: Sultan Mahmud, Tajrin Tahrin Tonmon, Ariful Islam.
Investigation: Sultan Mahmud, Md Mohsin.
Methodology: Sultan Mahmud, Md. Abu Sayed, Nabil Murshed.
Project administration: Sultan Mahmud.
Resources: Sultan Mahmud, Abdul Muyeed.
Software: Sultan Mahmud.
Supervision: Sultan Mahmud.

Validation: Sultan Mahmud, Md Mohsin.
Visualization: Sultan Mahmud, Md Mohsin, Abdul Muyeed, Shaila Nazneen, Md. Abu Sayed, Tajrin Tahrin Tonmon, Ariful Islam.
Writing – original draft: Sultan Mahmud, Md Mohsin, Abdul Muyeed, Shaila Nazneen, Md. Abu Sayed, Nabil Murshed, Tajrin Tahrin Tonmon, Ariful Islam.
Writing – review & editing: Sultan Mahmud, Md Mohsin, Abdul Muyeed, Shaila Nazneen, Md. Abu Sayed, Nabil Murshed, Tajrin Tahrin Tonmon, Ariful Islam.

References

[1] Organization WH. Coronavirus Disease (COVID-19); 2020. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19 [access date March 30, 2022].
[2] Mahmud S, Hossain S, Muyeed A, et al. The global prevalence of depression, anxiety, stress, and, insomnia and its changes among health professionals during COVID-19 pandemic: a rapid systematic review and meta-analysis. Heliyon. 2021;7:e07393.
[3] Brooks SK, Webster RK, Smith LE, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet. 2020;395:912–20.
[4] Hossain MM, Sultana A, Purohit N. Mental health outcomes of quarantine and isolation for infection prevention: a systematic umbrella review of the global evidence. Epidemiol Health. 2020;42:e2020038.
[5] Mahmud S, Mohsin M, Dewan M, et al. The global prevalence of depression, anxiety, stress, and insomnia among general population during COVID-19 pandemic: a systematic review and meta-analysis. Trends Psychol. 2023;31:143–70.
[6] Bruffaerts R, Demyttenaere K, Hwang I, et al. Treatment of suicidal people around the world. Br J Psychiatry. 2011;199:64–70.
[7] World Health Organization. Preventing suicide: a global imperative. World Health Organization. 2014. Available at: https://apps.who.int/iris/handle/10665/131056.
[8] Banna H. Minimising the economic impact of Coronavirus in Bangladesh. Business Standard. 2020. Available at: https://www.tbsnews.net/thoughts/minimising-economic-impact-coronavirus-bangladesh-56449 [access date March 30, 2022].
[9] Sultana M, Khan AH, Hossain S, et al. The association between financial hardship and mental health difficulties among adult wage earners during the COVID-19 pandemic in Bangladesh: findings from a cross-sectional analysis. Front Psychiatry. 2021;1364.
[10] Mohsin M, Mahmud S, Mian AU, et al. Side Effects of COVID-19 Vaccines and perceptions about COVID-19 and its vaccines in Bangladesh: a cross-sectional study. Vaccine X. 2022;12:100207.
[11] Tribune D. 70% more suicides than COVID-19 deaths in the last year. Dhaka Tribune. 2021. Available at: https://www.dhakatribune.com/bangladesh/2021/03/13/70-more-suicides-than-covid-19-deaths-in-the-lastyear [access date March 30, 2022].
[12] Mamun MA, Rayhan I, Akter K, et al. Prevalence and predisposing factors of suicidal ideation among the university students in Bangladesh: a single-site survey. Int J Ment Health Addict. 2020;1:14.
[13] Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. Annu Rev Clin Psychol. 2016;12:307–30.
[14] Cheng Q, Li TM, Kwok C-L, et al. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. J Med Internet Res. 2017;19:e243.
[15] Ream GL. The Interpersonal–psychological theory of suicide in college student suicide screening. Suicide Life Threat Behav. 2016;46:239–47.
[16] Macalli M, Navarro M, Orri M, et al. A machine learning approach for predicting suicidal thoughts and behaviours among college students. Sci Rep. 2021;11:1–8.
[17] Mamun MA, Sakib N, Gozal D, et al. The COVID-19 pandemic and serious psychological consequences in Bangladesh: a population-based nationwide study. J Affect Disord. 2021;279:462–72.
[18] Mamun MA, Akter T, Zohra F, et al. Prevalence and risk factors of COVID-19 suicidal behavior in Bangladeshi population: are healthcare professionals at greater risk? Heliyon. 2020;6:e05259.
[19] Rahman ME, Al Zubayer A, Bhuiyan MRAM, et al. Suicidal behaviors and suicide risk among Bangladeshi people during the COVID-19 pandemic: an online cross-sectional survey. Heliyon. 2021;7:e05937.
[20] Tasnim R, Islam MS, Sujan MSH, et al. Suicidal ideation among Bangladeshi university students early during the COVID-19 pandemic: prevalence estimates and correlates. Children Youth Serv Rev. 2020;119:105703.
[21] Khan M, Ali M, Rahman M, et al. Suicidal behavior among school-going adolescents in Bangladesh: findings of the global school-based student health survey. Soc Psychiatry Psychiatr Epidemiol. 2020;55:1491–502.
[22] Bzdok D, Varoquaux G, Steyerberg EW. Prediction, not association, paves the road to precision medicine. JAMA Psychiatry. 2021;78:127–8.
[23] Ryan EP, Oquendo MA. Suicide risk assessment and prevention: challenges and opportunities. Focus. 2020;18:88–99.
[24] Shen Y, Zhang W, Chan BSM, et al. Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm. J Affect Disord. 2020;273:18–23.
[25] Rezapour M, Hansen L. A machine learning analysis of COVID-19 mental health data. Sci Rep. 2022;12:14965.
[26] Chen Q, Zhang-James Y, 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.
[27] Kusuma K, Larsen M, Quiroz JC, et al. The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: a meta-analysis and systematic review. J Psychiatr Res. 2022;155:579–88.
[28] Rahman ME, Saiful Islam M, Mamun MA, et al. Prevalence and factors associated with suicidal ideation among university students in Bangladesh. Arch Suicide Res. 2022;26:975–84.
[29] Adnan AM. Universities in Bangladesh – a fun project. 2016. Available at: https://rstudio-pubs-static.s3.amazonaws.com/169329_95ad57e86a74439d8a2e18b0636ac2de.html [access date March 30, 2022].
[30] Chu C, Hsu A-L, Chou K-H, et al. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage. 2012;60:59–70.
[31] Vabalas A, Gowen E, Poliakoff E, et al. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14:e0224365.
[32] Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the beck depression and anxiety inventories. Behav Res Ther. 1995;33:335–43.
[33] Sadiq MS, Morshed NM, Rahman W, et al. Depression, anxiety, stress among postgraduate medical residents: a cross sectional observation in Bangladesh. Iran J Psychiatr. 2019;14:192.
[34] Alim SAHM, Kibria SME, Uddin MZ, et al. Translation of DASS 21 into Bangla and validation among medical students. Bangladesh J Psychiatr. 2014;28:67–70.
[35] Mamun MA, Hossain MS, Griffiths MD. Mental health problems and associated predictors among Bangladeshi students. Int J Ment Health Addict. 2022;20:657–71..
[36] Thakral M, Von Korff M, McCurry SM, et al. ISI-3: evaluation of a brief screening tool for insomnia. Sleep Med. 2021;82:104–9.
[37] Mamun MA, Alimoradi Z, Gozal D, et al. Validating insomnia severity index (ISI) in a Bangladeshi population: using classical test theory and Rasch analysis. Int J Environ Res Public Health. 2021;19:225.
[38] Aloba O, Ojeleye O, Aloba T. The psychometric characteristics of the 4-item Suicidal Behaviors Questionnaire-Revised (SBQ-R) as a screening tool in a non-clinical sample of Nigerian university students. Asian J Psychiatr. 2017;26:46–51.
[39] Osman A, Bagge CL, Gutierrez PM, et al. The Suicidal Behaviors Questionnaire-Revised (SBQ-R): validation with clinical and nonclinical samples. Assessment. 2001;8:443–54.
[40] Amini-Tehrani M, Nasiri M, Jalali T, et al. Validation and psychometric properties of suicide behaviors questionnaire-revised (SBQ-R) in Iran. Asian J Psychiatr. 2020;47:101856.
[41] Hasan MK, Alam MA, Das D, et al. Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access; 2020;8:76516–31.
[42] Raju MA, Mia MS, Sayed MA, et al. Predicting the outcome of English premier league matches using machine learning. IEEE; 2020:1–6.
[43] Senan EM, Al-Adhaileh MH, Alsaade FW, et al. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. J Healthc Eng. 2021;2021:1–10.
[44] Sanchez-Pinto LN, Venable LR, Fahrenbach J, et al. Comparison of variable selection methods for clinical predictive modeling. Int J Med Inform. 2018;116:10–7.
[45] Zhou Y, Zhang R, Wang S, et al. Feature selection method based on high-resolution remote sensing images and the effect of sensitive features on classification accuracy. Sensors. 2018;18:2013.
[46] van Mens K, de Schepper C, Wijnen B, et al. Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study. J Affect Disord. 2020;271:169–77.

[47] Zeng X, Martinez TR. Distribution-balanced stratified cross-validation for accuracy estimation. J Exp Theor Artif Intell. 2000;12:1–12.
[48] Sufriyana H, Husnayain A, Chen Y-L, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: systematic review and meta-analysis. JMIR Med Inform. 2020;8:e16503.
[49] Lloyd-Williams M. Case studies in the data mining approach to health information analysis. 1998.
[50] Mahmud S, Islam MA. Predictive ability of covariate-dependent markov models and classification tree for analyzing rainfall data in Bangladesh. Theor Appl Climatol. 2019;138:335–46.
[51] Ben-David A. About the relationship between ROC curves and Cohen's kappa. Eng Appl Artif Intell. 2008;21:874–82.
[52] Ryu S, Lee H, Lee D-K, et al. Use of a machine learning algorithm to predict individuals with suicide ideation in the general population. Psychiatr Investig. 2018;15:1030–6.
[53] Rasheduzzaman M, Al-Mamun F, Hosen I, et al. Suicidal behaviors among Bangladeshi university students: prevalence and risk factors. PLoS One. 2022;17:e0262006.
[54] Žilinskas E, Žulpaitė G, Puteikis K, et al. Mental health among higher education students during the COVID-19 pandemic: a cross-sectional survey from Lithuania. Int J Environ Res Public Health. 2021;18:12737.
[55] Oner S, Yenilmez C, Ozdamar K. Sex-related differences in methods of and reasons for suicide in Turkey between 1990 and 2010. J Int Med Res. 2015;43:483–93.
[56] Oner S, Yenilmez C, Ayranci U, et al. Sexual differences in the completed suicides in Turkey. Eur Psychiatry. 2007;22:223–8.
[57] Gouda MN, Rao SM. Factors related to attempted suicide in Davanagere. Indian J Community Med. 2008;33:15.
[58] Song J, Song TM, Seo D-C, et al. Data mining of web-based documents on social networking sites that included suicide-related words among Korean adolescents. J Adolesc Health. 2016;59:668–73.
[59] Zeng K, Le Tendre G. Adolescent suicide and academic competition in East Asia. Comp Educ Rev. 1998;42:513–28.
[60] Mueller AS. Does the media matter to suicide?: examining the social dynamics surrounding media reporting on suicide in a suicide-prone community. Soc Sci Med. 2017;180:152–9.
[61] Sun RC, Hui EK. Psychosocial factors contributing to adolescent suicidal ideation. J Youth Adolesc. 2007;36:775–86.
[62] Thompson EC, Thomas SA, Burke TA, et al. Suicidal thoughts and behaviors in psychiatrically hospitalized adolescents pre-and post-COVID-19: a historical chart review and examination of contextual correlates. J Affect Disord Rep. 2021;4:100100.
[63] Golberstein E, Wen H, Miller BF. Coronavirus disease 2019 (COVID-19) and mental health for children and adolescents. JAMA Pediatr. 2020;174:819–20.
[64] Ly H-B, Pham BT, Le LM, et al. Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models. Neural Comput Appl. 2021;33:3437–58.
[65] Asteris PG, Apostolopoulou M, Armaghani D, et al. On the metaheuristic models for the prediction of cement-metakaolin mortars compressive strength. Techno Press Services. 2020;1:63.
[66] Mahmud S, Sumana FM, Mohsin M, et al. Redefining homogeneous climate regions in Bangladesh using multivariate clustering approaches. Nat Hazards. 2022;111:1863–84.
[67] Schafer KM, Kennedy G, Gallyer A, et al. A direct comparison of theory-driven and machine learning prediction of suicide: a meta-analysis. PLoS One. 2021;16:e0249833.
[68] 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.
[69] Jacobucci R, Littlefield AK, Millner AJ, et al. Evidence of inflated prediction performance: a commentary on machine learning and suicide research. Clini Psychol Sci. 2021;9:129–34.