Current Research in Behavioral Sciences 2 (2021) 100044

Contents lists available at ScienceDirect

Current Research in Behavioral Sciences

journal homepage: www.elsevier.com/locate/crbeha

An in-depth analysis of machine learning approaches to predict depression

Md. Sabab Zulfiker[a,*], Nasrin Kabir[b], Al Amin Biswas[a], Tahmina Nazneen[c], Mohammad Shorif Uddin[b]

[a] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[b] Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh
[c] Department of Psychology, University of Dhaka, Dhaka, Bangladesh

ARTICLE INFO

Keywords:
Depression
Prediction
Machine learning
Classifier
Feature selection
SMOTE

ABSTRACT

Among all the forms of psychological and mental disorders, depression is the most common form. Nowadays a large number of youths and adults around the world suffer from depression. Depression can cause severe problems in case of failing to detect it at an early stage or failing to ensure the timely counseling of a depressed person. It is one of the major reasons to raise suicidal cases. But ironically, our society still does not want to acknowledge depression as a mental disorder causing a significant number of depressed persons to remain unidentified and untreated. This study has investigated six different machine learning classifiers using various socio-demographic and psychosocial information to detect whether a person is depressed or not. Besides, three different feature selection methods, such as Select K-Best Features (SelectKBest), Minimum Redundancy and Maximum Relevance (mRMR), and Boruta feature selection algorithm have been used for extracting the most relevant features from the dataset. To achieve better accuracy in predicting depression, Synthetic Minority Oversampling Technique (SMOTE) has been used that reduces the class imbalance of the training data. The AdaBoost classifier with the SelectKBest feature selection technique has outperformed all other approaches with an accuracy of 92.56%. Moreover, other evaluation metrics, namely sensitivity, specificity, precision, F1-score, and area under the curve (AUC) of different models have been calculated to identify the most efficient model for predicting depression.

1. Introduction

Depression is a psychological disorder that can be characterized by the existence of persistent sadness for at least two weeks. It creates an inability to perform daily activities, and the depressed persons lose their interests and pleasures in doing those things that they usually enjoy (World Health Organization, 2021). In this Covid-19 pandemic situation, depression has become a major health concern in the world, and 322 million people all over the world are living with depression. About 50% of these depressed people live in the Western Pacific and South-East Asian regions. Between 2005 and 2015, there was an increase in the number of depressed persons by 18.40% (Vos et al., 2016; World Health Organization, 2017). Depression is responsible for developing numerous chronic diseases, such as diabetes, heart disease, etc. among depressed persons. It is the second major cause of developing chronic diseases (Whooley et al., 2013; Otte et al., 2016). Severe depression can trigger suicidal cases. Every year about 0.8 million suicide cases take place worldwide. Among them, half of the cases occur due to depression (Otte et al., 2016). Depression affects the social life of a person. Depressed persons often feel a dilemma to interact with their co-workers, friends, and families. And their reluctance towards the relationship with others hampers their social life. Depression is responsible for forming an obsession with drugs and smoking. It also accelerates obesity (Thapar et al., 2012). Depressed persons often feel fatigued. As they do not enjoy their daily activities and often feel exhausted, their productivity deteriorates, which affects their economic condition. So, unless depressed persons are treated timely, it can cause severe socio-economic problems to them. In the long run, it creates a negative impact on the socio-economic status of the country also. Before counseling a person for depression, identifying whether the person is depressed or not is the most crucial task.

With the massive advancement of information and technology, the need for using machine learning algorithms to infer significant patterns from data of different sectors is increasing rapidly. Although machine learning algorithms have been widely used in the medical and health sectors, their use in psychological sectors is relatively low. For decades, statistical inferences have been used in psychological analysis and psy-

* Corresponding author.
E-mail addresses: sabab.rumc@gmail.com (Md.S. Zulfiker), etykabir16@gmail.com (N. Kabir), alaminbiswas.cse@gmail.com (A.A. Biswas), nazneen291572@gmail.com (T. Nazneen), shorifuddin@gmail.com (M.S. Uddin).

https://doi.org/10.1016/j.crbeha.2021.100044
Received 17 February 2021; Received in revised form 22 April 2021; Accepted 27 April 2021

chometrics. Machine learning has largely drawn the attention of the media for its use in psychometrics after the Cambridge Analytica Affair. Due to the replicability problems of statistical inference, researchers are tending to machine learning from statistical inferences in psychometrics and psychological analysis nowadays (Orrù et al., 2020).

Bangladesh is a densely populated country situated in South Asia. Due to socio-economic problems, a large number of people in this country face different psychological disorders. According to the World Health Organization (WHO), about 6.4 million people of Bangladesh suffer from depressive disorders with a prevalence of 4.10% of the total population (World Health Organization, 2017). Depression is the most notable reason for suicidal cases in Bangladesh. Here, the mortality rate due to suicide is 39.6 per 0.1 million population (Shah et al., 2017). So, it is necessary to screen and identify depressed persons at an early stage.

So far, very few researches have been performed to identify and predict depression using machine learning techniques. This research tries to fill this void and has proposed a state-of-the-art solution in this context. This study aims to identify whether a person is depressed or not, to determine the key factors that induce depression, and to determine the best machine learning approach to identify the depressed persons. This study also intends to minimize the required time for screening depression.

The major contributions of this work are as follows:

I. Extracting the most important socio-demographic and psychosocial factors responsible for forming depression.
II. Generation of a dataset including the socio-demographic and the psychosocial information of the persons to predict depression.
III. Exploring different machine learning and feature selection algorithms to screen the existence of depression efficiently.
IV. Due to the simplicity of the required demographic and psychosocial information used in this study, a suspected depressed person will feel less hassle to give the required information of this study to detect depression rather than answering the questions of different authentic depression screening scales.

The remaining part of the paper is arranged according to the following structure: the related works are described in Section 2. The entire methodology is described in Section 3. Section 4 exhibits the results of the study. The comparative study of this work with other existing works is presented in Section 5. And, finally, Section 6 ends this paper by drawing the conclusion and introducing some potential scopes of future study.

2. Related works

The aspects of this kind of research problem are notably complex and need to do an in-depth investigation. This section has gone through several related research articles to find out the tools and techniques used in the existing works and determine the research gaps.

Cvetković (2017) performed the prediction of depression among breast cancer patients. This study was performed by collecting the information of 84 patients with the age range of 30–78 years through a two-phase interview. In the first phase, the socio-demographic information of the patients was accumulated. The standardized Beck Depression Inventory (BDI) test of the patients was performed in the second phase. The actual depression range of the patients was measured by the BDI test. Here, features acquired from the socio-demographic information were the predictor variables, and the depression range derived from the BDI test was the target variable. This study compared the efficacy of three different algorithms, namely: Artificial Neural Network (ANN) with extreme learning algorithm, ANN with backpropagation learning algorithm, and Fuzzy with genetic algorithm. Here, ANN with extreme learning algorithm showed the best performance.

A data preprocessing technique was proposed by Iliou et al. (2019) to predict depression. They compared the efficiency of the Principal Component Analysis (PCA) method with their proposed preprocessing method and found that the proposed method surpassed the PCA.

Na et al. (2020) predicted the commencement of future depression in the context of the Republic of Korea. The predictive model was constructed using the Random Forest classifier. SMOTE was used to cope with the issues of class imbalances. This study achieved an accuracy of 86.20%. This study also indicated that satisfaction towards socio-familial relationship and satisfaction for health are the key factors to influence the onset of depression.

Sau and Bhakta (2017) performed their study on the geriatric population. They predicted the appearance of depression among elderly patients using ten different machine learning classifiers. The socio-demographic and health-related factors of the geriatric patients were acquired for the classification purpose. Among the ten classifiers, the Random Forest showed the best performance.

Hatton et al. (2019) utilized the psychometric and demographic information of 284 elderly patients to predict the prevalence of depression. They applied the Extreme Gradient Boosting algorithm for predicting the persistence of depression and compared its performance with the Logistic Regression model. They stated that the Extreme Gradient Boosting showed superior performance than the Logistic Regression.

After giving birth, a new mother often experiences a form of depression known as Postpartum Depression (PPD). Natarajan et al. (2017) intended to diagnose PPD. This study was conducted by using the data of 173 new mothers of different ethnicity. They compared the performance of the Functional Gradient Boosting algorithm with other classical machine learning algorithms and affirmed that the Functional Gradient Boosting algorithm showed the best performance in the case of diagnosing PPD.

A mobile-based application was developed by Jiménez-Serrano et al. (2015) to anticipate PPD. They evaluated the performance of Naive Bayes, Logistic Regression, SVM, and ANN using the socio-economic, clinical, and psychometric parameters of 1397 new mothers. They asserted that this study was the first valid Clinical Decision Support System capable of estimating the existence of PPD during the first week after delivery.

A soft computing-based approach incorporating neural network, fuzzy logic, and case-based reasoning was proposed by Ekong and Onibere (2015). Using different physiological and psychological factors, they categorized the severity of depression into five quality classes, namely: Near Absent, Mild, Moderate, Severe, and Very Severe. The proposed method demonstrated greater efficiency than the Diagnostic and Statistical Manual for Mental Disorder (DSM).

Type-2 fuzzy logic was used by Zarandi et al. (2019) to detect the level of depression. For enhancing the accuracy of the study and predicting the level of depression of the patients with fewer questions, they utilized the Mutual Information Feature Selection (MIFS) method. Their suggested approach used just fifteen questions to forecast the level of depression and achieved an accuracy of 84.00%.

Islam et al. (2018) intended to predict the existence of depression in a person from his Facebook posts and comments. Extracting the psycholinguistic features from the person's Facebook posts and comments, they built models with different machine learning classifiers like Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision trees (DT), and Ensemble classifier. Here, DT outperformed other classifiers.

The seafarers often suffer from a variety of psychiatric illnesses. Sau and Bhakta (2019) aimed to determine the presence of anxiety and depression among seafarers using machine learning approaches. They applied five different machine learning classifiers to predict anxiety and depression among the seafarers and found that the CatBoost classifier outperformed the other classifiers.

Priya et al. (2020) collected the data related to anxiety, stress, and depression for identifying these mental disorders by applying machine learning algorithms. They found that to predict depression, the Naive Bayes classifier showed the best performance with an accuracy of 85.50%.

Choudhury et al. (2019) performed their study to predict the existence of depression in the undergraduate students of Bangladesh. Initially, they collected the data of 935 students. After data cleansing and applying different data preprocessing techniques, they used the data of 577 students for conducting their study. Using the Random Forest classifier they achieved the highest accuracy of 75.00%.

From the above discussions, it can be said that most of these existing studies have been performed to predict depression among a certain group of people, like: among the people of a certain age range, patients diagnosed with a certain disease, etc. This study tries to overcome this restriction by taking into concern different socio-demographic information of the people of diverse age ranges, health conditions, and socioeconomic statuses.

3. Methodology

This section is divided into seven subsections. The entire methodology of this study has been described in the following subsections.

3.1. Data acquisition

A survey was performed for collecting the data from Bangladeshi citizens of different age ranges. A questionnaire consisting of 55 questions was designed. The first 30 questions were designed for gathering complex psychosocial, and socio-demographic information of the participants, and the last 25 questions were included from the Burns Depression Checklist (BDC). The survey was conducted in the period between April 2020 and August 2020. The dataset consists of the responses of 604 individuals.

BDC was used to assess the actual depression level of each participant during the survey. This study used the revised version of BDC with 25 questions. The 25 questions of the revised version of BDC are divided into four categories. The first ten questions are about the recent thoughts and feelings of the participants, the next seven questions inquire about the participants' recent activities and relationships. The next five questions ask about the participants' physical states and symptoms. Finally, the last three questions inquire about the participants' suicidal urges (Burns, 1999; Jabbar and Zaza, 2019). For screening depression using BDC, the participants had to give the intensity of different symptoms of depression that they were facing during the past week, including the day of the survey. For the revised version of BDC, the symptom intensity varies from 0 to 4. BDC is one of the highly accepted rating scales for diagnosing depression developed by David Burns. It is consistent and concentrates more on the specific symptoms of depression rather than the non-specific symptoms (Burns et al., 2013). The Cronbach's alpha value of BDC is 0.89, which indicates that it is one of the highly reliable scales to determine depression (Holtz et al., 2014).

A person's overall BDC score is assessed by adding the intensity of each symptom that the person has given. In BDC, if a person's overall score is higher than 10, then the person is considered depressed. Otherwise, the person is considered as not depressed. According to the BDC score, this study categorized each participant as depressed or not depressed. The developed dataset is available at: https://github.com/Sabab31/Depression-Repository.git.

3.2. Data description

The dataset obtained from the survey has thirty predictor variables and one target variable. The target variable was derived by applying the Burns Depression Checklist (BDC) to each of the participants. The possible value, variable description, variable type of each of the variables are shown in Table 1.

3.3. Data analysis

Among the 604 participants of the accumulated dataset, 397 participants have been found depressed. The prevalence of depression in the participants of the survey is 65.73%, which is much greater than the actual prevalence of depression in the total population of Bangladesh. As the dataset has been collected during the time of COVID-19 and the pandemic has created a psychosocial and socio-economic crisis all over the world, it may have triggered the increase of the prevalence of depression among the participants. 34.27% of the participants of the survey have been found as not depressed. Table 2 shows the distribution of depressed and not depressed participants in the dataset.

Table 3 shows the distribution of depressed and not depressed participants in the dataset according to different criteria. The female participants are more depressed than male participants. The prevalence of depression among the female participants is 69.80%, whereas the prevalence of depression among the male participants is 64.40%. According to the profession, the participants having government jobs are the least affected by depression. Only 40% of government job holder participants are depressed. The appearance of depression is higher in the participants doing business and in the unemployed. The prevalence of depression in the participants doing business is 90%. 75% of unemployed participants are depressed. The percentages of depression among the students, private jobholders, and the participants holding other occupations are 64.85%, 64.63%, and 70.59%, respectively. The divorced participants are the worst victims of depression. The rate of depression is the least in the married participants. The percentages of depression among the married, unmarried, and divorced participants are 63.22%, 65.95%, and 100%, respectively. The participants dwelling in villages are less likely to be affected by depression. The prevalence of depression among the village dwellers is 57.14%. The percentages of depression among the participants living in town and city are 68.75% and 67.85%, respectively.

3.4. Feature selection techniques

While building a machine learning model, only necessary features should be selected. Selecting irrelevant features may cause degradation of the performance of the model. Feature selection helps to eliminate the unnecessary and redundant features that do not have any contribution to the performance of the model.

3.4.1. Select K-best features (SelectKBest)

Select K-Best (SelectKBest) feature selection technique is univariate in nature. Using different univariate statistical tests, it selects K-best features from the feature set (1.13.Feature selection, 2021). This study used chi-square (χ^2) test-based method to select K-best features. The χ^2 test can be performed on non-negative features only. It computes the χ^2 score for each non-negative feature and the target feature. The χ^2 score for *n* pairs of expected and observed frequencies can be derived by Eq. (1).

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(OF_i - EF_i\right)^2}{EF_i} \tag{1}$$

Here, OF_i is the frequency which is observed for the *i*-th value of the feature F and EF_i is the frequency which is expected for the *i*-th value of the feature F.

Besides χ^2 test, other univariate statistical tests like Analysis of Variance (ANOVA) *F*-value, mutual information methods can be used to select *K*-best features.

3.4.2. Minimum redundancy and maximum relevance (mRMR)

Peng et al. (2005) proposed a feature selection algorithm naming Minimum Redundancy and Maximum Relevance (mRMR) algorithm. In the mRMR feature selection method, the features are ranked based on their relevance to the target variable. While ranking the features, the redundancy of the features is also considered. The feature having the maximum relevance to the target variable and minimum redundancy within the features gets the highest rank in mRMR. Both redundancy and relevance are measured using Mutual Information (MI).

Table 1
Variables for predicting depression.

Variable Name	Variable Type	Variable Description	Possible Values
AGERNG	Predictor	Age range (in years) of the participant	16–20, 21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61+
GENDER	Predictor	Gender of the participant	Male, Female
EDU	Predictor	Educational qualification of the participant	SSC, HSC, Graduate, Post Graduate
PROF	Predictor	The profession of the participant	Student, Service holder (Private), Service holder (Government), Businessman, Unemployed, Other
MARSTS	Predictor	Marital status of the participant	Unmarried, Married, Divorced
RESDPL	Predictor	Type of the residing place of the participant	Village, Town, City
LIVWTH	Predictor	It depicts whether the participant lives with his family or not	With Family, Without Family
ENVSAT	Predictor	Whether the participant is satisfied with his living environment or not	Yes, No
POSSAT	Predictor	Whether the participant is satisfied with his current position/ academic achievements or not	Yes, No
FINSTR	Predictor	Whether or not the participant has any financial stress	Yes, No
DEBT	Predictor	Whether the participant has any debt or not	Yes, No
PHYEX	Predictor	The frequency of taking physical exercises of the participant	Never, Sometimes, Regularly
SMOKE	Predictor	Whether the participant smokes or not	Yes, No
DRINK	Predictor	Whether the participant drinks alcohol or not	Yes, No
ILLNESS	Predictor	Whether the participant is suffering from any serious illness or not	Yes, No
PREMED	Predictor	Whether the participant takes any prescribed medication or not	Yes, No
EATDIS	Predictor	Whether the participant is suffering from eating disorders like overeating/ loss of appetite or not	Yes, No
AVGSLP	Predictor	Average hours that the participant sleeps at night	Below 5 h, 5 h, 6 h, 7 h, 8 h, More than 8 h
INSOM	Predictor	Whether or not the participant suffers from insomnia	Yes, No
TSSN	Predictor	Average hours that the participant spends in social network (in a day)	Less than 2 h, 2–4 h a day, 5–7 h a day, 8–10 h a day, More than 10 h a day
WRKPRE	Predictor	Current work or study pressure of the participant	Severe, Moderate, Mild, No Pressure
ANXI	Predictor	Whether the participant recently feels anxiety for something or not	Yes, No
DEPRI	Predictor	Whether or not the participant has recently felt that he/she has been deprived of something that he/she deserves	Yes, No
ABUSED	Predictor	Whether the participant has recently felt abused (physically, sexually, emotionally) or not	Yes, No
CHEAT	Predictor	Whether or not the participant has felt cheated by someone recently	Yes, No
THREAT	Predictor	Whether or not the participant has faced any life-threatening event recently	Yes, No
SUICIDE	Predictor	Whether the participant has any suicidal thought recently or not	Yes, No
INFER	Predictor	Whether the participant recently suffers from inferiority complex or not	Yes, No
CONFLICT	Predictor	Whether or not the participant has recently engaged himself in any kind of conflicts with his friends or family	Yes, No
LOST	Predictor	Whether or not the participant has recently lost someone close to him	Yes, No
DEPRESSED	Target	It is the target variable that portrays whether the participant is depressed or not	0 (Not depressed), 1 (Depressed)

Table 2
Distribution of depressed and not depressed participants in the dataset.

Class	Number of Participants	Percentage (%)
Depressed	397	65.73%
Not Depressed	207	34.27%

Consider two random variables m and n. Their mutual information can be described by the following equation:

$$I(m;n) = \iint p(m,n)\, log \frac{p(m,n)}{p(m)\; p(n)} dm\, dn \tag{2}$$

Here, $p(m,n)$ is the joint probabilistic density, and $p(m)$, $p(n)$ are the marginal probabilistic densities. The feature importance of a feature X_i can be represented by the following equation:

$$f^{mRMR}(X_i) = I(Y, X_i) - \frac{1}{|S|} \sum_{X_s \epsilon S} I(X_s,\ X_i) \tag{3}$$

Here, Y is the target variable, the set of selected features are represented by S. X_s is one of the selected features of set S, and X_i is a feature that is not selected yet.

3.4.3. Boruta feature selection algorithm

The Boruta algorithm works as a wrapper around the Random Forest classification algorithm (Kursa and Rudnicki, 2010). This method aims to select important features by iteratively removing the irrelevant attributes.

The steps of Boruta algorithm are given below:

- Step I: The dataset is extended by creating duplicates of all the features.
- Step II: Shadow features are created by shuffling the values of the duplicated features. Shuffling is performed for removing their correlations with the target variable.
- Step III: A random forest algorithm is applied to the extended dataset, and Z-scores are computed.
- Step IV: The Maximum Z-score among the Shadow Attributes (MZSA) is then detected.
- Step V: If a feature's importance is remarkably less than MZSA, then it is permanently removed from the dataset. On the other hand, if a feature's importance is remarkably greater than MZSA, then it is kept in the dataset.
- Step VI: Discard the shadow features from the dataset.
- Step VII: Repeat the steps until there is no unimportant feature in the dataset or for a predefined number of iterations.

3.5. Synthetic minority oversampling technique (SMOTE)

To enhance the predictive accuracy of the minority class, the dataset must be balanced. Synthetic Minority Oversampling Technique (SMOTE) is used to tackle the class imbalance problem. SMOTE generates synthetic samples of the minority class by operating in feature

Table 3
Distribution of depressed and not depressed participants in the dataset according to different criteria.

Criteria	Category	Total Participants	Number of Depressed Participants	Number of Not Depressed Participants	Depressed (%)	Not Depressed (%)
Gender	Male	455	293	162	64.40%	35.60%
	Female	149	104	45	69.80%	30.20%
Profession	Businessman	10	9	1	90.00%	10.00%
	Service holder (Government)	10	4	6	40.00%	60.00%
	Service holder (Private)	82	53	29	64.63%	35.37%
	Student	441	286	155	64.85%	35.15%
	Unemployed	44	33	11	75.00%	25.00%
	Other	17	12	5	70.59%	29.41%
Marital Status	Married	87	55	32	63.22%	36.78%
	Unmarried	514	339	175	65.95%	34.05%
	Divorced	3	3	0	100.00%	0%
Residing Place	Village	133	76	57	57.14%	42.86%
	Town	160	110	50	68.75%	31.25%
	City	311	211	100	67.85%	32.15%

space. The synthetic samples are introduced along the line adjoining each minority class sample, and its any or all of the K-nearest minority class sample neighbors. For creating a synthetic instance, firstly the difference between the feature vector of the minority class instance under consideration and its nearest neighbor is multiplied by a random number between 0 and 1. Then the multiplied result is added to the feature vector under consideration, and thus, a synthetic instance of minority class is produced (Chawla et al., 2002). Consider that f_i is the feature vector of the minority class sample under consideration and f_{near} is one of the K-nearest neighbors of f_i. The generated synthetic sample f_{new} can be represented by Eq. (4).

$$f_{new} = f_i + \left(f_i - f_{near}\right) \times R \tag{4}$$

Here, R is a random number between 0 and 1.

3.6. Machine learning techniques for depression detection

In order to predict the existence of depression, this study has used six different machine learning classifiers, namely: K-Nearest Neighbor (KNN), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), Bagging, and Weighted Voting classifier. The details of these classifiers are described below.

3.6.1. K-nearest neighbor (KNN)

KNN, which is non-parametric in nature, is a commonly used distance-based algorithm for solving both classification and regression problems. As KNN is also an instance-based learning algorithm, a new instance is classified by comparing it with the most similar instances in the training set. The similarity between the instances can be estimated using distance measures. The primary determinant of this classifier is the K value, which refers to the number of nearest neighbors. The new instance is simply allocated to the class of its nearest neighbor in case the value of K is 1 (Ali et al., 2019). Several distance measurements can be used, such as Manhattan Distance, Minkowski Distance, Hamming Distance, Euclidean Distance, etc. in KNN. This study has used the Minkowski Distance function. Consider two points, $E(e_1, e_2, \dots, e_n)$ and $F(f_1, f_2, \dots, f_n)$. Then the Minkowski Distance between these two points can be expressed as the following equation:

$$distance(E, F) = \left(\sum_{i=1}^{n} \left(\left|e_i - f_i\right|\right)^q \right)^{\frac{1}{q}} \tag{5}$$

Here, q is the order of the Minkowski distance.

3.6.2. Adaptive boosting (AdaBoost)

Adaptive Boosting classifier, or simply known as AdaBoost classifier, is an ensemble learning method. To construct a strong classifier, it utilizes a collection of weak classifiers. In this approach, a weak classifier learns from the errors of its prior classifier. Consider a dataset with n samples. Initially, each sample is assigned a weight of $(1/n)$. And with this dataset, a weak classifier is constructed. The total error ε of this classifier is computed. And using this total error, the influence of this classifier α in classifying the data samples is measured by the following equation:

$$\alpha = \frac{1}{2} \ln \left(\frac{1 - \varepsilon}{\varepsilon} \right) \tag{6}$$

Here, α is used for changing the weight of the samples of the dataset, resulting in creating a new dataset. Later this new dataset is used to construct the next weak classifier. Finally, the majority votes of the weak classifiers decide the class label of a sample (Vasilić et al., 2018).

3.6.3. Gradient boosting (GB)

Gradient Boosting (GB) classifier ensembles a set of weak models to sequentially create new models. Each new model tries to minimize the loss function. GB uses the gradient descent method to compute the loss function. To avoid overfitting problems, boosting should be stopped timely using stopping criteria. A maximum number of models created or a threshold on the predictive accuracy can be used as the stopping criteria (Rahman et al., 2020).

GB constructs a function , $f : R^n \to R$ using the linear combination of the weak models $g : R^n \to R$ as the following equation:

$$f(x) = \sum_{i=1}^{M} w_i g_i\left(x; \theta_i\right) \tag{7}$$

Here, $x \epsilon R^n$ represents the input vector. The number of weak models is represented by $M . w_i \epsilon R$ exhibits the weight. The function is constructed by selecting iteratively the weight w_i and the parameter θ_i of a weak learner to minimize the loss function (Son et al., 2015).

3.6.4. Extreme gradient boosting (XGBoost)

In terms of scalability and speed, Extreme Gradient Boosting improves the performance of Gradient Boosting Decision Tree (GBDT). The objective function of XGBoost is computed by adding the regularization with the loss function. The following equation shows the objective function of XGBoost:

$$I(\theta) = L(\theta) + \Omega(\theta), \ where\ L(\theta) = l(z, y),\ \Omega(\theta) = \alpha T + \left(\frac{1}{2}\right)\lambda ||\omega||^2 \tag{8}$$

In the above equation, $L(\theta)$ represents the loss function and the regularization is represented by $\Omega(\theta)$. θ represents the parameters trained from the given data (Chen et al., 2018; Zhang et al., 2018). The learning rate, number of leaves, weight of the leaves, and regularized parameter are represented by α, T, ω, and λ, respectively. If z is the predicted

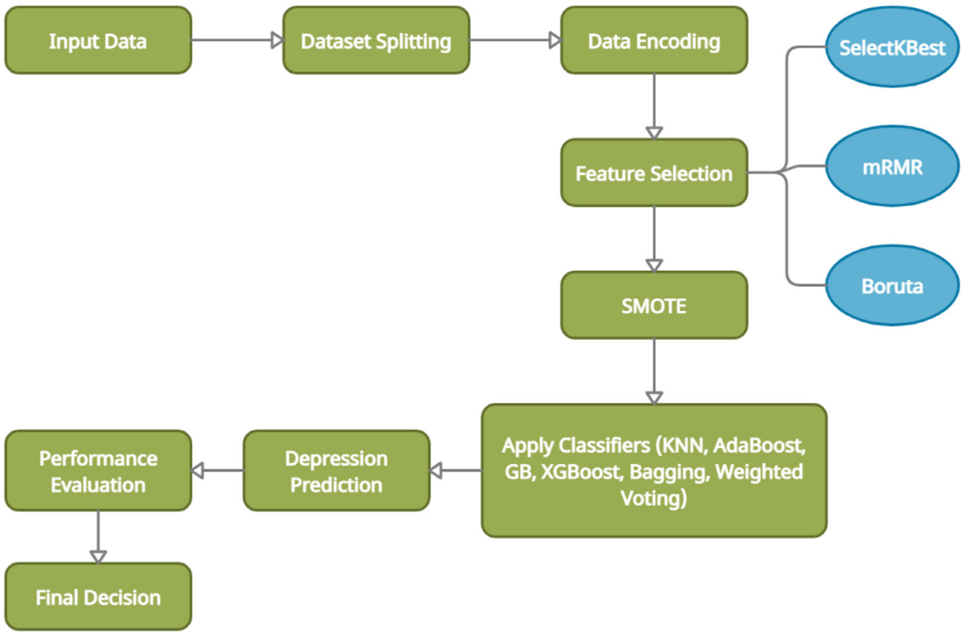


Fig. 1. Implementation steps for predicting the appearance of depression.

output and y is the actual output, then $L(\theta)$ can be computed by either mean square loss function or logistic loss function. Using mean square loss function, $L(\theta)$ can be represented by the following equation:

$$L(\theta) = l(z, y) = (z - y)^2 \tag{9}$$

And, using the logistic loss function, $L(\theta)$ can be represented by the following equation:

$$L(\theta) = l(z, y) = y \ln(1 + e^{-z}) + (1 - y) \ln(1 + e^{z}) \tag{10}$$

3.6.5. *Bagging classifier*

Bagging has been introduced based on the principles of bootstrapping and aggregating. In the Bagging classifier, bootstrap datasets are constructed from the training dataset. Each of these bootstrap datasets is then used to train different classifiers. And finally, the results of these classifiers are aggregated to get the final prediction. In the bootstrap dataset, misleading training objects are often avoided. And, aggregating the classifiers often gives better performance than a single classifier (Skurichina and Duin, 2002). As both of these features are combined in the Bagging classifier, Bagging often shows superior performance than the other classifiers. Bagging works in the following steps:

- Step I: For $n = 1, 2, \dots, N$:
  a. Create a bootstrap replicate X^n, from the training dataset X.
  b. Using X^n, construct a model $C^n(x)$.
- Step II: Aggregate the models $C^n(x)$ where $n = 1, 2, \dots, N$ by majority voting or averaging.

This study has performed the Bagging of Multilayer Perceptron (MLP) classifiers.

3.6.6. Weighted voting classifier

Voting classifier aggregates the output of different base classifiers, and it is one of the simplest ensemble learning techniques (Kumar et al., 2017). The classifiers can be heterogeneous or homogeneous. In the hard voting approach, the class label of a sample is decided based on the majority voting of the classifiers. On the contrary, in the soft voting approach, the predicted probabilities of the base classifiers decide the class label of the sample. In the weighted voting approach, a weight is assigned to each of the base classifiers. For soft voting with weighted classifiers, the predicted class label y_p of a sample can be derived by the following equation:

$$y_p = arg \max_i \sum_{k=1}^{m} w_k p_{ik} \tag{11}$$

Here, m is the number of the base classifiers and w_k represents the assigned weight for the $k-th$ classifier. The probability of the $i-th$ label of the $k-th$ classifier is represented by p_{ik}.

3.7. *Implementation procedures*

This work has been implemented using python and the Scikit-learn library. The acquired dataset contains data of 604 participants. It consists of one target variable and thirty predictor variables. The target variable indicates whether or not the participant is depressed. After preparing the dataset, it is fed into the proposed system. Fig. 1 gives the pictorial representation of the implementation procedures of this study.

The stepwise procedures for implementing this study are described in the next subsections.

3.7.1. *Dataset splitting*

Firstly, the obtained dataset has been split into training and test data. This study has used 80% data of the dataset as training data. And the rest 20% data of the dataset has been used for testing purposes.

3.7.2. *Data encoding*

After the completion of the Dataset Splitting technique, Data Encoding is performed on the obtained training and test datasets. Using numeric data, the majority of machine learning algorithms demonstrate better results. In the Data Encoding step, the categorical data of the training and test datasets have been converted into their numeric counterpart using the Label Encoder of the Scikit-learn library.

3.7.3. *Modifying the training and testing dataset using feature selection*

The presence of irrelevant features degrades the classifiers' efficiency. For extracting the relevant and necessary features from the dataset, three feature selection techniques have been used in this study separately.

Both the SelectKBest and mRMR feature selection techniques have chosen fifteen predictor variables separately for performing classification efficiently. And using the Boruta feature selection algorithm, thirteen most relevant predictor variables have been extracted from the total

Table 4
Selected features using feature selection techniques.

Feature Selection Technique	Total Features	Selected Features
SelectKBest	15	DEPRI, INFER, POSSAT, ANXI, ABUSED, CHEAT, CONFLICT, FINSTR, SUICIDE, ENVSAT, INSOM, THREAT, LOST, DEBT, EATDIS
mRMR	15	DEPRI, POSSAT, ANXI, INFER, ENVSAT, CHEAT, FINSTR, ABUSED, CONFLICT, SUICIDE, LOST, INSOM, THREAT, WRKPRE, DEBT
Boruta	13	ENVSAT, POSSAT, FINSTR, INSOM, ANXI, DEPRI, ABUSED, CHEAT, THREAT, SUICIDE, INFER, CONFLICT, LOST

Table 5
Percentage of depressed and not depressed participants in the training datasets.

Scenario	Depressed (%)	Not Depressed (%)	Total (%)
Before SMOTE	66.87%	33.13%	100.00%
After SMOTE	50.00%	50.00%	100.00%

predictor variables of the dataset. Table 4 shows the list of the selected features using these three feature selection techniques.

For avoiding information leakage, feature selection algorithms have been applied only to the training dataset. Considering the features selected by these three feature selection algorithms, three pairs of new training and test datasets have been generated from the original training and test datasets, respectively.

3.7.4. Applying SMOTE on the training data

Training a classifier with an imbalanced dataset leads to biased and inaccurate predictions. In the training datasets, the percentages of depressed and not depressed participants are 66.87% and 33.13%, respectively. As the training datasets are highly imbalanced, SMOTE has been used to remove their class imbalance problem. Table 5 shows the percentage of depressed and not depressed participants in the training datasets, before and after performing SMOTE.

3.7.5. Training and testing for predicting depression

In this step, the classifiers namely, KNN, AdaBoost, GB, XGBoost, Bagging, and Weighted Voting classifiers are trained with the training datasets. Following the training of these classifiers, each of them has been used to predict the depression of the participants of the test datasets.

3.7.6. Performance evaluation and final decision

In this step, different performance metrics like accuracy, precision, sensitivity, specificity, F1-score, and area under the curve (AUC) have been determined for all of the models constructed in the previous subsection. Efficacies of these models are evaluated based on these performance metrics. Finally, the best model for predicting depression has been chosen, according to the outcomes of these performance metrics.

4. Results and discussion

The responses of 121 participants have been used for testing purposes. Among them, 61.16% of the participants are originally depressed. And 38.84% of the participants are not depressed. For evaluating the performance of the models, accuracy, sensitivity, specificity, precision, and F1-score of the models have been computed by using the following formulas:

$$Accuracy\ (\%) = \frac{True\ Positive(TP) + True\ Negative(TN)}{True\ Positive(TP) + True\ Negative(TN) + False\ Positive(FP) + False\ Negative(FN)} \times 100 \quad (12)$$

$$Sensitivity\ (\%) = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Negative(FN)} \times 100 \quad (13)$$

$$Specificity\ (\%) = \frac{True\ Negative(TN)}{True\ Negative(TN) + False\ Positive(FP)} \times 100 \quad (14)$$

$$Precision\ (\%) = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \times 100 \quad (15)$$

$$F1-score\ (\%) = \frac{2\ \times True\ Positive(TP)}{2\ \times True\ Positive(TP) + False\ Positive(FP) + False\ Negative(FN)} \times 100 \quad (16)$$

Here,

- True Positive (TP): When a depressed participant is predicted as depressed by the classifier, then the outcome is known as True Positive (TP).
- True Negative (TN): When a participant without depression is predicted as not depressed by the classifier, then the outcome is known as True Negative (TN).
- False Positive (FP): When a participant without depression is predicted as depressed by the classifier, then the outcome is known as False Positive (FP).
- False Negative (FN): When a participant with depression is predicted as not depressed by the classifier, then the outcome is known as False Negative (FN).

Table 6 shows the confusion matrices of the predicted results for each of the classifiers using different feature selection techniques.

The measured accuracy, sensitivity, specificity, precision, F1-score, and area under the curve (AUC) of the classifiers for the constructed models are shown in Table 7.

It is observed from Table 7 that the Bagging classifier has shown the best result with an accuracy of 89.26% without performing feature selection. Here, the least accuracy has been shown by KNN. It has attained an accuracy of 66.94%. The achieved accuracies of AdaBoost, GB, XGBoost, and Weighted Voting classifiers are 87.60%, 86.78%, 85.95%, and 86.78%, respectively.

By applying different feature selection techniques, accuracies of all of these classifiers have been increased dramatically. While using the SelectKBest feature selection technique, the AdaBoost has outperformed the other classifiers in terms of accuracy. It has achieved an accuracy of 92.56%. By applying the SelectKBest feature selection technique, the accuracies of the other classifiers namely, KNN, GB, XGBoost, Bagging, and Weighted Voting classifiers are 85.12%, 91.74%, 86.78%, 90.91%, and 91.74%, respectively.

In the case of using the mRMR feature selection technique, KNN, AdaBoost, GB, XGBoost, Bagging, and Weighted Voting classifiers have attained accuracies of 84.30%, 91.74%, 90.08%, 90.08%, 90.08%, and 90.08%, respectively.

Using the Boruta feature selection technique, AdaBoost has shown superior performance than the other classifiers in terms of accuracy. Here, the achieved accuracies of KNN, AdaBoost, GB, XGBoost, Bagging,

Table 6
Confusion matrices of the classifiers using different feature selection techniques.

Classifier Name	Feature Selection Technique	TP	TN	FP	FN
KNN	Without Feature Selection	40	41	6	34
AdaBoost		68	38	9	6
GB		67	38	9	7
XGBoost		67	37	10	7
Bagging		69	39	8	5
Weighted Voting (With GB, Bagging, and AdaBoost)		67	38	9	7
KNN	SelectKBest	59	44	3	15
AdaBoost		68	44	3	6
GB		68	43	4	6
XGBoost		67	38	9	7
Bagging		67	43	4	7
Weighted Voting (With GB, Bagging, and AdaBoost)		68	43	4	6
KNN	mRMR	59	43	4	15
AdaBoost		67	44	3	7
GB		68	41	6	6
XGBoost		68	41	6	6
Bagging		67	42	5	7
Weighted Voting (With GB, Bagging, and AdaBoost)		68	41	6	6
KNN	Boruta	61	42	5	13
AdaBoost		69	42	5	5
GB		68	42	5	6
XGBoost		66	39	8	8
Bagging		68	41	6	6
Weighted Voting (With GB, Bagging, and AdaBoost)		68	42	5	6

Table 7
Performance of the classifiers using different feature selection techniques.

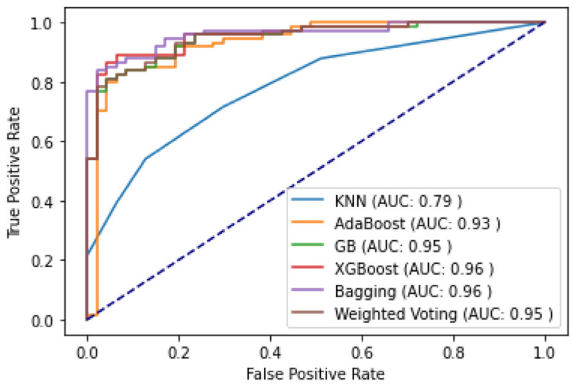
Classifier Name	Feature Selection Technique	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)	AUC
KNN	Without Feature Selection	66.94%	54.05%	87.23%	86.96%	66.67%	0.79
AdaBoost		87.60%	91.89%	80.85%	88.31%	90.07%	0.93
GB		86.78%	90.54%	80.85%	88.16%	89.33%	0.95
XGBoost		85.95%	90.54%	78.72%	87.01%	88.74%	0.96
Bagging		89.26%	93.24%	82.98%	89.61%	91.39%	0.96
Weighted Voting (With GB, Bagging, and AdaBoost)		86.78%	90.54%	80.85%	88.16%	89.33%	0.95
KNN	SelectKBest	85.12%	79.73%	93.62%	95.16%	86.76%	0.93
AdaBoost		92.56%	91.89%	93.62%	95.77%	93.79%	0.96
GB		91.74%	91.89%	91.49%	94.44%	93.15%	0.96
XGBoost		86.78%	90.54%	80.85%	88.16%	89.33%	0.94
Bagging		90.91%	90.54%	91.49%	94.37%	92.41%	0.96
Weighted Voting (With GB, Bagging, and AdaBoost)		91.74%	91.89%	91.49%	94.44%	93.15%	0.96
KNN	mRMR	84.30%	79.73%	91.49%	93.65%	86.13%	0.91
AdaBoost		91.74%	90.54%	93.62%	95.71%	93.06%	0.96
GB		90.08%	91.89%	87.23%	91.89%	91.89%	0.96
XGBoost		90.08%	91.89%	87.23%	91.89%	91.89%	0.95
Bagging		90.08%	90.54%	89.36%	93.06%	91.78%	0.96
Weighted Voting (With GB, Bagging, and AdaBoost)		90.08%	91.89%	87.23%	91.89%	91.89%	0.96
KNN	Boruta	85.12%	82.43%	89.36%	92.42%	87.14%	0.91
AdaBoost		91.74%	93.24%	89.36%	93.24%	93.24%	0.96
GB		90.91%	91.89%	89.36%	93.15%	92.52%	0.96
XGBoost		86.78%	89.19%	82.98%	89.19%	89.19%	0.94
Bagging		90.08%	91.89%	87.23%	91.89%	91.89%	0.96
Weighted Voting (With GB, Bagging, and AdaBoost)		90.91%	91.89%	89.36%	93.15%	92.52%	0.96

and Weighted Voting classifiers are 85.12%, 91.74%, 90.91%, 86.78%, 90.08%, and 90.91%, respectively.

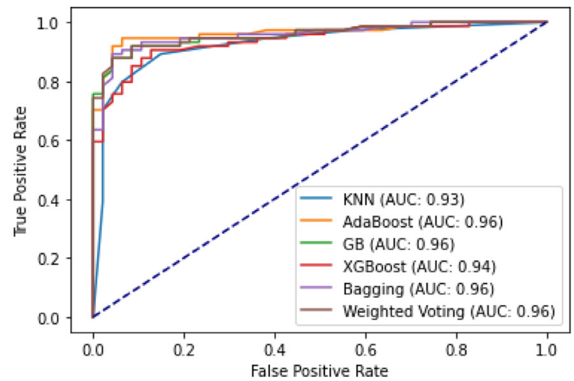
Sensitivity and specificity play a great role in evaluating the performance of a model. A model with a higher sensitivity has a higher ability to identify participants with depression, and a model with a higher specificity has a higher ability to identify participants without depression.

Without applying feature selection techniques, the Bagging classifier has achieved the highest sensitivity of 93.24%. But the attained specificity of the Bagging classifier is only 82.98%. On the other hand, the KNN classifier has shown the highest specificity of 87.23% without performing feature selection. But in terms of sensitivity, KNN has shown the worst performance. Here, the achieved sensitivity of KNN is 54.05%.

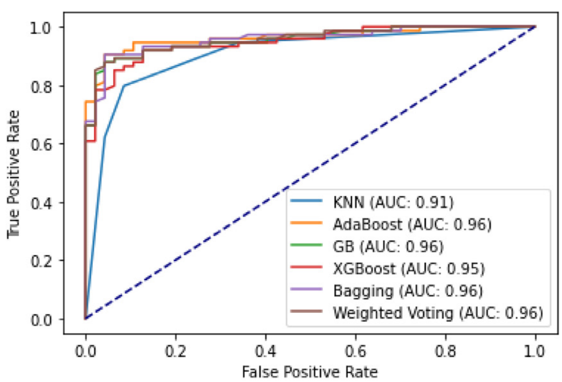
While using the SelectKBest feature selection technique, the AdaBoost classifier has obtained the highest sensitivity and specificity of 91.89% and 93.62%, respectively. AdaBoost has shown the highest specificity and sensitivity of 89.36% and 93.24%, respectively, in the case of using the Boruta feature selection technique also. While applying mRMR as a feature selection technique, GB, XGBoost, and Weighted Voting classifiers have gained the highest sensitivity of 91.89%. They have achieved specificities of 87.23%. Here, AdaBoost has gained the



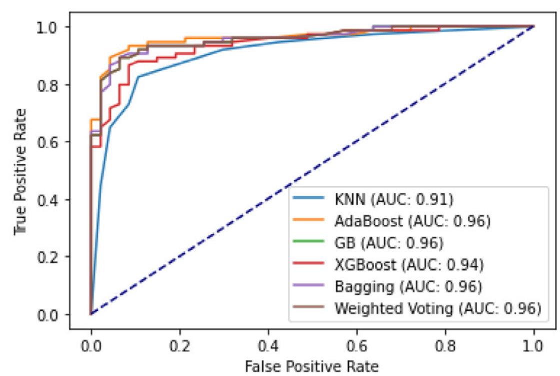
(a) ROC Curves of the Classifiers Without Feature Selection



(b) ROC Curves of the Classifiers with SelectKBest Feature Selection



(c) ROC Curves of the Classifiers with mRMR Feature Selection



(d) ROC Curves of the Classifiers with Boruta Feature Selection

Fig. 2. ROC curves of the classifiers for different feature selection techniques.

highest specificity of 93.62%. On the other hand, it has achieved a sensitivity of 90.54%. So, by analyzing the performance of the classifiers in terms of sensitivity and specificity, it can be ensured that by applying the above feature selection techniques, both the sensitivity and specificity of the classifiers have been improved.

Also, the Area Under Curve (AUC) values of these models have been calculated. If a model's AUC value is 1, then it is assumed to be a perfect model or classifier. When the AUC value of a model is 0.5, it cannot distinguish between the samples of different classes. So, a higher AUC value of a model is always desired. The mentioned feature selection techniques have enhanced the performance of the classifiers in terms of AUC, Precision, and F1-score also.

To demonstrate the trade-off between the sensitivity and specificity of the models, the Receiver Operator Characteristic (ROC) curve is used. It is a two-dimensional graph where the x-axis represents the False Positive Rate, and the y-axis represents the True Positive Rate. The closer the ROC curve of a classifier is to the top-left corner of the graph, the better the performance of the classifier is. Fig. 2 (a)–(d) shows the ROC curves of these classifiers using different feature selection techniques. Fig. 2 reveals that the ROC curves of the classifiers have moved closer to the graph's upper left corner after applying feature selection techniques.

Reducing the number of features using different feature selection techniques have made the classifiers faster. From Fig. 3, it is revealed that different feature selection techniques have reduced the training time of the classifiers significantly.

Fig. 3 shows that for minimizing the training time of the KNN classifier, mRMR has outperformed other feature selection techniques, and for minimizing the training time of the AdaBoost, GB, and XGBoost classifiers, SelectKBest feature selection technique has shown the best performance. Boruta feature selection technique has minimized the training time of Bagging and Weighted Voting classifiers more than other feature selection techniques.

So, from the above discussions, it is found that the SelectKBest feature selection technique has outperformed other feature selection techniques in most of the cases in order to minimize the training time of the classifiers.

Comparing the results of different models, it can be concluded that the AdaBoost classifier with the SelectKBest feature selection algorithm has surpassed the other models in terms of accuracy, AUC value, ROC-curves, and other measured performance metrics. Adaboost classifier with SelectKBest feature selection technique has shown the highest accuracy of 92.56%. This model has also attained the highest AUC value of 0.96.

Boruta feature selection algorithm discards the irrelevant features based on the Maximum Z-score. In this study, this algorithm has discarded seventeen predictor variables based on the Maximum Z-score and found the remaining thirteen predictor variables to be significant. The SelectKBest algorithm has ranked the predictor variables based on the χ^2 score. On the other hand, the mRMR algorithm has ranked the predictor variables based on their relevance and redundancy. This study has selected the top fifteen predictor variables from the rankings of both SelectKBest and mRMR algorithms, as in both cases using the top fifteen predictor variables as input variables for the classifiers yields the best results. In Table 4, the top fifteen features selected by the SelectKBest and the mRMR algorithms are arranged according to their rankings. Table 8 shows the accuracy of different classifiers by taking a different number of features from the rankings for both SelectKBest and mRMR algorithms.

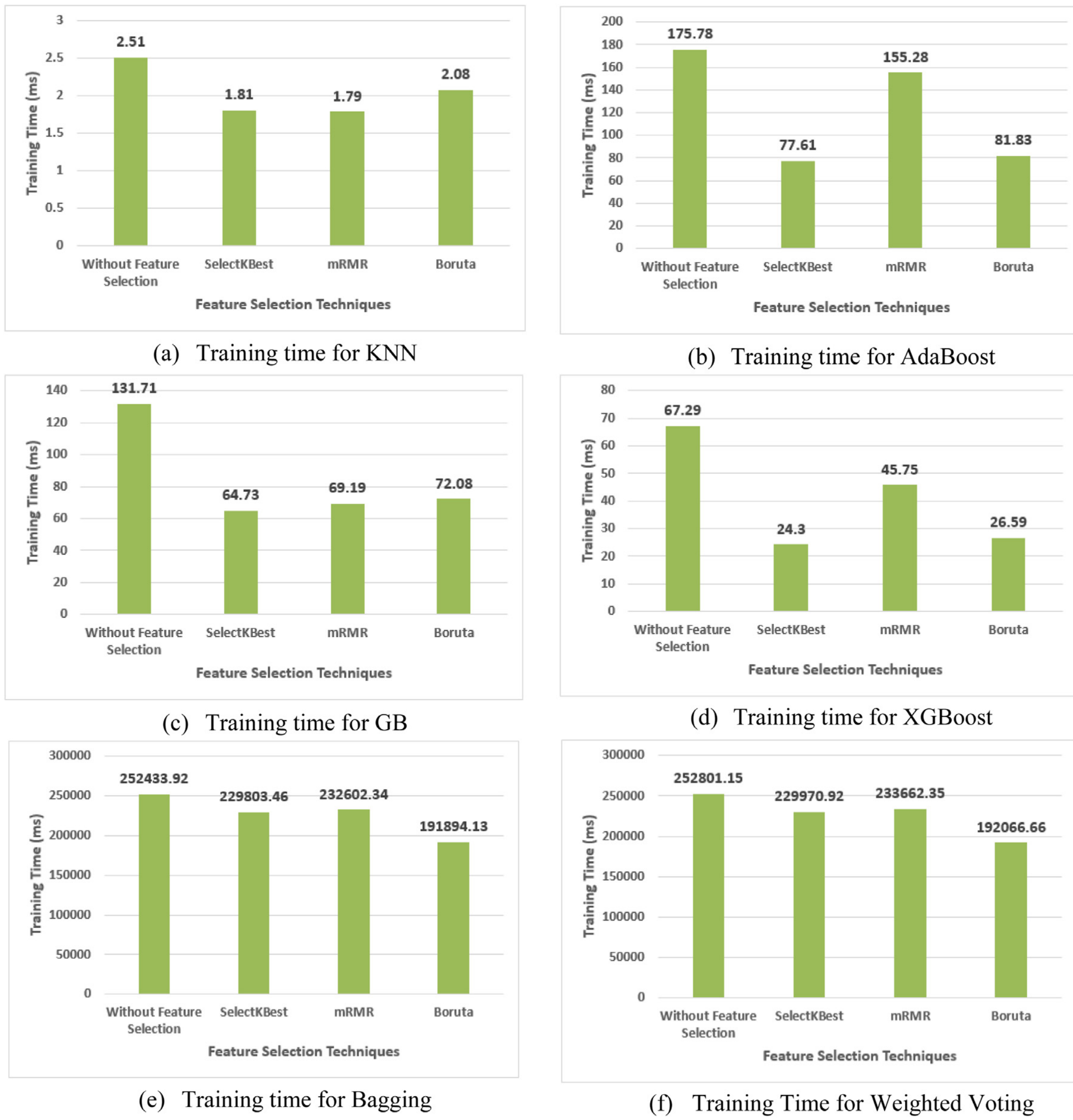


Fig. 3. Training time of the classifiers for different feature selection techniques.

Table 8
Accuracy of the classifiers for different number of features.

Feature Selection Technique	Classifier Name	Accuracy (%)					
		Top 10 Features	Top 11 Features	Top 12 Features	Top 13 Features	Top 14 Features	Top 15 Features
SelectKBest	KNN	80.17%	80.99%	80.99%	85.12%	85.95%	85.12%
	AdaBoost	86.78%	89.26%	87.60%	91.74%	92.56%	92.56%
	GB	85.12%	85.95%	85.95%	90.91%	89.26%	91.74%
	XGBoost	82.64%	85.12%	85.95%	86.78%	87.60%	86.78%
	Bagging	85.12%	85.12%	86.78%	90.08%	90.91%	90.91%
	Weighted Voting (With GB, Bagging, and AdaBoost)	85.12%	85.95%	85.95%	90.91%	90.08%	91.74%
mRMR	KNN	80.99%	84.30%	80.17%	83.47%	83.47%	84.30%
	AdaBoost	86.78%	87.60%	90.08%	91.74%	91.74%	91.74%
	GB	83.47%	86.78%	90.08%	89.26%	89.26%	90.08%
	XGBoost	84.30%	88.43%	90.08%	87.60%	89.26%	90.08%
	Bagging	85.12%	87.60%	90.08%	90.08%	89.26%	90.08%
	Weighted Voting (With GB, Bagging, and AdaBoost)	83.47%	86.78%	90.08%	89.26%	89.26%	90.08%

Table 9
Comparative analysis of this work with other existing works.

Works Done	Total Number of Participants	Participants Dealt with	Collected Information Type	Extracted the Most Relevant Factors Causing Depression	Depression Screening Scale	Accuracy (%)
Cvetković (2017)	84	Cancer patients	Socio-demographic information	✖	Standardized Beck Depression Inventory (Standardized BDI)	Not mentioned
Na et al. (2020)	6588	Korean participants of different age ranges and occupations	Socio-demographic, economic, clinical information	✔	Center for Epidemiologic Studies Depression Scale (CES-D-11)	86.20%
Sau and Bhakta (2017)	520	Geriatric patients	Socio-demographic, socio-economic, and health related information	✔	Hospital Anxiety and Depression Scale (HADS)	89.00%
Hatton et al. (2019)	284	Geriatric patients	Socio-demographic, psychometric information	✖	Patient Health Questionnaire (PHQ-9)	74.00%
Natarajan et al. (2017)	173	Mothers in their postpartum period	Demographic information and postpartum depression related risk factors	✖	Not applicable	Not mentioned
Jiménez-Serrano et al. (2015)	1397	Mothers in their postpartum period	Socio-economic, clinical, and psychometric information	✖	Edinburgh Postnatal Depression Scale (EPDS)	79.00%
Ekong and Onibere (2015)	80	Not mentioned	Psychological and physiological information	✖	Modified Patient Mental Health Questionnaire	92.40%
Islam et al. (2018)	Not mentioned	Facebook users	Facebook posts and comments	Not applicable	Not applicable	Not mentioned
Sau and Bhakta (2019)	470	Seafarers	Socio-demographic, occupational information	✔	Hamilton Depression Rating Scale (HAM-D)	89.30%
Priya et al. (2020)	348	Males and females aged between 20 and 60 years of different occupations	Psychological information	✖	Depression, Anxiety and Stress Scale (DASS-21)	85.50%
Choudhury et al. (2019)	577	Bangladeshi undergraduate students	Not mentioned	✖	Beck Depression Inventory-II (BDI-II), Depression, Anxiety and Stress Scale-Bengali Version (DASS-21-BV)	75.00%
This work	604	Bangladeshi citizens of different age ranges, occupations, and socio-economic backgrounds	Socio-demographic and psychosocial information	✔	BDC	92.56%

So, from Table 8, it can be undoubtedly stated that the majority of the classifiers shows the best performance while using the top fifteen features as input variables for both SelectKBest and mRMR algorithms.

5. Comparative study with existing works

For understanding the quality of this study, it is necessary to compare the results and contribution of this study with other existing works. Most of the previous works have been performed to predict depression among people of a certain age group, occupation, or health condition. A few of them have extracted the most relevant socio-demographic and psychosocial factors that cause depression. But this study has been conducted to predict depression among people of different age ranges, professions, and socioeconomic backgrounds. This study has also identified the most significant socio-demographic and psychosocial factors that cause depression. Table 9 shows a comparative analysis of this study with other existing works.

Table 9 depicts that the work presented in this study has filled the gaps of the previous works significantly. The achieved 92.56% accuracy of this study is both decent and promising enough.

6. Conclusion

Various factors can play roles in forming depression in a person. This study has tried to find out the most common factors that cause depression. Firstly, a dataset has been created, consisting of thirty socio-demographic, and psychosocial factors of 604 participants to screen depression. Different feature selection techniques have extracted the most important demographic, and psychosocial factors responsible for forming depression. These feature selection techniques have not only boosted the training speed of the classifiers but also helped the classifiers to screen depression more precisely. To ascertain the presence of depression, this research has used six different machine learning classifiers. By observing the outcomes of various models presented in this study, it can be confirmed that the AdaBoost classifier with the SelectKBest feature selection technique is almost the perfect model to predict depression among the participants. It has obtained an accuracy of 92.56%.

This study has considered BDC as the ground truth to diagnose depression, which is indeed a limitation of this research. No biological marker was included in the dataset for predicting depression. Different biological factors play a significant role in predicting depression in an individual. By including these biological factors, the model could predict depression more efficiently.

This work has only predicted the presence of depression in individuals. In the future, this study can be extended to identify the severity of depression in a person. As various biological factors have a remarkable impact on forming depression among the persons, different biological aspects of the participants can be included in the latter study. Several studies indicate that using different dimensionality reduction algorithms on the data preprocessing steps improves the models' performances. These approaches can be applied, and their outcomes can be compared with this current study's results in the future.

Data availability

The dataset related to this article are available via the GitHub repository (https://github.com/Sabab31/Depression-Repository.git).

Code availability

All source codes are available upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1.13. Feature selection, 2021. Scikit-Learn. Retrieved from https://scikit-learn.org/stable/modules/feature_selection.html#univariate-feature-selection. Accessed January 29, 2021.

Ali, N., Neagu, D., Trundle, P., 2019. Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Appl. Sci. 1 (12), 1–15.

Burns, D.D., 1999. Feeling Good: The New Mood Therapy. Avon, New York (revised and updated).

Burns, D., Westra, H., Trockel, M., Fisher, A., 2013. Motivation and changes in depression. Cogn. Ther. Res. 37 (2), 368–379.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, Z., Jiang, F., Cheng, Y., Gu, X., Liu, W., Peng, J., 2018. XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud. In: Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (Bigcomp). IEEE, pp. 251–256.

Choudhury, A.A., Khan, M.R.H., Nahim, N.Z., Tulon, S.R., Islam, S., Chakrabarty, A, 2019. Predicting depression in Bangladeshi undergraduates using machine learning. In: Proceedings of the 2019 IEEE Region 10 Symposium (TENSYMP). IEEE, pp. 789–794.

Cvetković, J., 2017. Breast cancer patients' depression prediction by machine learning approach. Cancer Investig. 35 (8), 569–572.

Ekong, V.E., Onibere, E.A., 2015. A soft computing model for depression prediction. Egypt. Comput. Sci. J. 39 (4).

Hatton, C.M., Paton, L.W., McMillan, D., Cussens, J., Gilbody, S., Tiffin, P.A., 2019. Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. J. Affect. Disord. 246, 857–860.

Holtz, C., Sowell, R., VanBrackle, L., Velasquez, G., Hernandez-Alonso, V., 2014. A quantitative study of factors influencing quality of life in rural Mexican women diagnosed with HIV. J. Assoc. Nurses AIDS Care 25 (6), 555–567.

Iliou, T., Konstantopoulou, G., Ntekouli, M., Lymperopoulou, C., Assimakopoulos, K., Galiatsatos, D., Anastassopoulos, G., 2019. ILIOU machine learning preprocessing method for depression type prediction. Evol. Syst. 10 (1), 29–39.

Islam, M.R., Kabir, M.A., Ahmed, A., Kamal, A.R.M., Wang, H., Ulhaq, A, 2018. Depression detection from social network data using machine learning techniques. Health Inf. Sci. Syst. 6 (1), 1–12.

Jabbar, S.A., Zaza, H.I., 2019. Post-traumatic stress and depression (PSTD) and general anxiety among Iraqi refugee children: a case study from Jordan. Early Child Dev. Care 189 (7), 1114–1134.

Jiménez-Serrano, S., Tortajada, S., García-Gómez, J.M., 2015. A mobile health application to predict postpartum depression based on machine learning. Telemed. e-Health 21 (7), 567–574.

Kumar, U.K., Nikhil, M.S., Sumangali, K., 2017. Prediction of breast cancer using voting classifier technique. In: Proceedings of the 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM). IEEE, pp. 108–114.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. J. Stat. Softw. 36 (11), 1–13.

Na, K.S., Cho, S.E., Geem, Z.W., Kim, Y.K., 2020. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. Neurosci. Lett. 721, 134804.

Natarajan, S., Prabhakar, A., Ramanan, N., Bagilone, A., Siek, K., Connelly, K., 2017. Boosting for postpartum depression prediction. In: Proceedings of the 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE, pp. 232–240.

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., Sartori, G., 2020. Machine learning in psychometrics and psychological research. Front. Psychol. 10, 2970.

Otte, C., Gold, S.M., Penninx, B.W., Pariante, C.M., Etkin, A., Fava, M., Mohr, D.C., Schatzberg, A.F., 2016. Major depressive disorder. Nat. Rev. Dis. Primers 2 (1), 1–20.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal. Mach. Intell. 27 (8), 1226–1238.

Priya, A., Garg, S., Tigga, N.P., 2020. Predicting anxiety, depression and stress in modern life using machine learning algorithms. Procedia Comput. Sci. 167, 1258–1267.

Rahman, S., Irfan, M., Raza, M., Moyeezullah Ghori, K., Yaqoob, S., Awais, M., 2020. Performance analysis of boosting classifiers in recognizing activities of daily living. Int. J. Environ. Res. Public Health 17 (3), 1082.

Sau, A., Bhakta, I., 2017. Predicting anxiety and depression in elderly patients using machine learning technology. Healthc. Technol. Lett. 4 (6), 238–243.

Sau, A., Bhakta, I., 2019. Screening of anxiety and depression among seafarers using machine learning technology. Inf. Med. Unlocked 16, 100228.

Shah, M., Ali, M., Ahmed, S., Arafat, S.M., 2017. Demography and risk factors of suicide in Bangladesh: a six-month paper content analysis. Psychiatry J. 2017.

Skurichina, M., Duin, R.P., 2002. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal. Appl. 5 (2), 121–135.

Son, J., Jung, I., Park, K., Han, B., 2015. Tracking-by-segmentation with online gradient boosting decision tree. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3056–3064.

Thapar, A., Collishaw, S., Pine, D.S., Thapar, A.K., 2012. Depression in adolescence. The Lancet 379 (9820), 1056–1067.

Vasilić, P., Vujnović, S., Popović, N., Marjanović, A., Đurović, Ž., 2018. Adaboost algorithm in the frame of predictive maintenance tasks. In: Proceedings of the 23rd International Scientific-Professional Conference on Information Technology (IT). IEEE, pp. 1–4.

Vos, T., Allen, C., Arora, M., Barber, R.M., Bhutta, Z.A., Brown, A., Boufous, S., 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. The Lancet 388 (10053), 1545–1602.

Whooley, M.A., Wong, J.M., 2013. Depression and cardiovascular disorders. Annu. Rev. Clin. Psychol. 9, 327–354.

World Health Organization, 2021. Eastern Mediterranean Region. World Health Organization. Retrieved from http://www.emro.who.int/mnh/what-you-can-do/index.html#accordionpan4. Accessed January 29.

World Health Organization, 2017. Depression and Other Common Mental Disorders: Global Health Estimates (No. WHO/MSD/MER/2017.2). World Health Organization.

Zarandi, M.F., Soltanzadeh, S., Mohammadi, A., Castillo, O., 2019. Designing a general type-2 fuzzy expert system for diagnosis of depression. Appl. Soft Comput. 80, 329–341.

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., Si, Y., 2018. A data-driven design for fault detection of wind turbines using random forests and XGboost. IEEE Access 6, 21020–21031.