SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I6P103

Volume 11 Issue 6, 20-39, June 2024

Original Article

A Comparative Analysis of Machine Learning Algorithms with Bagging and Boosting for Mental Health Assessment

Ravita Chahar[1], Ashutosh Kumar Dubey[2], Sushil Kumar Narang[3]

[1,2,3]*Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India*

[2]*Corresponding author : ashutosh.dubey@chitkara.edu.in*

Received: 08 April 2024 Revised: 20 May 2024 Accepted: 05 June 2024 Published: 29 June 2024

Abstract *- This study explores machine intelligence algorithms for evaluating mental health conditions on distinct datasets. The primary goal is to identify effective algorithms for anticipating mental health concerns. The central objective is to determine the most productive machine learning algorithms in predicting mental health issues using provided datasets. Various algorithms – Logistic Regression (LR), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), bagging, and boosting – were employed with diverse parameter settings. The OSMI (Ds1) and the District Mental Health Program dataset (Ds2) were used for experimentation. Among the algorithms tested, LR with an Optimal Threshold (LR-OT) achieved notable performance with accuracy, precision, recall, specificity, and F1-score ranging from 0.87 to 0.89 across both datasets. kNN with Optimized parameters (kNNO) achieved 90% and 91% accuracy rates for Ds1 and Ds2, respectively. SVM obtained average accuracies of 91% and 93% for Ds1 and Ds2, respectively, with specific variations showing superior outcomes. Notably, DT, RF, bagging, and boosting models exhibited R2 scores exceeding 0.70, with bagging and DT yielding the highest R^2 score. The study's findings demonstrate that DT and RF models enhanced by bagging and boosting techniques outperform other algorithms in predicting mental health concerns using the provided datasets. This highlights the significance of employing these models for effective mental health assessment through machine learning.*

Keywords *- Machine Learning, Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest.*

1. Introduction

As per the World Bank's report, close to 1 billion individuals are currently livings with a mental disorder [1]. The World Health Organization (WHO) reveals that the COVID-19 pandemic has severely disrupted or even halted critical mental health services in 93% of countries globally, while the demand for mental health support is steadily rising [2]. This situation is of great concern due to the chronic nature of mental health conditions, and it holds significant economic implications on a global scale. In fact, projections indicate that depression will impose a greater burden on nations than any other disease in the coming decade [2]. Mental health assessment plays a crucial role in understanding and diagnosing various psychological conditions [3-6]. Traditionally, mental health assessments heavily relied on subjective evaluations conducted by professionals in the field [7-10]. However, with the emergence of intelligent learning algorithms, a new era of mental health assessment has emerged, offering promising advancements and opportunities for more accurate and efficient evaluations [11, 12]. By leveraging this technology, mental health assessments can now benefit from objective analysis and data-driven insights.

Machine learning algorithms can process vast amounts of data, identify complex patterns, and generate predictive models that aid in diagnosing mental health disorders [13, 14].

One primary advantage of utilizing these models and algorithms for mental health assessment is their ability to analyze a wide range of variables simultaneously [15-18]. Traditional assessments often rely on limited data inputs and subjective interpretations, which can introduce potential biases and inaccuracies [19]. In contrast, machine learning algorithms can incorporate diverse sources of information, including patient demographics, genetic data, electronic health records, and even social media activities. This comprehensive approach allows for a more holistic understanding of an individual's mental health state. Another notable benefit of using machine intelligence approaches for mental health assessment is the potential for early detection and intervention [20, 21]. By analyzing large datasets and recognizing subtle patterns, these algorithms can identify

indicators of mental health disorders before they become prominent. Furthermore, machine learning algorithms can help personalize mental health assessments, tailoring them to individual needs [22, 23]. By analyzing an individual's unique data profile, including physiological markers, behavioral patterns, and treatment history, algorithms can generate personalized risk assessments and treatment recommendations [24, 25].

This individualized approach has the potential to optimize treatment plans and improve patient outcomes. Numerous researchers are involved in the betterment of the assessment ways. They have employed various machine intelligence methods such as k-Nearest Neighbor (kNN), Decision Tree (DT), Naive Bayes (NB), random forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and more [18-21]. The selection of these algorithms is based on their effectiveness in predicting, detecting, extracting, and analyzing data, enabling a wide range of applications in mental health evaluation.

The major problem with the traditional approaches considered is the limited perspective of mental state and impacting factors. Suppose may influence the personal bias. If the datasets used are not diverse or representative of the entire population, the algorithms may develop biases. These biases may lead to inaccurate classification, particularly for specific or diverse groups. In the majority of the works, limited diversity was considered.

The objectives and contribution of this paper are as follows:

- To evaluate the mental health status using computational learning algorithms. This entails employing various machine learning techniques, including LR, kNN, SVM, DT, RF, bagging, and boosting, to assess mental health. The assessment involves classification tasks with the goal of determining the status of mental health based on the features available in the datasets.
- To compare the performance of these algorithms based on performance metrics such as accuracy, precision, recall, F1-score, and the coefficient of determination (R2). Through the experimentation and assessment of these algorithms, valuable insights into their effectiveness and suitability for mental health prediction tasks can be obtained.

The experimental cases include L1, L2, C, and epochs as the parameters and hyperparameters in the case of LR. In the case of kNN, the parameters and the hyperparameters are k, distance metric, K-Dimensional (KD)-tree, ball-tree, and voting method. C, degree, as the parameters and hyperparameters and different kernel types are in the case of SVM. Other combinations are DT, RF, and bagging with DT, RF, and SVM. Finally, AdaBoost with DT, RF, and SVM are considered. The motivation for comparing various machine learning algorithms, including LR, kNN, SVM, DT, and RF, in the context of mental health diseases, is rooted in the complexity of mental health datasets and the imperative to develop effective predictive models. Mental health data often exhibits diverse and intricate patterns, and the choice of algorithm can significantly impact the model's ability to capture these nuances. Understanding the factors contributing to mental health conditions is crucial, and comparing algorithms provides insights into their interpretability and explainability in comprehending model predictions.

The comparison also facilitates the selection of the most suitable algorithm for mental health prediction tasks based on their respective strengths and weaknesses. Additionally, hyperparameter tuning allows for the optimization of model configurations, enhancing predictive capabilities. Given the limited size of mental health datasets, assessing how well each algorithm generalizes to new cases is essential. The ultimate goal is to contribute to the development of accurate and interpretable models that can assist in the diagnosis and treatment of mental health disorders. This research endeavor aims to provide a valuable approach that not only improves prediction accuracy but also offers insights into the complex relationships within mental health data, fostering advancements in the field.

The organizational structure of this paper is divided into distinct sections. Section 2 provides a review related to mental health assessment. Section 3 covers the methodological evaluation and exploration. Results have been investigated and explored in Section 4. Section 5 explores the impacts of the results, discusses key findings and implications of this study, and investigates the limitations. Finally, the conclusion is presented in Section 6.

2. Literature Review

In this section, a comprehensive review has been conducted. It explores the studies that utilize various machine learning models for the analysis of mental health. A sample of 2400 individuals who received Applied Behavior Analysis (ABA) treatment across various states were examined by Stevens et al. (2019) [26]. They utilized hierarchical clustering and Gaussian mixture models. The findings are preliminary and require replication in further samples of individuals with Autism Spectrum Disorder (ASD). In 2019, Sau and Bhakta [27] conducted a study on 470 seafarers at the Haldia Dock Complex, collecting sociodemographic, occupational, and health-related data. They used the hospital anxiety and depression scale. They applied LR, SVM, RF, Catboost, and NB to assess anxiety and depression. Catboost had an accuracy and precision of 82.6% and 84.1%, respectively, making it the most effective classifier. The manifestation of heterogeneity and multivariate pattern recognition techniques were discussed for the assessment of schizophrenia [28].

They suggested that these strategies could improve forecasts and understanding of the neurological basis of the disorder. Schnack [28] concluded that future studies should incorporate multi-center, multi-modal predictor data to model more complex relationships between biomarkers and clinical presentation.

In 2019, Zhao et al. [29] studied 179 students who completed the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder (GAD-7) questionnaires, recorded by Kinect cameras. They extracted features using fast fourier transforms and used different regression algorithms, including Gaussian processes, to train regression models that recognized anxiety and depression levels and classification models that detected specific depressive symptoms. Gaussian processes achieved a correlation coefficient of 0.51 between predicted and questionnaire scores for anxiety and depression, and prediction accuracies for depression and anxiety were 0.64 and 0.74, respectively. Epsilon-support vector regression was also used. Ay et al. [30] proposed a hybrid model that combines Long-Short Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures.

The model was utilized to identify sadness using Electroencephalogram (EEG) signals. The CNN layers learn the temporal characteristics, while the LSTM layers learn the sequences. EEG data from the left and right hemispheres of the brain were used. According to the findings, the CNN-LSTM model demonstrates both speed and accuracy in detecting depression using EEG signals. Bhak et al. [31] investigated various mental health conditions and developed a model to predict the risk. The RF classifiers demonstrated superior performance compared to the other models. In 2020, Tennenhouse et al. [32] utilized RF, Neural Network (NN), and LR techniques to forecast the occurrence of anxiety and major depressive disorder.

The researchers discovered that LR had superior performance compared to RF and NN. Nevertheless, to address the problems of overfitting and non-convergence, they aggregated data from patients with various immune-mediated inflammatory illness conditions. Two multiple regression analyses to determine the effectiveness of sociodemographic and clinical factors as predictors of psychological outcomes in medical professionals working in COVID-19 wards were conducted by Di et al. [33]. The dataset included 145 healthcare workers, comprising 72 medical doctors and 73 nurses. The multiple regression model was found to strongly predict the overall score on the Beck Depression Inventory (BDI) in relation to depression symptoms. A depression classification system was developed by Narziev et al. [34] using a wearable device and smartphone, achieving 96% accuracy across four PHQ-9 groups, showcasing its effectiveness. In 2021, Na et al. [35] developed machine learning algorithms to distinguish panic disorder from other anxiety disorders using Heart Rate Variability (HRV) data from 61 individuals with different anxiety disorders and 60 patients with panic disorder. LR achieved the best accuracy of 78.4%, followed by artificial NN and SVM with 73%, Gradient Boosting Machines (GBM) with 67.6%, and RF with 64.9%. LR also had the highest Matthews correlation coefficient of 57.2%, F1-score of 79%, specificity of 73.7%, and sensitivity of 83.3%. The study included participants aged 20 to 65 years.

In 2021, Andrade et al. [36] aimed to enhance the International Classification of Diseases (ICD-10) protocol and enable a fast diagnosis of ASD based on a single, inconspicuous symptom. Their analysis indicated that only nine out of the eighty characteristics evaluated were sufficient to indicate the diagnosis of ASD accurately. In 2021, Walambe et al. [37] proposed a method for rapidly detecting workload-related stress using an artificial neural network. They utilized a late and early fusion model and attained accuracies of 90.45% and 96.67%. Jacobson et al. [38] utilized a deep autoencoder to predict anxiety disorder symptoms in participants. They studied 265 subjects who had taken part in Midlife in the United States.

The outcome suggests that wearable movement data could effectively anticipate individuals likely to undergo symptom deterioration. In 2021, Zhu et al. [39] conducted a study involving 48 cases of schizophrenia and 50 healthy cases. They utilized diverse machine learning algorithms, such as RF, DT, SVM, etc., for classifying schizophrenia cases. Among all the models, the SVM model exhibited superior performance. In 2021, Shah et al. [40] developed a method to predict depressed mood over 1 month.

Among the computational models used, the voting regressor performed the best on average among subjects. In 2022, Han [41] developed a questionnaire based on the Center for Epidemiologic Studies Depression Scale (CES-D) to assess depression and proposed a systematic approach for data cleaning, pre-processing, and mental health assessment using a multi-model. The approach included sentiment analysis using LSTM for text-based data and Visual Geometry Group (VGG) 16 for image-based data. The experimental results showed high accuracy rates, providing support for psychological education and counseling for students in higher vocational education. In 2022, Hanif et al. [42] analyzed 95,677 records from the year 2011-2012, which was conducted in the United States for children aged 2 to 17 years. They used computational algorithms to predict and classify the severity of ASD. The deep neural network achieved the highest accuracy of 87.11%. In 2022, Aleem et al. [43] proposed an approach for depression detection by using generic computation. Based on this approach, they have also suggested future ways of improvement in depression detection. In 2022, Ryu et al. [44] conducted a study on 31 individuals with Post-Traumatic Stress disorder (PTSD) and

34 without PTSD. SVM, voting ensemble models, kNN, RF, and statistical analysis utilising LR were used to predict PTSD. The study discovered that a radial basis function kernel with the SVM model accurately predicted PTSD. In 2023, Kuhathasan et al. [45] analyzed strainprint app data (Feb 2017 - Oct 2020) with 68,819 cannabis use observations from 1,307 individuals addressing mental health symptoms. They applied machine learning algorithms and SHapley Additive exPlanations to assess the features' impact.

In 2023, González-Pérez et al. [46] introduced the AwarNS Framework, a context-aware modular software development framework for android smartphones. It facilitates reliable, power-efficient, and ethical health app development, incorporating data sampling, analysis, and intervention capabilities. Advantages include transparency and reusability, with challenges in the complexity of mHealth app development. Real-life case studies showcase its effectiveness in implementing mobile interventions for diverse mental disorders. In 2023, Bhatnagar et al. [47] collected data using a survey given to engineering students at a university level, and 127 students answered that questionnaire. They applied different algorithms like DT, NB, SVM and RF to analyze the data. The RF method had the best accuracy at 78.9%, followed by SVM at 75.55%, NB at 71.05%, and DT.

In 2023, Shvetcov et al. [48] analyzed participant screening data from the 'Vibe Up' smartphone-based mental health intervention trial for Australian university students. They employed supervised (shrinkage discriminant analysis and classification and regression trees) and unsupervised (K-means clustering) learning algorithms. The study identifies user subgroups with stable and timepoint-dependent characteristics, underscoring the importance of tailored interventions addressing trauma, app usage reasons, and burnout for enhanced engagement. In 2023, Slima et al. [49] applied association rules mining to uncover cause-and-effect connections between mental health conditions and contributing factors.

Using an online survey during the first peak of COVID-19 in Tunisia, the study revealed associations, such as doctor consultation with anxiety and COVID test with depression, emphasizing the combined impact of multiple factors on mental health vulnerability for informed medical decisions. In 2024, Jayakumar and Reshma [50] reviewed methodologies for predicting mental health issues using machine learning, discussing its benefits, like timely intervention and challenges, like data biases, with recommendations for future research and development in the field. In 2024, Gupta et al. [51] investigated mobile apps from major app stores, analyzing 216 apps designed to support individuals with depression. Their study highlights the apps' roles in education, therapy, and mood monitoring, emphasizing mobile solutions to aid depression management globally.

3. Materials and Methods

This work utilized machine learning methods to assess mental wellness utilizing two datasets. The specific characteristics and information on the datasets utilized are outlined below:

OSMI 2014 dataset (Ds1): This dataset comprises responses from 1,259 participants, encompassing their answers and attributes, including 27 columns that incorporate a timestamp. The attributes include age, gender, country, and state. The questionnaires are based on self-employment, family history, treatment, work interference, number of employees, remote work, tech company affiliation, benefits, care options, wellness programs, seeking help, anonymity, leave policies, mental health consequences, physical health consequences, relationships with coworkers and supervisors, mental health interviews, physical health interviews, perceptions of mental versus physical health, and the presence of observable consequences. This dataset consists of survey results collected over multiple years, starting from 2014 [52]. Therefore, researchers have the option to select data from specific years, such as 2014, 2016, and 2019, or they can choose to combine data from all available years. The dataset is also accessible on the Kaggle repository. Approximately 80% of the participants in the dataset are male, while less than 1% has an unidentified gender. Specifically, there are 992 male participants, 250 female participants, and 12 participants with other gender identifications. The dataset consists of 26 categorical attributes and one numeric attribute.

The overall percentage of missing values in the dataset is 5.6%. District Mental Health Program, Kurukshetra (Ds2): This dataset represents the compilation of data from the Lok Nayak Jai Prakash Hospital, Kurukshetra, Haryana. It encompasses cases documented from April 2019 to March 2022, comprising a total of 15,687 records. The dataset covers various mental health conditions, such as dementia, stress, anxiety, depression, schizophrenia, bipolar disorder, autism, sexual disorders, and more. Within this dataset, special attention has been given to cases related to depression (7,654 records), anxiety (438 records), stress (1,027 records), and normal cases (987 records). These four categories collectively make up 10,106 records. For the analysis of these cases, common symptoms have been considered, including mood changes (S1), fatigue (S2), sleep disturbances (S3), difficulty concentrating (S4), irritability (S5), changes in appetite (S6), and physical symptoms (S7). The primary objective was multiclass classification, where the model distinguishes between normal cases and those with mental health problems such as anxiety, depression, and stress. This categorization represents instances without mental health issues (normal) and those with specific mental health problems like anxiety, depression, and stress, respectively.

3.1. Preprocessing and Data Preparation

The OSMI dataset, although abundant in information, posed challenges stemming from the inclusion of missing,

inconsistent, and unnecessary values. To ensure the dataset's reliability for machine learning models, a comprehensive data cleaning process was executed. This involved the exclusion of specific attributes, such as timestamp, country, and state, which were deemed irrelevant to the specific analysis at hand. The initial phase of the data cleaning process focused on the 'gender' column, which exhibited a plethora of values. To enhance model interpretability and simplify this variable, gender was categorized into three distinct parts: male, female, and others. This categorization aimed to streamline the feature and mitigate complexity in subsequent analyses. Addressing missing values constituted a pivotal step in data preprocessing. Each attribute with missing values underwent a unique treatment based on its inherent nature. To identify missing and duplicate values, Python's built-in functions, such as isnull().sum(), were used to identify any missing values, and duplicated().sum() was employed to detect any duplicate tuples for attributes. For example, missing values in the 'self-employed' attribute were replaced with 'no,' while those in the 'work interfere' attribute were replaced with 'don't know.' This meticulous approach ensured that the dataset remained robust and suitable for analysis. Following the data cleaning phase, preprocessing steps were implemented to ready the data for machine learning models. Label encoding was used for transforming categorical data into numerical form. This transformation was applied to all attributes except 'age,' given its inherent numerical nature. The primary objective was binary classification, where the model discerns between 0 and 1, representing cases without mental health problems and those with mental health problems, respectively. The training dataset, featuring predictor variables such as family history, care options, gender, age, and others, was employed to train the model. The target variable, 'treatment,' was selected to predict the presence or absence of mental health problems. In the experimental setup, a 70–30 split was implemented, allocating 70% of the data for training and the remaining 30% for testing. For experimentation, ratios of 75:25 and 80:20 were also considered. To improve the reliability of the model evaluation, a tenfold cross-validation approach was utilized.

3.2. Selection of Algorithms

Our primary focus on a specific set of algorithms, including LR, kNN, SVM, DT, RF, bagging, and boosting, highlights the careful consideration given to assessing mental health across diverse cases. LR, known for its simplicity and efficacy, played a key role in addressing both binary and multiclass classification challenges. kNN was chosen for its ability to detect subtle local patterns and relationships within the dataset. SVM was included for its ability to handle non-linear decision boundaries and high-dimensional data. The interpretability of DT made it an effective method for capturing complex decision boundaries in mental health assessment. RF, being an ensemble method, synergistically leverages the advantages of many decision trees to enhance accuracy and resilience. In addition, bagging and boosting approaches were employed to improve the performance of the base models by leveraging different combinations and hyperparameters. The complete framework of this study is shown in Figure 1, showcasing the strategic selection and application of these algorithms, each designed to address specific aspects of mental health assessment.

3.3. Hardware and Software Specifications

For the experimentation, Python 3.9 was used. The hardware for coding tasks included an Intel(R) Core(TM) i5-10210U CPU with a base clock speed of 1.60 GHz and a maximum boost clock speed of 2.11 GHz. The operating system used was Windows 10. To meet the computational demands of running Jupyter Notebook and handling data manipulation and model training processes, 8 GB of RAM was accommodated.

3.4. Imbalanced Classification

Class imbalance in machine learning occurs when the distribution of classes in the training dataset is uneven, posing challenges for models, particularly when the minority class is of interest. In the case of Ds2, addressing this imbalance is critical for accurate analysis of mental health conditions. The dataset covers diverse conditions such as dementia, stress, anxiety, depression, and others, with a focus on depression, anxiety, stress, and normal cases, totaling 10,106 records.

Notably, depression dominates with 7,654 records, while anxiety, stress, and normal cases are underrepresented (438, 1,027, and 987 records, respectively). The Synthetic Minority Over-Sampling Technique (SMOTE) process involves three steps: First, obtaining the k-nearest neighbors for each case in the minority class. Second, setting a sampling rate of N based on the imbalanced proportion, where N examples are randomly selected from the k-nearest neighbors for each example. Third, employing to generate new records, incorporating a random number factor (rand(0, 1)). Following oversampling, the Ds2 is reconstructed, enabling the application of various classification models for enhanced analysis.

3.5. Machine Learning Algorithms Logistic Regression (LR)

LR provides probabilistic interpretations, allowing for informed decision-making and setting decision thresholds based on desired trade-offs. Additionally, LR can help identify the importance of features in predicting outcomes, enabling feature selection, and understanding variable influence. It can handle outliers and missing values through techniques like robust regression and imputation. LR has a low risk of overfitting, particularly when the number of predictors is small, and regularization techniques can further mitigate overfitting. The penalty parameter determines the strength of regularization. The inverse of the regularization parameter C controls the amount of regularization applied (Equations 1 and 2). Smaller values of C increase the regularization strength, while larger values decrease it. The solver determines to optimize the LR parameters.

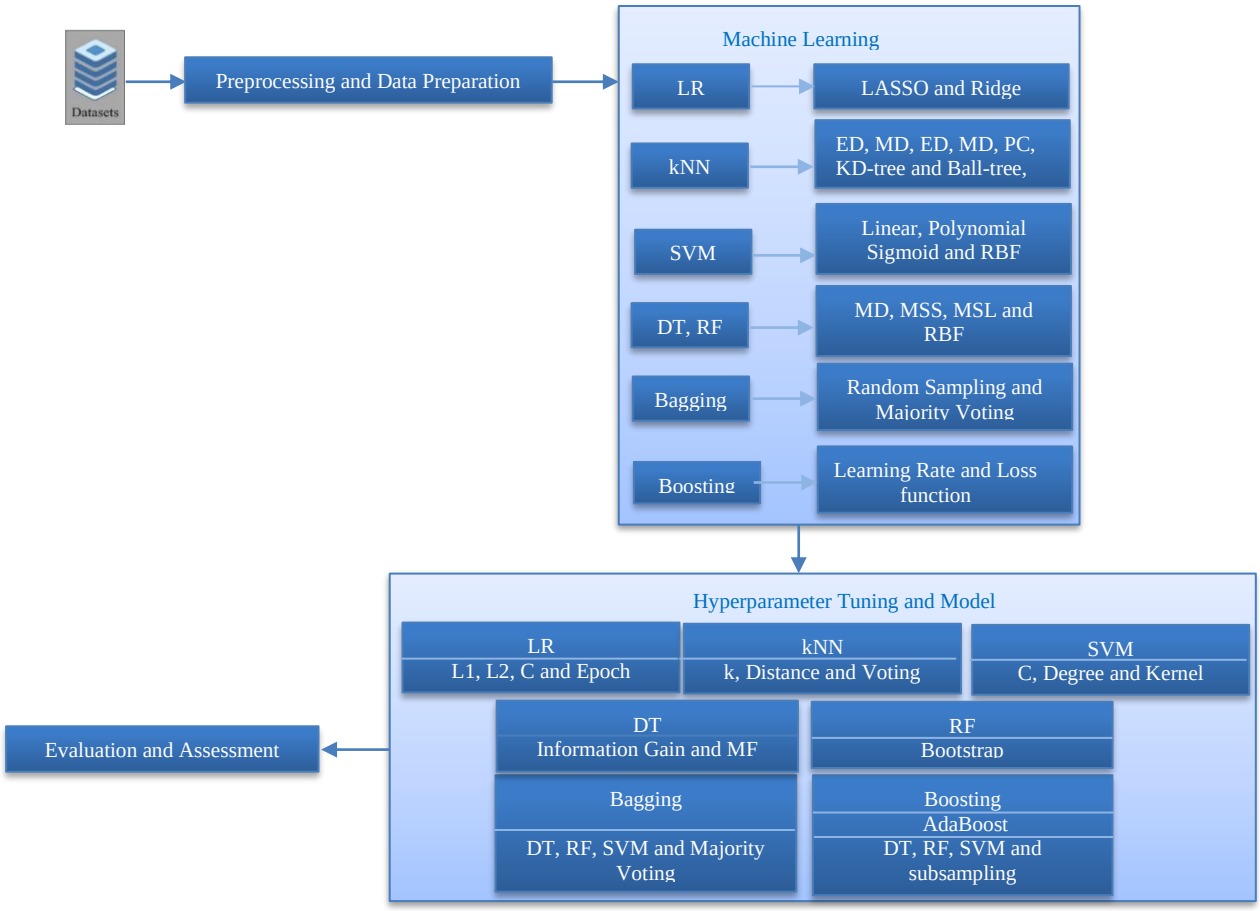


Fig. 1 Proposed framework to show the working mechanism considering parameters and hyperparameters

LR produces probability estimates, and a threshold can be set to classify instances of different predicted probabilities.

$$L1 = C \times \Sigma|\beta| \quad (1)$$

$$L2 = (C/2) \times \Sigma(\beta^2) \quad (2)$$

Where L1 and L2 are the regularization terms, C is a regularization parameter. β represents the coefficients of the LR. In LR, the loss function quantifies the uniqueness by gauging the difference between the predicted probabilities and the true labels. The goal is to minimize this loss function during model training. Logistic Loss (or cross-entropy loss (LL)) was used in our approach (Equation 3):

$$LL = -\Sigma\,[y \times \log(p) + (1-y) \times \log(1-p)] \quad (3)$$

Here, y represents the actual binary labels (0 or 1), and p represents the predicted probabilities. So, the objective function (F1) can be calculated as shown in Equation 4.

$$F1 = C \times (L1\ \&\ L2) + (LL) \quad (4)$$

The regularization term is typically the L2 norm (Euclidean norm) of the coefficients multiplied by 1/2. For the "lbfgs", "newton-cg", or "sag" solvers, C is directly used as the regularization strength in the objective function without inversion. The objective function is defined as (Equation 5):

$$F1 = C \times (L2) \quad (5)$$

The key parameters and hyperparameters of LR include penalty (Regularization) with options like L1 regularization (LASSO) and L2 regularization (Ridge) alongside C, which represents the inverse of regularization strength.

3.6. k-Nearest Neighbour (kNN)

It is a machine intelligence approach to classification. It assigns the class or value of a data point based on the average of its nearest neighbors. It calculates distances between the data points, considering training data points to find the closest neighbors. kNN is non-parametric, handles numerical and categorical features, and is easy to implement. kNN offers a simple yet effective approach for prediction based on the similarity of data points. In kNN, the voting method is used to determine the class or label of a new data point based on the classes or labels of its k nearest neighbors. Majority voting and weighted voting were used in this paper for experimentation. In the majority voting method, each neighbor has an equal vote, and the class with the highest number of votes among the k neighbors is assigned as the Predicted Class (PC) for the new data point (Equation 6).

$$PC=\max(CC) \quad (6)$$

Here, CC represents the count of occurrences of each class, and max () is a function that returns the class with the highest count. In the weighted voting method, each neighbor's vote is weighted based on its proximity to the new data point. The closer neighbors have a higher influence on the prediction than the farther ones (Equation 7).

$$PC = \max(\text{sum}(W \times CC)) \quad (7)$$

Here, W represents the weights assigned to each neighbor. In our approach, W was calculated using the following three distance measures (Equations 8, 9 and 10).

Euclidean distance (ED): $$d= \sqrt{\sum_{i=1}^{n}(X_i - Y_i)} \quad (8)$$

Manhattan distance (MD): $$d= \sum_{i=1}^{n} |X_i - Y_i| \quad (9)$$

Pearson Correlation (PC): $$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}} \quad (10)$$

3.7. Support Vector Machine (SVM)

It is a powerful machine learning algorithm that operates by finding an optimal hyperplane that separates data points into distinct classes. SVM's key strength lies in its ability to handle high-dimensional feature spaces and complex decision boundaries effectively. By utilizing a kernel function, it can implicitly map input data into higher dimensions, allowing for nonlinear classification. By prioritizing a wide margin, SVMs can effectively handle complex datasets and avoid overfitting (Equation 11). So hard as well as soft margins were considered.

$$\text{Minimize: } 0.5 \times ||w||^2 + C \times \Sigma(\max(0, 1 - y_i \times (w^T \times x_i + b))) \quad (11)$$

Subject to: $y_i \times (w^T \times x_i + b) \geq 1$, for all training samples (x_i, y_i)

Where

w: value based on the computed weight vector.

b: value based on the computed bias term.

x_i represents the feature vector.

y_i is the decision label (-1 or +1) (0 or 1).

The kernel function (K) is responsible for mapping the input data into a higher-dimensional feature space, where linear separation may be easier to achieve (Equation 12). The kernel functions are depicted in Equations 13-16.

$$K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j) \quad (12)$$

Here, xi and xj are input data points, and φ() represents the mapping to a higher-dimensional feature space. Linear Kernel (LK), Polynomial Kernel (PK), RBF Kernel (Gaussian Kernel) (RBFK) and Sigmoid Kernel (SK) are shown below (Equations 13-16).

$$LK= K(x_i, x_j) = x_i \cdot x_j \quad (13)$$

$$PK= K(x_i, x_j) = (\gamma \cdot x_i \cdot x^j + r)d \quad (14)$$

Here, γ is the coefficient of the kernel, r is an optional constant, and d is the degree of the polynomial.

$$RBFK= K(x_i, x_j) = \exp(-\gamma \cdot ||x_i - x_j||^2) \quad (15)$$

Here, $||x_i - x_j||$ represents the Euclidean distance between x_i and x_j.

$$SK= K(x_i, x_j) = \tanh(\gamma \cdot x_i \cdot x_j + r) \quad (16)$$

The gamma parameter (γ) determines the influence of each training sample. A higher gamma value emphasizes closer samples (Equation 17).

$$\gamma = 1 / (2 \times \sigma^2) \quad (17)$$

Here, σ is the standard deviation of the Gaussian function used in the RBFK. The smaller the value of σ, the larger the value of γ, resulting in a narrower RBFK.

3.8. Decision Tree (DT)

DT is a popular computational learning algorithm. It represents decisions and their consequences in a tree-like structure. DTs can handle both categorical and numerical features, and they can capture complex interactions between them. Additionally, DT can handle missing values and are robust to outliers. The uniqueness lies in the control of DT complexity through two parameters. The Max-Depth (MD) parameter governs the upper limit on tree depth, while the Min_Samples_Split (MSS) parameter determines the samples (minimum number) needed to split the node (internal). It ensures that nodes are not split if they contain a small number of samples, which can help prevent overfitting. The Min_Samples_Leaf (MSL) parameter sets the samples (minimum number) required to be a node (leaf). It prevents the creation of leaf nodes with too few samples, which can also aid in reducing overfitting. The max-Features (MF) parameter determines the features (maximum number) to consider when looking for the best split at each node. It helps control the randomness and complexity of the DT.

3.9. Random Forest (RF)

It is a widely used ensemble algorithm, combining DTs through bootstrapping and random feature selection to enhance predictions. The exceptional feature of this method lies in its capability to manage high-dimensional data and capture intricate relationships among variables effectively. Additionally, it can estimate feature importance, allowing for effective feature selection. One of its distinctive attributes is its computational efficiency, allowing it to parallelize efficiently and handle large datasets with ease. The parameter n_estimators (n) determines the number of DTs in RF. Increasing n can enhance performance but also adds

computational complexity. Bagging, or Bootstrap aggregating, is an ensemble learning technique aimed at increasing the stability and accuracy of learning models. It involves creating multiple sets through random sampling with replacement. Each subset, known as a bootstrap sample, trains models independently. The models' outputs are combined via averaging (for regression) or majority voting (for classification) for the final prediction, reducing overfitting by diversifying training subsets. This enhances performance by lowering variance and bolstering resilience to outliers.

Boosting is a machine learning method that builds a strong ensemble model by sequentially training models, each focusing on samples poorly predicted by predecessors. Each model's weighted predictions are combined for the final output. It emphasizes misclassified samples, enabling learning from prior errors. AdaBoost was used in this study. Boosting excels in managing complex datasets and elevating predictive accuracy.

4. Results and Discussion

The results obtained using the discussed methods are presented in this section, considering various cases. The experimental cases are outlined in Table 1. Evaluation metrics considered for results assessment are as follows:

Accuracy: It is calculated by dividing the number of correctly predicted instances by the total number of instances. The formula is shown in Equation 18:

$$\text{Accuracy} = \frac{\text{NCPI}}{\text{TNI}} \tag{18}$$

Where NCPI = Number of correctly predicted instances, TNI = Total number of instances.

Precision: It is calculated by dividing the number of true positives by the sum of true positives and false positives. The formula is shown in Equation 19:

$$\text{Precision} = \frac{\text{TPs}}{\text{TPs+FPs}} \tag{19}$$

Where TPs=True positives, FPs= False positives, FNs= False negatives, TNs=True negatives. Recall (Sensitivity): It is also known as sensitivity or the true positive rate, and it is calculated by dividing the number of true positives by the sum of true positives and false negatives. The formula is shown in Equation 20:

$$\text{Recall} = \frac{\text{TPs}}{\text{TPs+FNs}} \tag{20}$$

Specificity: It is also known as the true negative rate and is calculated by dividing the number of true negatives by the sum of true negatives and false positives. The formula is shown in Equation 21:

$$\text{Specificity} = \frac{\text{TNs}}{\text{TNs+FPs}} \tag{21}$$

F1-Score: It is the harmonic mean of precision and recall. The formula is shown in Equation 22:

$$\text{F1-Score} = 2\times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

The coefficient of determination (R^2) is used in this study to demonstrate the accuracy and validity of the model. The value of R^2 may be calculated with the help of Equation 23:

$$R^2 = 1 - \frac{SSR}{SST} \tag{23}$$

Where SSR represents the Sum of Squared Residuals, which corresponds to the summation of squared differences between predicted values and actual values, on the other hand, SST represents the total sum of squares, being the sum of squared differences between actual values and the mean of the dependent variable.

For ES1, the results were obtained by considering L1, L2, C, and epochs in a complete cycle, using tenfold cross-validation. The dataset was split into different ratios: 70:30, 75:25, and 80:20. The number of epochs ranged from 100 to 600. The results are presented on a scale of 0 to 1. LR-L1 corresponds to LASSO regularization, LR-L2 corresponds to ridge regularization and LR with optimal threshold (LR-OT). From Figure 2, it is evident that LR-OT achieved the best results for both the Ds1 and Ds2 datasets. In the case of Ds1, the precision, recall, specificity, F1-score, and accuracy achieved were 0.88, 0.89, 0.87, 0.87, and 0.88, respectively. In the case of Ds2, the precision, recall, specificity, F1-score, and accuracy achieved were 0.87, 0.87, 0.88, 0.88, and 0.89, respectively. LR-L1 outperformed LR-L2 in the case of Ds1, which can be attributed to L1 regularization being suitable when a subset of features has a significant impact on mental health outcomes.

Table 1. Different cases considered for experimentation and assessment

S. No	Experimental Scenario (ES)	Description
1	ES1	LR with D1 and D2 (Parameters and Hyperparameters (L1, L2, C, epochs))
2	ES2	kNN with D1 and D2 (Parameters and Hyperparameters (k, distance metric, KD-tree, Ball-tree, and Voting method)
3	ES3	SVM with C, degree, and different kernel
4	ES4	DT, RF, Bagging with DT, Bagging with RF and Bagging with SVM
5	ES5	AdaBoost with DT, AdaBoost with RF and AdaBoost with SVM

It automatically selects the most relevant features, leading to improved performance. On the other hand, LR-L2 outperformed LR-L1 in the case of Ds2 because, in Ds2, all features are considered relevant. This indicates that ridge regularization is useful when all features are relevant or when multicollinearity is present. Figure 3 depicts the analysis of accuracy results obtained from the Ds1 and Ds2 datasets, considering the variation in epochs. The aim was to observe how different epoch values influenced the accuracy of the applied models. The findings revealed only minor variations across the range of epochs.

For ES2, the selection of the value of k was initially performed using cross-validation and the square root method. Figure 4 depicts the selection of k based on the accuracy obtained during epochs, considering the Ds1 and Ds2 datasets. The results indicate that the highest accuracy was achieved within the range of k (4-7) for Ds1, while for Ds2, the highest accuracy was attained within the range of k (8-11). Since both ranges yielded the same best score, a smaller value for k was chosen after carefully examining the individual accuracies. This decision was made because higher values of k may involve the inclusion of data points that are farther away from the original. Consequently, for Ds1, k = 5 was selected, and for Ds2, k = 8 was chosen.

To improve the efficiency of the kNN algorithms by hierarchically organizing the data points, the KD tree data structure was employed. This structure allows for faster search times for the nearest neighbors. Additionally, the Ball tree data structure was utilized for optimizing the kNN algorithms. Subsequently, a voting method was applied to reduce bias and variance, with distance-based voting being considered for this purpose. This selection is referred to as the optimized version of kNN, denoted as kNNO.

Further assessment was conducted based on kNNO. Figure 5 presents the results of precision, recall, specificity, F1-score, and accuracy metrics considering different variations of kNN with the Ds1 and Ds2 datasets. kNNO-ED outperforms all other variations with an average accuracy of 90% for D1 and 91% for D2. The combination of the ED or MD metrics with the kNN algorithm has demonstrated a substantial improvement in accuracy. In contrast, when kNN is combined with the PC coefficient, the achieved results are less significant compared to the ED or MD. These findings suggest that utilizing either the ED or MD metrics in conjunction with the kNN algorithm leads to more substantial enhancements in accuracy, while the PC coefficient does not yield the same level of improvement.

For ES3, in the case of SVM, parameter combinations and hyperparameter tuning were performed considering various factors. Given that the dataset has been thoroughly cleaned and preprocessed, a higher value of C was chosen to prioritize the accurate classification of the training set. In the first case, SVMO1, the SVM model with the RBFK was employed. The RBFK is effective in capturing complex, non-linear relationships within the data. The decision boundary was constructed using Gaussian functions. The choice of gamma value is crucial, as a higher gamma makes the decision boundary more focused on individual training examples, which can potentially lead to overfitting. Conversely, a lower gamma value results in a smoother decision boundary that may generalize better.

Therefore, the selection of the optimal gamma value is crucial. In the second case, SVMO2, the degree parameter was considered in combination with the PK. The degree parameter controls the flexibility of the decision boundary and determines the complexity of the polynomial function used to transform the data. A higher degree allows for more complex decision boundaries, but it also increases the risk of overfitting. Finally, SVMO3 represents the combination of optimal parameter values obtained through the combined approach. Figure 6 shows the results of precision, recall, specificity, F1-score, and accuracy metrics considering SVM variations considering the Ds1 and Ds2 datasets. The average accuracy achieved from the final selection, SVMO3, is 91% for Ds1 and 93% for Ds2.

For ES4, the parameters considered are MD, MSS and MSL. The MD of a DT and RF limits the number of splits and controls the complexity of the tree. MSS establishes the minimum number of samples necessary to split an internal node. Raising this value can help prevent overfitting by ensuring an internal node has enough samples before it is split, but it might lead to a more generalized tree.

MSL, on the other hand, signifies the minimum number of samples that must be present in a leaf node. Setting a higher value for MSL can help prevent overfitting by ensuring that each leaf contains enough samples. By enforcing this constraint, the model avoids creating leaf nodes with very few samples, which could lead to capturing noise or outliers in the data. It encourages the model to make more generalized splits, resulting in a simpler and less complex tree structure.

To determine the optimal range of sub-models, we created bagging ensemble models for each base learner, with a varying number of component sub-models. Specifically, we generated bagging ensembles with 20 sub-models each, ranging from 10 to 200 sub-models in increments of 10. To identify the best structures, we examined the correlation coefficient values between the sub-models within each ensemble. High correlation coefficients indicated strong agreement among the sub-models and were used as a criterion for selecting the best structures.

By analysing these correlation coefficients, we were able to identify the optimal range of sub-models for each base learner. Figure 7 shows the R^2 score result for ensemble

models predicting compressive strength using different sub-models (Bagging with DT, Bagging with RF and Bagging with SVM). As depicted in Figure 7, the ensemble model utilizing DT outperforms other boosting models.

Notably, the ensemble model demonstrates strong performance with values exceeding 0.70, indicating favourable results. Figure 8 presents the results of precision, recall, specificity, F1-score, and accuracy metrics for different bagging variations, considering the D1 and D2 datasets. Bagging with DT achieves the highest average accuracy, reaching 94% for Ds1 and 96% for Ds2, showcasing its superior performance in both cases. For ES5, the implementation of AdaBoost with DT, AdaBoost with RF, and AdaBoost with SVM was considered. Figure 9 illustrates the performance of the ensemble models using Adaboost, with a particular focus on the R2 metric for prediction accuracy.

The ensemble model combining Adaboost with DT outperforms the other boosting models, yielding a high R^2 value. Notably, the ensemble model demonstrates strong performance with values exceeding 0.72, indicating favourable results. Figure 10 showcases the comprehensive results of precision, recall, specificity, F1-score, and accuracy metrics for various bagging variations, considering the Ds1 and Ds2 datasets. Boosting with DT exhibits remarkable performance, achieving the highest average accuracy of 95% for Ds1 and 96% for Ds2. These results highlight the superiority of the ensemble model employing boosting with DT, indicating its effectiveness in both cases.

Figure 11 depicts the accuracy of the Ds1 and Ds2 datasets in evaluating the impact and average impact of different split ratios. The aim was to investigate how varying split ratios, such as 70:30, 75:25, and 80:20, affected the accuracy of the applied models and determine their average impact. The observed differences in the split ratios were negligible, suggesting that any of the examined split ratios can be selected without causing a significant impact on the results. The Receiver Operating Characteristic (ROC) curve serves as a visual depiction of a classifier's efficacy across diverse discrimination thresholds, offering insights into the trade-offs between sensitivity and false positive rates. In contrast, the Area Under the Curve (AUC) is a singular numerical metric encapsulating the classifier's overall performance. This scalar value ranges from 0 to 1, with 0.5 denoting a classifier akin to random chance and 1 indicating perfection. The ROC curve materializes by plotting sensitivity against the false positive rate across distinct threshold settings. As illustrated in Figure 12, it becomes apparent that the AUC values for the ROC curves of RF surpass those of LR, kNN, SVM and DT in the case of Ds1. Notably, the RF exhibits superiority across all assessed metrics. This suggests that the RF model outperforms others. The consistently higher AUC values for RF indicate its robust performance in achieving a balanced trade-off between sensitivity and specificity across diverse classification thresholds. RF's proficiency in capturing complex relationships and handling intricate data patterns contributes to its superior discrimination ability.

In the case of Ds2, it is indicated that SVM outperforms LR, kNN, DT, and RF when it comes to multi-class classification (Figure 13). The superior performance of SVM in Ds2's multi-class classification scenario is due to handling cdecision boundaries and intricate relationships within the data effectively. SVM excels in finding optimal hyperplanes that separate different classes, making it particularly robust when dealing with diverse and overlapping class distributions. The data was divided into ten segments to align with the tenfold cross-validation approach. Cross-validation assesses each method's performance on simulated data. The mean cross-validation score, computed by averaging results across the ten folds, is presented in Tables 2 and 3. The outcomes highlight the superior performance of both RF and DT over all other algorithms.

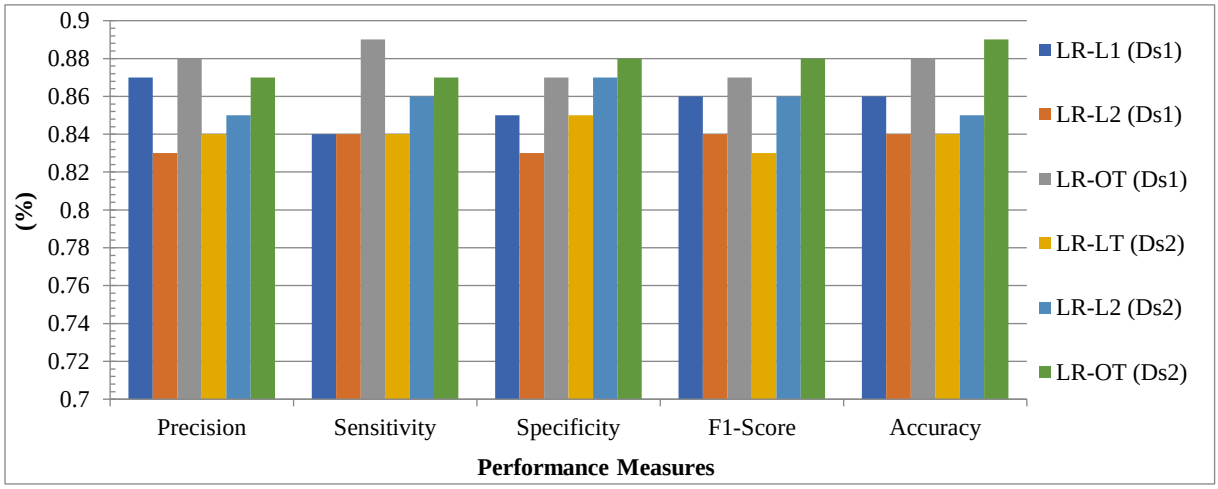


Fig. 2 The performance metrics (Precision, Sensitivity, Specificity, F1-Score, and Accuracy) were evaluated for both the Ds1 and Ds2 datasets in the case of LR

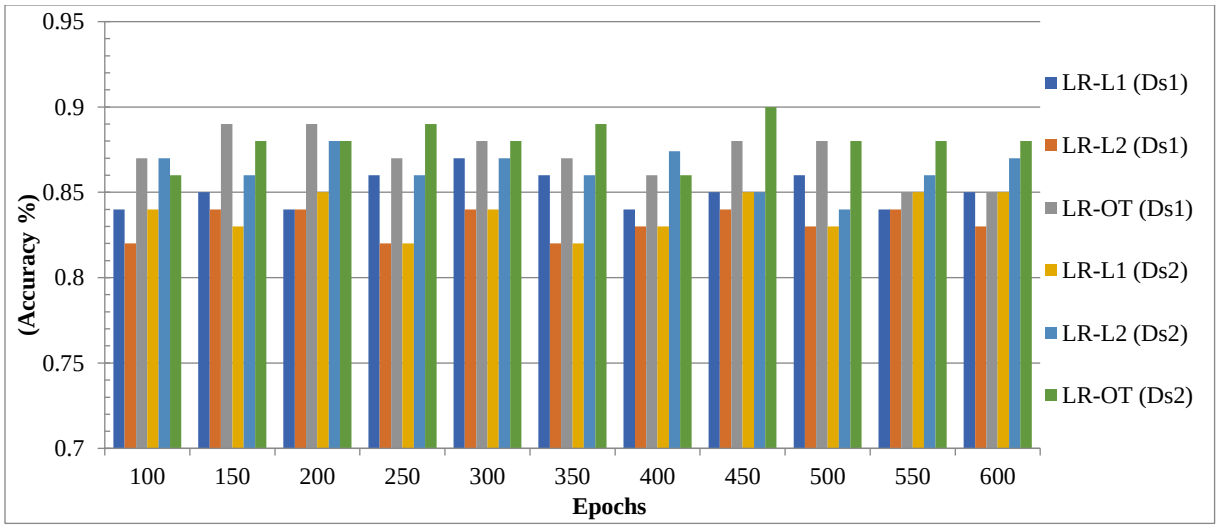


Fig. 3 Accuracy obtained from the Ds1 and Ds2 datasets was examined, considering the variation in epochs

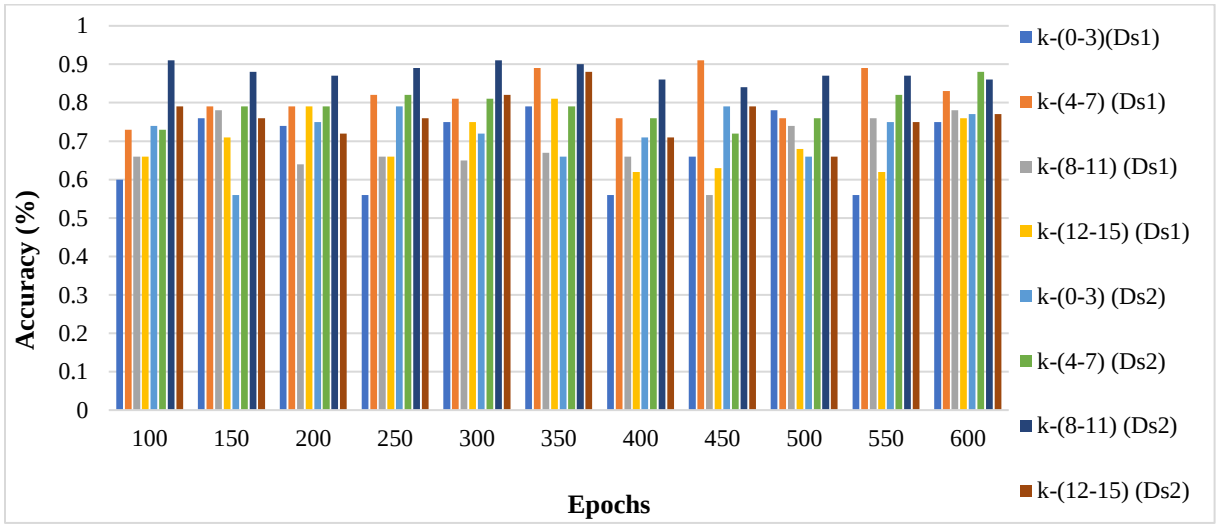


Fig. 4 Selection of k based on the accuracy obtained on epochs, considering the Ds1 and Ds2 datasets

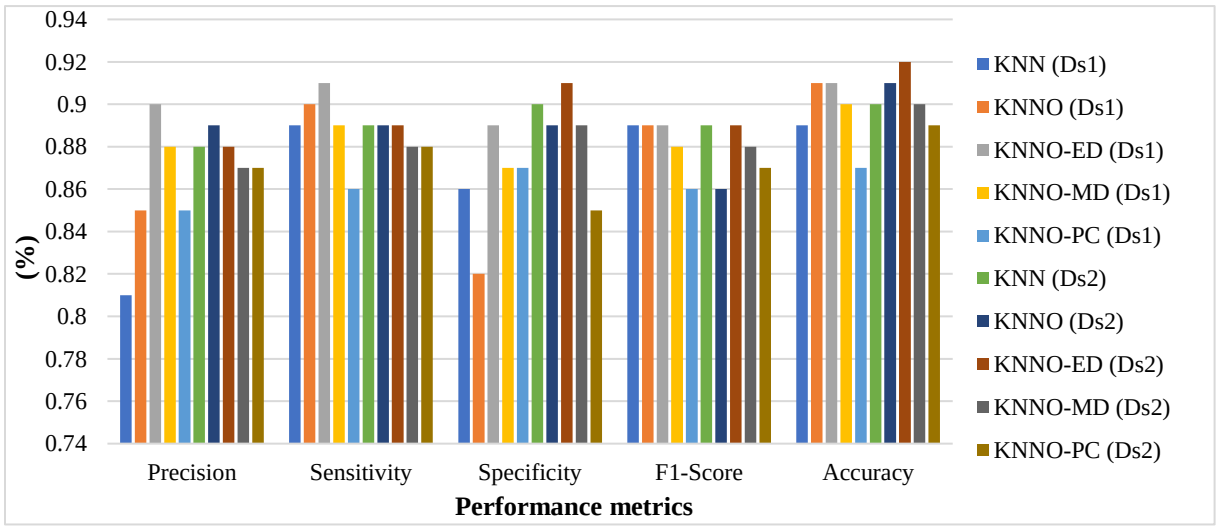


Fig. 5 The performance measures (Precision, Sensitivity, Specificity, F1-Score, and Accuracy) for various kNN variations were evaluated using the Ds1 and Ds2 datasets

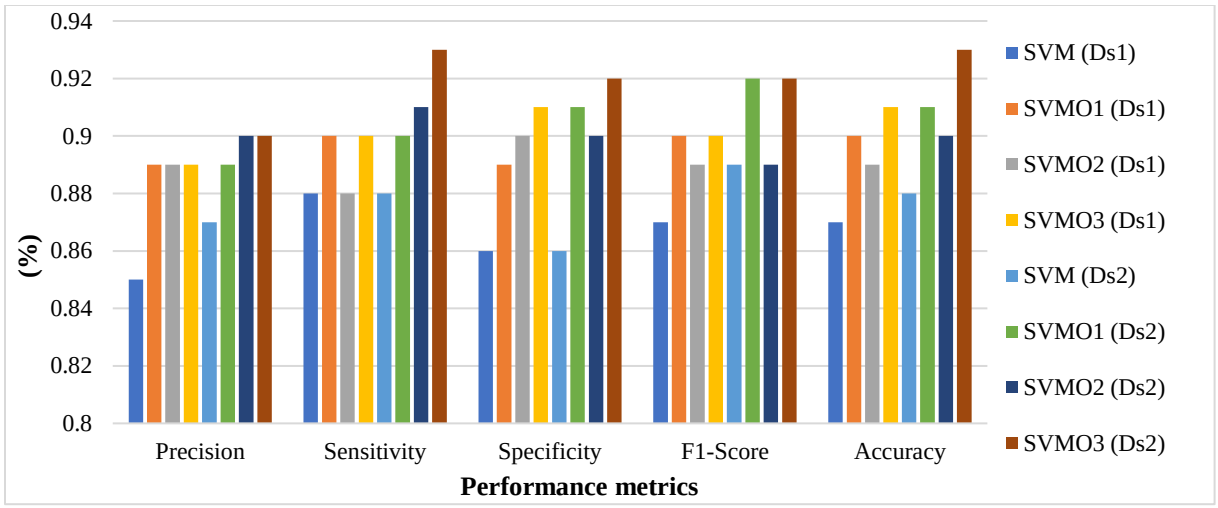


Fig. 6 The performance metrics (Precision, Sensitivity, Specificity, F1-Score, and Accuracy) for different variants of SVM utilizing Ds1 and Ds2 datasets.

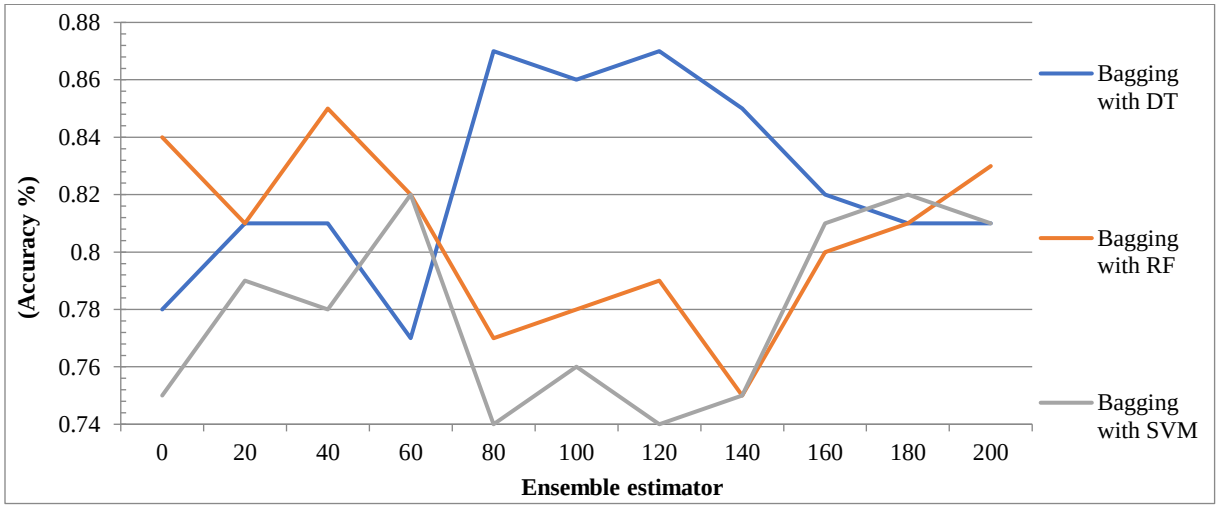


Fig. 7 R^2 score measure for ensemble models predicting compressive strength using different sub-models (DT, RF, Bagging with DT, Bagging with RF and Bagging with SVM)

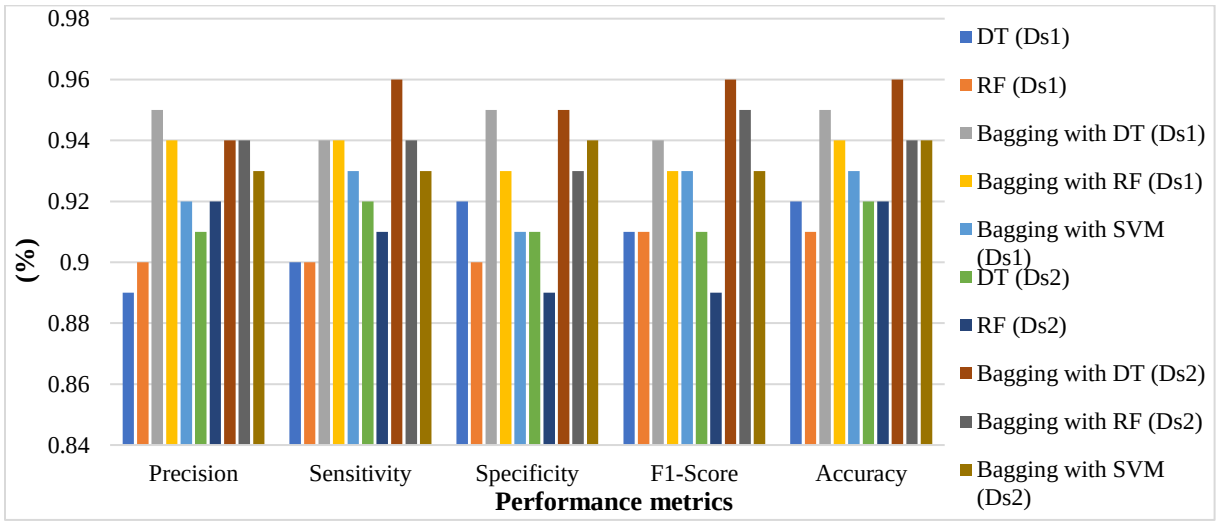


Fig. 8 The performance measures (Precision, Sensitivity, Specificity, F1-Score, and Accuracy) for bagging variations utilizing the Ds1 and Ds2 datasets.

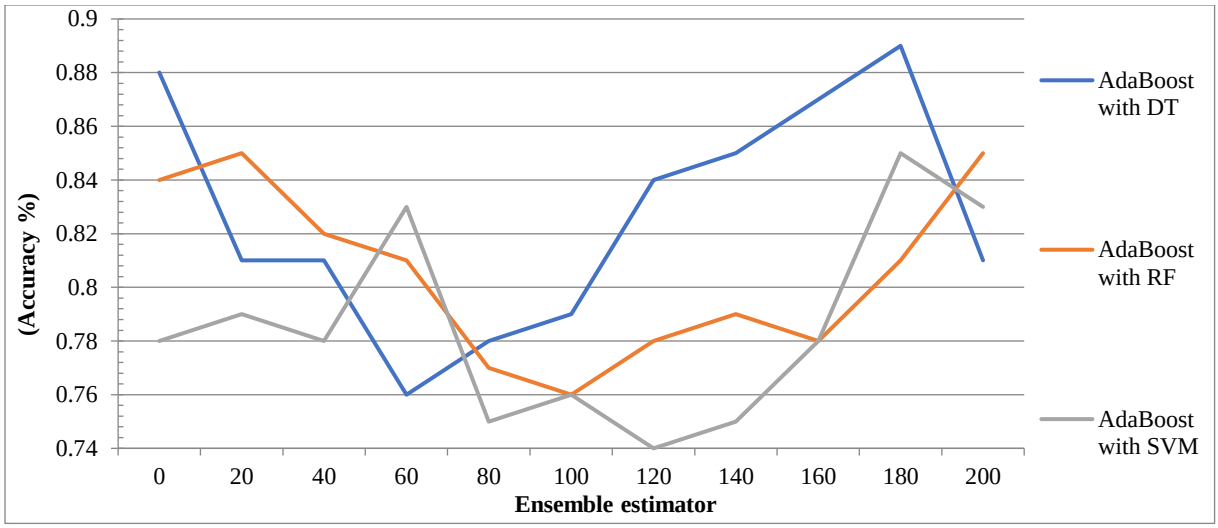


Fig. 9 R^2 score measure for ensemble models predicting compressive strength using different sub-models (AdaBoost with DT, AdaBoost with RF and AdaBoost with SVM)

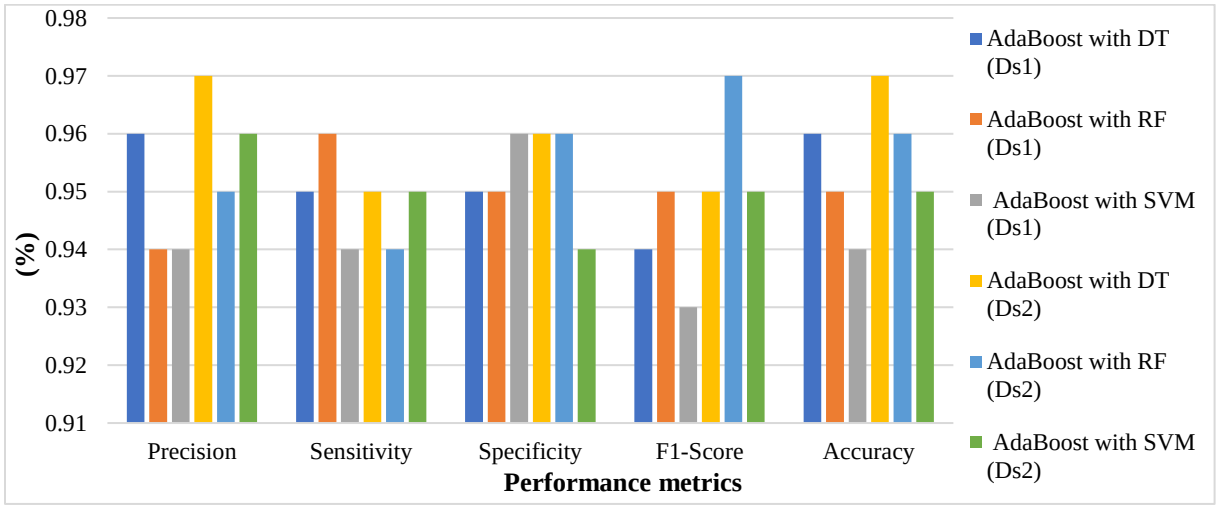


Fig. 10 The performance measures (Precision, Sensitivity, Specificity, F1-Score, and Accuracy) for different boosting variations were evaluated using the Ds1 and Ds2 datasets

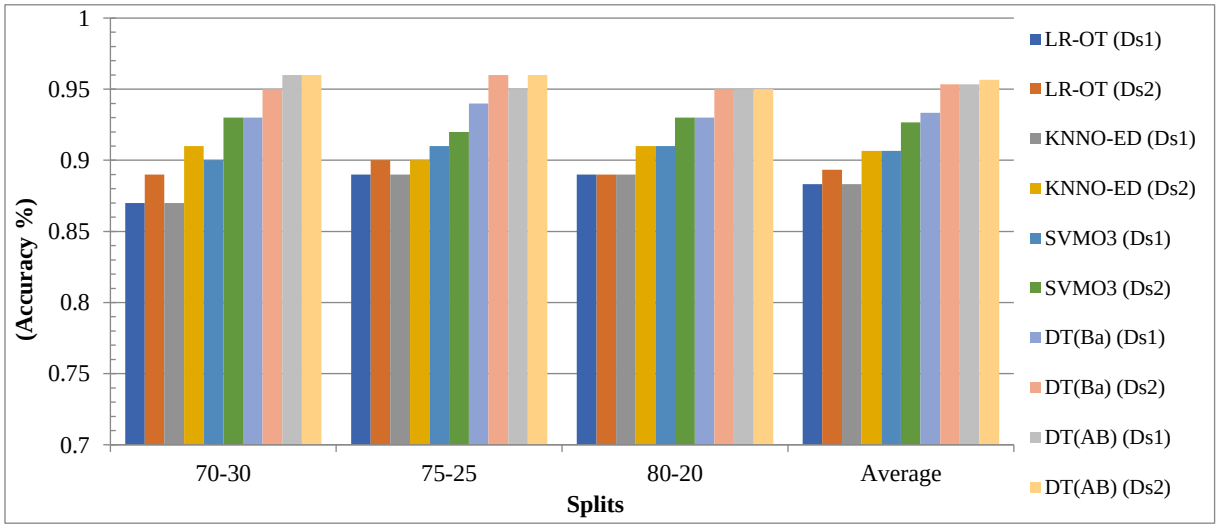
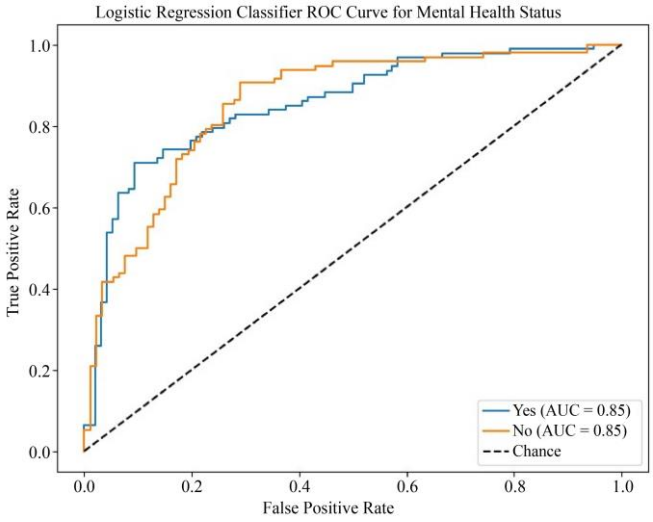
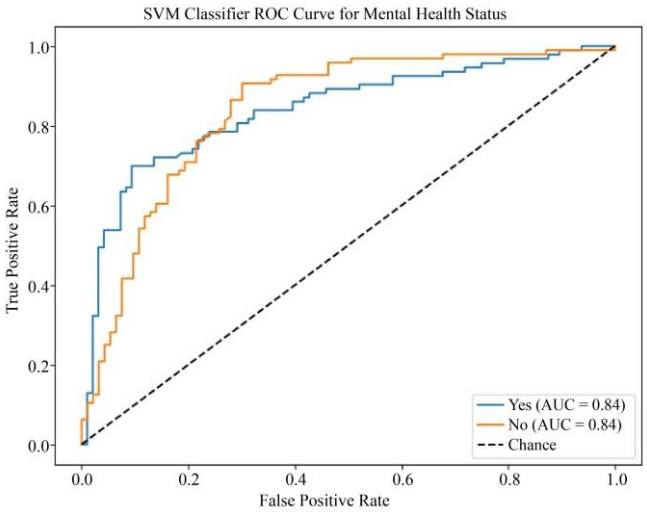


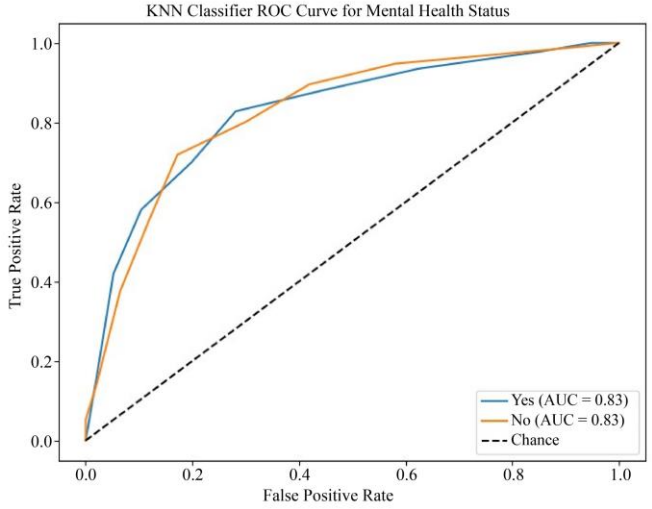
Fig. 11 Accuracy of the Ds1 and Ds2 datasets to assess the impact and average impact of varying split ratios



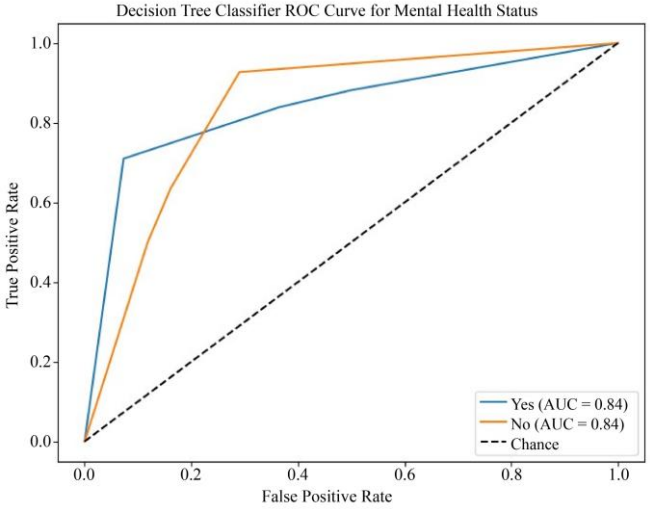
(a)



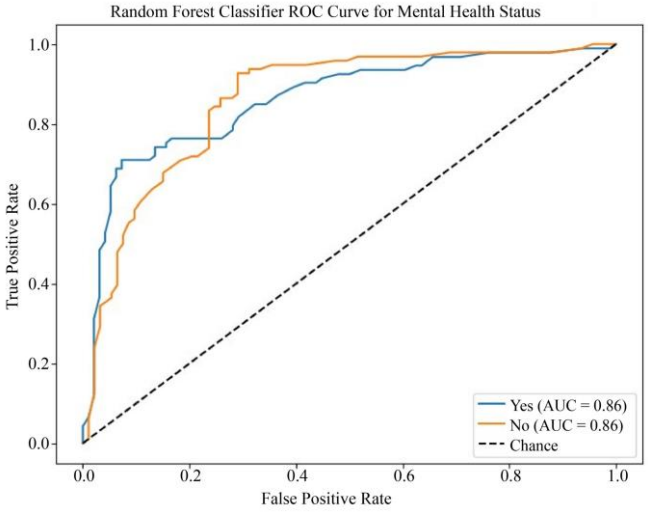
(c)



(b)

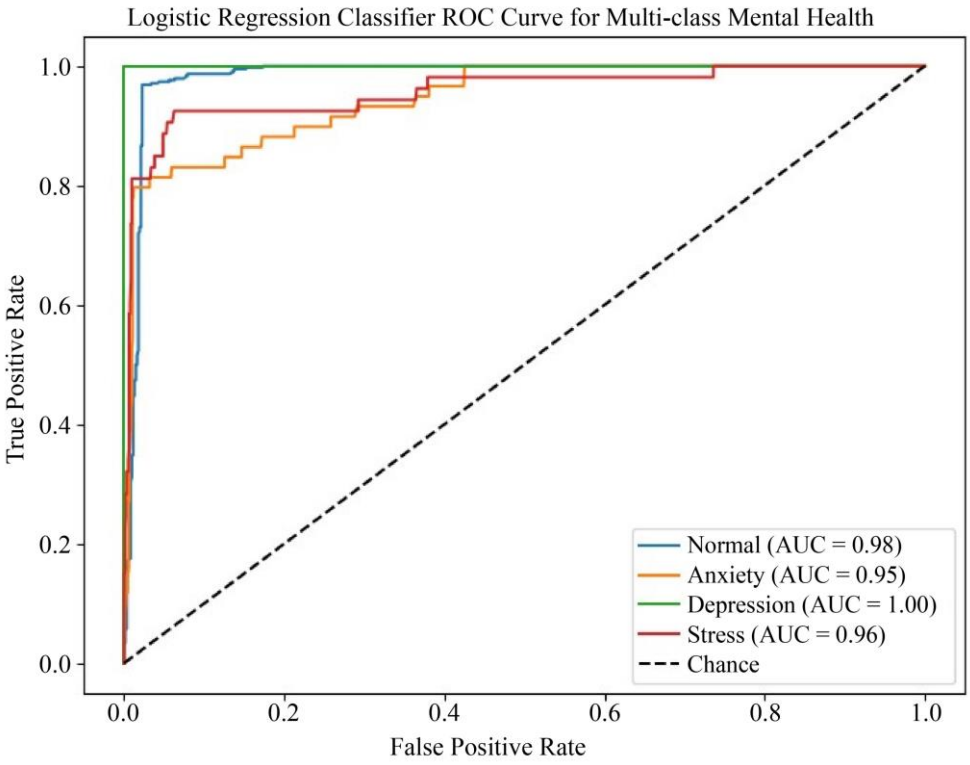


(d)

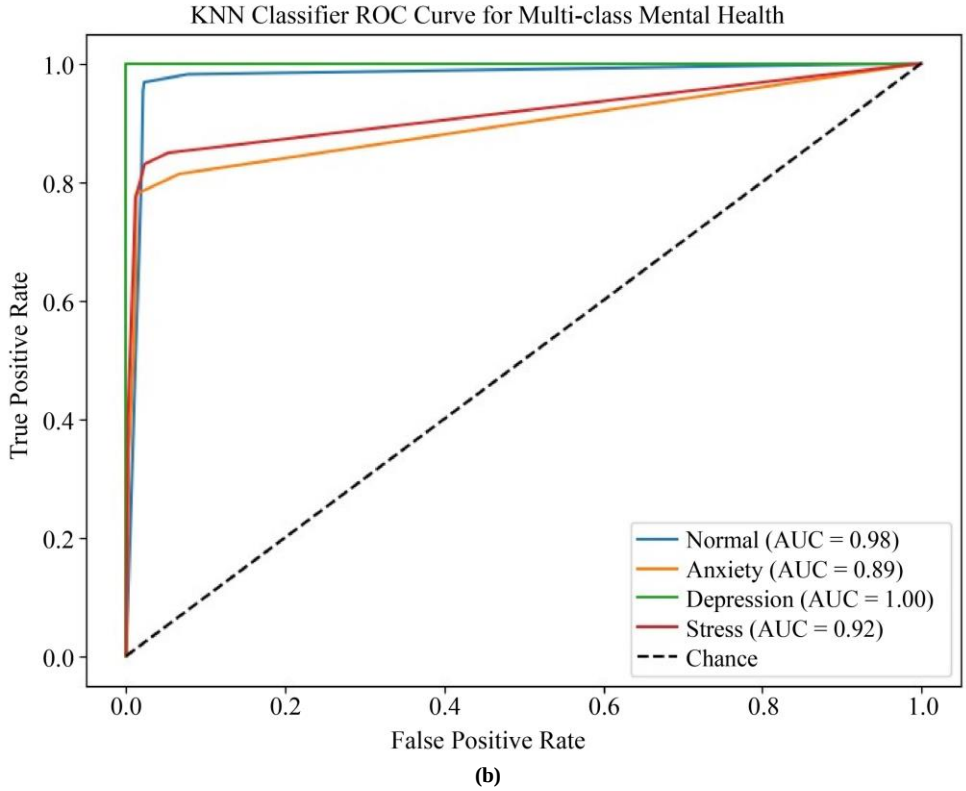


(e)

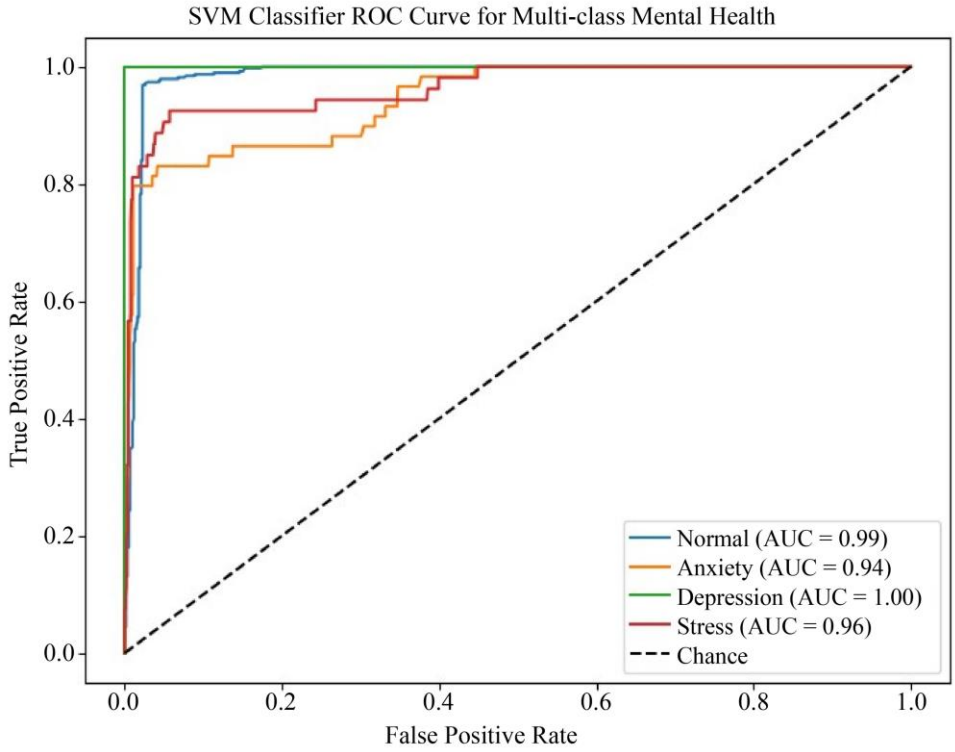
Fig. 12 ROC curve for (a) LR, (b) kNN, (c) SVM, (d) DT and (e) RF classifier on Ds1 dataset


Logistic Regression Classifier ROC Curve for Multi-class Mental Health
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Normal (AUC = 0.98)
Anxiety (AUC = 0.95)
Depression (AUC = 1.00)
Stress (AUC = 0.96)
Chance

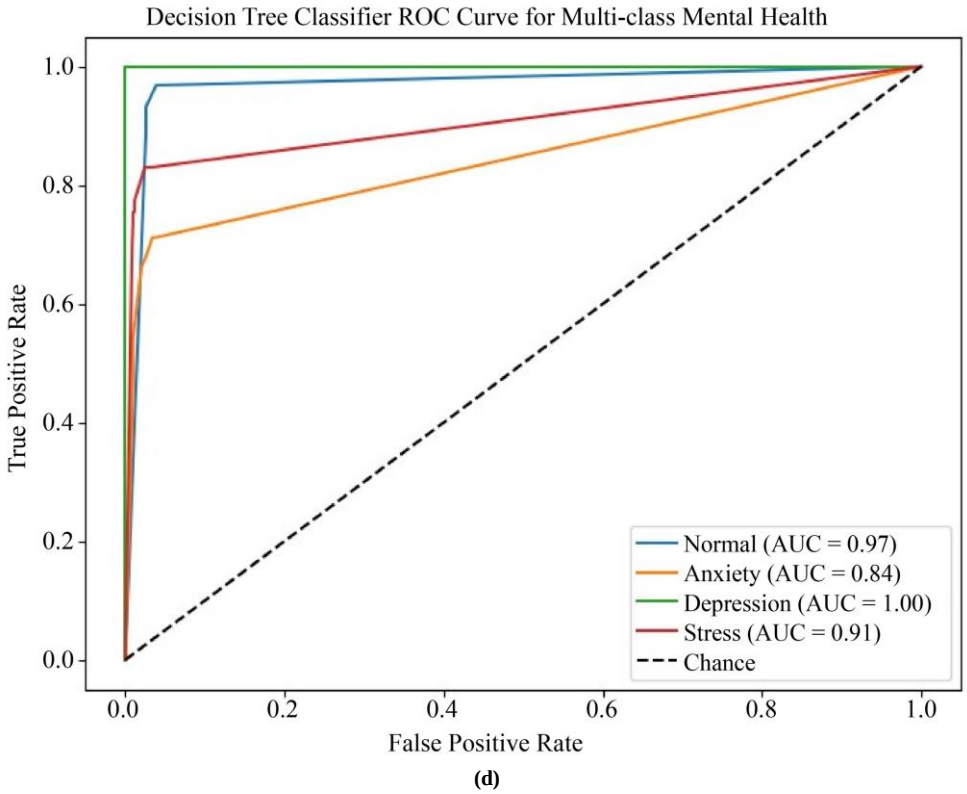
(a)


KNN Classifier ROC Curve for Multi-class Mental Health
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Normal (AUC = 0.98)
Anxiety (AUC = 0.89)
Depression (AUC = 1.00)
Stress (AUC = 0.92)
Chance

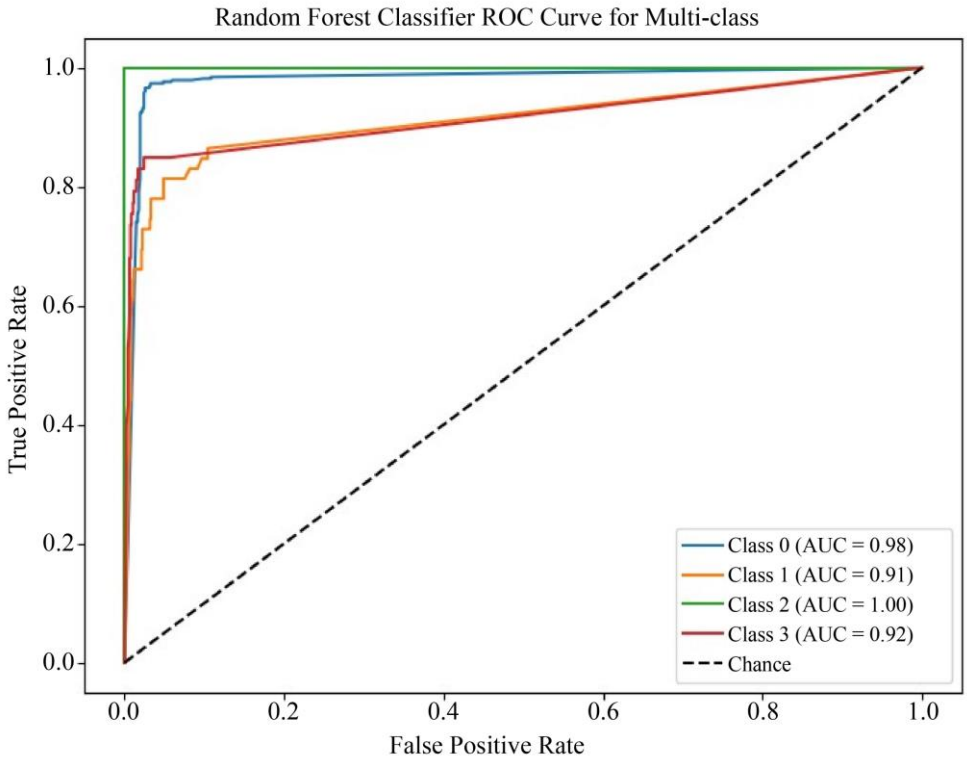
(b)


SVM Classifier ROC Curve for Multi-class Mental Health
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Normal (AUC = 0.99)
Anxiety (AUC = 0.94)
Depression (AUC = 1.00)
Stress (AUC = 0.96)
Chance

(c)


Decision Tree Classifier ROC Curve for Multi-class Mental Health
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Normal (AUC = 0.97)
Anxiety (AUC = 0.84)
Depression (AUC = 1.00)
Stress (AUC = 0.91)
Chance

(d)



(e)

Fig. 13 ROC curve for (a) LR, (b) kNN, (c) SVM, (d) DT and (e)RF classifier on Ds2 dataset

Table 2. K-fold cross-validation of the machine learning algorithms with best optimal values in the case of Ds1 (LR, kNN, SVM, DT, RF)

Algorithms	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	Mean accuracy
LR	0.88	0.83	0.87	0.85	0.87	0.88	0.86	0.84	0.87	0.89	0.864
kNN	0.91	0.91	0.9	0.87	0.9	0.91	0.92	0.9	0.89	0.91	0.902
SVM	0.92	0.93	0.9	0.93	0.9	0.92	0.89	0.92	0.89	0.93	0.913
DT	0.94	0.93	0.95	0.94	0.92	0.92	0.92	0.94	0.94	0.95	0.935
RF	0.93	0.94	0.94	0.94	0.92	0.92	0.92	0.94	0.93	0.93	0.931

I denote the iterations.

Table 3. K-fold cross-validation of the machine learning algorithms with best optimal values in the case of Ds2 (LR, kNN, SVM, DT, RF)

Algorithms	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	Mean accuracy
LR	0.93	0.92	0.93	0.92	0.91	0.93	0.93	0.93	0.91	0.93	0.924
kNN	0.94	0.96	0.94	0.94	0.95	0.95	0.94	0.96	0.95	0.95	0.948
SVM	0.96	0.95	0.95	0.95	0.96	0.95	0.97	0.96	0.95	0.95	0.955
DT	0.96	0.96	0.97	0.96	0.97	0.97	0.97	0.96	0.96	0.96	0.964
RF	0.98	0.96	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.97	0.968

I denote the iterations.

5. Limitations and Future Scope

1. The sample size of the Ds1, consisting of limited participants, may raise concerns about its representativeness. The predominance of male participants, approximately 80%, and the low percentage of unidentified gender (less than 1%) could potentially introduce biases in the results. Additionally, the dataset's composition, collected over multiple years, may pose challenges in maintaining consistency across different temporal contexts. The Ds2 dataset's regional focus may limit the generalizability of findings to a broader population.
2. Variability in mental health conditions and symptom presentation across diverse populations can affect the generalizability of algorithm effectiveness. The sensitivity of algorithms to the specific characteristics of mental health data introduces uncertainty in their relative effectiveness.
3. This study utilized a combined approach to determine the

optimal parameter values for the models. However, alternative parameter optimization processes can also be employed. Future research can consider alternative optimization techniques, such as grid search or Bayesian optimization, to enhance model performance and reproducibility.

4. While this study emphasizes predictive accuracy, it is equally important to consider the interpretability of the models. Alongside predictive accuracy, future studies can focus on the interpretability of models. Techniques such as feature importance ranking, model explanations, or decision rules can be employed to understand the factors driving predictions, thereby increasing the transparency and trustworthiness of the models.

6. Conclusion

In this study, a variety of machine learning algorithms (LR, kNN, SVM, DT, RF) were individually employed with different variable parameters. Additionally, bagging and boosting techniques were utilized with DT, RF, and SVM for mental health assessment. Two distinct datasets were utilized to evaluate the algorithms' performance, and the results revealed varying levels of effectiveness among them. LR-OT achieved high accuracy, considering different performance measures, highlighting its suitability for mental health assessment.

Among the different variations of kNN, the version with optimized parameters and ED outperformed the others. SVM with the RBFK and optimal parameter values demonstrated the best performance among all SVM variations. RF and DT with bagging and boosting techniques showcased remarkable results, surpassing other algorithms in terms of R2 score and accuracy. Boosting with DT exhibited remarkable performance, achieving the highest average accuracy of 95% for Ds1 and 96% for Ds2.

These findings emphasize the significance of ensemble methods in enhancing the accuracy and predictive power of mental health assessment models. The research outcomes provide valuable insights for leveraging machine learning algorithms to gain a deeper understanding of mental health assessment models and facilitate effective interventions across various domains, including the tech industry.The limited size of the datasets, particularly in certain mental health categories, introduces a level of uncertainty in the findings. Emphasizing the need for more rigorous evaluation and validation in future studies is crucial.

A larger and more diverse sample size would enhance the reliability and generalizability of the results. Future research can prioritize datasets that encompass a broader spectrum of mental health conditions, ensuring a comprehensive representation of the population. It is imperative to emphasize the necessity for exploring and testing other ensemble methods and deep learning techniques in future studies.

Acknowledgements

We want to express our gratitude to Dr. Surender Mandhan, Addl. SMO (Psychiatrist) at Lok Nayak Jai Prakash Hospital, Kurukshetra, Haryana

References

[1] Rialda Kovacevic, Mental Health: Lessons Learned in 2020 for 2021 and Forward, World Bank Blogs, 2021. [Online]. Available: https://blogs.worldbank.org/health/mental-health-lessons-learned-2020-2021-and-forward

[2] WHO Highlights Urgent Need to Transform Mental Health and Mental Health Care, World Health Organization, 2022. [Online]. Available: https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care

[3] Anna Z. Antosik-Wójcińska et al., "Smartphone as a Monitoring Tool for Bipolar Disorder: A Systematic Review Including Data Analysis, Machine Learning Algorithms and Predictive Modelling," *International Journal of Medical Informatics*, vol. 138, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Piyush Kumar et al., "A Machine Learning Implementation for Mental Health Care. Application: Smart Watch for Depression Detection," *2021 11th International Conference on Cloud Computing, Data Science & Engineering*, Noida, India, pp. 568-574, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ellen E. Lee et al., "Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 6, no. 9, pp. 856-864, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] Asha S. Manek et al., "Role of AI and Machine Learning in Mental Healthcare," *International Conference on Data Science and Big Data Analysis*, pp. 33-48, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Emma Nuraihan Mior Ibrahim, Nurulfitrie Jamali, and Ahmad Iqbal Hakim Suhaimi, "Exploring Gamification Design Elements for Mental Health Support," *International Journal of Advanced Technology and Engineering Exploration*, vol. 8, no. 74, pp. 114-125, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Chang Su et al., "Deep Learning in Mental Health Outcome Research: A Scoping Review," *Translational Psychiatry*, vol. 10, no. 1, pp. 1-26, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Md Abdur Rahaman et al., "Deep Multimodal Predictome for Studying Mental Disorders," *Human Brain Mapping*, vol. 44, no. 2, pp. 509-522, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] B. Anishfathima et al., "The Impact of Mental Health due to Covid 19–A Mental Health Detector Using Machine Learning," *Second International Conference on Artificial Intelligence and Smart Energy*, pp. 248-252, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Sarah Graham et al., "Artificial Intelligence for Mental Health and Mental Illnesses: An Overview," *Current Psychiatry Reports*, vol. 21, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[12] Bharti Thakur, Nagesh Kumar, and Gaurav Gupta, "Machine Learning Techniques with ANOVA for the Prediction of Breast Cancer," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 87, pp. 232-245, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Jing Sui et al., "Neuroimaging-Based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises," *Biological Psychiatry*, vol. 88, no. 11, pp. 818-828, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Simon D'Alfonso, "AI in Mental Health," *Current Opinion in Psychology*, vol. 36, pp. 112-117, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Tine Kolenik, *Methods in Digital Mental Health: Smartphone-Based Assessment and Intervention for Stress, Anxiety, and Depression*, Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector, pp. 105-128, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Bens Pardamean et al., "Quantified Self-Using Consumer Wearable Device: Predicting Physical and Mental Health," *Healthcare Informatics Research*, vol. 26, no. 2, pp. 83-92, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Adrian B.R. Shatte, Delyse M. Hutchinson, and Samantha J. Teague, "Machine Learning in Mental Health: A Scoping Review of Methods and Applications," *Psychological Medicine*, vol. 49, no. 9, pp. 1426-1448, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ensar Arif Sağbaş, Serdar Korukoglu, and Serkan Balli, "Stress Detection Via Keyboard Typing Behaviors by Using Smartphone Sensors and Machine Learning Techniques," *Journal of Medical Systems*, vol. 44, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Érico de Moura Silveira Júnior et al., "Decoding Rumination: A Machine Learning Approach to a Transdiagnostic Sample of Outpatients with Anxiety, Mood and Psychotic Disorders," *Journal of Psychiatric Research*, vol. 121, pp. 207-213, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Purude Vaishali Narayanrao, and P. Lalitha Surya Kumari, "Analysis of Machine Learning Algorithms for Predicting Depression," *2020 International Conference on Computer Science, Engineering and Applications*, Gunupur, India, pp. 1-4, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Shankar Shambhu et al., "Binary Classification of Covid-19 CT Images Using CNN: Covid Diagnosis Using CT," *International Journal of E-Health and Medical Communications*, vol. 13, no. 2, pp. 1-3, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Jiang Qiao, "A Systematic Review of Machine Learning Approaches for Mental Disorder Prediction on Social Media," *2020 International Conference on Computing and Data Science*, Stanford, CA, USA, pp. 433-438, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Ashley E. Tate et al., "Predicting Mental Health Problems in Adolescence using Machine Learning Techniques," *PloS One*, vol. 15, no. 4, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Ziyi Zhang, "Early Warning Model of Adolescent Mental Health Based on Big Data and Machine Learning," *Soft Computing*, vol. 28, no. 1, pp. 811-828, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Tarun Jain et al., "Machine Learning Techniques for Prediction of Mental Health," *2021 Third International Conference on Inventive Research in Computing Applications*, Coimbatore, India, pp. 1606-1613, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[26] Elizabeth Stevens et al., "Identification and Analysis of Behavioral Phenotypes in Autism Spectrum Disorder via Unsupervised Machine Learning," *International Journal of Medical Informatics*, vol. 129, pp. 29-36, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[27] Arkaprabha Sau, and Ishita Bhakta, "Screening of Anxiety and Depression among the Seafarers Using Machine Learning Technology," *Informatics in Medicine Unlocked*, vol. 16, pp. 1-11, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[28] Hugo G. Schnack, "Improving Individual Predictions: Machine Learning Approaches for Detecting and Attacking Heterogeneity in Schizophrenia (And Other Psychiatric Diseases)," *Schizophrenia Research*, vol. 214, pp. 34-42, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Nan Zhao et al., "See Your Mental State from Your Walk: Recognizing Anxiety and Depression through Kinect-Recorded Gait Data," *PLoS One*, vol. 14, no. 5, pp. 1-13, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[30] Betul Ay et al., "Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals," *Journal of Medical Systems*, vol. 43, no. 7, pp. 1-2, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[31] Youngjune Bhak et al., "Depression and Suicide Risk Prediction Models Using Blood-Derived Multi-Omics Data," *Translational Psychiatry*, vol. 9, no. 1, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[32] Lana G. Tennenhouse et al., "Machine-Learning Models for Depression and Anxiety in Individuals with Immune-Mediated Inflammatory Disease," *Journal of Psychosomatic Research*, vol. 134, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[33] Marialaura Di Tella et al., "Mental Health of Healthcare Workers during the COVID-19 Pandemic in Italy," *Journal of Evaluation in Clinical Practice*, vol. 26, no. 6, pp. 1583-1587, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[34] Nematjon Narziev et al., "STDD: Short-Term Depression Detection with Passive Sensing," *Sensors*, vol. 20, no. 5, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Kyoung-Sae Na, Seo-Eun Cho, and Seong-Jin Cho, "Machine Learning-Based Discrimination of Panic Disorder from Other Anxiety Disorders," *Journal of Affective Disorders*, vol. 278, pp. 1-4, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[36] Evandro Andrade et al., "A Protocol for the Diagnosis of Autism Spectrum Disorder Structured in Machine Learning and Verbal Decision Analysis," *Computational and Mathematical Methods in Medicine*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[37] Rahee Walambe et al., "Employing Multimodal Machine Learning for Stress Detection," *Journal of Healthcare Engineering*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[38] Nicholas C. Jacobson et al., "Deep Learning Paired with Wearable Passive Sensing Data Predicts Deterioration in Anxiety Disorder Symptoms Across 17–18 Years," *Journal of Affective Disorders*, vol. 282, pp. 104-111, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[39] Lulu Zhu et al., "The Machine Learning Algorithm for the Diagnosis of Schizophrenia on the Basis of Gene Expression in Peripheral Blood," *Neuroscience Letters*, vol. 745, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[40] Rutvik V. Shah et al., "Personalized Machine Learning of Depressed Mood Using Wearables," *Translational Psychiatry*, vol. 11, no. 1, pp. 1-8, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[41] Yun Han, "Application of Computer Digital Technology and Group Psychological Counseling in Higher Vocational Education Mental Health Education Curriculum under Complex Network Environment," *Mathematical Problems in Engineering*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[42] Muhammad Kashif Hanif et al., "Employing Machine Learning-Based Predictive Analytical Approaches to Classify Autism Spectrum Disorder Types," *Complexity*, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[43] Shumaila Aleem et al., "Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions," *Electronics*, vol. 11, no. 7, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[44] Yeong Hwan Ryu et al., "Prediction of Poststroke Depression Based on the Outcomes of Machine Learning Algorithms," *Journal of Clinical Medicine*, vol. 11, no. 8, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[45] Nirushi Kuhathasan et al., "Predictors of Perceived Symptom Change with Acute Cannabis Use for Mental Health Conditions in a Naturalistic Sample: A Machine Learning Approach," *Comprehensive Psychiatry*, vol. 122, pp. 1-8, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[46] Alberto González-Pérez et al., "AwarNS: A Framework for Developing Context-Aware Reactive Mobile Applications for Health and Mental Health," *Journal of Biomedical Informatics*, vol. 141, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[47] Shaurya Bhatnagar, Jyoti Agarwal, and Ojasvi Rajeev Sharma, "Detection and Classification of Anxiety in University Students through the Application of Machine Learning," *Procedia Computer Science*, vol. 218, pp. 1542-1550, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[48] Artur Shvetcov et al., "Machine Learning Identifies a COVID-19-Specific Phenotype in University Students Using a Mental Health App," *Internet Interventions*, vol. 34, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[49] Ilef Ben Slima et al., "COVID-19 Pandemic's Effect on the Mental Health among the Tunisian General Population: Associated Factors Mining via Machine Learning," *Scientific African*, vol. 21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[50] Nithya Jayakumar, and N. Reshma, "Modeling Mental Health: Advances in Predictive Science towards Proactive Health Care," *2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship*, Bangalore, India, pp. 202-206, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[51] Meenu Gupta et al., "MHA: Mobile Application-Based Analysis of Mental Disorders," *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing*, Bhubaneswar, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[52] Lakshmi Prasanna, and Shashi Mehrotra, "Comparative Analysis of Machine Learning Algorithms on Mental Health Dataset," *International Conference on Information and Communication Technology for Intelligent Systems*, vol. 719, pp. 599-606, 2023. [CrossRef] [Google Scholar] [Publisher Link]