Received: 21 August 2023 | Revised: 21 February 2024 | Accepted: 26 March 2024
DOI: 10.1002/hsr2.2037

ORIGINAL RESEARCH

Health Science Reports Open Access WILEY

Predicting anxiety, depression, and insomnia among Bangladeshi university students using tree-based machine learning models

Arman Hossain Chowdhury[1] | Dana Rad[2] | Md. Siddikur Rahman[1]

[1]Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh

[2]Center of Research Development and Innovation in Psychology, Aurel Vlaicu University of Arad, Arad, Romania

Correspondence
Md. Siddikur Rahman, Department of Statistics, Begum Rokeya University, Rangpur, Rangpur 5404, Bangladesh.
Email: siddikur@brur.ac.bd

Abstract

Background and Aims: Mental health problem is a rising public health concern. People of all ages, specially Bangladeshi university students, are more affected by this burden. Thus, the objective of the study was to use tree-based machine learning (ML) models to identify major risk factors and predict anxiety, depression, and insomnia in university students.

Methods: A social media-based cross-sectional survey was employed for data collection. We used Generalized Anxiety Disorder (GAD-7), Patient Health Questionnaire (PHQ-9) and Insomnia Severity Index (ISI-7) scale for measuring students' anxiety, depression and insomnia problems. The tree-based supervised decision tree (DT), random forest (RF) and robust eXtreme Gradient Boosting (XGBoost) ML algorithms were used to build the prediction models and their predictive performance was evaluated using confusion matrix and receiver operating characteristic (ROC) curves.

Results: Of the 1250 students surveyed, 64.7% were male and 35.3% were female. The students' ages ranged from 18 to 26 years old, with an average age of 22.24 years (SD = 1.30). Majority of the students (72.6%) were from rural areas and social media addicted (56.6%). Almost 83.3% of the students had moderate to severe anxiety, 84.7% had moderate to severe depression and 76.5% had moderate to severe insomnia problems. Students' social media addiction, age, academic performance, smoking status, monthly family income and morningness-eveningness are the main risk factors of anxiety, depression and insomnia. The highest predictive performance was observed from the XGBoost model for anxiety, depression and insomnia.

Conclusion: The study findings offer valuable insights for stakeholders, families and policymakers enabling a more profound comprehension of the pressing mental health disorders. This understanding can guide the formulation of improved policy strategies, initiatives for mental health promotion, and the development of effective

Arman Hossain Chowdhury and Md. Siddikur Rahman contributed equally to this study.

counseling services within university campus. Additionally, our proposed model might play a critical role in diagnosing and predicting mental health problems among Bangladeshi university students and similar settings.

KEYWORDS
mental health, morningness-eveningness, predictive models, risk factors, social-media addiction

1 | INTRODUCTION

The most important aspect of human health is mental well-being. Different mental disorders among university students are one of the serious rising public health threats particularly in Bangladesh.[1,2] Transitioning from their higher secondary school to university life, students may experience several challenges in their admission test and new academic environments, such as mess/hostel life, difficulties in study, poor time maintenance, poor diet, financial burden, complicated relationships and many more.[3–5] These challenges, pivotal for individual and professional development, frequently put students under emotional and mental strain which can exacerbate their levels of anxiety, depression, insomnia, and other mental health problems.[3]

The three most prevalent global mental health disorders are anxiety, depression and insomnia, which are the main determinants of disability and a significant portion of the total burden of diseases.[6] According to estimates, 5% of the adults worldwide experience depression,[7] while anxiety affects 4.6% of males and 2.6% of females.[8] In Bangladesh, the intensity of mental health problems among students has been extensively investigated and the findings highlight that these issues are common, with rates of depression and anxiety reaching as high as 54.3% and 64.8%, respectively.[9–12] On the other hand, insomnia has become a rising global health threat among university students, especially in South Asia. A previous study revealed that the pooled prevalence rate of insomnia among university students in South Asian countries was 52.1%.[13]

Exploring the sociodemographic and socioeconomic factors influencing students mental health is crucial. Age, gender, marital status, place of residence, educational level, monthly family income and addiction to social media platforms (e.g., Facebook, YouTube, Instagram, etc.) can significantly influence on their mental health.[14,15] While several statistical techniques[16–18] have been used in prior studies to examine these links,[19–22] the evaluation of insomnia among Bangladeshi university students remains unexplored. On the other hand, machine learning (ML) models are more robust and efficient for the prediction and classification problem.[23–27] The ML models performed admirably in many previous studies, such as dengue,[24,28–31] COVID-19[26,32–36] and many other diseases and achieved high accuracy. However, the application of ML models for predicting anxiety, depression and insomnia is scanty.[37,38] Predicting specific clinical outcomes using a variety of factors is the overall goal of ML in the field of healthcare.[39,40] Healthcare is one area where ML has enormous promise. Almost all clinical specialties have reported using ML-based technologies to diagnose and predict at or beyond the human level.[41] Therefore, this study was designed to find the risk factors and predict the anxiety, depression and insomnia among university students in Bangladesh using the tree-based ML models. Our study findings will help the decision makers and communities better understand and avert this grave danger through improved decision-making techniques and community settings.

2 | MATERIALS AND METHODS

2.1 | Study participants and data collection

We conducted a cross-sectional study using social media using convenience sampling among Bangladeshi university students from November to December 2022. The required sample size for the analysis was computed using Cochran's formula[38] after considering the acceptable margin of error, $d = 0.02$ with 95% confidence interval and $p = 0.846$ (since 84.6% response rate in our pilot study). The mathematical expression of Cochran's formula is as follows:

$$n = \frac{z_{\alpha/2}^{2} p(1 - p)}{d^{2}} \tag{1}$$

Therefore, the required size of the sample was found, $n = 1251.24 \approx 1250$. As a result, 1250 participants' data were gathered from 24 universities comprising 20 public and 4 private institutions (Supporting Information 1: Table 1) utilizing a well-designed Google form.[42] All students gave their informed permission before participating in the study. The questionnaire used in the study was interpreted into the local language, “Bangla,” by the researcher, reviewed and checked its validity for cultural relevance. The questionnaire consisted of a total of 43 items, which split into six sections: (i) sociodemographics, (ii) Bergen Social Media Addiction Scale (BSMAS), (iii) Morningness-eveningness, (iv) Generalized Anxiety Disorder-7 (GAD-7), (v) Patient Health Questionnaire-9 (PHQ-9) and (vi) Insomnia Severity Index-7 (ISI-7) (Supporting Information 2). A pilot study is a small-sized preliminary inquiry carried out before the major study project. It is used to evaluate and improve the protocols, techniques for collecting data, and any required instruments, such as questionnaires or surveys.[43] The questionnaire was pretested through a pilot study with 50 participants and the validity was measured by Cronbach's α value. The pilot study confirmed that the pooled

Cronbach's α value of the sociodemographic, social media addiction, morningness-eveningness, anxiety, depression and insomnia was 0.901 (Supporting Information 1: Table S2).

2.2 | Sociodemographic information

The sociodemographic data collected contains the student's age, sex, education level, place of residence, birthplace, monthly family income, academic performance, smoking status, and marital status. The education level was split into two categories: undergraduate and graduate. The place of residence was categorized as campus and on campus. Birthplace was divided into rural and urban categories. This study labeled the families as upper class, with a monthly income of 20,000 BDT or more. Middle-class families were those having monthly income ranges between 10,000–20,000 BDT and monthly income less than 10,000 BDT were considered as lower class. We labeled the students' academic performance as good; those with a CGPA greater than or equal to 3.50 and the CGPA of less than 3.5 as poor. Smokers and non-smokers categorized smoking status. Marital status was classified as married and unmarried. The acceptable Cronbach's α was found to be 0.725.

2.3 | BSMAS

This study applied the BSMAS to evaluate the experiences with social media addiction among university students over the past year.[44] Six questions on the BSMAS highlight the fundamental aspects of addiction, including salience, mood-altering behaviors, endurance, abstinence, disagreement, and recurrence. Each question asks about situations that occurred over the previous 12 months, and responses are graded on a 5-point Likert scale of 1, indicating very rarely and 5 showing very often, with scores hovering from 6 to 30. A higher score indicates the chance of being addicted to social media. A cutoff BSMAS score of 19 or greater indicates the likelihood of developing social media addiction.[45] Earlier studies revealed that the BSMAS has demonstrated strong validity and reliability[46,47] and in our research, we found that Cronbach's α was excellent (0.906).

2.4 | Morningness-eveningness

The term "morningness-eveningness" describes the unique variations in nocturnal preferences, activity-related sleep-wake patterns, and attentiveness in the morning and evening.[48] The morningness-eveningness of the university students was evaluated using the Spanish version of the reduced scale of the Morningness-Eveningness Questionnaire (rMEQ).[49] The Spanish rMEQ is a trustworthy scale with great sensitivity when categorizing subjects. It has five attributes, including daily physical activity and preference for being awake or asleep. The direct cutoff score is used to determine the type of students, with values of 4–11 designating an evening type (ET), 12–17 a neutral type (NT), and 18–25 a morning type (MT). The acceptable Cronbach's α was found to be 0.752.

2.5 | GAD-7

GAD-7 consists of seven questions that are used to measure the intensity of anxiety among people. Each question contains four criteria that are scored as 0, 1, 2, and 3, respectively: "not at all," "several days," "more than half the days," and "nearly every day."[50,51] Since each of the 7 items is graded from 0 to 3, the GAD-7 scale score hovers from 0 to 21. Based on scores of 0–4, 5–9, 10–14, and 15–21, the corresponding anxiety level was minimal, mild, moderate, and severe. This study used a cutoff score ≥ 10 as the presence of anxiety.[51] The measure's validity was determined using Cronbach's α value and was found to be excellent (0.956). The GAD-7 is a well-established questionnaire for assessing anxiety problems and prior studies conducted in Bangladesh have yielded satisfactory internal consistency for anxiety.[52,53]

2.6 | PHQ-9

The PHQ-9 is a scalable instrument for evaluating the presence of depression.[54] It is a 9-item depression questionnaire with four criteria for each question, "not at all," "several days," "more than half the days," and "nearly every day," each of which is scored as 0, 1, 2, and 3, respectively.[55] Since each of the 9 items is graded from 0 to 3, the PHQ-9 scale score hovers from 0 to 27. A required cut-off point of 10 and above represents the presence of depression.[55] Based on the scores of 0–4, 5–9, 10–14, 15–19, and 20–27, the corresponding level of depression was minimal, mild, moderate, moderately severe, and severe. In our study, we used a cutoff score ≥ 10 as the presence of depression among the students[56] and Cronbach's α was found to be excellent (0.943). The PHQ-9 is a widely used tool for evaluating depression problems, and previous studies conducted in Bangladesh have yielded acceptable internal consistency results for depression.[52,57]

2.7 | ISI-7

The ISI-7 is made up of seven items that assess (a) the intensity of falling asleep, (b) staying asleep, (c) waking up early in the morning (terminal), (d) contentment with the present sleep pattern, (e) disruption of day-to-day activities, (f) observable impairment attributable to the sleep problem, and (g) degree of discomfort brought on by the sleep problem by the sleep problem. Each question is evaluated on a 5-point Likert scale.[58,59] Since each of the 9 items is a 5-point Likert scale scored from 0 to 4, the ISI-7 scale score ranges from 0 to 28. A required cut-off point of 8 and above represents the presence of insomnia. The categorized levels of insomnia are No clinically significant Insomnia, Subthreshold insomnia, Clinical insomnia (moderate severity), and Clinical insomnia (severe) based on the scores 0–7, 8–14,15–21, and 22–28, respectively.[60] This study used a cutoff score of >7 as the presence of insomnia among the students.[61] The Cronbach's α values indicate that the questionnaire is reliable (0.959).

2.8 | Model selection and parameter tuning

The eXtreme Gradient Boosting (XGBoost) is a data-driven ML technique commonly used in regression and classification problems. It is an optimized implementation of the gradient boosting (GB) technique, which uses decision trees (DT). It was first established by Chen Tianqi and Carlos Gestrin.[62] The Distributed ML Community (DMLC) includes XGboost as a package. The GB uses additive modeling that is stage-wise.[63] The XGBoost algorithm uses splits, that is, follows the iterative process of selecting features that classify the data best into two groups. The model can classify data and achieve high accuracy without requiring preprocessing and feature engineering. ML models can be trained for robust performance using various training and testing methods, including hold-out, k-fold cross-validation (cv), and so on.[64] In a k-fold cv technique, the sample set is split into several partitions (splits) equal in size at random, where k represents the number of folds. For 10-fold cv, the data sets are divided into 10 samples. In this case, nine samples are utilized as training data, while the final sample is employed as test data. The results are then summarized by their average.[65] To remove overfitting or underfitting issues and to assess how well the model generalizes to the data, this study employed the robust k-fold cv technique. We applied a 10-fold cv procedure for the training data sets to accomplish this process. The predictive accuracy of ML models can be improved by tuning their several hyperparameters. Regarding statistics, hyperparameter tuning provides an overview of how a model performs and contrasts it with earlier pictures.[66] The tree-based XGBoost model contains several hyperparameters such as min_child_weight, max_depth, eta, gamma, etc. The performance of the model can be improved by tuning these parameters. Most of these parameters are specific to ML and rely on its nature, while others regulate how well the boosting algorithm performs.[67] Gamma is essential in keeping the model from overfitting and handling the cost function's regularization portion. In contrast, eta holds how quickly the model picks up on the patterns in the data. To manage the L2 regularizations on weights, lambda is adjusted throughout the cv. Another hyperparameter, min child weight, is crucial for preventing any possible feature interactions that might lead to overfitting.[67,68] In this study, we applied the grid search method for hyperparameter optimization and calculated the classification accuracy of the models. The details of the ML model building are shown in Supplement (Supporting Information 1: Tables S3-S7).

2.9 | Evaluation metrics

We assess the effectiveness of the suggested approach by using assessment criteria including classification accuracy, sensitivity measure, specificity, recall, precision and area under the receiver operating characteristics AUC (ROC) curves. The proportion of individuals correctly identified overall is known as accuracy. Sensitivity is the percentage of individuals who test positive and have the disease.[38] The mathematical formulation of these metrics is as follows:

$$\text{Accuracy} = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \tag{2}$$

$$\text{Sensitivity} = \text{Recall} = \frac{\sum TP}{\sum TP + \sum FN} \tag{3}$$

$$\text{Specificity} = \frac{\sum TN}{\sum TN + \sum FP} \tag{4}$$

$$\text{Precision} = \frac{\sum TP}{\sum TP + \sum FP} \tag{5}$$

The ROC curve represents the True Positive Rate versus False Positive Rate in a plot at various categorization thresholds.

$$\text{TPR} = \frac{\sum TP}{\sum TP + \sum FN} \tag{6}$$

$$\text{FPR} = \frac{\sum FP}{\sum TN + \sum FP} \tag{7}$$

AUC represents the Area under the ROC curve. The AUC = 1 value indicates that the anomaly classifier is excellent compared to the AUC of 0.5, which means that the model's result is worse than a random guess.[69]

2.10 | Data analysis

After preprocessing and coding the explanatory features, the final data were split into training and test data. We used three tree-based ML models, including supervised DT model, random forest (RF), and the ensemble XGBoost for modeling the training data and the test data were applied to evaluate the predictive performance of the models. Finally, using the best model, the risk factors were identified (Figure 1). The performance was assessed using the model evaluation metrics, including classification accuracy, sensitivity measure, specificity, recall, precision, and the AUC (Area Under the ROC Curve). All the summary analyses were generated using the IBM SPSS Statistics (Version 25.0).[70] The tree-based ML models were developed using RStudio (Version 4.2.1).[71] The "ggplot2" package was used for graphical visualizations.

3 | RESULTS

3.1 | Assessment of anxiety

Among the 1250 participants, 64.7% of the students were male and 35.3% were female, with an average age of 22.24 years (SD = 1.30)

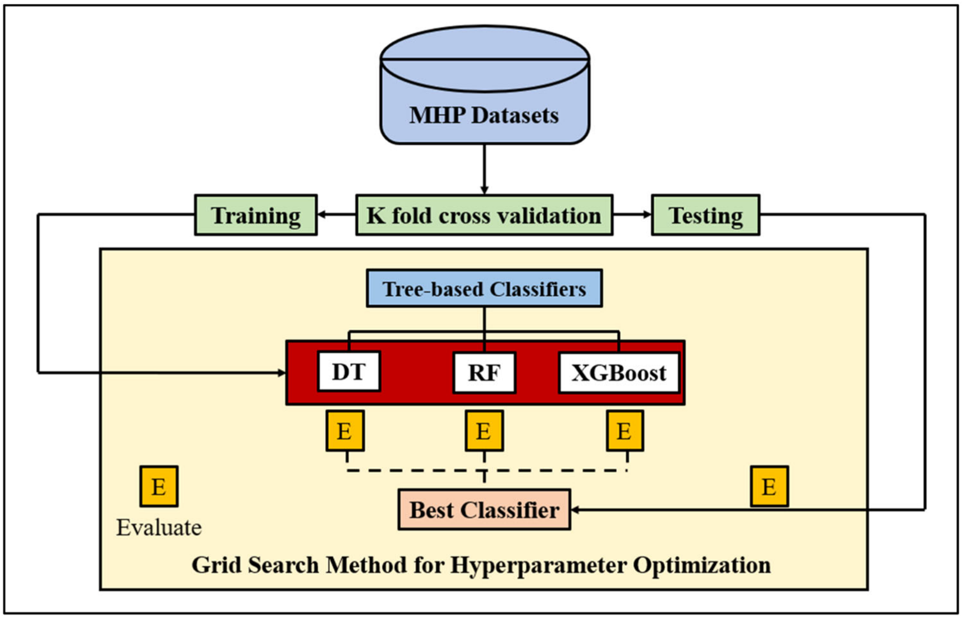

FIGURE 1 The framework of tree-based methods for predicting mental health problems. DT, decision tree; MHP, mental health problem; RF, random forest; XGBoost, ensemble eXtreme Gradient Boosting.

ranging from 18 to 26. The majority of the students were undergraduates (96.9%), from the lower class (59%), came from a rural site (72.6%) and more than half of them were social media addicted (56.6%) and MT (52.4%). The incidence rate of anxiety among the students was 83.3%, with no significant gender differences. Anxious students were significantly more likely to be undergraduate, unmarried, from off-campus, had lower family income, poor academic performance, nonsmoker, social media addicted and MT than nonanxious students (Table 1).

3.2 | Assessment of depression

Almost 84.7% of the students experienced moderate to severe depression with a significant gender difference. Depressed students were significantly more likely to be male, unmarried, from off campus, came from rural sites, had lower family income, poor academic performance, social media addiction and MT than nondepressed students (Table 2).

3.3 | Assessment of insomnia

The incidence of moderate to severe insomnia among university students was 76.5%, with significant gender differences. Students having mild to severe insomnia were significantly more likely to be male, undergraduates, unmarried, from off campus, came from rural sites, had lower family income, poor academic performance, nonsmoker, social media addicted, and MT than those students having no insomnia (Table 3).

3.4 | ML model interpretation

The three tree-based ML models (DT, RF, and XGBoost) were fitted and their performance was presented in Figure 2. The assessed performance revealed that the XGBoost model is more efficient compared to DT and RF at predicting anxiety disorder, depression disorder and insomnia among Bangladeshi university students. For example, the XGBoost provided 86% accurate prediction for anxiety (accuracy = 0.86) and 96% of the cases of positivity that were anticipated to be positive (sensitivity = 0.96), 84% accurate prediction for depression (accuracy = 0.84) and 93% of the cases of positivity that were anticipated to be positive (sensitivity = 0.93), as well as 81% of the accurate prediction of insomnia (accuracy = 0.81) and 87% of the cases of positivity that were anticipated to be positive (sensitivity = 0.87).

For predicting the depression, anxiety and insomnia among Bangladeshi university students during November to December 2022, the AUC measure of DT, RF and XGBoost models was estimated. The estimated AUC values were 0.79, 0.87, 0.94 (for anxiety), 0.86, 0.86, 0.88 (for depression) and 0.82, 0.86, 0.89 (for insomnia) respectively. Hence, the XGBoost model performed better than the DT and RF model in predicting anxiety, depression and insomnia (Figure 3).

Figure 4 reveals that using the XGBoost model as best, the 10 features among the 11 surveyed features were selected for anticipating depression, anxiety and insomnia among Bangladeshi university students. Students' social media addiction, age, academic performance and smoking status were the main determinants and marital status, education level, monthly family income and morningness-eveningness were the tentative features for predicting

TABLE 1 Frequency distribution of sociodemographic, social media addiction, morningness-eveningness, and their association with anxiety among 1250 university students in Bangladesh during November to December 2022.

Variables	Yes (n = 1042; 83.3%) n	% (95% CI)	No n	% (95% CI)	Total N (%)	χ^2	p Value
Sex							
Male	665	63.8 (60.9-66.7)	144	69.2 (62.7-75.2)	809 (64.7)	0.136	0.153
Female	377	36.2 (33.3-39.1)	64	30.8 (24.8-37.3)	441 (35.3)		
Education level							
Undergraduate	1018	97.7 (96.6-98.5)	193	92.8 (88.7-95.7)	1211 (96.9)	13.819	<0.01*
Graduate	24	2.3 (1.5-3.4)	15	7.2 (4.3-11.3)	39 (3.1)		
Marital status							
Married	476	45.7 (42.7-48.7)	9	4.3 (2.2-7.7)	485 (38.8)	124.877	<0.01*
Unmarried	566	54.3 (51.3-57.3)	199	95.7 (92.3-97.8)	765 (61.2)		
Place of residence							
Off-campus	956	91.7 (90.0-93.3)	171	82.2 (76.6-86.9)	1127 (90.2)	17.769	<0.01*
On campus	86	8.3 (6.7-10.0)	37	17.8 (13.1-23.4)	123 (9.8)		
Birthplace							
Rural	758	72.7 (70.0-75.4)	150	72.1 (65.7-77.9)	908 (72.6)	0.035	0.865
Urban	284	27.3 (24.6-30.0)	58	27.9 (22.1-34.3)	342 (27.4)		
Monthly family Income (BDT)							
Lower	658	63.1 (60.2-66.0)	80	38.5 (32.0-45.2)	738 (59)	45.291	
Middle	263	25.2 (22.7-27.9)	81	38.9 (32.5-45.7)	344 (27.5)		<0.01*
Upper	121	11.6 (9.8-13.7)	47	22.6 (17.3-28.6)	168 (13.4)		
Academic performance							
Poor	598	57.4 (54.4-60.4)	20	9.6 (6.2-14.2)	618 (49.4)	158.315	<0.01*
Good	444	42.6 (39.6-45.6)	188	90.4 (85.8-93.8)	632 (50.6)		
Smoking status							
No	1019	97.8 (96.8-98.6)	172	82.7 (77.1-87.4)	1191 (95.3)	87.913	<0.01*
Yes	23	2.2 (1.4-3.2)	36	17.3 (12.6-22.9)	59 (4.7)		
Social media addiction							
No	367	35.2 (32.4-38.2)	176	84.6 (79.2-89.0)	543 (43.4)	172.180	<0.01*
Yes	675	64.8 (61.8-67.6)	32	15.4 (11.0-20.8)	707 (56.6)		
Morningness-eveningness							
Evening type	260	25.0 (22.4-27.6)	26	12.5 (8.5-17.5)	286 (22.9)	51.867	
Neutral	218	20.9 (18.5-23.5)	91	43.8 (37.1-50.5)	309 (24.7)		<0.01*
Morning type	564	54.1 (51.1-57.1)	91	43.8 (37.1-50.5)	655 (52.4)		

Note: Asterisk (*) indicates the statistically significant at 0.05 level.
Abbreviations: BDT, Bangladeshi Taka; χ^2, Chi-square.

anxiety. On the other hand, social media addiction, age, monthly family income, morningness-eveningness and academic performance were the main determinants and marital status, sex, education level, birthplace and place of residence were the tentative features of predicting depression, and insomnia among university students in Bangladesh (Figure 4).

4 | DISCUSSION

Mental health illness is a severe global health burden and university students in Bangladesh are primarily vulnerable to this disease. Students' health and academic activities are severely disrupted due to this burden.[72–74] Therefore, mental health problems among

TABLE 2 Frequency distribution of socio-demographic, social media addiction, morningness-eveningness and their association with depression among 1250 university students in Bangladesh during November to December 2022.

Variables	Yes (n = 1059; 84.7%) *n*	% (95% CI)	No *n*	% (95% CI)	Total N (%)	χ^2	*p* Value
Sex							
Male	699	66.0 (63.1-68.8)	110	57.6 (50.5-64.4)	809 (64.7)	5.017	<0.01*
Female	360	34.0 (31.2-36.9)	81	42.4 (35.6-49.5)	441 (35.3)		
Education level							
Undergraduate	1030	97.3 (96.1-98.1)	181	94.8 (90.9-97.3)	1211 (96.9)	3.338	0.07
Graduate	29	2.7 (1.9-3.9)	10	5.2 (2.7-9.1)	39 (3.1)		
Marital status							
Married	480	45.3 (42.3-48.3)	5	2.6 (1.0-5.6)	485 (38.8)	124.295	<0.01*
Unmarried	579	54.7 (51.7-57.7)	186	97.4 (94.4-99.0)	765 (61.2)		
Place of residence							
Off campus	968	91.4 (89.6-93.0)	159	83.2 (77.5-88.0)	1127 (90.2)	12.148	<0.01*
On campus	91	8.6 (7.0-10.4)	32	16.8 (12.0-22.5)	123 (9.8)		
Birthplace							
Rural	783	73.9 (71.2-76.5)	125	65.4 (58.5-71.9)	908 (72.6)	5.872	<0.01*
Urban	276	26.1 (23.5-28.8)	66	34.6 (28.1-41.5)	342 (27.4)		
Monthly family Income (BDT)							
Lower	680	64.2 (61.3-67.1)	58	30.4 (24.2-37.1)	738 (59)	82.643	<0.01*
Middle	264	24.4 (22.4-27.6)	80	41.9 (35.1-49.0)	344 (27.5)		
Upper	115	10.9 (9.1-12.8)	53	27.7 (21.8-34.4)	168 (13.4)		
Academic performance							
Poor	602	56.8 (53.8-59.8)	16	8.4 (5.1-12.9)	618 (49.4)	152.077	<0.01*
Good	457	43.2 (40.2-46.2)	175	91.6 (87.1-94.9)	632 (50.6)		
Smoking status							
No	1010	95.4 (94.0-96.5)	181	94.8 (90.9-97.3)	1191 (95.3)	0.133	0.711
Yes	49	4.6 (3.5-6.0)	10	5.2 (2.7-9.1)	59 (4.7)		
Social media addiction							
No	365	34.5 (31.6-37.4)	178	93.2 (89.0-96.1)	543 (43.4)	227.143	<0.01*
Yes	694	65.5 (62.6-68.4)	13	6.8 (3.9-11.0)	707 (56.6)		
Morningness-eveningness							
Evening type	260	24.6 (22.0-27.2)	26	13.6 (9.3-19.0)	286 (22.9)	36.058	
Neutral	230	21.7 (19.3-24.3)	79	41.4 (34.6-48.4)	309 (24.7)		<0.01*
Morning type	569	53.7 (50.7-56.7)	86	45.0 (38.1-52.1)	655 (52.4)		

Note: Asterisk (*) indicates the statistically significant at 0.05 level.
Abbreviations: BDT, Bangladeshi Taka; χ^2, Chi-square.

university students have become a severe concern to policymakers and university campus health professionals across the globe.[75–77] That's why this study was designed to predict depression, anxiety, and insomnia among Bangladeshi university students.

This study found the highest prevalence rate of anxiety (83.3%), depression (84.7%) and insomnia (76.5%) among university students compared to the earlier surveyed studies conducted in Bangladesh.[9–11] One major determinant behind these high incidences of anxiety, depression and insomnia can be attributed to social media addiction (e.g., Facebook, YouTube, Instagram, etc.), as indicated in our study, since increased social media usage may lead to sleep problems, lack of exercise, and peer pressure. This findings aligns with some previous studies.[78,79]

TABLE 3 Frequency distribution of socio-demographic, social media addiction, morningness-eveningness, and their association with insomnia among 1250 university students in Bangladesh during November to December 2022.

Variables	Yes (*n* = 957; 76.5%) *n*	% (95% CI)	No *n*	% (95% CI)	Total N (%)	χ^2	*p* Value
Sex							
Male	635	66.4 (63.3-69.3)	174	59.4 (53.7-64.9)	809 (64.7)	4.769	<0.01*
Female	322	33.6 (30.7-36.7)	119	40.6 (35.1-46.3)	441 (35.3)		
Education level							
Undergraduate	940	98.2 (97.2-98.9)	271	92.5 (89.1-95.1)	1211 (96.9)	24.385	<0.01*
Graduate	17	1.8 (1.1-2.8)	22	7.5 (4.9-10.9)	39 (3.1)		
Marital status							
Married	474	49.5 (46.4-52.7)	11	3.8 (2.0-6.4)	485 (38.8)	197.949	<0.01*
Unmarried	483	50.5 (47.3-53.6)	282	96.2 (93.6-98.0)	765 (61.2)		
Place of residence							
Off campus	880	92.0 (90.1-93.6)	247	84.3 (79.8-88.1)	1127 (90.2)	14.812	<0.01*
On campus	77	8.0 (6.4-9.9)	46	15.7 (11.9-20.2)	123 (9.8)		
Birthplace							
Rural	715	74.7 (71.9-77.4)	193	65.9 (60.3-71.3)	908 (72.6)	8.825	<0.01*
Urban	242	25.3 (22.6-28.1)	100	34.1 (28.9-39.7)	342 (27.4)		
Monthly family Income (BDT)							
Lower	657	68.7 (65.7-71.5)	81	27.6 (22.8-33.0)	738 (59)	172.994	<0.01*
Middle	183	19.1 (16.7-21.7)	161	54.9 (49.2-60.6)	344 (27.5)		
Upper	117	12.2 (10.3-14.4)	51	17.4 (13.4-22.1)	168 (13.4)		
Academic performance							
Poor	595	62.2 (59.1-65.2)	23	7.8 (5.2-11.3)	618 (49.4)	264.827	<0.01*
Good	362	37.8 (34.8-40.9)	270	92.2 (88.7-94.8)	632 (50.6)		
Smoking status							
No	918	95.9 (94.5-97.0)	273	93.2 (89.9-95.6)	1191 (95.3)	4.774	<0.01*
Yes	39	4.1 (3.0-5.5)	20	6.8 (4.4-10.1)	59 (4.7)		
Social media addiction							
No	293	30.6 (27.8-33.6)	250	85.3 (80.9-89.0)	543 (43.4)	273.255	<0.01*
Yes	664	69.4 (66.4-72.2)	43	14.7 (11.0-19.1)	707 (56.6)		
Morningness-eveningness							
Evening type	258	27.0 (24.2-29.8)	28	9.6 (6.6-13.3)	286 (22.9)	74.950	
Neutral	187	19.5 (17.1-22.1)	122	41.6 (36.1-47.3)	309 (24.7)		<0.01*
Morning type	512	53.5 (50.3-56.6)	143	48.8 (43.1-54.5)	655 (52.4)		

Note: Asterisk (*) indicates the statistically significant at 0.05 level.
Abbreviations: BDT, Bangladeshi Taka; χ^2, Chi-square.

Another major reason for the high incidence of anxiety among students can be attributed to poor academic performance, as indicated in our study, since it may lead to increased worrying and emotional distress.[80] Although poor academic performance contributed as a leading determinant for anxiety in our study, it was not as major for insomnia due to the intricate interplay of various factors. Moreover, this study identified no significant gender disparities in anxiety, which is consistent with similar earlier investigation carried out in Bangladesh.[11] Conversely, the second major risk factor for the high incidence of depression among students can be attributed to age, due to the developmental difficulties and challenges

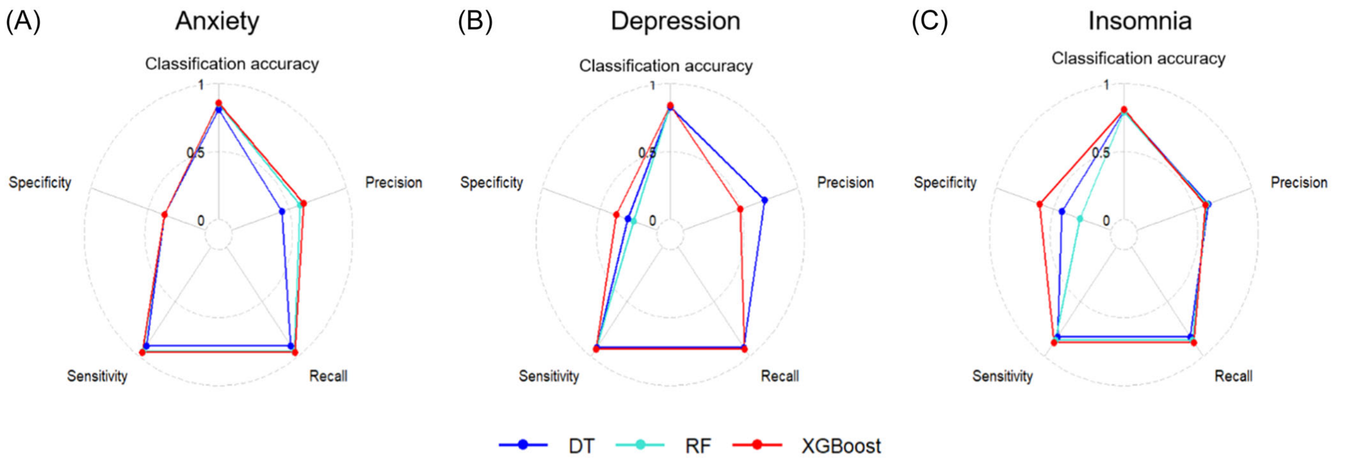

FIGURE 2 Performance evaluation with different metrics of the tree-based models that predict (A) anxiety, (B) depression, and (C) insomnia among Bangladeshi university students during November to December 2022.

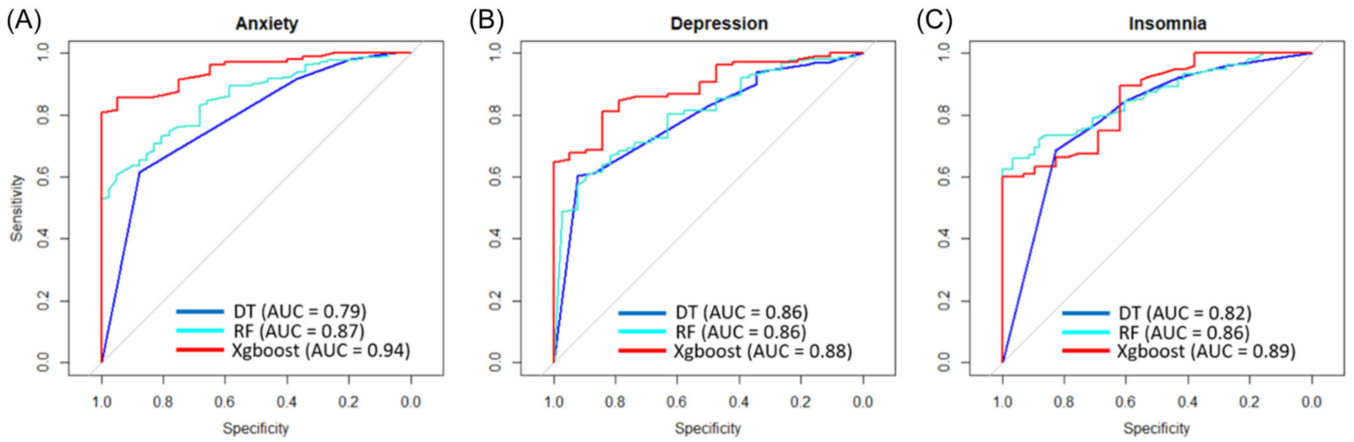

FIGURE 3 ROC curve to predict (A) anxiety, (B) depression, and (C) insomnia among university students in Bangladesh during November to December 2022. ROC, receiver operating characteristics.

younger students face in adapting to university life. This finding aligns with a similar previous study.[81] Age also plays a crucial role in insomnia since younger people—especially university students—often go through major life transitions, have higher stress levels, and have unpredictable schedules, all of which can cause sleep disorders by interfering with sleep rhythms and routines. This finding aligns with previous related study.[82] Our study found that majority of the students came from lower income families. Lower family income can be attributed to another major determinant for depression and insomnia as it can lead to financial hardship, limited access to healthcare, difficult living situations, emotional distress and poor nutrition.[83] Different people have different biological cycles and prefer the morning or evening for work. The MT always prefers the morning hour and the ET prefers the evening hour to accomplish their work.[84] In our research, we found that the majority of the students (52.4%) were MT and had a significant association with depression and insomnia, which is consistent with previous studies.[85,86] According to earlier studies, the incidence and seriousness of mental health issues are rising problems.[87] In addition to increasing the risk of suicide actions, mental health issues can cause significant psychological distress.[72] It is, therefore, crucial to comprehend this potentially vulnerable population and provide appropriate, efficient, accessible counseling and support.

We evaluated the predictive performance of three tree-based ML algorithms, for example, DT model, RF and ensemble XGBoost to predict mental health outcomes among university students in Bangladesh. The XGBoost model performed best in predicting anxiety, depression, and insomnia among the university students for exhibiting higher prediction accuracy, that is, 86% accuracy and 96% sensitivity and 94% of AUC for anxiety, 84% accuracy, 93% sensitivity, and 88% of AUC for depression as well as 81% accuracy, 87% sensitivity and 89% of AUC for insomnia. These findings demonstrate how ML can effectively detect and predict mental health issues in students, providing stakeholders and policymakers with a valuable tool.

The study findings elucidate that various factors, including students' social media addiction, age, academic performance, monthly family

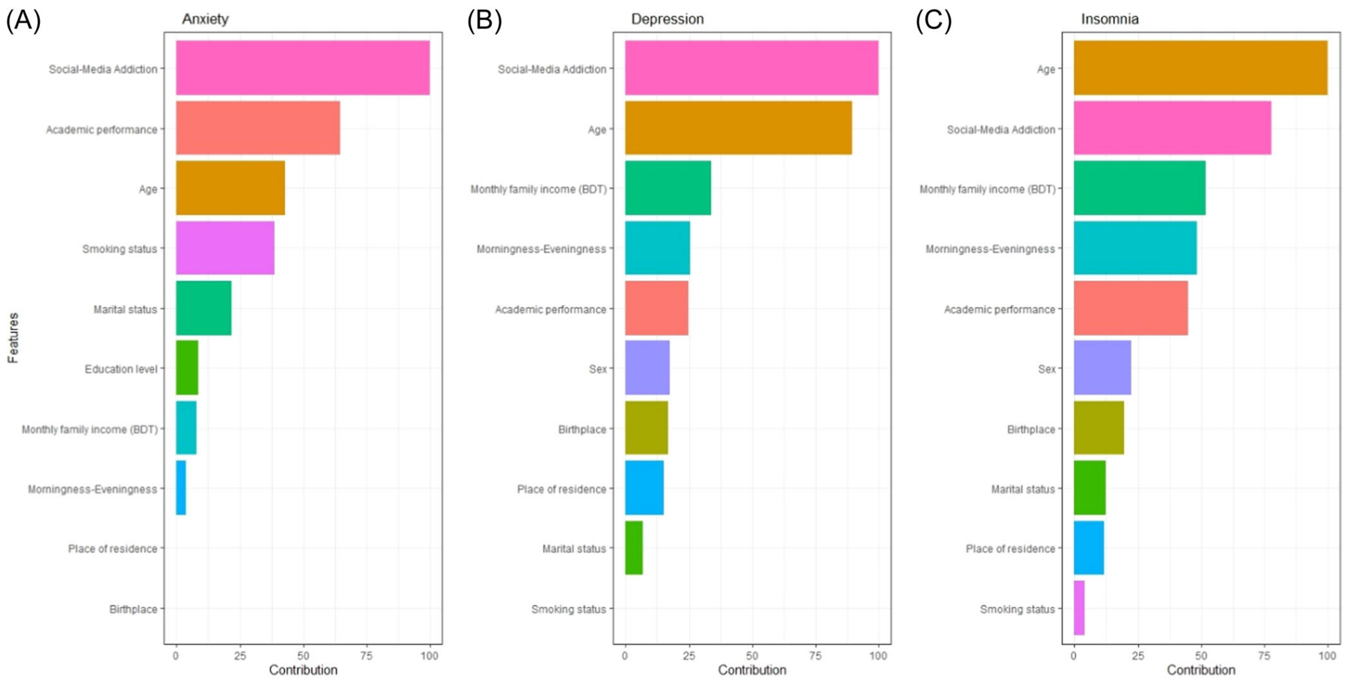

FIGURE 4 Important features selection using the XGBoost model. (A) Anxiety, (B) depression, and (C) insomnia.

income, morningness-eveningness, marital status, sex, education level, and birthplace, were identified as significant contributors to their levels of anxiety, depression, and insomnia. The predictive accuracy of these determinants was demonstrated through the application of the XGBoost model, showcasing its superior performance compared to other models. Concurrently, conventional statistical analysis, such as the chi-square test, reinforced the role of these factors as statistically significant correlates. However, it is essential to clarify that the term "main determinants" refers to the variables incorporated into the model rather than an exhaustive list of all possible determinants of poor mental health. The intention was to assess the predictive capacity of the chosen variables within the context of the study, acknowledging that other unexamined factors may also influence mental health outcomes among university students. Regarding the uniqueness of the findings, the study contributes by emphasizing the efficacy of the XGBoost model in identifying and predicting mental health outcomes, thereby providing a nuanced understanding of the role of various socio-demographic and personal factors. However, it is duly noted that the identified determinants align with existing literature on contributors to poor mental health. The focus on ML models aims to underscore the methodological advancements and precision achieved in predicting mental health outcomes, rather than introducing entirely novel determinants.

5 | LIMITATIONS

The study has certain limitations that should be acknowledged to provide a more comprehensive understanding of its scope and potential implications. Firstly, the sample size was not suffiecient enough and universities were not categorized based on provincial or national targeting.

Moreover, it is essential to acknowledge that the study exclusively employed tree-based ML models and did not explore other methodological approaches. For instance, the decision to have a single person translate existing measures may introduce potential biases in the interpretation of mental health indicators. This choice could impact the validity and reliability of the collected data, serving as a limitation in the study's overall design.

Additionally, the study's cross-sectional nature is a noteworthy limitation that warrants discussion. The use of a cross-sectional design restricts our ability to establish causality between participant characteristics and the assessed mental health outcomes. It is crucial to recognize that the observed associations do not imply a cause-and-effect relationship, and other temporal factors or confounding variables may contribute to the identified patterns. This limitation underscores the need for caution in drawing definitive conclusions about the directionality of the relationships observed in this study.

Furthermore, the exclusion of chronic diseases, such as heart disease, diabetes, hypertension, and arthritis, from the investigation poses another limitation. These health conditions are known to exert significant influences on mental health, and their omission restricts the holistic understanding of the interplay between physical and mental well-being among university students. Future studies should consider incorporating a broader spectrum of psychological disorders and chronic diseases, employing diverse methodologies, and addressing potential translation biases to enhance the robustness and applicability of findings in this field.

6 | RECOMMENDATION

This study found a high prevalence of anxiety, depression and insomnia among Bangladeshi university students. Therefore, considering the high prevalence of anxiety, depression and insomnia among university students, mental health interventions such as mental health promotion and the establishment of efficient university counseling facilities are recommended to reduce the grave danger of this disease. In addition, educating students on the significance of leading a balanced lifestyle that includes enough sleep, a healthy diet, and regular exercise can be another effective technique. Furthermore, creating a conducive learning ambiance and educating parents, teachers, and administrators about the value of mental health are crucial. To further promote a comprehensive approach to pupil well-being, parents should be included in mental health awareness campaigns, faculty and staff can receive mental health training, and open communication can be encouraged.

7 | CONCLUSION

This study extends to the mounting evidence that Bangladeshi university students experience elevated levels of anxiety, depression, and insomnia. In contrast with previous research, this study discovered a high incidence of mental health disorders (e.g., depression, anxiety, and insomnia). It highlighted a variety of risk variables, including social media addiction, age, academic performance, smoking status, monthly family income and morningness-eveningness, that are strongly linked to these disorders. Moreover, the XGBoost ML model was found to do better over the DT and RF. Our proposed model and study findings can assist stakeholders, families, and policymakers in better understanding and averting this crisis through improved policy-making strategies. In addition, the ML based prediction model can enhance the digital healthcare system and mental health conditions by allowing more precise diagnoses, individualized treatment plans, predictive analytics and remote tracking.

AUTHOR CONTRIBUTIONS

Arman Hossain Chowdhury: Data curation; formal analysis; investigation; methodology; resources; software; validation; visualization; writing—original draft; writing—review and editing. **Dana Rad**: Writing—review and editing. **Md. Siddikur Rahman**: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

The authors are also very thankful to the university students participating in the study and volunteering their time to help gather data during the survey.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data and codes are available at https://github.com/siddikur2022/mental-health-prediction-using-ML-models.

ETHICS STATEMENT

The authors explained the purpose of the study to the participants and written consent was obtained from the study participants before interviews. All data collected were kept confidential. This study was approved by Department of Statistics, Begum Rokeya University, Rangpur (EA/05/23).

TRANSPARENCY STATEMENT

The lead author Md. Siddikur Rahman affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ORCID

Arman Hossain Chowdhury http://orcid.org/0000-0003-1498-287X

Md. Siddikur Rahman https://orcid.org/0000-0001-8925-6544

REFERENCES

1. Lund C, Brooke-Sumner C, Baingana F, et al. Social determinants of mental disorders and the sustainable development goals: a systematic review of reviews. *Lancet Psych*. 2018;5:357-369. doi:10.1016/S2215-0366(18)30060-9
2. Mohammad M, Chowdhury MAB, Islam MN, et al. Health awareness, lifestyle and dietary behavior of university students in the northeast part of Bangladesh. *Int J Adolesc Med Health*. 2018;33(2). doi:10.1515/ijamh-2018-0105
3. Siddique MAB, Ovi MR, Ahammed T, Chowdhury MAB, Uddin MJ. Mental health knowledge and awareness among university students in Bangladesh. *Heliyon*. 2022;8:e11084. doi:10.1016/J.HELIYON.2022.E11084
4. Afsar N, Kulsoom B. Stress, anxiety, and depression among medical students in a multiethnic setting. *Neuropsychiatr Dis Treat*. 2015;11:1713. doi:10.2147/NDT.S83577
5. ul Haqul MA, Dar IS, Aslam M, Mahmood QK. Psychometric study of depression, anxiety and stress among university students. *J Public Health (Bangkok)*. 2018;26:211-217. doi:10.1007/S10389-017-0856-6/METRICS
6. Ferrari A. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet Psych*. 2022;9:137-150. doi:10.1016/S2215-0366(21)00395-3
7. WHO. *Depressive disorder (depression)*. WHO; 2023. https://www.who.int/news-room/fact-sheets/detail/depression
8. The Recovery Village. *Anxiety disorders facts and statistics*. The Recovery Village; 2023. https://www.therecoveryvillage.com/mental-health/anxiety/anxiety-disorder-statistics/
9. Mamun MA, Rafi MA, Al Mamun AHMS, et al. Prevalence and psychiatric risk factors of excessive Internet use among Northern Bangladeshi Job-Seeking graduate students: a pilot study. *Int*

J Mental Health Addict. 2021;19:908-918. doi:10.1007/S11469-019-00066-5/TABLES/2

10. Hossain MD, Ahmed HU, Chowdhury WA, Niessen LW, Alam DS. Mental disorders in Bangladesh: a systematic review. *BMC Psychiatry*. 2014;14:216. doi:10.1186/S12888-014-0216-9/TABLES/3
11. Alim SAHM, Rabbani MG, Karim E, et al. Assessment of depression, anxiety and stress among first year MBBS students of a public medical college, Bangladesh. *Bangladesh J Psych*. 2017;29:23-29. doi:10.3329/BJPSY.V29I1.32748
12. Alim SAHM, Kibria SME, Islam MJ, et al. Translation of DASS 21 into bangla and validation among medical students. *Bangladesh J Psych*. 2017;28:67-70. doi:10.3329/BJPSY.V28I2.32740.
13. Chowdhury AI, Ghosh S, Hasan MF, Khandakar K, Azad F. Prevalence of insomnia among university students in south asian region: a systematic review of studies. *J Prev Med Hyg*. 2020;61:E525. doi:10.15167/2421-4248/JPMH2020.61.4.1634
14. Azizi SM, Soroush A, Khatony A. The relationship between social networking addiction and academic performance in Iranian students of medical sciences: a cross-sectional study. *BMC Psychol*. 2019;7(1):28. doi:10.1186/s40359-019-0305-0
15. Pellegrino A, Stasi A, Bhatiasevi V. Research trends in social media addiction and problematic social media use: A bibliometric analysis. *Front Psychiatry*. 2022;13:1017506. doi:10.3389/FPSYT.2022.1017506/BIBTEX
16. Shamsuddin K, Fadzil F, Ismail WSW, et al. Correlates of depression, anxiety and stress among Malaysian university students. *Asian J Psych*. 2013;6:318-323. doi:10.1016/J.AJP.2013.01.014
17. Steptoe A, ardle J, Tsuda A, Tanaka Y. Depressive symptoms, socio-economic background, sense of control, and cultural factors in university students from 23 countries. *Int J Behav Med*. 2007;14:97-107. doi:10.1007/BF03004175
18. Kader Maideen SF, Mohd Sidik S, Rampal L, Mukhtar F. Prevalence, associated factors and predictors of anxiety: a community survey in selangor, Malaysia. *BMC Psychiatry*. 2015;15(1):262. doi:10.1186/s12888-015-0648-x
19. Mamun MA, Hossain MS, Siddique AB, Sikder MT, Kuss DJ, Griffiths MD. Problematic Internet use in Bangladeshi students: the role of socio-demographic factors, depression, anxiety, and stress. *Asian J Psychiatr*. 2019;44:48-54. doi:10.1016/j.ajp.2019.07.005
20. McMichael AJ, Hetzel BS. Mental health problems among university students, and their relationship to academic failure and withdrawal. *Med J Aust*. 1975;1:499-504. doi:10.5694/j.1326-5377.1975.tb111533.x
21. Uddin MS, Mamun AA, Iqbal MA, et al. Internet addiction disorder and its pathogenicity to psychological distress and depression among university students: a cross-sectional pilot study in Bangladesh. *Psychology*. 2016;07:1126-1137. doi:10.4236/psych.2016.78113
22. Bashir H, Bhat SA. Effects of social media on mental health: a review. *Int J Indian Psychol*. 2017;4(3). doi:10.25215/0403.134
23. Maniruzzaman M, Kumar N, Menhazul Abedin M, et al. Comparative approaches for classification of diabetes mellitus data: machine learning paradigm. *Comput Meth Prog Biomed*. 2017;152:23-34. doi:10.1016/j.cmpb.2017.09.004
24. Salim NAM, Wah YB, Reeves C, et al. Prediction of dengue outbreak in selangor Malaysia using machine learning techniques. *Sci Rep*. 2021;11:939. doi:10.1038/s41598-020-79193-2
25. Alim M, Ye GH, Guan P, Huang DS, Zhou BS, Wu W. Comparison of ARIMA model and XGBoost model for prediction of human brucellosis in mainland China: a time-series study. *BMJ Open*. 2020;10:e039676. doi:10.1136/bmjopen-2020-039676
26. Rahman MS, Chowdhury AH. A data-driven extreme gradient boosting machine learning model to predict COVID-19 transmission with meteorological drivers. *PLoS One*. 2022;17:e0273319. doi:10.1371/journal.pone.0273319
27. Noorunnahar M, Chowdhury AH, Mila FA. A tree based extreme gradient boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *PLoS One*. 2023;18:e0283452. doi:10.1371/journal.pone.0283452
28. Salami D, Sousa CA, Martins MRO, Capinha C. Predicting dengue importation into Europe, using machine learning and model-agnostic methods. *Sci Rep*. 2020;10:9689. doi:10.1038/s41598-020-66650-1
29. Rahman MS, Pientong C, Zafar S, et al. Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. *One Health*. 2021;13:100358. doi:10.1016/J.ONEHLT.2021.100358
30. Yavari Nejad F, Varathan KD. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. *BMC Med Inform Decis Mak*. 2021;21(1):141. doi:10.1186/S12911-021-01493-Y/TABLES/6
31. Sarma D, Hossain S, Mittra T, Bhuiya MAM, Saha I, Chakma R. Dengue Prediction using Machine Learning Algorithms. *IEEE Reg 10 Humanit Technol Conf R10-HTC*, 2020. doi:10.1109/R10-HTC49770.2020.9357035
32. Alotaibi A, Shiblee M, Alshahrani A. Prediction of severity of COVID-19-Infected patients using machine learning techniques. *Computers*. 2021;10:31. doi:10.3390/COMPUTERS10030031
33. Arslan H. Machine learning methods for COVID-19 prediction using human genomic data. *Proc*. 2021;74:20. doi:10.3390/PROCEEDINGS2021074020
34. Villavicencio CN, Macrohon JJE, Inbaraj XA, Jeng JH, Hsieh JG, et al. COVID-19 prediction applying supervised machine learning algorithms with comparative analysis using WEKA. *Algorithms*. 2021;14:201. doi:10.3390/A14070201
35. Suma LS, Anand HS, Vinod chandra SS. Nature inspired optimization model for classification and severity prediction in COVID-19 clinical dataset. *J Ambient Intell Humaniz Comput*. 2021;1:1-13. doi:10.1007/S12652-021-03389-1/FIGURES/7
36. Marcos M, Belhassen-García M, Sánchez-Puente A, et al. Development of a severity of disease score and classification model by machine learning for hospitalized COVID-19 patients. *PLoS One*. 2021;16:e0240200. doi:10.1371/JOURNAL.PONE.0240200
37. Choudhury AA, Khan MRH, Nahim NZ, Tulon SR, Islam S, Chakrabarty A. Predicting Depression in Bangladeshi Undergraduates using Machine Learning. *Proc 2019 IEEE Reg 10 Symp TENSYMP 2019*, 2019:789-794. doi:10.1109/TENSYMP46218.2019.8971369
38. Rois R, Ray M, Rahman A, Roy SK. Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms. *J Health Popul Nutr*. 2021;40(1):50. doi:10.1186/s41043-021-00276-5
39. Mateen BA, Liley J, Denniston AK, Holmes CC, Vollmer SJ. Improving the quality of machine learning in health applications and clinical research. *Nat Mach Intell*. 2020;2:554-556. doi:10.1038/s42256-020-00239-1
40. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell*. 2021;3:199-217. doi:10.1038/s42256-021-00307-0
41. Topol EJ. High-performance Medicine: the convergence of human and artificial intelligence. *Nature Med*. 2019;25:44-56. doi:10.1038/s41591-018-0300-7
42. Ha M. Data collection using online questionnaires in marketing. In: *Sage Research Methods: Doing Research Online*. SAGE Publications, Ltd.; 2022. doi:10.4135/9781529603569
43. Lowe NK. What Is a Pilot Study. *Journal of Obstetric, Gynecologic & Neonatal Nursing*. Elsevier B.V. 2019;48:117-118. doi:10.1016/j.jogn.2019.01.005.

44. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a facebook addiction scale. *Psychol Rep*. 2012;110:501-517. doi:10.2466/02.09.18.PR0.110.2.501-517
45. Bányai F, Zsila Á, Király O, et al. Problematic social media use: results from a Large-Scale nationally representative adolescent sample. *PLoS One*. 2017;12:e0169839. doi:10.1371/JOURNAL.PONE.0169839
46. Brailovskaia J, Teismann T, Margraf J. Physical activity mediates the association between daily stress and facebook addiction disorder (FAD)–A longitudinal approach among German students. *Comput Human Behav*. 2018;86:199-204. doi:10.1016/J.CHB.2018.04.045
47. Atroszko PA, Balcerowska JM, Bereznowski P, Biernatowska A, Pallesen S, Schou Andreassen C. Facebook addiction among polish undergraduate students: validity of measurement and relationship with personality and well-being. *Comput Human Behav*. 2018;85:329-338. doi:10.1016/J.CHB.2018.04.001
48. Susman EJ, Dockray S, Schiefelbein VL, Herwehe S, Heaton JA, Dorn LD. Morningness/eveningness, morning-to-afternoon cortisol ratio, and antisocial behavior problems during puberty. *Dev Psychol*. 2007;43:811-822. doi:10.1037/0012-1649.43.4.811
49. Adan A, Almirall H. Horne & östberg morningness-eveningness questionnaire: a reduced scale. *Pers Individ Dif*. 1991;12:241-253. doi:10.1016/0191-8869(91)90110-W
50. Mohd Sidik S, Arroll B, Goodyear-Smith F. Validation of the GAD-7 (Malay version) among women attending a primary care clinic in Malaysia. *J Pri Health Care*. 2012;4:5-11. doi:10.1071/hc12005
51. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166:1092-1097. doi:10.1001/ARCHINTE.166.10.1092
52. Islam S, Akter R, Sikder T, Griffiths MD. Prevalence and factors associated with depression and anxiety among first-year university students in Bangladesh: a cross-sectional study. 2020.
53. Dhira TA, Rahman MA, Sarker AR, Mehareen J. Validity and reliability of the generalized anxiety Disorder-7 (GAD-7) among university students of Bangladesh. *PLoS One*. 2021;16:e0261590. doi:10.1371/JOURNAL.PONE.0261590
54. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *Gen Hosp Psychistry*. 2006;28:71-77. doi:10.1016/J.GENHOSPPSYCH.2005.07.003
55. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613. doi:10.1046/j.1525-1497.2001.016009606.x
56. Kroenke K, Spitzer RL. The PHQ-9: A new depression diagnostic and severity measure. *Psychiatr Ann*. 2002;32:509-515. doi:10.3928/0048-5713-20020901-06
57. Rahman MA, Dhira TA, Sarker AR, Mehareen J. Validity and reliability of the patient health questionnaire scale (PHQ-9) among university students of Bangladesh. *PLoS One*. 2022;17:e0269634. doi:10.1371/journal.pone.0269634
58. Bastien C. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med*. 2001;2:297-307. doi:10.1016/S1389-9457(00)00065-4
59. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. 18-20.
60. Smith MT, Wegener ST. Measures of sleep: the insomnia severity index, medical outcomes study (MOS) sleep scale, Pittsburgh sleep diary (PSD), and Pittsburgh sleep quality index (PSQI). *Arth Rheumat*. 2003;49:S184-S196. doi:10.1002/art.11409
61. Mulette P, Ravoninjatovo B, Guguen C, et al. Insomnia in adults with cystic fibrosis: strong association with anxiety/depression and impaired quality of life. *BMC Pulm Med*. 2021;21:108. doi:10.1186/S12890-021-01473-Y/FIGURES/1
62. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*, 2023. doi:10.1145/2939672
63. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29:1189-1232. doi:10.1214/aos/1013203451
64. Farooq Z, Rocklöv J, Wallin J, et al. Artificial intelligence to predict West Nile virus outbreaks with eco-climatic drivers. *Lancet Reg Health Euro*. 2022;17:100370. doi:10.1016/j.lanepe.2022.100370
65. Stone M. Cross-Validatory choice and assessment of statistical predictions. *J Royal Stat Soc*. 1974;36:111-133. doi:10.1111/J.2517-6161.1974.TB00994.X
66. Elgeldawi E, Sayed A, Galal AR, Zaki AM. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*. 2021;8:79. doi:10.3390/INFORMATICS8040079
67. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*. 2022. doi:10.1145/2939672
68. Zeng X, An J, Lin R, et al. Prediction of complications after paediatric cardiac surgery. *Eur J Cardiothorac Surg*. 2019;57:350-358. doi:10.1093/EJCTS/EZZ198
69. Budholiya K, Shrivastava SK, Sharma V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J King Saud Univer Comp Informat Sci*. 2022;34:4514-4523. doi:10.1016/J.JKSUCI.2020.10.013
70. Meyers LS, Glenn CG, Guarino AJ. *Performing data analysis using IBM SPSS*. John Wiley & Sons; 2013. https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25
71. Team R. *RStudio: integrated development environment for R*. RStudio Team; 2018. https://www.rstudio.com/
72. Eisenberg D, Gollust SE, Golberstein E, Hefner JL. Prevalence and correlates of depression, anxiety, and suicidality among university students. *Am J Orthopsychiatry*. 2007;77:534-542. doi:10.1037/0002-9432.77.4.534
73. Nicky Stanley JM. Responding to students' mental health needs: Impermeable systems and diverse users. *J Ment Health*. 2009;10:41-52. doi:10.1080/2-09638230020023606
74. Benton SA, Robertson JM, Tseng WC, Newton FB, Benton SL. Changes in counseling center client problems across 13 years. *Profess Psychol Res Pract*. 2003;34:66-72. doi:10.1037/0735-7028.34.1.66
75. Bayram N, Bilgel N. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Soc Psychiatry Psychiatr Epidemiol*. 2008;43:667-672. doi:10.1007/S00127-008-0345-X/METRICS
76. Stewart-Brown S, Evans J, Patterson J, et al. The health of students in institutes of higher education: an important and neglected public health problem? *J Public Health (Bangkok)*. 2000;22:492-499. doi:10.1093/PUBMED/22.4.492
77. Viñas Poch F, Villar E, Caparros B, Juan J, Cornella M, Perez I. Feelings of hopelessness in a Spanish university population - descriptive analysis and its relationship to adapting to university, depressive symptomatology and suicidal ideation. *Soc Psychiatry Psychiatr Epidemiol*. 2004;39:326-334. doi:10.1007/S00127-004-0756-2/METRICS
78. Malaeb D, Salameh P, Barbar S, et al. Problematic social media use and mental health (depression, anxiety, and insomnia) among Lebanese adults: any mediating effect of stress? *Perspect Psychiatr Care*. 2021;57:539-549. doi:10.1111/PPC.12576
79. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. *J Affect Disord*. 2017;207:163-166. doi:10.1016/J.JAD.2016.08.040
80. Vitasari P, Wahab MNA, Othman A, Herawan T, Sinnadurai SK. The relationship between study anxiety and academic performance among engineering students. *Procedia Soc Behav Sci*. 2010;8:490-497. doi:10.1016/J.SBSPRO.2010.12.067

81. Cheung DK, Tam DKY, Tsang MH, Zhang DLW, Lit DSW. Depression, anxiety and stress in different subgroups of first-year university students from 4-year cohort data. *J Affect Disord*. 2020;274:305-314. doi:10.1016/J.JAD.2020.05.041
82. Gallego-Gómez JI, González-Moro MTR, González-Moro JMR, et al. Relationship between sleep habits and academic performance in university nursing students. *BMC Nurs*. 2021;20:100. doi:10.1186/S12912-021-00635-X/TABLES/4
83. Wang J, Yang C, Wang J, Sui X, Sun W, Wang Y. Factors affecting psychological health and career choice among medical students in eastern and Western region of China after COVID-19 pandemic. *Front Public Health*. 2023;11:1081360. doi:10.3389/FPUBH.2023.1081360/BIBTEX
84. Randler C. Association between morningness-eveningness and mental and physical health in adolescents. *Psychol Health Med*. 2011;16:29-38. doi:10.1080/13548506.2010.521564
85. Koo DL, Yang KI, Kim JH, et al. Association between morningness-eveningness, sleep duration, weekend catch-up sleep and depression among Korean high-school students. *J Sleep Res*. 2021;30:e13063. doi:10.1111/JSR.13063
86. Mozafari A, Tabaraie M, Tahrodi MHM. Morningness-Eveningness chronotypes, insomnia and sleep quality among medical students of qom. *J Sleep Sci*. 2016;1:67-73. https://jss.tums.ac.ir/index.php/jss/article/view/31.
87. Hunt J, Eisenberg D. Mental health problems and Help-Seeking behavior among college students. *J Adolesc Health*. 2010;46:3-10. doi:10.1016/J.JADOHEALTH.2009.08.008

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Chowdhury AH, Rad D, Rahman MS. Predicting anxiety, depression, and insomnia among Bangladeshi university students using tree-based machine learning models. *Health Sci Rep*. 2024;7:e2037. doi:10.1002/hsr2.2037