Classification of Psychological Disorders by Feature Ranking and Fusion using Gradient Boosting

Classification of Psychological Disorders

Saba Tahseen[1], Ajit Danti[2]
Ph.D. Scholar[1], Professor[2]
Department of Computer Science and Engineering, Christ University,
Bangalore, India[2]

***Abstract*—Negative emotional regulation is a defining element of psychological disorders. Our goal was to create a machine-learning model to classify psychological disorders based on negative emotions. EEG brainwave dataset displaying positive, negative, and neutral emotions. However, negative emotions are responsible for psychological health. In this paper, research focused solely on negative emotional state characteristics for which the divide-and-conquer approach has been applied to the feature extraction process. Features are grouped into four equal subsets and feature selection has been done for each subset by feature ranking approach based on their feature importance determined by the Random Forest-Recursive Feature Elimination with Cross-validation (RF-RFECV) method. After feature ranking, the fusion of the feature subset is employed to obtain a new potential dataset. 10-fold cross-validation is performed with a grid search created using a set of predetermined model parameters that are important to achieving the greatest possible accuracy. Experimental results demonstrated that the proposed model has achieved 97.71% accuracy in predicting psychological disorders.**

Keywords—Electroencephalograph (EEG); psychological disorders; negative state emotions; gridSearchCV; gradient boosting classifier

I. Introduction

A clinically severe issue with the capacity of an individual to think straight or rule out their emotions or behavior characterizes a psychological disorder. It is typically accompanied by anxiety or impairment in critical areas of functioning. There are many different sorts of mental diseases. Psychological problems are another term for mental problems. The latter is a larger phrase that includes mental disorders, psychological disabilities, and (additional) states of mind that cause severe distress, functional impairment, or the risk of self-harm. This preliminary report is about psychiatric illnesses. The 11th Revised Version of the Disease Classification System (ICD-11) estimates that 960 million people, or one out of every eight people, will live on the planet in 2019 and suffer from a mental disorder, with anxiety and depression being the most common. Depression disorders harmed 280 million people in 2019, including 23 million children and adolescents. Anxiety disorders afflict 301 million people, with 58 million of them being children and adolescents. Schizophrenia affects roughly 24 billion people worldwide or one in every 300. In 2019, 40 million people were impacted by bipolar disorder [1]. As a result of the COVID-19 disease outbreak, the number of individuals living with anxiety and depression increased significantly in 2020. Estimations reveal a 26% spike in anxiety disorders and a 28% spike in serious depression disorders in only one year [2]. Many patients with psychiatric problems don't have access to suitable treatment or preventative options. Many people face stereotypes, marginalization, and infractions of their human rights. EEG is inexpensive and useful for assessing resting-state activity in the brain in natural environments, allowing for large amounts of data to be collected quickly. Furthermore, as the acquisition of technology improves and calculations improve, EEG is gathering steam as a foundational technology for brain-computer interfaces [3]. Relatively low cost, ease of use, and adaptable territory setup Echocardiography has been widely used in uncovering the aetiologies of various mental illnesses (e.g., depression [4], Alzheimer's [5], epilepsy [6], schizophrenia [7], autism spectrum disorder [8], anxiety [9], and so on). Typical brain activity and emotional swings are always present in depression, a mental disorder with clinical signs like severe depression and impaired thinking. EEG can therefore identify these aberrant events as a technique for monitoring brain activity.

Metadata and computational scientific research advances are being made in transforming mental healthcare. The scope of evidence that can be measured in terms of neural mechanisms and objective markers has expanded. Furthermore, the use of machine learning, also known as artificial intelligence, has grown. Machine learning can evaluate the effectiveness of forecasts on previously unseen (test) data that wasn't used prior, fitting the model for training data, utilizing out-of-sample forecasts, and providing individualized, and possibly high levels of therapeutic applications [10]. Machine learning is expected to aid or even replace physician judgments such as diagnostic testing, prognosis, and patient experience [11]. Individually, Cross-validation learning and psychological disorders had previously been investigated. Collaborations between these two fields have recently been combined, and machine learning has been used by researchers to identify psychological disorders using EEG data. Negative emotions like anxiety and depression encourage a series of psychological and physiological changes that put one's long-term health at risk. So for this, we sought to develop a new classification model for considering negative emotional features from EEG brainwave emotion features (positive, negative, or neutral) in patients with severe psychological disorders. The contributions to this study are as follows:

- It is difficult but also necessary to accurately extract EEG features since the success of the classifying depends on this extraction. We extract negative state features from EEG brainwave data from emotional features (positive, negative, neutral) for psychological disorders.
- For negative state features we have applied the “Divide and Conquer” approach to four equal parts for finding more optimal results from each feature subset.
- We present the K-Means cluster technique to obtain labeled EEG signal features for each feature subset.
- Ensemble methods (RF-RFECV) techniques for feature selection have been employed to build and determine the highest-ranking features from each feature subset and merge them into an appropriate feature set for classifying psychological disorders.
- To obtain an optimal hyperparameter of the Gradient boosting, the tree classifier GridSearchCV has been used.
- Performance of hyperparameters of gradient boosting trees such as the number of trees, tree depth, several learning rates, and the number of subsamples are utilized.

This article is broken up into five sections. Beginning with the introduction. Most related works were described in Section II. Additionally, Section III provides materials and methods, and resources. In Section IV, Experimental Results are discussed. Finally, the conclusions are given in Section V.

II. Related Work

Most of the recent imaging research (i.e., employing magnetic resonance) has relied on supervised machine learning. To differentiate patients from health controls (HCs). Studies have primarily concentrated on Alzheimer's disease, schizophrenia, and depression, but have rapidly been expanded to include other diagnosing topics [12]. Literary work suggests that certain machine learning in EEG might predict significant psychological disorders and serve as an unbiased index of psychological disorders, according to the findings. The comprehension of the support vector machine, elastic net, and random forest machine learning methods was highlighted. The elastic net model with intelligence quotient adjustment performed the best with different bands of the EEG dataset [13]. According to various cross-validation experiments, the probability of diagnosis achieved employing the provided approach in a training dataset among 207 entities, such as 64 patients with severe depression, 40 patients with severe schizophrenia, 12 bipolar depression patients, and 91 healthy or normal subjects, is over 85% with some further advances [14]. The proposed methodology has the possibility of being a beneficial adjunct treatment tool for healthcare practitioners. EEG acquisition and pre-processing were adequate, discovered that numerous of them lacked thorough clinical characteristic identification Moreover, numerous studies employed parameters of the model or testing techniques that were flawed. Indeed, it's suggested that future researchers of psychological disorders using Deep Learning enhance the accuracy of clinical evidence and use cutting-edge model choice and test procedures to improve research standards and progress toward diagnostic value[15]. Using techniques for extracting multivariate EEG features and algorithms, the goal of this research is to provide an analytical framework for robotic GAD identification. Resting-state EEG data were collected from 45 GAD patients and 36 health control (HC) with 97.83 percent accuracy aberrated features helped classification performance [16]. This study classifies the EEGs of 43 VHS and 53 MDD patients using data mining techniques. This includes cleaning and normalizing the data beforehand, using Linear Discriminant Analysis (LDA) to map information into a brand-new feature space, and Genetic Algorithm (GA) to determine the much more features [17]. Methods are used in mental health to predict the likelihood of mental diseases and, thus, to execute prospective treatment outcomes. This review article lists many machine-learning techniques for identifying and diagnosing depression. The three classes of ML-based depression detection methods include deep learning, classification, and ensemble. The authors describe a generic paradigm for diagnosing depression that includes raw data, which was before, ML training; exposure, detection classifications, and performance assessment are all part of the process [18].

In [19], six channels (FT7, FT8, T7, T8, TP7, and TP8) are used to extract features from the frontal area of the brain. The following band powers—delta, theta, alpha, beta, gamma1, and gamma2—along with their related asymmetry and paired asymmetry—are employed as linear characteristics. The classifiers used are bagging and three different kernel functions of the Support Vector Machine (SVM) (polynomial, gaussian, and sigmoidal). Relief is the creation of predictive models utilizing the Decision Tree (DT) technique to find rules and relevant features; the feature selection method is applied. The feature selection methods SVM (Gaussian Kernel Function) and Relief were employed to achieve the best classification accuracy of 96.02% and 79.19% for the identification and severity rating of depression, respectively [19]. A widespread EEG system with three electrodes in the prefrontal lobe was used to record all electroencephalogram (EEG) signals from subjects during the sound stimulus and resting state at just the Fp1, Fp2, and Fpz electrolytic positions. A maximum of 270 linear and nonlinear features were recovered after denoising with the Finite Moment Generating Filter, which incorporates the Kalman derivation method, Discrete Wavelet Transformation, and an Adaptive Predictor Filter. The dimensionality of the feature space was then decreased using the minimal-redundancy-maximum-relevance feature extraction strategy. The depressed individuals were separated from the healthy controls using four classification techniques (SVM, KNN, Classification Trees, and ANN) [20].

The spatial frequency data of a few chosen EEG channels is used to extract features. Theta and beta bands were chosen as EEG frequency bands for this investigation using a technique called "choosing a frequency range". The characteristics of the chosen frequency ranges of EEG are subject to feature selection. As for limitations of existing methods, it may be that they struggle with handling large datasets, lack effective feature selection algorithms, or have difficulty generalizing to new data. The proposed method may be designed to address these limitations and provide a more effective solution to the problem at hand. Finally, a variety of machine learning methods were used to categorize the chosen subset of characteristics from the statistically relevant EEG networks' proper frequencies. A random forest classification model with either nine or ten attributes is used. It is possible to classify anxiety on two or four levels with an accuracy of 94.90% and 92.74%, respectively [21]. The carefully chosen main studies were used in comprehensive mapping research. The objectives were to present a comprehensive picture of the most important research areas in the diagnosis and forecast of mental diseases by combining EEG with DL. [22]. SVM was used to categorize the stress levels using the labeled data from the k-means clustering method. Using only the beta-band ultimate power feature in the right (Fp2) prefrontal region, the achievement of the classification model was endorsed using the 10-fold cross-validation method. This result confirmed the excellent efficiency of 98% accuracy because of the significant adjustments in beta activity all through pre- and post-stimuli latent patterns using localized and reduced features and evaluating model accuracy and false positive findings on EEG data from people with MDD and HV. The motivation to write this research to solve the following issue came from considering the above kinds of literature and using abilities in this field [23]. Most of the research for psychological disorders has been done with resting states, eye open and closed states of EEG dataset.

- To find the best solution for a problem, it is important to consider all possible options and evaluate them based on relevant criteria. This involves a systematic and analytical approach to identifying the optimal solution.
- How can unsupervised learning data be effectively handled or processed, given the lack of labeled information, to improve the quality of the resulting models or insights?
- How to solve the overfitting issue in machine learning refers to identifying and implementing techniques that can prevent a model from becoming too complex and fitting too closely to the training data, which can result in poor performance on new, unseen data.
- Which parameter is used to develop the best model?
- Lack of standardization: Many studies in this field use different EEG acquisition protocols, pre-processing methods, feature extraction techniques, and machine learning algorithms, which make it challenging to compare results and generalize findings.
- Limited diagnostic focus: While some studies have investigated a range of mental health conditions, others have primarily focused on a few specific disorders, such as depression, anxiety, schizophrenia, or Alzheimer's disease. Further research is needed to evaluate the usefulness of machine learning with EEG for diagnosing other psychological disorders.

III. Proposed Methodology

Psychological disorders are complex and can manifest in many ways, making diagnosis challenging for healthcare professionals. Machine learning algorithms have the potential to improve the accuracy and efficiency of diagnosis by identifying patterns in large datasets that may be difficult for humans to detect. The use of feature ranking and fusion in combination with gradient boosting is a promising approach for improving the performance of machine learning algorithms in the context of psychological disorder classification. Feature ranking techniques can help identify the most relevant features for classification, while feature fusion can combine different sources of information to improve the overall accuracy of the algorithm. Therefore, the motivation behind this research is to explore how these techniques can be applied to improve the accuracy and efficiency of psychological disorder classification using machine learning. The goal is to develop a more effective diagnostic tool that can assist healthcare professionals in accurately identifying and treating psychological disorders.

In this work, we have comprehensively analyzed the positive, negative, and neutral states of the publicly accessible EEG brainwave Dataset. In this research, for the recognition of psychological disorders, we extract the negative state of this dataset. The proposed method for the classification of psychological disorders using feature ranking and fusion with gradient boosting is an appropriate solution for the problem due to its ability to handle large datasets, and improve generalization, and flexibility. The method uses a feature selection algorithm to select the most important features for classification, which can enhance the accuracy and efficiency of the model. The proposed method also employs techniques to prevent overfitting, such as grid search for hyperparameter tuning, which can enable the model to generalize better to new and unseen data This section outlines the configuration for the classification procedure, as well as the methodology used to conduct the research, and discusses a potential strategy for minimizing features in an EEG analysis by establishing the RF-RFECV method. The proposed architecture diagram shows the steps of this research in Fig. 1.

A. *Data Preprocessing*

Divide and Conquer strategy has been applied to the negative state features of the EEG brainwave dataset to find an optimal solution to a problem. Four distinct feature sets were compared to see how changing the measurements would affect the results and which feature set performed better [25].

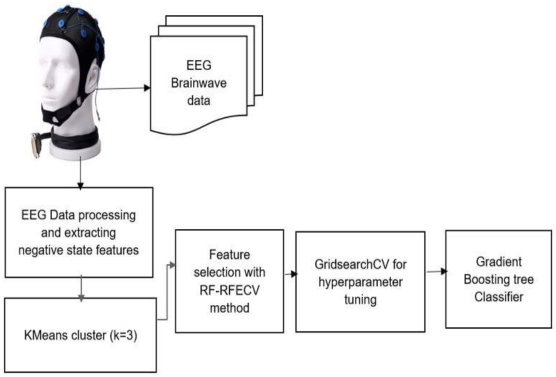

Fig. 1. Proposed architecture of psychological disorder classification

Steps of divide and conquer approach:

1) Divide: Divide the dataset D into 4 Feature subsets (FS)

$$D = \frac{(f0, f1, f2, \ldots, fn)}{4} \quad (1)$$

$$= FS_1, FS_2, FS_3, FS_4$$

2) Conquer: Applied feature selection technique for each FS.

3) Combine: Selected features from each FS have been fusion into the final features set (FFS).

Further, the K-Means clustering technique [26] is applied to label unlabeled datasets to group the features into 3 similarity clusters.

The steps of K-Means are as follows:

- Initially, we generate random k points, referred to as means or cluster centroids.
- Every feature is categorized according to the nearest mean, and the precise location of that mean, which represents the average values of the features categorized in that cluster so far.
- Clusters are the result of repeating the process for a predetermined number of iterations.

The algorithm's final goal is to minimize the squared error function, which is represented by:

$$\boldsymbol{J(v)} = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\| \boldsymbol{x_i} - \boldsymbol{v_j} \|)^2 \quad (2)$$

Where,

"$\| \boldsymbol{x_i} - \boldsymbol{v_j} \|$" is a measure of how far the n data points are from each cluster's center [27]. In this research, the three clusters are chosen as 0, 1, and 2 for labeling EEG negative data as shown in Fig. 2

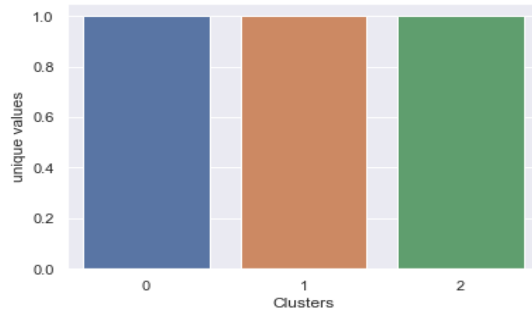

Fig. 2. Negative state features with 3 clusters i.e., 0, 1, 2

B. Feature Selection

Selecting features from thousands of features is the most challenging research problem. RF-RFECV is used for feature selection. The algorithm is trained by Random Forest (RF) to generate the importance of features. The importance of each feature is calculated by equation 3.

$$\boldsymbol{fi_i} = \frac{\sum_{j:node\ j\ split\ on\ feature\ i} ni_j}{\sum_{k \in all\ modes} ni_k} \quad (3)$$

RF classifier can be trained to produce feature importance values [28] that represent the relative importance of each feature. Following that, features are ranked in order of importance value. The component with the lowest importance value is eliminated. The classifier is then retrained using the remaining features until it runs out of features to train with. Finally, the complete ranking of the features can be obtained using the feature-importance-based RFE method i.e., (RF-RFECV). It has been demonstrated that, an embedded feature selection method performs well and makes up for the drawbacks of the filter and wrapper methods. The following pseudocode represents the proposed feature selection algorithm.

Algorithm 1: Feature Selection with RF-RFECV

Input: Feature Subset FS: {f0, f1, f2,….,fn}
Output: Rank features according to smallest feature importance value, R
Step 1: set R= {}
Step 2: Repeat steps 3-9 until *FS* is not empty
Step 3: Train the RF using *FS*.
Step 4: Compute the importance of the feature with an equation (3).
Step 5: Determine the ranking method, Rank= fi^2
Step 6: Rank the features in sorted orders.

$$N_{rank} = sort(Rank)$$

Step 7: modification of the feature rank list

$$modify\ R = R + FS(N_{rank})$$

Step 8: Delete the features with the lowest rank

$$modify\ FS = FS - FS(N_{rank})$$

Step 9: Fusion the highest-rank features of FS_1, FS_2, FS_3, FS_4 into the final dataset.
Return final dataset Fusion feature subset (FFS)

C. Machine Learning Model

Gradient-boosting tree classifier is applied to predict the test results of psychological disorders. Gradient boosting is a sequenced method based on the ensemble principle. It integrates a group of weak learners and generates higher prediction accuracy. The model results are weighted based on the results of the initial instant *t-1* at any instant *t*. The working procedure gradient boosting is as follows.

Step 1**:** Create a fundamental model to predict the dataset. Take the total of the cluster column and presume that represents the expected value. The simple mathematical calculation behind these first steps.

$$F_0(x) = arg\min_y \sum_{i=1}^{n} L(yi, y) \quad (4)$$

Step 2***:*** Determine the Residuals

$$e = (Y - \hat{Y}) \quad (5)$$

Step 3***:*** Determine the decision tree's leaf output values

$$Y_m = arg\min_Y \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + Y h_m(x_i) \quad (6)$$

Step 4: Update the prediction

$$F_m(x) = F_{m-1(x)} + V_m h_m(x) \quad (7)$$

D. Hyperparameter Optimization by GridSearchCV

A method for identifying the best hyperparameters in a grid out of a set of parameters is called GridSearchCV. Gradient boosting trees could be a challenge to set up as an algorithm [29]. Grid search common ranges are produced because of the gradient boosting technique's key hyperparameters, which serve as the starting point for one's work. This might be done by assigning a dictionary for links, the names of the model hyperparameters, to the values to search for in the GridSearchCV [30]. The following is the procedure for finding the best hyperparameters with GridSeachCV.

Algorithm 2: GridSearchCV for hyperparameter optimization

Input: Dataset *FFS* = {f0, f1, f2, …. , fn}
Output: Best parameter with the highest accuracy
Step 1: Create the gradient boosting tree model and parameter
Step 2: Create a dictionary using the model's parameters.

- Develop an *estimator* of gradient boosting tree classifier
- Develop a *Param_grid* with key hyperparameters
  The efficiency of model evaluation metrics.
  Score=f (Key Parameter)
- Develop the *CV* for iterations

Step 3: Repeat the process step 2, going through each possible set of the grid's values one at a time.
Step 4: Fit the data set in the object function
Step 5: Run the objective functions multiple times per each possible pair of hyperparameter values.
Return the most accurate hyper-parameters available with the highest accuracy.

The GridSearchCV key hyperparameter for gradient-boosting trees is shown below in Table I.

TABLE I. KEY HYPERPARAMETERS FOR GRADIENT BOOSTING TREE WITH GRIDSEARCHCV

Model	**n_estimators**	**Learning rate**	**subsample**	**Max depth**
Gradient Boos tree	[10,50, 500,1000]	[0.0001, 0.001,0.01, 1.0]	[0.5, 0.7,1.0]	[0.5, 0.7, 1

E. Performance Evaluations

Before the prediction model is constructed, a model must be assessed using several evaluation criteria [31]. To date, we have evaluated our prediction models using means and accuracy scores. However, the accuracy score and mean alone aren't always sufficient to assess a model adequately because it doesn't specify whether a class (positive or our models incorrectly forecast a negative) in the event of a poor accuracy rating this is clarified by precision score.

$$\text{Mean } \bar{x} = \left(\frac{(\sum x)}{n}\right) \quad (8)$$

$$\text{Std deviation SD} = \sqrt{\sum \left(\frac{x - \bar{x}}{n-1}\right)} \quad (9)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (12)$$

$$\text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

$$\text{Confidence Score} = CI = \bar{x} \pm z\frac{s}{\sqrt{n}} \quad (14)$$

IV. EXPERIMENTAL RESULTS

The study is significant in that it demonstrates the usefulness of an affordable and straightforward approach to diagnosing several psychological disorders using EEG. The proposed method employs a feature selection algorithm to select the most important features for classification, which can improve the accuracy and efficiency of the model. Additionally, the gradient boosting algorithm used in the method can handle large and complex datasets, making it well-suited for problems that involve many features and samples. The EEG brainwave dataset has been used, which has three states of emotion: positive, negative, and neutral. From these, three emotional states negative emotional states features were extracted for the recognition of psychological disorders for this work. 708 rows and 2548 columns of unlabeled data were used in the negative state. Applied the "Divide and Conquer" approach into 4 equal parts FS_1, FS_2, FS_3, FS_4 to find more accurate results from each feature subset. The ensemble approach has been used for feature selection, followed by random forest feature importance with recursive feature elimination with cross-validation technique. Based on this procedure, selected the rank features for each subset, and fusion these new features into a new feature subset or new dataset.

An applied new dataset for the classification process used a gradient-boosting tree with GridSearchCV hyperparameter tuning. The best parameter combination is kept after the GridSearchCV evaluates all potential parameter value combinations. This research primarily focuses on four parameters. GridSearchCV uses *max_depth* for denting the number of depths of a tree, *n_estimators* for several sequential trees *learning_rate* is used to determine how each tree will affect the predictions and, the *sub_sample* for several analyses that will be chosen for every tree for the strongest impact on prediction accuracy. Several performance metrics are used for choosing the best parameter such as mean score, standard deviation score as well as accuracy.

The configuration that showed the best performance, achieving a means score of approximately 94.6 %, had a *learning_rate* of 0.1, a *max_depth* of 9 levels, 1000 *n_estimators*, and a *sub_sample* of 50% has presented in Table II.

Table II also presents the performance of parameters with different combinations of parameter values like *learning_rate* of 1.0, 1000 of *n_estimators, Subsample* of 40%, and *max_depth* of 7 achieved means score is 92%.

learning_rate of 0.2, 500 of *n_estimators, Subsample* of 90% and *max_depth* of 6 achieved means score is 91%.

learning_rate of 1.0, 1000 of *n_estimators, Subsample* of 90% and *max_depth* of 7 achieved means score is 89%.

Accuracy of each parameter combination 91%, 89%, 92%, 92%, 96.71%.

TABLE II. BEST PARAMETERS FROM THE GRIDSEARCHCV METHOD

Sl.no	Best parameter	Means Score	Std score	Accuracy
1	n_estimators=1000, Subsample=0.4, learning_rate=1.0, max_depth=7	92.00%	0.028	91.00%
2	n_estimators=500, Subsample=0.9, learning_rate=0.2, max_depth=6	91.00%	0.26	89.00%
3	n_estimators=1000, Subsample=0.9, learning_rate=1.0, max_depth=7	92.00%	0.028	92.00%
4	n_estimators=1000, Subsample=0.9, learning_rate=1.0, max_depth=7	89.00%	0.24	92.00%
5	n_estimators=1000. subsample=0.5, learning_rate=0.1, max_depth=9	94.93%	0.27	96.71%

Additionally, the accuracy of each cluster is 0, 1, 2, and has achieved 95.5%, 96.8%, and 100% accuracy. The classification results in representation in Table III for clusters of 0, 1 a, and 2, Precision, Recall, and f1-Score achieved the highest accuracy in cluster 2 compared to other clusters for the proposed gradient boosting Classifier with GridSearchCV for a new dataset.

TABLE III. CLASSIFICATION REPORT OF EACH CLUSTER

Clusters	precision	recall	F1 score
0	94%	96%	95%
1	98%	97%	97%
2	100%	100%	100%

Fig. 3 represents the accuracy analysis of four algorithms with the x-axis being proposed and existing algorithms and the y-axis being proposed as the accuracy value. The accuracy of the proposed classifier without ranked features is 94% and with ranked features is 96.71%.

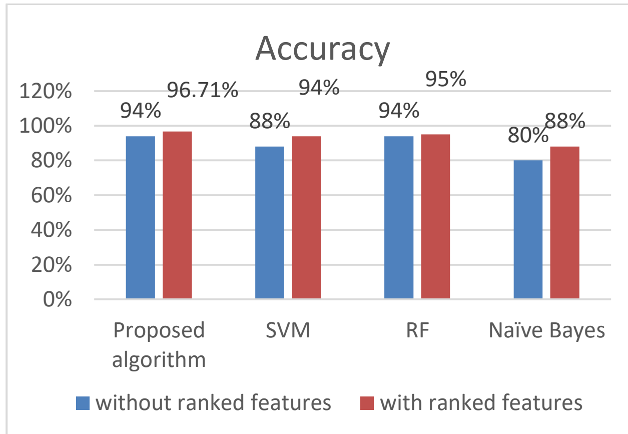

Fig. 3. Comparative analysis with another existing algorithm

The accuracy of SVM without ranked features is 92% and with ranked features is 94%. The accuracy of RF with ranked features is 95% and without ranked features is 94%. The accuracy of Naïve Bayes classifier achieved 80% accuracy without ranked features and 88%accuracy with ranked features. It can be concluded that the proposed algorithm with ranked features (new dataset) produces effective results as compared to existing algorithms.

Table IV displays the Confidence score of various psychological disorders (stress, depression, bipolar disorder, anxiety, autism, schizophrenia, mood disorder, and personality disorder) in an EEG negative emotional state with clusters 0, 1, and 2, which is represented. Schizophrenia disorders achieved an 85% confidence score as compared to other disorders.

TABLE IV. THE CONFIDENCE SCORE OF PSYCHOLOGICAL DISORDER

Psychological disorder	Accuracy	Negative Emotional State (0,1,2)
Depression	65.00	Cluster-0
Anxiety	35.00	Cluster -0
Stress	77.00	Cluster -0
Bipolar Disorder	50.50	Cluster -1
Personality	58.00	Cluster -1
Schizophrenia	85.00	Cluster -2
Autism	50.00	Cluster-2

Table V compares the state-of-the-art methods of other machine learning methods with the same dataset our proposed algorithm achieved the highest accuracy.

TABLE V. COMPARATIVE ANALYSIS OF PROPOSED WORK WITH OTHER WORK

Study	Classifiers	Datasets	Accuracy
The proposed method	Gradient Boosting Tree with GridSearchCV	EEG brainwave dataset	96.71%
[24]	Adaptive Boosted LSTM and DevoMLP	EEG brainwave dataset	85%
[31]	RNN	EEG brainwave dataset	95%
[32]	XGBoost	EEG brainwave Dataset	95%

V. CONCLUSION

The study has demonstrated the usefulness of an affordable and straightforward approach to the brain utilizing EEG for the diagnosis of several psychological disorders, including stress, bipolar disorder, autism, mood, personality, anxiety, and depression. One significant advantage of the proposed method is its ability to handle large and complex datasets using gradient boosting, which is a powerful algorithm for handling such data. In addition, the techniques used to prevent overfitting, such as grid search for hyperparameter tuning, can help the model generalize better to new and unseen data. This is an important consideration when dealing with medical data where the model's ability to generalize to new data is crucial for accurate diagnosis. This study presents a unique method of feature selection with different feature subsets and makes a new dataset with 1300 features with the RF-RFECV algorithm with labeling using the K-Means Cluster technique. To address overfitting and optimize the parameters of the gradient boosting classifier, we employed the GridSearchCV algorithm to find the optimal hyperparameters for predicting psychological disorders from the EEG brainwave dataset, which has achieved 96.71% accuracy in classifying psychological disorders using negative states of emotion. The future will be, to calculate the severity of the psychological disorder and develop a web application for clinical diagnostics.

ACKNOWLEDGMENT

None of the funding sources for this study were public, private, or nonprofit. We sincerely thank the professors of Christ University's department of computer science and engineering, Bangalore, India.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx), (https://vizhub.healthdata.org/gbd-results/, accessed 14 May 2022).

[2] Mental Health and COVID-19: Early evidence of the pandemic's impact. Geneva: World Health Organization; 2022.

[3] G. Li, B. Li, Y. Jiang, W. Jiao, H. Lan, and C. Zhu, "A new method for automatically modeling brain functional networks," *Biomedical Signal Processing and Control*, vol. 45, pp. 70–79, Aug. 2018, doi: 10.1016/j.bspc.2018.05.024.

[4] A. Khalid, B. S. Kim, M. K. Chung, J. C. Ye, and D. Jeon, "Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology," *NeuroImage*, vol. 101, pp. 351–363, 2014.

[5] M. Cejnek, O. Vysata, M. Valis, and I. Bukovsky, "Novelty detection-based approach for alzheimer's disease and mild cognitive impairment diagnosis from EEG," *Medical & Biological Engineering & Computing*, vol. 59, no. 11-12, pp. 2287–2296, 2021.

[6] S. C. Ponten, F. Bartolomei, and C. J. Stam, "Small-world networks and epilepsy: Graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures," *Clinical Neurophysiology*, vol. 118, no. 4, pp. 918–927, 2007.

[7] S. Micheloyannis, E. Pachou, C. J. Stam, M. Breakspear, P. Bitsios, M. Vourkas, S. Erimaki, and M. Zervakis, "Small-world networks and disturbed functional connectivity in schizophrenia," *Schizophrenia Research*, vol. 87, no. 1-3, pp. 60–66, 2006.

[8] S. Carlier, S. Van der Paelt, F. Ongenae, F. De Backere, and F. De Turck, "Empowering children with ASD and their parents: Design of a serious game for anxiety and stress reduction," *Sensors*, vol. 20, no. 4, p. 966, 2020.

[9] H. Byeon, "Exploring factors for predicting anxiety disorders of the elderly living alone in South Korea using Interpretable Machine Learning: A population-based study," *International Journal of Environmental Research and Public Health*, vol. 18, no. 14, p. 7625, 2021.

[10] Q. J. Huys, T. V. Maia, and M. J. Frank, "Computational psychiatry as a bridge from neuroscience to clinical applications," *Nature Neuroscience*, vol. 19, no. 3, pp. 404–413, 2016.

[11] G. E. Simon, "Big data from Health Records in Mental Health Care," *JAMA Psychiatry*, vol. 76, no. 4, p. 349, 2019.

[12] D. B. Dwyer, P. Falkai, and N. Koutsouleris, "Machine learning approaches for clinical psychology and psychiatry," *Annual Review of Clinical Psychology*, vol. 14, no. 1, pp. 91–118, 2018.

[13] S. M. Park, B. Jeong, D. Y. Oh, C.-H. Choi, H. Y. Jung, J.-Y. Lee, D. Lee, and J.-S. Choi, "Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach," *Frontiers in Psychiatry*, vol. 12, 2021.

[14] A. Khodayari-Rostamabad, J. P. Reilly, G. Hasey, H. de Bruin, and D. MacCrimmon, "Diagnosis of psychiatric disorders using EEG data and employing a statistical decision model," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010.

[15] M. de Bardeci, C. T. Ip, and S. Olbrich, "Deep learning applied to electroencephalogram data in mental disorders: A systematic review," *Biological Psychology*, vol. 162, p. 108117, 2021.

[16] Z. Shen, G. Li, J. Fang, H. Zhong, J. Wang, Y. Sun, and X. Shen, "Aberrated multidimensional EEG characteristics in patients with generalized anxiety disorder: A machine-learning based analysis framework," *Sensors*, vol. 22, no. 14, p. 5420, 2022.

[17] M. Mohammadi, F. Al-Azab, B. Raahemi, G. Richards, N. Jaworska, D. Smith, S. de la Salle, P. Blier, and V. Knott, "Data mining EEG signals in depression for their diagnostic value," *BMC Medical Informatics and Decision Making*, vol. 15, no. 1, 2015.

[18] S. Aleem, N. ul Huda, R. Amin, S. Khalid, S. S. Alshamrani, and A. Alshehri, "Machine learning algorithms for depression: Diagnosis, insights, and Research Directions," *Electronics*, vol. 11, no. 7, p. 1111, 2022.

[19] S. Mahato, N. Goyal, D. Ram, and S. Paul, "Detection of depression and scaling of severity using six channels EEG Data," *Journal of Medical Systems*, vol. 44, no. 7, 2020.

[20] H. Cai, J. Han, Y. Chen, X. Sha, Z. Wang, B. Hu, J. Yang, L. Feng, Z. Ding, Y. Chen, and J. Gutknecht, "A pervasive approach to EEG-based depression detection," *Complexity*, vol. 2018, pp. 1–13, 2018.

[21] F. Muhammad and S. Al-Ahmadi, "Human State Anxiety Classification framework using EEG signals in response to exposure therapy," *PLOS ONE*, vol. 17, no. 3, 2022.

[22] M. J. Rivera, M. A. Teruel, A. Maté, and J. Trujillo, "Diagnosis and prognosis of mental disorders by means of EEG and Deep Learning: A Systematic Mapping Study," *Artificial Intelligence Review*, vol. 55, no. 2, pp. 1209–1251, 2021.

[23] T. Y. Wen and S. A. Mohd Aris, “Hybrid approach of EEG stress level classification using K-means clustering and support vector machine,” *IEEE Access*, vol. 10, pp. 18370–18379, 2022.

[24] J. J. Bird, D. R. Faria, L. J. Manso, A. Ekárt, and C. D. Buckingham, “A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction,” *Complexity*, vol. 2019, pp. 1–14, 2019.

[25] S. E. Sánchez-Hernández, R. A. Salido-Ruiz, S. Torres-Ramos, and I. Román-Godínez, “Evaluation of feature selection methods for classification of epileptic seizure EEG signals,” *Sensors*, vol. 22, no. 8, p. 3066, 2022.

[26] A. Bablani, D. R. Edla, V. Kuppili, and D. Ramesh, “A multi stage EEG data classification using K-means and feed forward neural network,” *Clinical Epidemiology and Global Health*, vol. 8, no. 3, pp. 718–724, 2020.

[27] A. Bi, W. Ying, and L. Zhao, “Fast enhanced exemplar-based clustering for incomplete EEG signals,” *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–11, 2020.

[28] H. AlSagri and M. Ykhlef, “Quantifying feature importance for detecting depression using Random Forest,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, 2020.

[29] L. Hertel, P. Baldi, and D. L. Gillen, “Reproducible hyperparameter optimization,” *Journal of Computational and Graphical Statistics*, vol. 31, no. 1, pp. 84–99, 2021.

[30] Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, and M. Tu, “Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances,” *Journal of Petroleum Science and Engineering*, vol. 160, pp. 182–193, 2018.

[31] M. K. Chowdary, J. Anitha, and D. J. Hemanth, “Emotion Recognition from EEG Signals Using Recurrent Neural Networks,” *Electronics*, vol. 11, no. 15, p. 2387, Jul. 2022, doi: 10.3390/electronics11152387.

[32] S. Klibi, M. Mestiri and I. R. Farah, "Emotional behavior analysis based on EEG signal processing using Machine Learning: A case study," *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen, 2021, pp. 1-7, doi: 10.1109/ICOTEN52080.2021.9493537.