Proceedings of the Second International Conference on Automation, Computing and Renewable Systems (ICACRS-2023)
IEEE Xplore Part Number: CFP23CB5-ART; ISBN: 979-8-3503-4023-5

2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) | 979-8-3503-4023-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACRS58579.2023.10405040

AI-based Depression Detection using Profile Information

Shaik Rasheeda Begum
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu Dt.,
Tamil Nadu, Chennai, India
rs3614@srmist.edu.in
rashee22ameen@gmail.com

Saad Yunus Sait
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu Dt.,
Tamil Nadu, Chennai, India
saady@srmist.edu.in

Arul Saravanan Ramachandran
Department of Psychiatry,
SRM Medical College Hospital & RC,
SRM Institute of Science and Technology
Kattankulathur, Chengalpattu Dt.,
Tamil Nadu, Chennai, India
arulpsg@gmail.com

***Abstract*— Depression is a severe mental health problem for people around the world, regardless of age, gender, or race. It is a cause of psychological disability, and these disorders can have an impact on a person's interpersonal connections, such as work environment and family life, as well as their overall health and routines, such as irregular eating and sleeping patterns. However, unfortunately, the majority of cases of depression go undiagnosed and, therefore, untreated. Depression, when not detected at an earlier stage, can become a severe illness and may lead to suicide at later stages. Consequently, it becomes crucial to identify and prevent depression at an earlier stage. The data for this study are collected through a survey from undergraduates in consultation with psychiatrists and professors.**

Further, Natural Language Processing(NLP) techniques and Machine learning methodologies were used to train the data and evaluate the efficiency of the proposed model. This study looked at various feature selection (FS) techniques such as filter method Maximum Relevance and Minimum Redundancy-mRMR, wrapper method Recursive Feature Elimination-RFE, Boruta, Embedded method: Least Absolute Shrinkage and Selection Operator-LASSO were used to extract the most significant features from profile information of the user responsible for forming a depression. The Adaboost model produced an accuracy of 94% when considering all elements in the dataset. The different feature selection techniques, when applied, we found Adaboost with mRMR FS using Optuna Hypertuning produced an accuracy of 96%.

Keywords- Depression; Machine learning; Feature selection; Filter; mRMR; wrapper method; Boruta; RFE; Embedded Method; Lasso; SMOTE .

I. INTRODUCTION

Mental health is considered a global epidemic. Approximately 450 million people grieve from depression at some point in their lives worldwide. Modern lifestyle is the leading cause of various mental health disorders, such as anxiety, stress, and depression [1]. Depression is a complex mental illness that affects an individual's day today activities resulting in loss of concentration, sleep disturbances, inability to think, and lack of social integration. The people who are suffering from this disease are less aware of the presence of the illness and even if they come to know there is an avoidance in acceptance and treatment, due to the association of social stigma and misconceptions with the treatment of mental disorders [2]. Various other chronic illnesses, including diabetes, heart disease, etc., are also linked to the development of depressed individuals. It is the second most common reason people get chronic conditions [3]. Severe depression can trigger suicidal cases and thus needs to be recognized earlier and treated with utmost attention and care.

Depression, in particular, is common among young people who are susceptible to severe difficulties. According to a survey, 24% of University and college students suffer from depression [4]. It's critical to treat this illness in students as soon as possible to avoid any long-term, debilitating consequences. Our work focuses on using artificial intelligence methods and measurement systems such as PHQ-9 to identify depression.

The rapid advancement in Information technology benefits people by using machine learning concepts to understand the data patterns and analyze the requirements. Nowadays, the usage of artificial intelligence can be seen in different areas like education [5], Health, psychology [6], Internet addiction, etc., to understand the presence of depression. Machine learning algorithms are widely used in the field of medicine and healthcare, but they are less applied in the field of psychology. In our work, Machine learning concepts created precise, fast, and automated screening procedures to identify individuals at high risk of depression by utilizing their specific predictors.

The study sought to build an algorithm which helps in early prediction of depression which is crucial for prompt treatment and elimination of serious health problems prevailing in college students, both physical and cognitive consciousness. If not identified leads to low motivation, pessimism, lack of enjoyment of life, Loss of appetite and memory, to the point of suicidal thoughts, as well as emotional consequences, including irritable, sad and crying [7]. In the present study, we worked with a real time dataset, collected by our own, where the classification model is developed to compare the performance of different standard feature selection techniques and without feature selection. Feature selection is used to reduce the number of relevant features in a dataset. This process is crucial in machine learning to improve model performance, reduce overfitting, and enhance interpretability.

Contributions of this research are as follows:

- Dataset creation that includes demographic, economic, and social information of the persons to predict depression.
- Mining economic, social, and personal attributes accountable for the cause of depression.
- Examining various machine learning and conventional feature selection algorithms for effective diagnosis of depression.

II. RELATED WORK

Researchers have done numerous studies on analyzing and identifying depressed Individual based on different types of data such as EEG, predictor variables and social networks data. This paper focuses on analyzing the predictor variables for effective diagnosis for depression.

A. Using only predictor variables without feature selection

In [8], the author collected datasets from 210 Bangladeshi students using the DASS-21 scale online and implemented Logistic regression to analyze depression and anxiety among students using different variables like eating, bonding with family and friends, bedtime habits, economic status, etc. Nemesure et al.[9] identified significant predictors of MDD from 4184 students of Nice Sophia-Antipolis. The XGBoost model analyzed 59 features, considering demographic and biomedical, and produced for the validation set- .67 AUC.

Su et al.[10] developed an automated system to identify depression in elderly Chinese people. The dataset used was CLHLS survey data. Several depression risk factors were predicted using six machine learning (ML) models and long short-term memory (LSTM). Among ML models, logistic regression with lasso regularization had the greatest area under the ROC curve (AUC-0.629) values. Ebert et al.[11] implemented a bivariate model to assess the major depressive disorder prevalence in 958 college students during the first year of their studies.

B. Using predictor variables with feature selection

Nayan et al.[12] developed a model that predicts the intensity level of depression among 2121 students studying college by employing different basic ML models. The RF model achieved a higher accuracy of 88.66%. Narkbunnum et al.[13] developed 5 machine learning model that uses the bootstrap sampling technique and SMOTE technique model compared with different filter feature selection technique to see the performance of imbalanced data of college students. Lee et al.[14]developed a machine-learning framework to identify depression signs among US adults. They included feature selection techniques such as LASSO and Boruta in the NHANES dataset, which involved PHQ-9 of 8628 people. The highest accuracy was achieved by SVM (77.1%) and ANN (81.3%). To find depression in stroke survivors, Ryu and colleagues [15] created a machine learning-based system. At a medical center in Korea, 623 individuals participated in the Hamilton Depression Prediction Index and the NIHSS survey. The SVM and KNN reported an accuracy of 77.5% and 73.3%, respectively.

III. DATA COLLECTION

The data has been collected from college students of SRM University through a Google survey form consisting of general questions related to their personal (studies, work, social, and health) and financial factors around 11 months. The student's data are kept confidential in an encrypted format.

Table 1: Characteristics/aspects of Depression

Descriptions (Features)
Age of the student (age)
Gender of Student: Male, Female (gender)
Town, City, Village (reside)
With family or Alone (living)
Satisfaction with Environment: Yes/No (envsat)
Transport Mode -By Walk, MTC, College bus (transmode)
Course chosen by Students: Yes/No (choice)
Satisfaction on Studies: Yes/No (acasat)
Socializing with friends: Yes/No (outing)
Any financial worries: Yes/No (finprob)
Student suffering from inferiority complex: Yes/No (inf)
Pressure on Studies: Yes/No (acapressure)
Relationship Breakup with loved ones: Yes/No (breakup)
Smoke / Drink : Yes/No (habit)
Suffering from sleeping problems: Yes/No (medical)
Practicing Yoga / physical exercises: Yes/No (exercise)
The average time you spend on social media: Yes/No (avgnw)
Teaching in virtual mode : Yes/No (online)
Feeling sad for not getting what you want: Yes/No (deprived)
Extracurricular Interested: Yes/No (extra)
Conflicts of any type with your friends or family: Yes/No (conflicts)
Died Closed Relation: Yes/No (lost)
No confidence in onesself: Yes/No (confi_dence)
Suicidal Urge: Yes/No (suicide)
Emotional, Physical, sexual : Yes/No (abused)
Restrictions on Mobile Phones: Yes/No (phoneuse)
Avoiding socializing with neighbours: Yes/No (avoid)
interested in mental health counseling: Yes/No (counsel)

IV. METHODOLOGY

A. The Pre-Processing

The Data encoding method of preprocessing technique is used to convert the categorical data into their numeric format for further processing of data. The Label Encoder of the Scikit-learn library is used for the preprocessing techniques [16].

B. Architecture Diagram

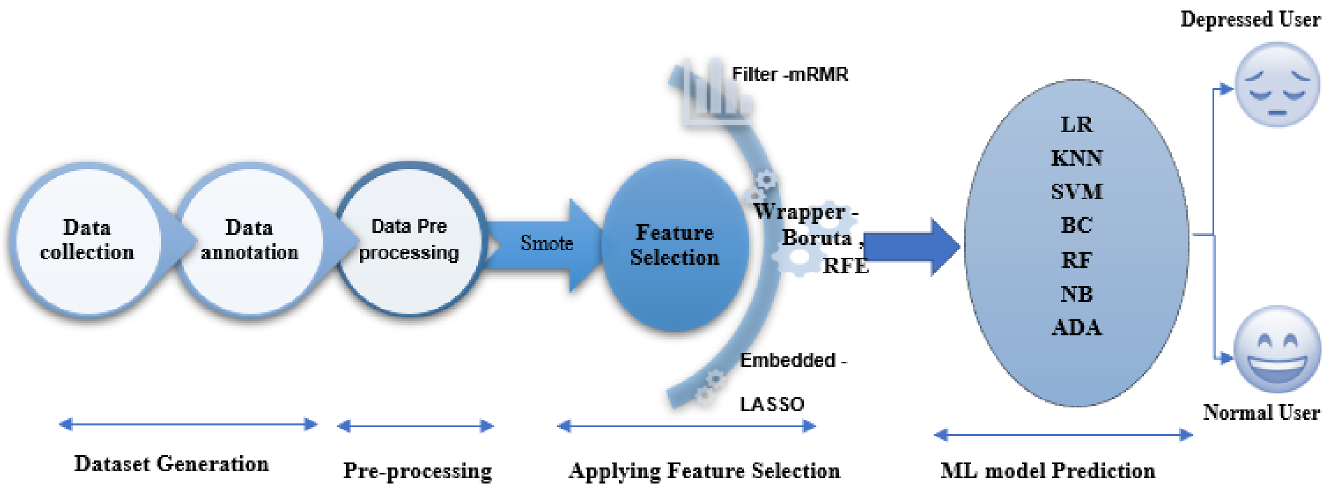

Figure 1: Flow of detection of depression using different Feature Selection and Machine Learning

C. Feature Selection

Feature selection is a procedure of automatically selecting pertinent features and eliminating irrelevant data during the building of the model [17]. The feature selection aims to provide Improved data quality, reduced computation time with predictive models, and improved predictive performance. By doing this, the model will perform better when compared to including all features. In this study, the different feature selection techniques that come under the filter, wrapper and embedded method are employed to find the various essential factors that affect whether a person is suffering from depression.

1) *Filter Method for Classification of Depression*

Filter-based methods are a class of feature selection techniques that selects features based on statistical measures. Different statistical measures such as correlation, mutual information, or chi-square can be used to rank the features and select the top-k parts for the model. It does not rely on the classifier to assess the features. The statistical measures are selected based on the variable data types. These approaches are usually faster and less computationally costly than wrapper or embedded methods as they do not require training a model to evaluate the usefulness of the features.

a) Minimum Redundancy and Maximum Relevance

mRMR aims to select features that are relevant meaning that are highly correlated with the output/response and not correlated with other features that have been selected, to avoid redundancy in the selection of features.

$$f^{mRMR}(X_i) = I(Y, X_i) - \frac{1}{|Se|}\sum_{X_{se} \in Se} I(X_{se}, X_i) \quad (1)$$

Here, Y is the Output

Se refers to all selected features.

Xse is one of the features that are selected from set Se, and Xi represents features that are not selected.

The most important features that are selected for predicting depression are ‘choice’, ‘envsat’, ‘acasat’, ‘confi_dence’, ‘out’, ‘practise’, ‘conflict’, ‘deprived’, ‘infer’, ‘surround’, ‘medication’, ‘counsel’, ‘abused’, ‘finprob’, ‘reside’, ‘lost’, ‘mode’.

2) *Wrapper Method for Classification of Depression*

Wrapper methods are feature selection methods that are employed in machine learning to find the most pertinent characteristics in a dataset that can be utilized to create prediction models. The wrapper method employs models to be built with an initial set of features, based on performance metrics evaluated, a new subset of features is generated by adding/removing a feature till the optimal set of features is selected [18].

a) Boruta feature selection algorithm

The essential features are selected by repeatedly removing unwanted attributes in an iterative manner. It acts as a wrapper for the random forest classification algorithm. It works by creating a shadow feature set, which is a randomized version of the original feature set. The features that are considered as important are ‘age’, ‘reside’, ‘envsat’, ‘mode’, ‘choice’, ‘acasat’, ‘out’, ’finprob’, ‘infer’, 'pressure', 'medical','practise','deprived','conflict','confi_dence', 'alive', 'abused','surround','counsel'.

.

b) Recursive feature elimination (RFE)

This procedure starts by considering each feature and then, reclusively, eliminates the features that aren't needed. At each step, the model's performance is evaluated to determine the most important features. The features that are considered important are 'age', 'living', 'envsat', 'choice', 'out', 'pressure', 'break', 'medical', 'practise', 'avgnw', 'conflict', 'confi_dence', 'alive', 'surround', 'counsel'.

3) Embedded Method for Classification of Depression

This method chooses features that are integrated as part of learning or building the model.

a) Least Absolute Shrinkage and Selection Operator

The LASSO regression is a linear regression that is mainly used to select the most relevant feature by reducing the irrelevant feature coefficient to zero. It can also be used for regularization. It shall minimize the sum of the residuals squared provided that the sum of the absolute values of the regression coefficients is less than a fixed value.

$$\text{Lasso} = \text{sum of squared errors} + \lambda|\text{coefficients}| \quad (2)$$

where λ signifies the shrinkage amount.

The features selected by this technique are 'age', 'reside', 'living', 'envsat', 'mode', 'choice', 'acasat', 'out', 'finprob', 'infer', 'pressure', 'medical', 'practise', 'avgnw', 'deprived', 'extra', 'conflict', 'lost', 'confi_dence', 'alive', 'surround', 'counsel'.

D.Synthetic Minority Oversampling Techniques

SMOTE(Synthetic Minority Over-sampling Technique), is a technique used in the context of imbalanced datasets in machine learning. Imbalanced datasets occur when the number of instances of one class (the minority class) is significantly lower than the number of instances of the other class (the majority class).A classifier when trained with an imbalanced dataset will lead to incorrect and biased results. In order to improve the minority class's predictive accuracy, the SMOTE operates in the feature and generates synthetic samples of the minority class. For each instance in the minority class, SMOTE selects k nearest neighbors. It then creates synthetic instances along the line segments connecting each instance to its k nearest neighbors. This helps to fill in the gaps in the feature space of the minority class. In our training samples, the ratio of depressed and normal participants is 338:763. Since our training dataset is not balanced, the SMOTE technique was utilized to remove their class Imbalanced problem [19]. SMOTE allows the model to learn more effectively from the minority class, reducing the risk of false negatives and improving overall predictive accuracy.

V. DETECTION OF DEPRESSION USING ML

The Depression detection model is a supervised binary classification problem. A binary-valued variable Yi = [1] is introduced for every user based on the predictor variables. The system is developed using Python language with required libraries from Scikit. The total dataset (1101) size is split into training - 15% (770), Validation- 15% (166) and-Test -15% (165) samples.

The Python Sklearn library train_valid_test_split is used to partition the total samples into training, test, and validation subsets for performance analysis of the model. Table 2 describes the dataset splitting size of our model. To predict the presence of depression, the different machine learning models used are as follows.

A. Logistic Regression (LR)

LR is an algorithm for classification that is used to predict binary outcomes based on demographic, behavioural, and social factors given as input to the model and output as depressed or non-depressed users. The hyper tuning parameters considered in our model are {'C': 0.1, 'solver': 'newton-cg', 'penalty': 'l2'}.

B. K-nearest neighbor (KNN)

KNN in nature is non-parametric also called an instance-based algorithm which is used for giving solutions to both regression and classification problems. In our model, we have used hyper tuning parameters as {'metric': ' Minkowski', 'n_neighbors': 15, 'weights': 'distance'}. The distance metric is computed as shown in Eq. (6).

$$\textit{Minkowski Distance} = \left[\sum_{i=1}^{k}\left(\left|x_i - y_i\right|\right)^q\right]^{\frac{1}{q}}$$

(3)

C. Support Vector Machine (SVM)

It is a linear binary classifier, non-probabilistic which examines data to determine its classification. The process creates a hyperplane in high-dimensional feature space that separates each class into the one that most closely matches the training data objective. The model used hyper tuning parameters such as kernel= 'rbf ', C = 1, and gamma =1.

D. *Naive Bayes Classifier(NB)*

Naïve Bayes (NB) is a statistical machine learning based on Bayes Theorem that treats features within a class as independent of other features. The classifier used is the Gaussian Naive Bayes. The Best Parameters used in our model is {'priors': None, 'var_smoothing': 1e-09}

E. Bagging classifier (BC)

The concepts of aggregation and bootstrapping were used to develop bagging. In bagging classifiers, the training datasets are used to produce a bootstrap dataset to train several classifiers and the final prediction is generated by averaging the output of various classifiers. Additionally, aggregated classifiers frequently perform better than single classifiers. The parameter used as hyper tuning for our model is {'n_estimators': 1000}.

F. Random Forest (RF)

In an effort to increase predicted accuracy and reduce overfitting, a random forest is a meta-estimator which offers average results from many decision tree classifier fits. The parameter used as hyper tuning for our model is {'max_features': 'sqrt', 'n_estimators': 1000}.

G. *AdaBoost (ADA)*

AdaBoost is a boosting technique that creates a strong learner out of several weak ones. This approach uses the error of the prior classifier to construct a new one. The classifier α using error Ɛ in classifying the data samples is measured by the following equation.

$$\alpha = \frac{1}{2}\ln\left(\frac{1-\epsilon}{\epsilon}\right) \quad (4)$$

Here, α is used to change the sample weights of the dataset, resulting in a new dataset. To tune the AdaBoost algorithm, the Optuna hyperparameter optimization library was used. Optuna requires the definition of an objective function that evaluates a variety of model hyperparameter values and fits the model to the training data. The Best trial Value is 0.9327, parameters are n_estimators:142 and learning_rate:.430876.

VI. RESULTS AND DISCUSSIONS

The model's performance is evaluated without applying any feature selection techniques. Next, to examine the effect of the feature selection method on classifier accuracy, we evaluated the performance of the model for the selected feature set only.

A. Performance Evaluation

Our dataset consists of a total of 28 features as independent variables and 1 target dependent variable. The training dataset was fed to the learning classifiers for the model to be trained. The model after training has been used to classify the student's status on depression based on the test datasets. The following formulas were used to determine the model's precision, recall, accuracy, and F1 score to assess its performance.

$$Accu = \frac{\left(T_{Positive} + T_{Negative}\right)}{\left(T_{Positive} + T_{Negative} + F_{Positive} + F_{Negative}\right)} \quad (5)$$

$$Rec = \frac{T_{Positive}}{\left(T_{Positive} + F_{Negative}\right)} \quad (6)$$

$$Prec = \frac{T_{Positive}}{\left(T_{Positive} + F_{Positive}\right)}$$

(7)

$$F - Mea = \frac{2T_{Positive}}{\left(2T_{Positive} + F_{positive} + F_{Negative}\right)} \quad (8)$$

P (Actual Positive) indicates a depressed user which is an actual positive Instance,

N (Actual Negative) indicates not depressed user which is an actual negative Instance,

$T_{Negative}$ (True Negative) represents actual instance is not depressed, and the prediction values are not depressed as well,

$F_{Negative}$ (False Negative) represents the actual instance is depressed, but the prediction values are not depressed,

$F_{positive}$ (False Positive) represents the actual instance is not depressed, but the prediction values are depressed and

$T_{positive}$ (True Positive) represents the actual instance is depressed, and the predicted value is also depressed.

B. *Classification Results without Applying Feature Selection(WF)*

Table 2 displays the results of ML models without applying feature selection, to predict the depression outcome. The AdaBoost classifier has outperformed all other models with an accuracy of 94% SVM, RF and NB performed better results with an accuracy of 92%. The LR, and Bagging Classifier achieved an accuracy of 91%. The Least accuracy model KNN achieved an accuracy of 87%.

Table 2: Performance Metrics without Feature Selection

Model	Without Feature Selection (Metrics)					
	Accu	Prec	Rec (RE)	RE(0)	RE(1)	F1
LR	0.91	0.9	0.93	0.9	0.93	0.91
KNN	0.87	0.86	0.86	0.87	0.87	0.86
SVM	0.92	0.91	0.92	0.91	0.93	0.92
BC	0.91	0.92	0.89	0.93	0.89	0.9
RF	0.92	0.94	0.89	0.95	0.89	0.91
NB	0.92	0.89	0.93	0.89	0.94	0.91
ADA	0.94	0.94	0.93	0.95	0.94	0.94

C. Classification Results after Applying Feature Selection

Applying different feature selection techniques improves the model accuracy of all these classifiers that are displayed on Table3. Using mRMR algorithm, ADA, and LR classifier has shown the best result with an accuracy of 96% and 94% . The SVM and NB, model achieved an accuracy of 94%. RF classifier achieved an accuracy of 93%. The Least accuracy model are KNN achieved an accuracy of 87%. When applying the filter selection method to the model, the mRMR algorithm with the Adaboost classifier performs well and produces an accuracy of 96%.

In the RFE algorithm, the Bagging Classifier, RF and ADA have shown the best result with an accuracy of 95%. The NB and SVM models achieved an accuracy of 94%, and 92% respectively. The Least accurate model is KNN achieved an accuracy of 87%. Using the Boruta algorithm, ADA classifier has shown the best result with an accuracy of

93%. The LR, SVM, and RF produced an accuracy of 91 %, the CAT, NB and Bagging classifiers achieved 88 and 89%. The Least accuracy model is KNN achieved an accuracy of 84%. When applying the Wrapper selection method to the model, the RFE algorithm with the Bagging Classifier, RF and ADA performs well and produces an accuracy of 95%.

Table 3:The performance metrics (Recall) for the Filter and Wrapper Feature selection techniques using Machine Learning models

Model	FILTER METHOD		WRAPPER METHOD				EMBEDDED	
	mRMR		RFE		Boruta		Lasso	
	0	1	0	1	0	1	0	1
LR	0.94	0.97	0.92	0.95	0.89	0.93	.88	0.93
KNN	0.92	0.81	0.93	0.86	0.91	0.77	0.82	0.65
SVM	0.92	0.97	0.94	0.95	0.93	0.89	0.88	0.81
BC	0.92	0.95	0.94	0.88	0.92	0.87	0.88	0.87
RF	0.92	0.94	0.95	0.92	0.93	0.9	0.89	0.86
NB	0.93	0.94	0.92	0.96	0.88	0.89	0.85	0.91
ADA	0.94	0.95	0.95	0.97	0.92	0.94	0.88	0.92

In the RFE algorithm, the Bagging Classifier, RF and ADA have shown the best result with an accuracy of 95%. The NB and SVM models achieved an accuracy of 94%, and 92% respectively. The Least accurate model is KNN achieved an accuracy of 87%. Using the Boruta algorithm, ADA classifier has shown the best result with an accuracy of 93%. The LR, SVM, and RF produced an accuracy of 91 %, the CAT, NB and Bagging classifiers achieved 88 and 89%. The Least accuracy model is KNN achieved an accuracy of 84%. When applying the Wrapper selection method to the model, the RFE algorithm with the Bagging Classifier, RF and ADA performs well and produces an accuracy of 95%.

When applied LASSO method, LR and ADA achieved an accuracy of 90%. The Least accuracy model are LR achieved an accuracy of 74%.Overall, Filter Method – mRMR , wrapper method–RFE , Boruta embedded method applied to the classifier , mRMR of Filter method gave best results of 96% in terms of different metrics such as Accuracy ,Recall ,Precision, F1 score and ROC -AUC . AdaBoost with mRMR FS reduces bias by training weak learners iteratively, which improves prediction performance. It also learns the correlation between attributes very quickly. Table 4 shows the comparative results of the proposed work with the present state-of-the-art works

.

Table 4 shows the comparative results of the proposed work with the present state-of-the-art works.

Author Year	Participants	Collected Information Type	Depression Screening Scale	Feature Selection	Machine learning models	Accuracy (%)
Proposed work	1101 College Students	sociodemographic and psychosocial information	PHQ-9	Mrmr (Filter Method)	Adaboost	Acc -96% AUC-99.161.
M.S. Zulfiker et al [20]	604 -Bangladeshi citizens of different age ranges, occupations, socio-economic	Socio-demographic and psychosocial information	BDC	SelectKBest (Filter Method)	Adaboost	Accuraccy-92.56% AUC-96%
Na.et.al ,[21]	6588 Korean participants of different age ranges and occupations	Socio-demographic, economic, clinical information	CES-D-11	All Features	RF	Accuracy – 86.2% AUC-87.0%
Sau and Bhakta,[22]	470 Seafarers	Socio-demographic, occupational information	HAM-D	Recursive feature elimination (RFE) method	CatBoost	Accuracy-89.3%
Sau and Bhakta [23]	520 Geriatric patients	Socio-demographic, socio-economic, and health related information	Geriatric Depression Scale (30 item)	Correlation (CFS),One R (OR),PCA, Gain ratio(GR) , symmetrical uncertainty	Random forest (RF)	AUC – 87.0% Accuracy-86.2%

CES-D-11 :- Center for Epidemiologic Studies Depression Scale , HAM-D :- Hamilton Depression Rating Scale , BDC :- Burn's Depression Checklist

D. Training Time of Different Feature selection algorithms:

The model training time is faster when the number of features given to the model is reduced. From Table 5, Logistic regression takes more training time in BORUTA selection algorithms. KNN, when trained without feature selection techniques takes less time when compared to the reduced feature set given to the model. SVM except for LASSO FS techniques greatly reduces the training time of all other FS techniques. Overall, Bagging take more training time when feature selection techniques are applied compared to all other Machine learning algorithms. Lasso feature selection methods take more time when compared to the overall features passed to the model.

Table 5 Comparison of training time (in sec) for different feature selection techniques.

Model	**WF**	RFE	Boruta	**mRMR**	LASSO
LR	**1.95**	0.78	2.02	**1.56**	2.08
KNN	**0.9**	1.06	0.81	**0.9**	1.46
SVM	**11.13**	9.23	9.12	**7.07**	13.99
BC	**4. 15**	3.29	0.2	**3.42**	8.9
RF	**2.5**	16.5	0.15	**2.43**	3. 12
NB	**0.06**	0.05	0.05	**0.05**	0.11
ADA	**0.48**	0.54	0.29	**0.47**	0.61

VII. CONCLUSION

Depression is the fastest-growing illness and one of the foremost causes of death globally. Thus, an early and appropriate diagnosis of depression can help a person recover from self-harming behavior or lead to suicide. In this research, after performing different standard FS techniques, we observe that depression among college students is increased due to various factors such as age, gender, course choice, academic pressure, inferiority complex, behavioral habits, social interactions, financial stress, and internet addiction. Additionally, classification experiments were performed considering all features and a reduced feature set to analyze the impact of the selected elements on the prediction accuracy of different ML prediction models. The Adaboost model produced an accuracy of 94% when considering all features in the dataset. The mRMR feature selection technique, when applied to the Adaboost model, had an accuracy of 96 %. In future work, clinical parameters and social media data can also be considered to predict depression with different intensity levels.

REFERENCES

[1] P. Kumar, S. Garg, and A. Garg, "Assessment of anxiety, depression and stress using machine learning models," *Procedia Computer Science,* vol. 171, pp. 1989-1998, 2020.
[2] S. R. Begum and S. Y. Sait, "Effective Techniques for Depression Detection on Social Media: A Comprehensive Review," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-9, 2022.
[3] Y. Ma, Q. Xiang, C. Yan, H. Liao, and J. Wang, "Relationship between chronic diseases and depression: the mediating effect of pain," *BMC psychiatry,* vol. 21, pp. 1-11, 2021.
[4] R. Hoque, "Major mental health problems of undergraduate students in a private university of Dhaka, Bangladesh," *European Psychiatry,* vol. 30, p. 1880, 2015.
[5] A. Daza, C. Guerra, N. Cervera, and E. Burgos, "A stacking based hybrid technique to predict student dropout at universities," *J Theor Appl Inf Technol,* vol. 100, pp. 1-12, 2022.
[6] J. Chang, Y. Yuan, and D. Wang, "Mental health status and its influencing factors among college students during the epidemic of COVID-19," *Nan Fang Yi Ke Da Xue Xue Bao,* pp. 171-176, 2020.
[7] A. D. Vergaray, J. C. H. Miranda, J. B. Cornelio, A. R. L. Carranza, and C. F. P. Sánchez, "Predicting the depression in university students using stacking ensemble techniques over oversampling method," Informatics in Medicine Unlocked, p. 101295, 2023.
[8] A. R. Arusha and R. K. Biswas, "Prevalence of stress, anxiety and depression due to examination in Bangladeshi youths: A pilot study," *Children and youth services review,* vol. 116, p. 105254, 2020.
[9] M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, "Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence," *Scientific reports,* vol. 11, p. 1980, 2021.
[10]D. Su, X. Zhang, K. He, and Y. Chen, "Use of machine learning approach to predict depression in the elderly in China: A longitudinal study," *Journal of affective disorders,* vol. 282, pp. 289-298, 2021.
[11]D. D. Ebert, C. Buntrock, P. Mortier, R. Auerbach, K. K. Weisel, R. C. Kessler, *et al.*, "Prediction of major depressive disorder onset in college students," *Depression and anxiety,* vol. 36, pp. 294-304, 2019.
[12]M. I. H. Nayan, M. S. G. Uddin, M. I. Hossain, M. M. Alam, M. A. Zinnia, I. Haq, *et al.*, "Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among University Students in Bangladesh: A result of the first wave of the COVID-19 pandemic," *Asian Journal of Social Health and Behavior,* vol. 5, p. 75, 2022.
[13]W. Narkbunnum and K. Wisaeng, "Prediction of Depression for Undergraduate Students Based on Imbalanced Data by Using Data Mining Techniques," *Applied System Innovation,* vol. 5, p. 120, 2022.
[14]C. Lee and H. Kim, "Machine learning-based predictive modeling of depression in hypertensive populations," *PLoS One,* vol. 17, p. e0272330, 2022.
[15]Y. H. Ryu, S. Y. Kim, T. U. Kim, S. J. Lee, S. J. Park, H.-Y. Jung, *et al.*, "Prediction of poststroke depression based on the outcomes of machine learning algorithms," *Journal of Clinical Medicine,* vol. 11, p. 2264, 2022.
[16]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research,* vol. 12, pp. 2825-2830, 2011.
[17]I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research,* vol. 3, pp. 1157-1182, 2003.
[18]R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence,* vol. 97, pp. 273-324, 1997.
[19]D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Information Sciences,* vol. 505, pp. 32-64, 2019.
[20] Zulfiker, M.S., Kabir, N., Biswas, A.A., Nazneen, T., and Uddin, M.S.: 'An in-depth analysis of machine learning approaches to predict depression', Current research in behavioral sciences, 2, pp. 100044, 2021.
[21] Na, K.-S., Cho, S.-E., Geem, Z.W., and Kim, Y.-K.: 'Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm', Neuroscience Letters, 721, pp. 134804, 2020.
[22] Sau, A., and Bhakta, I.: 'Screening of anxiety and depression among seafarers using machine learning technology', Informatics in Medicine Unlocked, 2019, 16, pp. 100228
[23] Sau, A., and Bhakta, I.: 'Predicting anxiety and depression in elderly patients using machine learning technology', Healthcare Technology Letters, 4, (6), pp. 238-243, 2017.