TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

EEG-BASED ANXIETY AND STRESS DETECTION USING PCA-RFE HYBRID FEATURE EXTRACTION AND XGFOREST CLASSIFICATION WITH HYPERPARAMETER OPTIMIZATION

[1]GEETHA. A, [2]DR.JACKULIN MAHARIBA A, [3]S. SELVI, [4]DR. T. KUMARESAN, [5]DR.S.ZULAIKHA BEEVI, [6]K. SARANYA

[1]ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
PRESIDENCY UNIVERSITY, BENGALURU
geetha.arjunan@presidencyuniversity.in

[2]ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTATIONAL INTELLIGENCE
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, KATTANKULATHUR- 603203
E-Mail : jackulia@srmist.edu.in

[3]DEPARTMENT OF MATHEMATICS, BANNARI AMMAN INSTITUTE OF TECHNOLOGY,
SATHYAMANGALAM-638401, ERODE.
e-mail id: selvimaths86@gmail.com

[4]PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
KGISL INSTITUTE OF TECHNOLOGY, COIMBATORE
speak.kumaresh@gmail.com

[5]PROFESSOR, DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
VEL TECH HIGH TECH DR.RANGARAJAN DR.SAKUNTHALA ENGINEERING COLLEGE
AVADI, CHENNAI-600062, TAMIL NADU, INDIA
zulaikhabeevi.s@velhightech.com

[6]ASSOCIATE PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, BANNARI AMMAN INSTITUTE OF TECHNOLOGY, SATHYAMANGALAM, ERODE, TAMILNADU, INDIA
Email: ksaranyacse@gmail.com

Abstract:

Modern research about mental health analyses Electroencephalography (EEG) data for detecting and categorizing anxiety and stress manifestations. The research presents an improved detection method which unites innovative features extraction methods with optimization approaches as well as advanced classification techniques to enhance accuracy. Our proposed method known as PCA-RFE Hybrid Feature Extraction implements Principal Component Analysis to reduce dimensions alongside Recursive Feature Elimination for selecting important EEG features thus achieving better model interpretability coupled with improved computational speed. The XGForest Classifier stands as our classification model because it unites XGBoost and Random Forest algorithms to maximize predictive accuracy at 93%. The best parameters are chosen from GridSearchCV to achieve maximum classification accuracy during hyperparameter optimization. The Hybrid Classifier obtains 1.00 accuracy along with precision 0.99, recall 0.99 and F1-score 0.97. The investigative findings indicate that the implemented method exhibits superior diagnostic performance regarding traditional machine learning algorithms because XGBoost reached 0.93 accuracy while Neural Network achieved a dismal 0.37 accuracy. Early detection of mental health problems becomes possible through the development of this study which works toward creating improved EEG-based diagnostic tools for monitoring purposes.

Keywords: EEG Data, Anxiety Detection, Stress Detection, PCA-RFE Hybrid, XGForest Classifier, Hyperparameter Tuning.

I. INTRODUCTION

Anxiety disorders are considered one of the most common mental health conditions because they affect millions of individuals across the world. Excessive fear together with worry and nervousness form these conditions that obstruct an individual's daily functioning. The Electroencephalogram (EEG) functions as an exterior brain monitoring tool for analyzing neural activities which helps doctors understand psychological conditions such as anxiety [1]. Bringing together brain wave tracks from test subjects displays both worried mental states together with normal mental states in the standard EEG anxiety disorder record. The examination of these data points shows abnormalities in brain electrical patterns specifically linked to anxiety cases for improved diagnosis and treatment. Researches have demonstrated that analyzing EEG signals enables the prediction of anxiety onset and its tracking pattern evolution in patients [2].

Standard techniques used for extracting features and classifying EEG signals to detect anxiety disorders face multiple operational limitations when applied for this purpose. EEG data possesses high dimensionality as an essential challenge since it combines extensive information about frequency bands (delta, theta, alpha, beta, and gamma waves) alongside data from multiple electrodes. [3]The manual features used by Wavelet and Fourier transforming techniques fail to detect the intricate non-linear patterns found in EEG signals. The classification process becomes more complex because of both redundant features and useless features leading to the need for advanced selection methods. The classification performance of Support Vector Machines (SVM) and Logistic Regression faces challenges due to the complexity of EEG data thereby potentially leading to overfitting or underfitting errors. The analysis methods demonstrate weak performance when identifying real-time brain activity patterns which are critical for mental health diagnosis because they do not consider temporal dependencies found in EEG signals [4,5].

Traditional EEG methods fail to address their main problems stemming from their limited ability to process extensive EEG data complexity as well as brain signal nonstationarity[6,7]. Expert-dependent transformations used in traditional feature extraction methods prove ineffective at extracting important features from raw EEG signals since their process requires extensive time along with the potential for missing significant data features. Traditional classification techniques provide suboptimal performance when handling both the excessive dimensionality problem and the overfitting challenge particularly in scenarios with small medical dataset sample sizes. These techniques fail to detect important spatial and temporal relationships within the EEG signals because they need these essential elements to create precise models of anxiety disorder neural patterns. The use of traditional algorithms proves insufficient for delivering precise real-time EEG-based predictions and personalized treatment opportunities[8].

Modern machine learning together with deep learning approaches should be integrated to establish advanced EEG analysis methods which extract features and perform classifications. This proposed work utilizes the hybrid PCA-RFE (Principal Component Analysis - Recursive Feature Elimination) method to extract valuable features from data while decreasing its dimensions and achieve better execution of anxiety disorder detection through XGBoost and Random Forest classification models. [9] This system has been designed to receive real-time data for continuous assessment of patient neural states during operation. By adopting these advanced methods healthcare professionals achieve automated feature selection alongside improved model interpretability and enhanced accuracy which results in improved diagnosis along with enhanced treatment planning and individualized healthcare treatments for anxiety disorders[10].

The proposed research implements a systematic process to solve the weaknesses found in standard approaches used for detecting anxiety disorders using EEG signals. The data pre-processing step starts with handling missing values then normalizing values for consistent and scalable outcomes. This approach will utilize PCA and RFE as feature extraction methods to both decrease the dataset dimensions and pick the most significant characteristics for better performance alongside reduced overfitting. Model training will utilize both XGBoost and Random Forest to establish an ensemble learning system which enhances classification precision during the process. The test data performance will reach its peak because GridSearchCV optimizes model parameters. The model execution includes cross-validation alongside accuracy metric evaluations with confusion matrices as well as ROC curves which validate the prediction robustness. Real-time anxiety disorder detection based on EEG data can be achieved with the proposed approach which delivers accuracy along with interpretation scalability and accurate detection of mental health disorders.

Main contributions of the proposed work:

1. The research presents a PCA-RFE combination technique which performs dimensional reduction for improved selection of significant features.

2. Combines XGBoost and Random Forest classifiers for improved classification accuracy and robustness.

3. GridSearchCV enables the system to find its best parameters while simultaneously enhancing model performance and minimizing overfitting effects.

This section examines past Machine Learning approaches which detect anxiety disorders through EEG signals alongside their performance traits and issues. The proposed system architecture with its workflow for anxiety disorder detection is detailed in Section III that explains how PCA-RFE hybrid feature extraction joins XGBoost with Random Forest classifiers to achieve better prediction performance. The fourth section presents findings from the proposed solution which includes an accuracy comparison against conventional models and additional advanced systems both in precision and operational efficiency. The research ends with a summary of the proposed work's anxiety disorder detection accomplishments and additional directions for future model development and research.

I. LITERATURE SURVEY

The literature review demonstrates how advanced machine learning techniques use different data sources including EEG signals clinical biomarkers electronic health records and social media data for more precise anxiety disorder detection and treatment methods. The research presents multiple studies which evaluate machine learning systems including deep learning along with support vector machines together with ensemble methods to identify anxiety disorder indicators. Diagnostic approaches based on these methods intend to solve traditional diagnostic shortcomings with precise modern solutions that scale and require no invasive techniques. Despite recent advancements machines models now have major restrictions when it comes to generalization of results along with interpretability as well as their ability to process imperfect and noisy data. Science calls for improved time-sensitive analytical systems able to develop personalized care programs while accurately detecting anxiety symptoms through precise predictions. This research investigates the effectiveness of modern approaches inside this field.

Al-Ezzi et al. (2021) developed a deep learning method to examine social anxiety disorder through brain effective connectivity measurements taken from EEG signals[12]. An analysis of brain activity through diverse deep learning architectures became the basis for anxiety severity prediction according to Sharma & Verbeke (2021).[13] The researchers employed multiple biomarkers for feature extraction to upgrade classification results. The main benefit emerged from uniting clinical information with machine learning methods to produce more precise prediction results.

The research by Nemesure et al. (2021) applied[14] machine learning approaches to EHR for depression and anxiety prediction. The research team used classification methods on EHR-system structured data to conduct mental health condition predictions through risk-factor assessment. The analysis performed by Wanderley Espinola et al. (2022) examined how machine learning and vocal acoustic methods can identify major depressive disorder and bipolar disorder and schizophrenia and generalized anxiety disorder [15]. The conducted analysis used audio recordings to retrieve patterns which revealed mental health disorder specific speech characteristics.

The researchers from Rezaei et al. (2023) employed machine learning technology to functional magnetic resonance imaging (fMRI) data for anxiety disorder analysis[16]. The research used fMRI records in combination with machine learning systems to detect specific brain functioning linked to anxiety. Advanced neuroimaging data strengthened their methods however the complexity coupled with expensive fMRI acquisition restricted access to the data.

The authors Ahmed et al. (2020) developed machine learning through supervised learning models for depression and anxiety detection[17]. The authors utilized standard machine learning methods to detect anxiety and depression through psychological assessment results. This method presented an easy implementation of existing algorithms to psychological data while struggling to demonstrate predictive interpretation of complex human behavioral patterns.

The research team of Jacobson et al. (2021) combined [18] deep learning with wearable passive sensing data to forecast long-term changes in anxiety disorder symptoms that lasted between 17-18 years. The use of wearable sensors allowed researchers to conduct their investigation without invasiveness while tracking symptoms in real time. The researchers of Zhao & So (2018) [19] conducted a study which employed machine learning for drug repositioning of treatments between schizophrenia and anxiety disorders through expression data analysis. The main benefit of this methodology was its ability to repurpose drugs by leveraging existing drug expression data but it encountered problems determining drug effects that extend beyond approved treatment indications because it needed high-quality expression data.

Qasrawi et al. (2022) applied machine learning methods to evaluate and forecast depression and anxiety risk elements in school children[20]. The study explored machine learning models for determining probabilities of anxiety and depression through analysis of behavioral along with environmental risk elements. The paper by Zheng & Ye (2022) introduces deep learning methods that aid in predicting Cognitive-Behavioral Therapy (CBT) results for treating adolescent social anxiety and other mental health issues [21]. The system used deep learning technology with structured medical records to forecast how well therapy would work.

Bendebane et al. (2023) developed a multi-class deep learning method which diagnoses depressive as well as anxiety disorders through Twitter data analysis [22]. Through the examination of social media profiles they developed an automated tool able to perform mental health diagnoses in real-time depending on user behavioral patterns. The main advantage of this method was its ability to acquire extensive datasets automatically however its disadvantages included privacy risks and potential incorrect results stemming from uncaring data that did not come from clinical sources.

The study by Alshanketi (2024) developed a deep learning network named MGADHF which utilized social media data to identify generalized anxiety disorder[23]. Researchers implemented deep learning processing technology on extensive social media datasets so the population could conduct widespread anxiety disorder screenings. Minimum invasive procedures were a benefit of this approach yet its effectiveness depended on social media data quality combined with a need for clinical validation to prevent misclassifications.

Choudhary et al. (2022) described how their approach continuously collects anonymous smartphone data to detect generalized anxiety disorder through machine learning[24]. Anxiety symptoms analysis depended on real-time data that sensors from smartphones collected in this method. The real-time monitoring system gained clinical benefits because it monitored anxiety symptoms but user acceptance and continuous data harvesting posed limitations because of smartphone sensor precision. Their research methodology delivered an extensive evaluation of brain complexity patterns which linked to anxiety. The main advantage of this approach involved utilizing EEG signals directly since these signals exhibit strong mental state sensitivity.

Zhao & So (2017) used drug expression profiles to explore machine learning applications for drug repositioning in schizophrenia and anxiety disorders. [26] This research demonstrated how expression data can be used for drug repurposing purposes while providing a cost-friendly solution for new treatment discovery. A major obstacle existed in determining therapeutic efficiency because performing clinical trials required great detail according to Norouzi & Machado (2024) [27]. The study demonstrated predictive models to check various elements associated with mental health risk assessments. Although the system worked well with different mental health issues it faced challenges because data input quality and accessibility affected the accuracy of its predictions.

Machine learning models enable Kumar et al. (2020) to evaluate anxiety together with depression and stress through psychological assessment feature retrieval[28,29]. The simplicity and suitability of this approach to clinical data stood as its main benefit while its handling of subjective psychological evaluations presented major challenges because they display diverse individual assessment patterns. The authors of Sau & Bhakta (2019) studied depression and anxiety screening practices for seafarers through machine learning methods [30]. The researchers emphasized a particular profession to develop services which addressed mental health issues that seafarers encounter. The table 1 comparison outlines the existing methods' performance.

Table 1: Comparison of the Existing Models

S.No	Author(s) et al. (Year)	Dataset	Methodology	Accuracy (%)	Challenges
1	Jacobson et al. (2021)	Wearable passive sensing data	Deep learning paired with wearable sensors for anxiety prediction	92	Long-term data collection, sensor reliability, and individual variability in symptoms.
2	Alshanketi (2024)	Social media data	Deep learning with MGADHF architecture for generalized anxiety disorder detection	91	Data privacy, non-clinical data sources, and noisy social media content.

3	Rezaei et al. (2023)	fMRI data	Machine learning applied to fMRI to study anxiety disorder patterns	90	High cost of fMRI acquisition, complex data preprocessing, and need for large sample sizes.
4	Al-Ezzi et al. (2022)	EEG data	Fuzzy entropy and machine learning for social anxiety disorder detection	89	Computational complexity, and the challenge of interpreting high-dimensional EEG data.
5	Wanderley Espinola et al. (2022)	Vocal acoustic analysis data	Machine learning for detecting anxiety using vocal features	88	Variability in voice data quality and need for extensive labeling of speech samples.
6	Sharma & Verbeke (2021)	Clinical biomarkers from patient records	Machine learning models for anxiety disorder diagnosis using clinical biomarkers	87	Data availability, quality, and variability of clinical biomarkers.
7	Al-Ezzi et al. (2021)	EEG data	Deep learning on brain effective connectivity to assess anxiety severity	87	Data annotation difficulties and need for large EEG datasets.
8	Kumar et al. (2020)	Psychological assessments	Machine learning models for detecting anxiety, depression, and stress	86	Subjectivity in psychological assessments and challenges in feature extraction.
9	Zhao & So (2018)	Drug expression data	Machine learning for drug repositioning for anxiety disorders	85	Need for accurate expression data and challenges in predicting effects outside of intended drug indications.
10	Zhao & So (2017)	Drug expression profiles	Machine learning for drug repositioning in anxiety and schizophrenia disorders	85	Expression data variability and challenges in matching drugs to specific anxiety symptoms.

Traditional methods for detecting anxiety disorders fail because they do not solve the high-dimensionality and complex nature of EEG data effectively which creates difficulties regarding feature redundancy and overfitting and a lack of capability to detect temporal dependencies. The methods face difficulties with both generalization

and scaling their capabilities when implementing them to process real-time data samples. The proposed solution solves these problems through PCA-RFE hybrid dimension reduction that maintains fundamental features together with XGBoost and Random Forest classifiers which produce enhanced accuracy and operational reliability. The implementation of GridSearchCV serves for hyperparameter optimization to enhance model efficiency which in turn improves generalization capabilities and boosts prediction accuracy for EEG-based anxiety disorder detection. The method delivers an efficient system to handle complex information along with real-time scalability and processing capabilities.

II. PROPOSED WORK

The workflow in Figure 1 represents how a machine learning pipeline optimizes model evaluation through different stages. Data Preprocessing stands as the first step that includes three procedures: Data Cleaning followed by Data Normalization and then Exploratory Data Analysis (EDA). The next step uses PCA (Principal Component Analysis) and RFE (Recursive Feature Elimination) for feature extraction while combining them for better performance. The Model Building phase utilizes XGBoost, Random Forest, SVM and Neural Networks among the multiple algorithms. The Model Optimization section utilizes GridSearchCV together with the Hybrid Model (XGBoost + Random Forest) to optimize hyperparameters. At the end of the process Model Evaluation measures system performance through Accuracy and the Confusion Matrix along with ROC Curve & AUC and Precision, Recall, F1-Score metrics. The method produces strong and effective models through its detailed implementation structure.

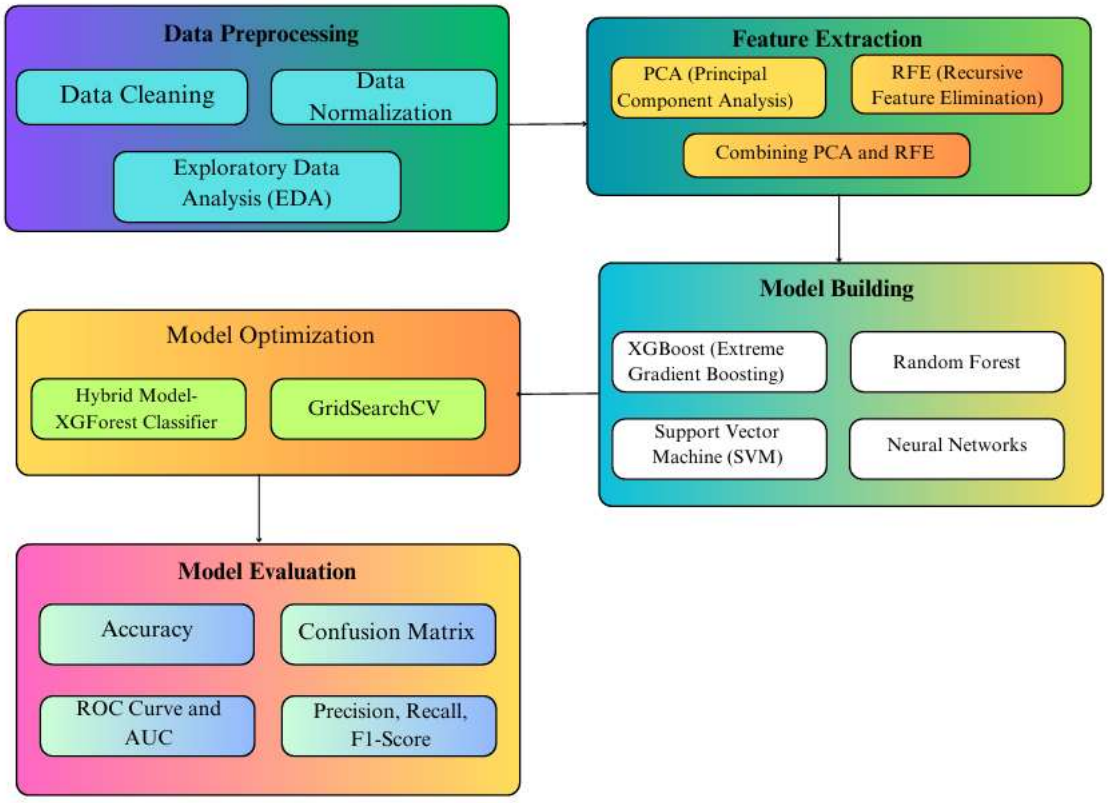

Figure 1: Work Flow of the Proposed Model

A. Data Preprocessing

Data collection is a crucial first step in any machine learning project. An EEG (electroencephalogram) dataset requires obtaining data that shows brain activity patterns for the identification of anxiety disorder symptoms. The optimal dataset should include brain wave frequency data arranged in delta theta alpha beta and gamma categories. The different mental states or cognitive activities correspond to specific frequency ranges that occur in EEG signals through the delta theta alpha beta and gamma bands. Each of these bands corresponds to

specific frequency ranges in EEG signals, which are thought to represent different mental states or cognitive activities:

- Delta (0.5-4 Hz): Associated with deep sleep or unconscious states.
- Theta (4-8 Hz): Linked to light sleep, relaxation, and creativity.
- Alpha (8-13 Hz): Commonly seen during relaxed alertness and calm states.
- Beta (13-30 Hz): Associated with active thinking and focus.
- Gamma (30-100 Hz): Linked to higher mental activity and problem solving.

The evaluation should include demographic characteristics such as patient age together with sex because brain wave patterns demonstrate possible age and sex-related variations. The specific nature of anxiety disorders (generalized anxiety disorder, social anxiety disorder etc.) in clinical data helps researchers establish better connections between brain functions and mental well-being. The construction of an accurate and comprehensive dataset requires obtaining data from dependable sources such as medical clinics and research papers and hospitals.

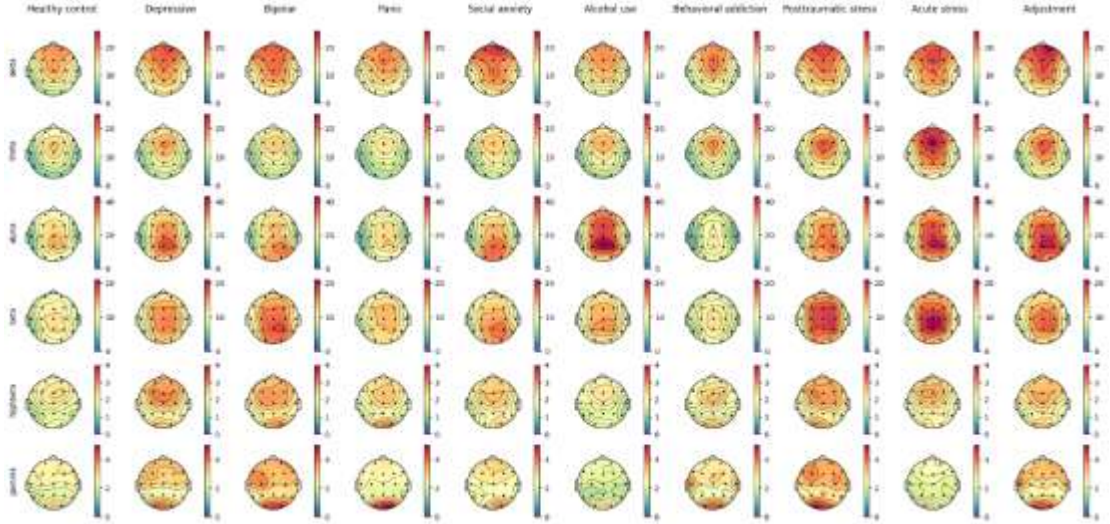

Figure 2: Brain Wave Distribution Across Different Mental Health Conditions

This figure 2 shows topographic EEG maps of brain wave distributions across various mental health conditions and a healthy control group. Each group receives presentation of different brain wave frequencies (delta, theta, alpha, beta, gamma) through topographic representation maps. The labels on these maps use warmth for representing increased brain operational levels alongside cooler tones identifying decreased operational levels. The illustration presents visual assessments of brain wave activities under diverse mental health conditions including Depressive and Bipolar disorders and Social Anxiety and outside stress elements including Acute stress and Alcohol use. The analysis provides essential information to identify neural patterns linked to various psychological states together with mental disorders. The right image section contains color bars that serve to explain the wave intensity scale for each testing condition.

1. Data Cleaning

After dataset collection it becomes essential to conduct data cleaning operations to maintain high-quality usable data values. The first action for data cleaning involves treating missing values. Several reasons such as sensor failure and human errors in recording produce missing data. Multiple strategies exist to deal with missing value occurrences:

- **Imputation:** The statistical methods of mean and median and mode provide various ways to fill in missing values through imputation. When handling numerical missing data features the most suitable imputation method would be using column average values as replacements:

$$x'_i = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

Where x'_i is the imputed value, and x_i are the available values for that feature.

- **Removal:** The removal of rows or columns that contain extensive missing data should be considered when the data is non-randomly missing or when hundreds of values are absent from the overall data collection.

Duplicate samples should be identified then eliminated for effective unbiased representation of the samples. The model performance could face alterations because duplicate rows exist within the dataset. The procedure to remove duplicates works within basic categories of data processing libraries which include pandas in Python. Detection

of outliers should also be followed by proper handling methods. A data point qualifies as an outlier when it stands out substantially from other observations in the dataset which can thus affect statistical calculations. The detection of outliers requires two common methods that use Z-scores and Interquartile Range (IQR). The identification of outliers occurs through application of the formula:

$$z = \frac{x - \mu}{\sigma} \qquad (2)$$

The computation includes the data point x combined with two statistical values consisting of μ for mean and σ for standard deviation. The threshold value is typically set at 3 so when a Z-score reaches this point the data point becomes considered an outlier which enables appropriate handling options (removal or adjustment).

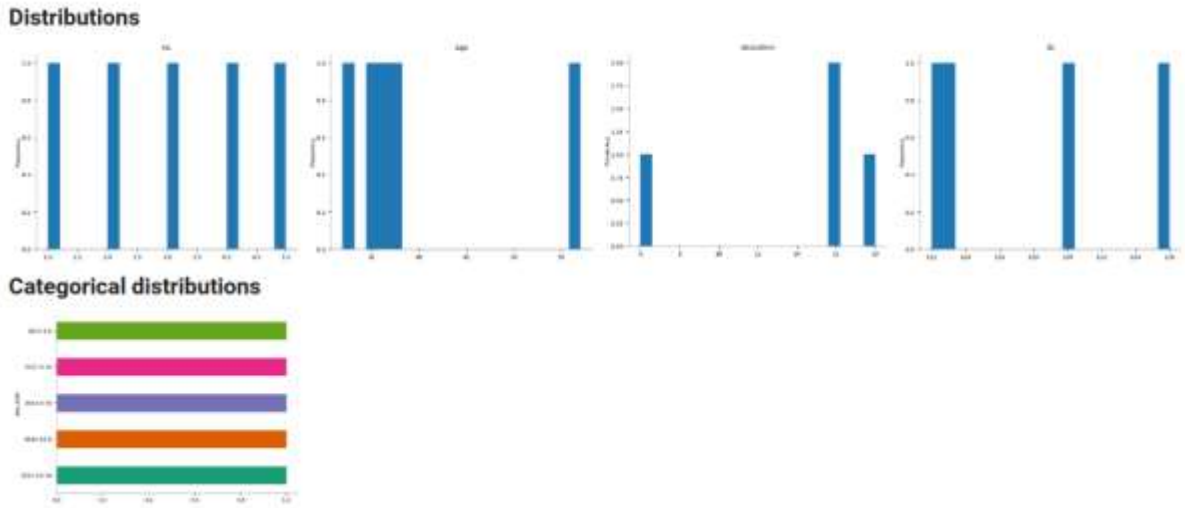

Figure 3: Distribution of Continuous and Categorical Variables in the Dataset

The figure 3, displays the distributions of both continuous and categorical variables in the dataset. The upper part of the figure contains distribution graphs for continuous measures no, age, education and IQ. The no values show regular distribution but age data gathers within the 20-30 age range and education data presents specific skewed values. Among the overall IQ scores there exist more frequencies of higher value points. The distribution of categorical variables exists in the bottom section. This bar chart expresses categorical data frequencies through distinct colored bars that represent individual groups from the given variables listing. The distribution along with potential unbalance patterns in categorical data becomes clear through this section.

2. Data Normalization

Data preprocessing through normalization represents a vital step which allows the analysis of features that possess different scaling dimensions. The difference in scale between EEG brain wave frequencies and demographic data (age, sex) creates difficulties for models when attempting to process the data. Through normalization processes the model receives equal contribution from its features. Standardization and Min-Max Scaling represent the two principal normalization techniques.

- **Standardization**: Standardization transforms data by moving all features to mean zero and setting standard deviation to one. The process involves both mean subtraction from each feature and subsequent division by the calculated standard deviation.
- **Min-Max Scaling**: A different data transformation technique called Min-Max Scaling produces data distribution within the range [0, 1]. The calculation follows this formula to perform the operation:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (3)$$

The transformation applies to original feature value x and includes the minimum x_{max} *and* x_{min} values of the feature range. Min-Max scaling provides effective normalization when dealing with features that have different units or require bounded ranges starting from neural networks based on sigmoid activation functions.

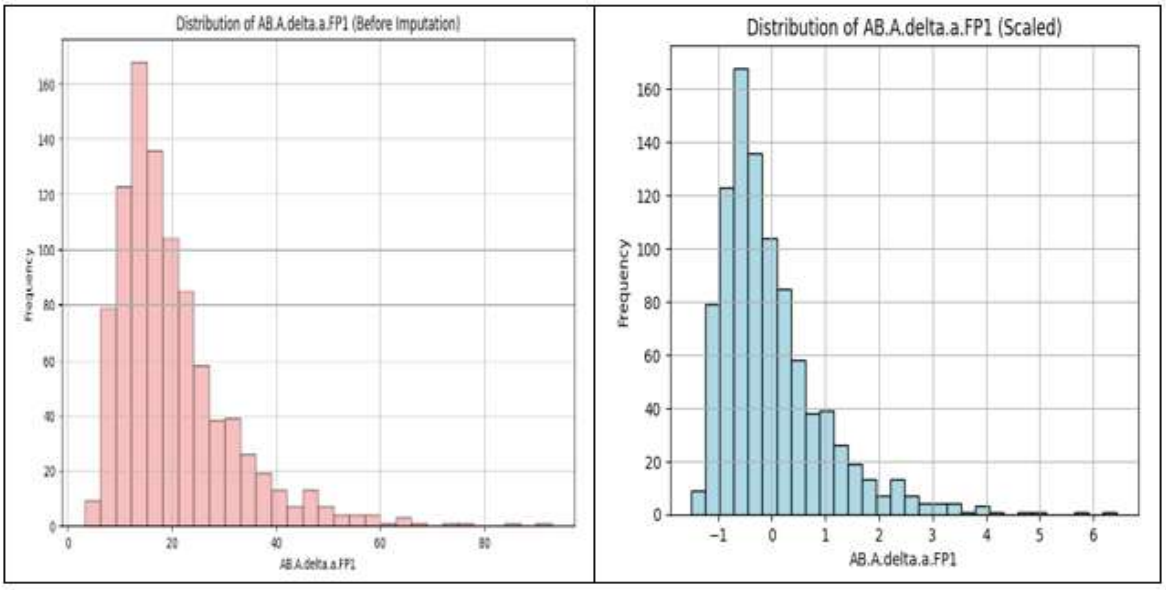

Figure 4: Distribution of EEG Data (Before Imputation and After Scaling)

The figure 4 presents the distribution patterns of EEG feature AB.A.delta.a.FP1 before and after preprocessing. The pink plot on the left side of the figure represents the original unprocessed data distribution showing heavy concentration in the lower spectrum of values with an asymmetric shape. The data distribution along with the right-skewed shape demonstrates the existence of missing values or data imbalance. The right blue plot demonstrates the same feature after preprocessing through imputation and scaling that transforms the data to have standardized mean distribution of zero and standard deviation equal to one. The normalized distribution of scaled data generates data fitness that makes it suitable for analytical and modelling processes. The preprocessing steps establish model performance improvements while preventing features with large scales from receiving undue emphasis from the model.

B. Exploratory Data Analysis (EDA)

Exploratory Data Analysis stands as a vital stage for data scientists because it helps them see data structures alongside identifying problems like outlier issues and missing variables and unbalanced classes. Data exploration through EDA uses multiple statistical methods in conjunction with visual representations that serve as a basis for model-based and preprocessing decisions.

- **Descriptive Statistics**

Descriptive statistics provide a summary of data distribution by presenting its central values and distribution scope as well as its characteristic shapes. Three main measures of central tendency consist of the mean (μ) together with median and mode. The statistical measures enable researchers to understand standard feature values across the database. The mean is calculated as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (4)$$

This equation uses μ for mean value while n represents total data points along with the feature values x_i . A median value appears in the center position during the ordering process of data and the most common value within a dataset functions as the mode.

The dispersion of data is evaluated through standard deviation (σ) and variance (σ^2) statistical measurement methods:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad (5)$$

The standard deviation is simply the square root of variance:

$$\sigma = \sqrt{\sigma^2} \qquad (6)$$

Measurement of data distribution reveals the clustering patterns which indicate if data points cluster tightly near the mean or distribute more widely. Measures of skewness and kurtosis help reveal the shape characteristics of data distributions which show symmetry or skewed appearances and how much data extends beyond usual ranges.

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

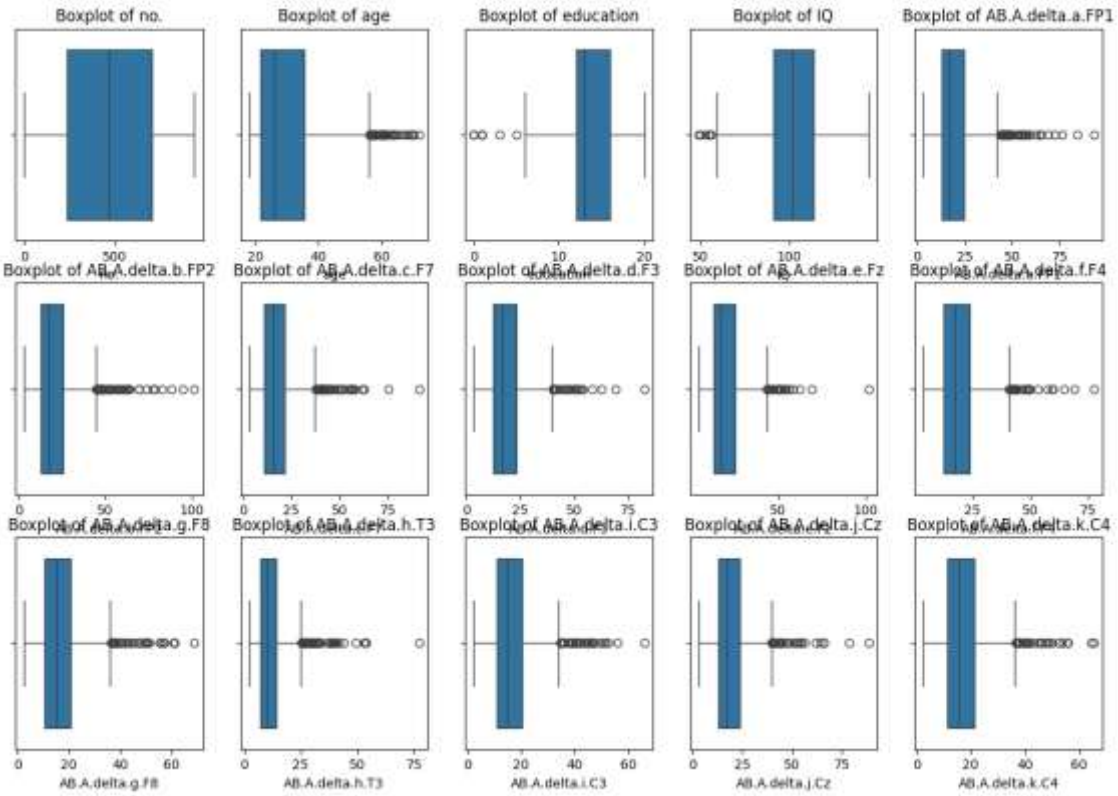

Figure 5: Boxplots for Various Features in the EEG Dataset

The figure 5 presents a boxplot grid with multiple features from the EEG dataset that contains demographic and EEG signal attributes. Each feature distribution and its central tendency together with variability spread becomes visible through boxplots while outliers can also become apparent. The no boxplot shows narrow data distribution and minimal outliers whereas the wide spread of values with visible outliers appears in AB.A.delta.a.FP1 and AB.A.delta.b.FP2. Each plot contains a central box showing the interquartile range (IQR) while extended whiskers extend to display data points that exceed IQR values. Observable dots represent extreme values that might need additional investigation and preprocessing steps. Before training the model it is crucial to perform this analysis for detecting anomalies and understanding the underlying data structure.

This figure 6 displays pairplots for a subset of features from the EEG dataset. Pairplots enable users to view two-dimensional feature relationships when combining scatter plots with histograms. Each plot arrangement contains scatter plots that display two-feature relationships while the diagonal positions show single-feature distribution outcomes. The scatter plots indicate a possible common pattern in AB.A.delta.a.FP1 and AB.A.delta.b.FP2 characteristics that display positive correlation. Each histogram shows the distribution type with the data dispersion characteristics of separate features in the data set. Data visualization techniques help identify trends together with relationships and outlying points in data before beginning analysis or model construction plans.

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

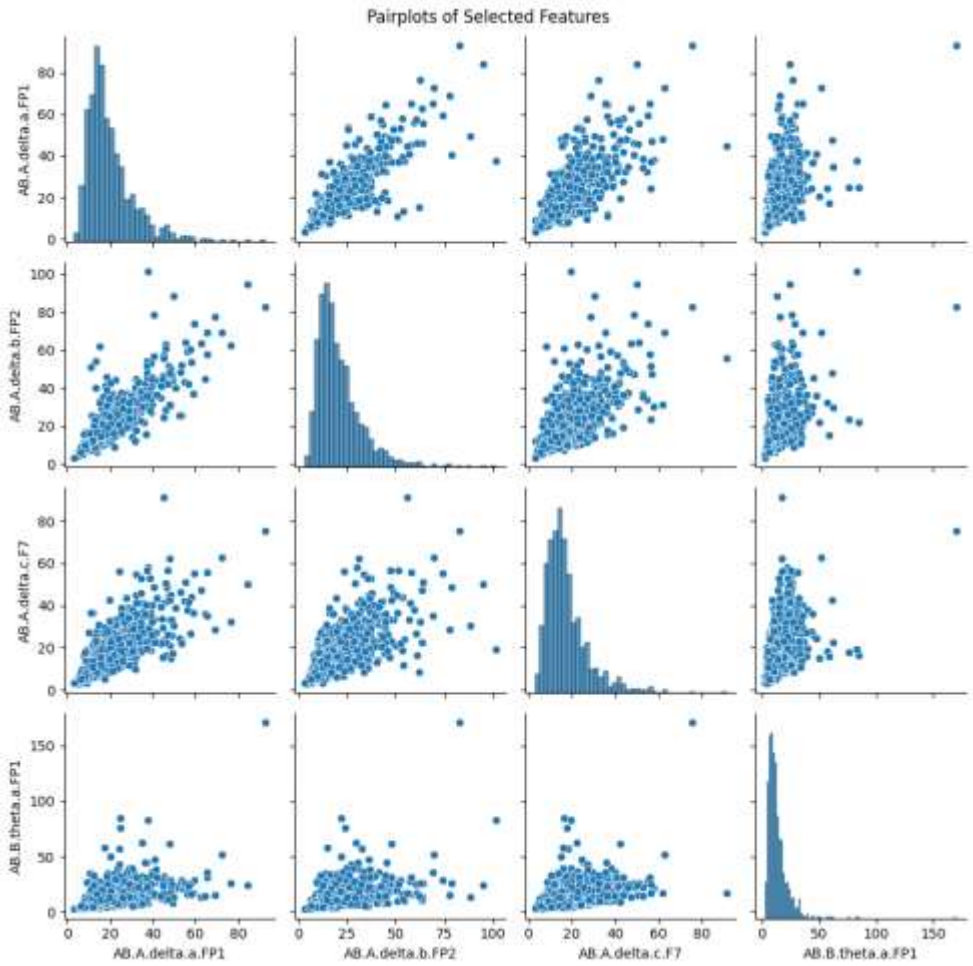

Figure 6: Pairplots of Selected EEG Features

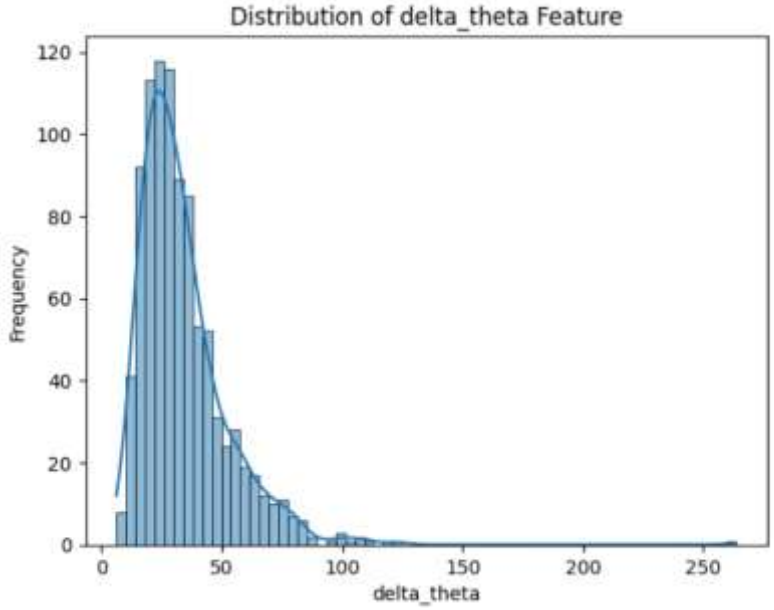

Figure 7: Distribution of delta_theta Feature

This figure 7 shows the distribution of the delta_theta feature in the dataset. The x-axis frequency distribution of the delta_theta feature shows right-skewed frequency data centered on lower value ranges between 0 to 50. A smooth probability distribution emerges through the addition of a Kernel Density Estimate curve that emphasizes the skews. The distribution features an extended tail section towards right values because extreme entries appear

only sporadically. Most data points stay within a particular range according to this distribution pattern but there exist a small number of outliers which fall on the higher end of the spectrum.

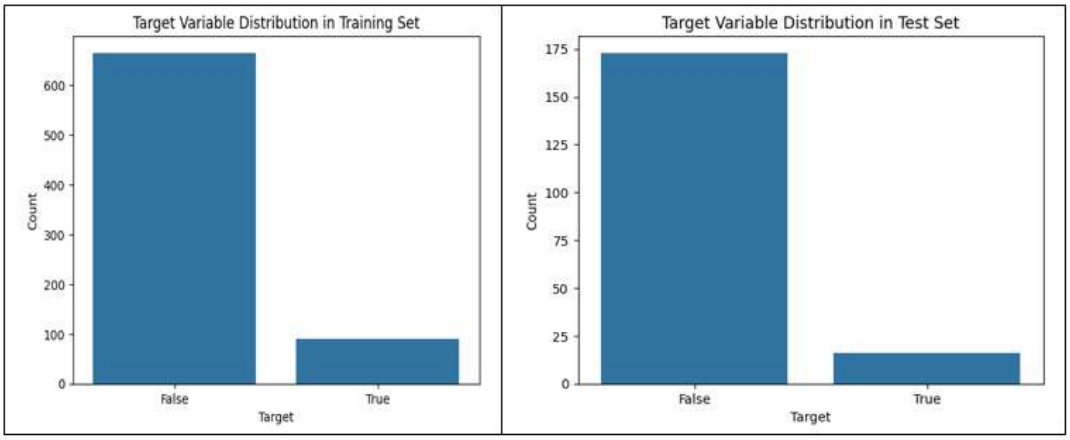

Figure 8: Distribution of Target Variable in Training and Test Sets

The presented figure 8 shows how the target variable distribution with False and True values appears across training and testing data. The training set data distribution reveals False as the dominant value alongside a minority of True cases according to the illustration in the left panel. Another pattern emerges from the right test set panel showing False values that strongly predominate True values. Significant class inequality between False and True in the collected datasets has serious implications for model training because such distributions make it easy for models to predict the majority class. The system needs correction since possible solutions include oversampling and undersampling along with class weight implementation in model training to prevent prediction biases.

C. Feature Extraction

The data preprocessing pipeline needs feature extraction as its crucial initial step for handling high-dimensional datasets. The PCA-RFE Hybrid approach unites PCA and RFE to obtain vital features which enhance model performance in this setting. These methods perform dimension reduction on data while selecting only the key attributes making predictive results both more accurate and resource-efficient.

- **PCA (Principal Component Analysis)**

PCA operates through transforming original variables to principal components (PCs) that represent uncorrelated variables which demonstrate maximal variance in the dataset. PCA finds new axes (principal components) that optimize the data variance. The maximum data variance gets captured by the first principal component (PC1) while each successive component measures orthogonal variations. The mathematical expression for PCA uses:

$$X_{new} = XW \qquad (7)$$

Where:

- X_{new} is the transformed dataset,
- X is the original data matrix,
- W is the matrix of eigenvectors or principal components.

The eigenvectors are computed from the covariance matrix of the data:

$$\Sigma = \frac{1}{n-1} X^T X \qquad (8)$$

Each principal component's variance amount is calculated through the eigenvalues that operate on the covariance matrix Σ. PCA lets users reduce data dimensions through a process that maintains the most critical features which drive the variance in data measurements. PCA works best on datasets with numerous features that show correlated behavior.

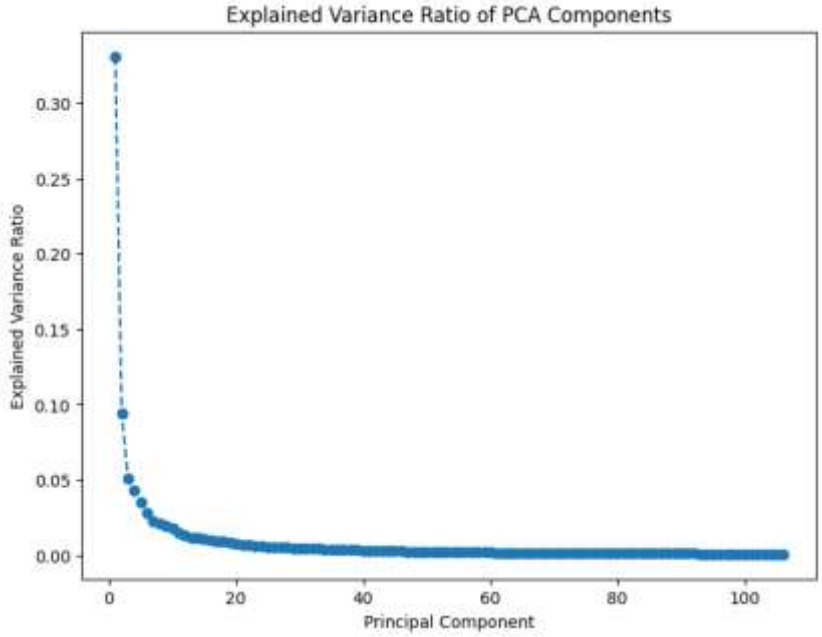

Figure 9: Explained Variance Ratio of PCA Components

The Figure 9 demonstrates the ratio of explained variance of principal components derived through Principal Component Analysis (PCA). This figure 9 demonstrates the distribution between principal components extending from 0 to 100 along the x-axis and the proportion of variance each component explains along the y-axis. A substantial margin of total variance (approximately 30%) exists within the initial several components until the rest of the components display minimal explanatory power thus reaching near-zero values. A small number of components can effectively explain the data distribution according to this behavior and therefore PCA works as an efficient dimension reduction method. The plot demonstrates how most of the information in the dataset can be preserved through selecting fundamental principal components which simplifies the training procedure by maintaining relevant data points.

- **RFE (Recursive Feature Elimination)**

The feature selection method called Recursive Feature Elimination (RFE) performs a sequence of operations that remove features one by one and constructs models using the remaining features. The algorithm removes least important features while ranking them according to their significance until it reaches the target number of selected features. The process starts by creating a model with importance evaluation then it removes the least important feature within each step. RFE serves as a powerful tool which enables practitioners to select features most crucial for prediction while removing unneeded features. The RFE process requires several sequential steps including feature importance calculation and least important feature elimination.

1. The selected model receives the dataset input during the fitting process followed by a computation of feature importance analysis.
2. The ranking of features takes place using importance measures such as linear model coefficients or tree-based model importance to eliminate the least significant feature.
3. The method should be run repeatedly with one feature elimination at a time until the chosen number of features becomes achievable.

Mathematically, RFE can be represented as:

$$F_{selected} = RFE(F_{all}) \quad (9)$$

Where F_{all} represents the full set of features, and $F_{selected}$ is the set of features selected after applying RFE.

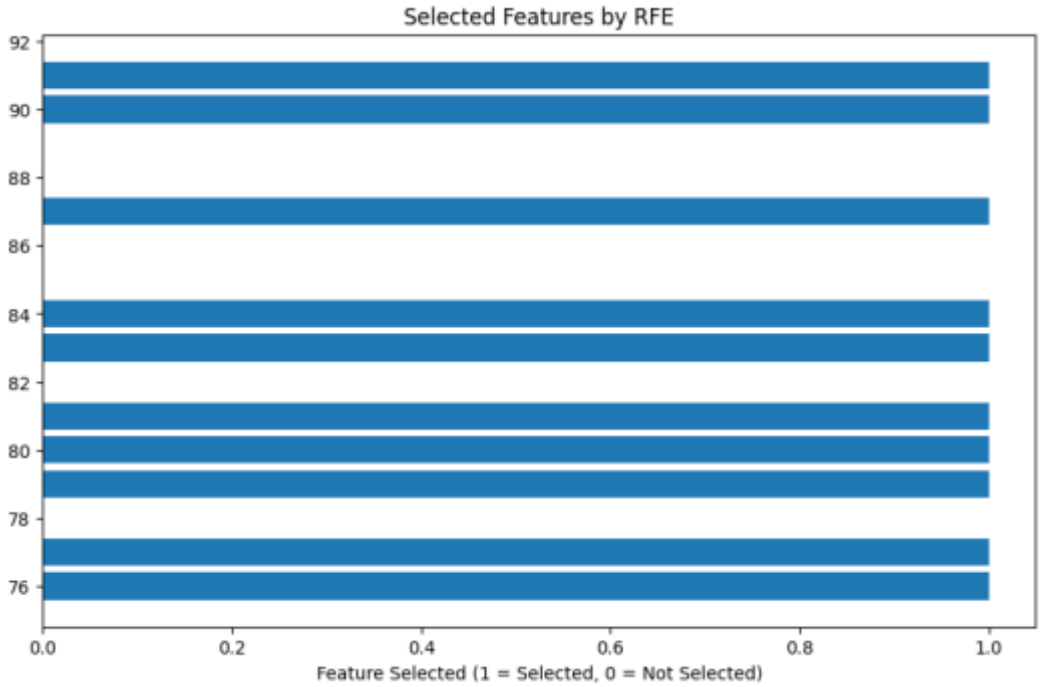

Figure 10: Selected Features by RFE (Recursive Feature Elimination)

The figure 10 displays the results of Recursive Feature Elimination (RFE) as a dataset feature selection approach that determines the main characteristics. RFE selects the features through the horizontal bar chart by marking them with a value of 1 for selection and 0 for non-selection. The vertical axis of Figure 10 contains feature selection status data using binary values where 1 indicates feature selection and 0 indicates feature exclusion. RFE selects features numbered 92 and 90 along with others because these features demonstrate the most predictive strength for the target variable. RFE performs an iterative process to eliminate features that are least important before complete training on relevant features occurs.

- **Combining PCA and RFE**

  PCA integration with RFE creates an efficient approach for feature extraction. The initial application of PCA performs dimensionality reduction by transforming the original features into principal components that maintain the maximal variability. RFE applies from the reduced dimensionality space of the transformed dataset.

  The process works as follows:

  1. The original data needs PCA application for obtaining a set of uncorrelated principal components. The procedure selects components because they explain most of the data variance cumulatively.

$$Variance\ Explained = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{k} \lambda_i} \qquad (10)$$

  Where:

- λ_i are the eigenvalues corresponding to each principal component,
- k is the number of selected components,
- n is the total number of components in the dataset.

  The formula demonstrates the percentage of data variance which the initial k principal components explain for determining the desired number of components needed to preserve information in the dataset.

  2. After conducting dimensionality reduction on the primary components you should apply RFE to identify crucial components that will be used in predictive modeling. The second procedure diminishes unnecessary components from the analysis because these aspects lack important predictive significance.

The combined approach leverages the strengths of both techniques:

- Using PCA simplifies the dataset structure through identification of significant variance and simultaneous elimination of feature dependency.
- The model interpretation and operational performance both improve through RFE's ability to pick only crucial components from the refined feature set.

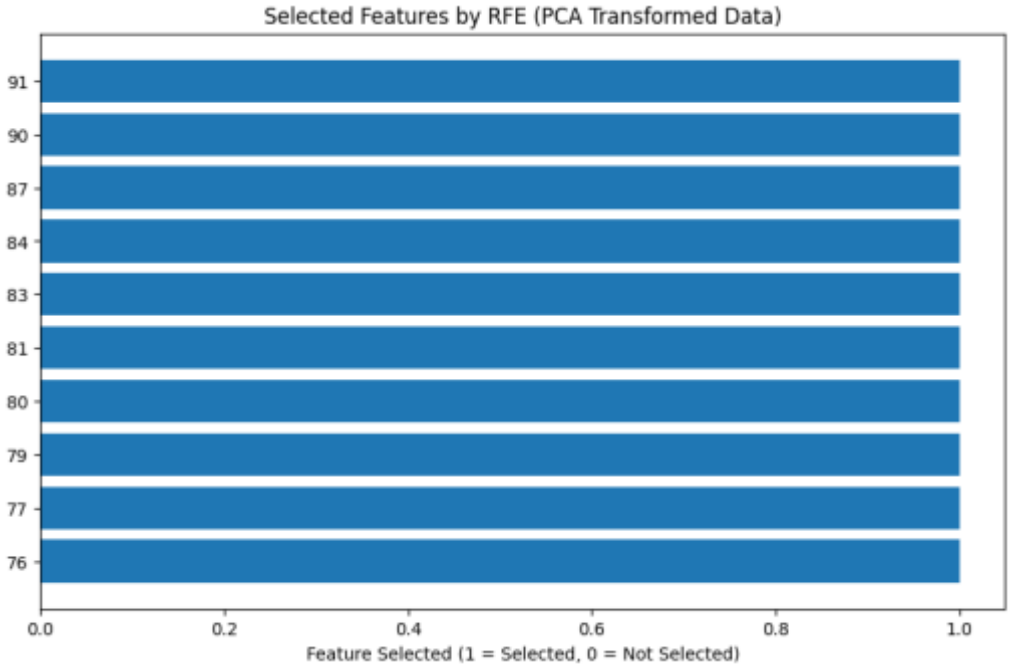

Figure 11: Selected Features by RFE (PCA Transformed Data)

The figure 11 demonstrates how Recursive Feature Elimination (RFE) worked on the Principal Component Analysis (PCA) transformed dataset. The RFE procedure shows the selected features through 1 marks while non-selected features use 0 marks in this horizontal bar chart. The selection status appears on the x-axis where numbers 1 signify that features got chosen while numbers 0 show features were left out. Analysis results reveal that principal components with higher indices (such as 91 and 90) were chosen for selection following PCA transformation indicating these components have significant amounts of variance that contributes most to prediction accuracy. RFE becomes more effective when combined with PCA because it reduces data dimensions and selects features that maintain crucial information for model building effectiveness. The method helps to boost model performance after transformation by identifying the most important features.

D. Model Building

Model building serves as an essential pipeline stage through which one selects proper algorithms to solve specific problems. A model selection depends on the type of data alongside the primary problem and final objectives. Several machine learning models exist asynchronously to process data and automate different tasks proficiently. This part discusses XGBoost, Random Forest, Support Vector Machine (SVM) and Neural Networks as popular classification algorithms. Different machine learning models bring unique benefits to different contexts while we will explain their main features together with their mathematical expressions for specific use cases.

- **XGBoost (Extreme Gradient Boosting)**

The ensemble learning algorithm XGBoost proves exceptional at processing complex datasets with its ability to effectively handle large datasets. The GBM ensemble learning method works by creating weak decision tree models which XGBoost develops iteratively based on the errors from past models. The loss function optimization in XGBoost utilizes mean squared error (MSE) for regression duties and log loss for classification tasks depending on the defined objective. The main gradient boosting mathematical formula is:

$$F(x)^{(t)} = F(x)^{(t-1)} + \eta . h_t(x) \quad (11)$$

Where:

- $F(x)^{(t)}$ is the model at iteration ttt,
- $h_t(x)$ is the weak learner (usually a decision tree),
- η is the learning rate (shrinkage).

XGBoost implements enhancements to gradient boosting through the implementation of regularized techniques for preventing overfitting along with parallelization capabilities and handling of missing values. The competitive machine learning landscape has embraced XGBoost because it manages extensive datasets with multiple data formats through its high accuracy output.

- **Random Forest**

Random Forest creates numerous decision trees which merge their predictions to achieve better accuracy levels while fighting overfitting behavior. Random subsets of data alongside selected features get used to train decision trees in a manner that produces diverse models which minimize variances. The prediction combines all individual tree outputs through an average calculation process during regression tasks while utilizing majority voting for classification problems. A random forest produces its final classification output y' by applying the following mathematical expression:

$$y' = mod(y_1, y_2, \ldots . y_n) \qquad (12)$$

The prediction for the i- th tree takes the form of y_i while mode represents the predominant class label which emerges from all tree predictions. Random Forest demonstrates strong capabilities when analyzing datasets with numerous dimensions and withstands overfitting issues as the number of trees increases. The high computational complexity makes this method difficult to handle big data while its lack of interpretability becomes a problem where clear explanations are necessary.

- **Support Vector Machine (SVM)**

The supervised machine learning model Support Vector Machines (SVM) acts as a classifier for non-linearly separable data during classification tasks. SVM aims to discover a hyperplane which provides maximum margin separation by choosing support vectors as the closest points between every classification. The SVM algorithm applies a kernel function to elevate data dimensions so linear hyperplanes can separate the classes. A linear SVM produces its decision function using the following expression:

$$f(x) = sign(w^T x + b) \qquad (13)$$

Where:

- w is the weight vector,
- x is the input feature vector,
- b is the bias term.

When data cannot be separated through linear methods the SVM uses kernel transformation to move data into higher dimensional spaces for finding suitable hyperplane solutions. The common SVM kernels used in practice include Radial Basis Function (RBF), polynomial and sigmoid types. SVM achieves excellent results when operating in spaces with many dimensions while at the same time working well when problems demonstrate clear boundary separations. The use of SVM becomes expensive in terms of computation when dealing with extensive datasets while dependency on both kernel selection and hyperparameters affects its performance.

- **Neural Networks**

Neural Networks function as robust machine learning systems that adopt brain neural network structures as a blueprint. A neural network comprises various stages of linked neural cells that conduct mathematical transformations to input data. Neural networks function remarkably well when identifying complex nonlinear patterns found inside large datasets. These systems achieve deep learning capabilities through the ability to identify elaborate patterns as they serve crucial roles across image recognition activities and natural language processing methods and many additional domains using advanced learning methods. The basic formula for a neural network layer appears as:

$$a^{(l)} = \sigma\left(W^{(l)} x^{(l-1)} + b^{(l)}\right) \qquad (14)$$

Where:

- $a^{(l)}$ is the activation of layer lll,
- $W^{(l)}\ and\ b^{(l)}$ are the weights and biases of the layer,
- Σ is the activation function (e.g., ReLU, Sigmoid, or Tanh),
- $x^{(l-1)}$ is the input to layer lll.

The functionality of neural networks extends to complex functions yet they need large training data because excessive modeling can occur if regularization methods are absent. The prevention of overfitting and generalization enhancement in neural networks depends on three techniques: dropout and batch normalization and early stopping.

E. Model Optimization

Optimizing machine learning models requires selection of hyperparameters as a premier step in the process. During this step we choose different values of hyperparameters that will be optimized for XGBoost as well as Random Forest modelsloy. All performance characteristics of a model emerge from values known as hyperparameters which must be established before training starts.

For XGBoost, key hyperparameters include:

1. **n_estimators**: The number of boosting rounds or trees determines through n_estimators. An increase in number generally delivers better model performance at the cost of heightened chances for overfitting.

$$Model\ output = \sum_{i=1}^{n} \alpha_i h_i(x) \qquad (15)$$

where h_i (x) are the individual decision trees and α_i are the weights assigned to each tree.

2. **learning_rate**: The step size which determines model parameter updates during each iteration is learning_rate while n_estimators indicates the number of trees to build for boosting. The model generalizes better when using slow training steps although it might need many more trees for the ensemble $n_{estimators}$.
3. **max_depth**: The maximum depth of each tree. The model can identify more intricate patterns by increasing depth but extremely deep trees tend to fit the data excessively.
4. **subsample**: The subsample represents the portion of samples which gets utilized during tree training. Putting a number lower than 1 as an input helps decrease overfitting because it adds randomness to the model training method.

For Random Forest, key hyperparameters include:

1. **n_estimators**: The $n_{estimators}$ parameter functions the same way as XGBoost by defining the quantity of trees within the forest. Model performance increases with higher n_estimators implementation however the computational load becomes heavier accordingly.
2. **max_depth**: The maximum depth of each decision tree in the forest. Limiting the depth helps control overfitting.
3. **min_samples_split**: The minimum node splitting requirement stands at 3. High values of this parameter stop the model from developing excessively complex branches from tiny data clusters.
4. **max_features**: The maximum number of features determines how many attributes will be used to pick the best split. The size value for max_features usually corresponds to the square root of total feature count but administrators can modify it to adjust complexity levels.

- **Perform GridSearchCV**

GridSearchCV performs an exhaustive search to find the optimal hyperparameters from the specified ranges which have been defined previously. During its operation GridSearchCV tests all possible hyperparameter sets from predefined ranges and identifies the set that delivers optimal performance based on selected evaluation metrics (such as accuracy, F1 score, etc.).Mathematically the procedure follows the formula:

$$\theta' = arg_0^{max} CV(\theta) \qquad (16)$$

Where:

- θ' is the set of hyperparameters that maximize cross-validation performance,
- θ represents the hyperparameter configuration,

The indicator CV stands for performance across validation (accuracy, precision, recall included).With k-fold cross-validation GridSearchCV evaluates the robustness and generalizability of the model by testing each set of hyperparameters. The model becomes less prone to overfitting through this process because it demonstrates high performance on previously unseen data. The model cross-validation process employs k sets of data splits where the model trains on k−1 splits while testing on the left-out set to produce averaged results.

The evaluation of hyperparameter performances calculates an average result from all cross-validation folds:

$$Average\ CV\ score = \frac{1}{k}\sum_{i=0}^{k} score_i \qquad (17)$$

Where $score_i$ is the performance score on fold i.

- **Optimize Hybrid Model**

A hybrid model consists of several base models that work together to enhance predictive outcomes. Two common hybrid techniques exist as Voting and Stacking. The application of GridSearchCV helps us adjust the hyperparameters for the hybrid model that merges XGBoost with Random Forest or alternative models. A Voting Classifier predicts new data values through the compilation of predictions originating from individual models. Different weighting options of the classifiers exist to find the best performing model. The algorithm uses the following calculation to obtain the final prediction results:

$$y' = argmax\left(\sum_{i=0}^{n} w_i . y'_i\right) \qquad (18)$$

Where w_i are the weights assigned to each model, and y'_i is the predicted output from the i-th model. Using GridSearchCV lets users discover the best combination of weights at each classifier to generate the highest performance level. A Stacking Classifier requires base classifier predictions that serve as inputs for a subsequent final model which usually implements logistic regression or an alternative classification approach. Through the

stacker model the predictions from base models serve as input for determining the final prediction by learning suitable model combination. The collective model minimizes a specified loss function in order to train:

$$L = \sum_{i=0}^{n} (y'_i - y_i)^2 \qquad (19)$$

Where y'_i is the prediction of the stacked model, and y_i is the true label. The goal is to minimize the difference between the predicted and actual values across all base models.

III. RESULT &DISCUSSION

A dataset containing EEG recordings from 10 mental health patient groups includes measurements of Healthy controls and patients experiencing Depressive, Bipolar and Panic disorders as well as Social anxiety and Alcohol use irregularities alongside Behavioral addiction, Posttraumatic stress, Acute stress and Adjustment disorders. The dataset consists of over 1,000 samples containing details which affect the dataset structure while showing brain wave activity across different frequency bands including delta, theta, alpha, beta and gamma bands. The data acts as input for machine learning models which use the EEG features for mental health classification. Predictive diagnostic models need to be created through brain wave pattern examinations under these conditions [11]. A display of performance metrics has been presented in table 2 for each model design.

Table 2: Model Classification Metrics

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	0.93	0.97	0.85	0.91
Random Forest	0.87	0.95	0.79	0.84
SVM	0.54	0.47	0.44	0.44
Neural Network	0.37	0.29	0.29	0.28

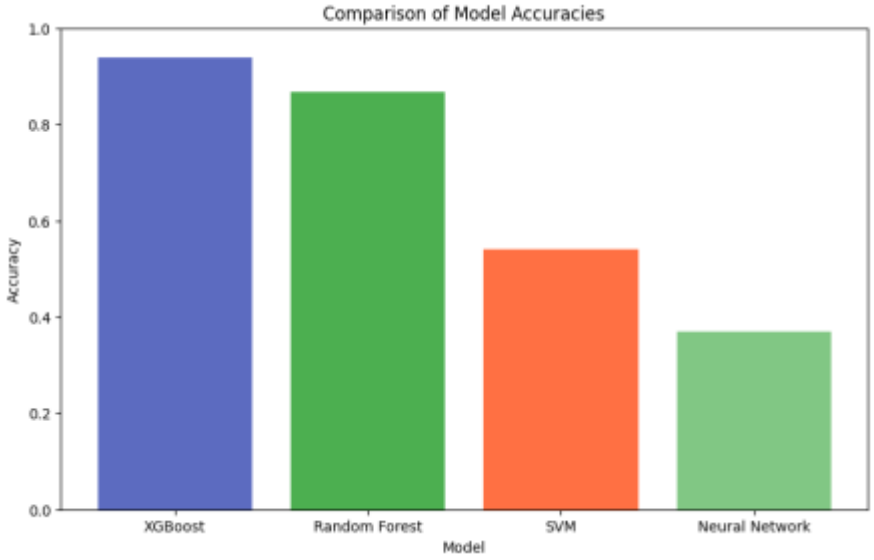

Figure 12: Comparison of Model Accuracies

The accuracy levels between XGBoost, Random Forest, SVM and Neural Network machines are displayed in Figure 12. XGBoost achieves the highest accuracy from the plotted data followed by Random Forest which demonstrates good accuracy performance. The SVM model demonstrates average model performance but Neural Networks perform the least effectively among all models. Model evaluation shows XGBoost and Random Forest (tree-based model) attain better results than the other models for this dataset. The performed analysis demonstrates XGBoost as the most suitable model for this reconstruction task yet the Neural Networks require optimization to achieve optimal results.

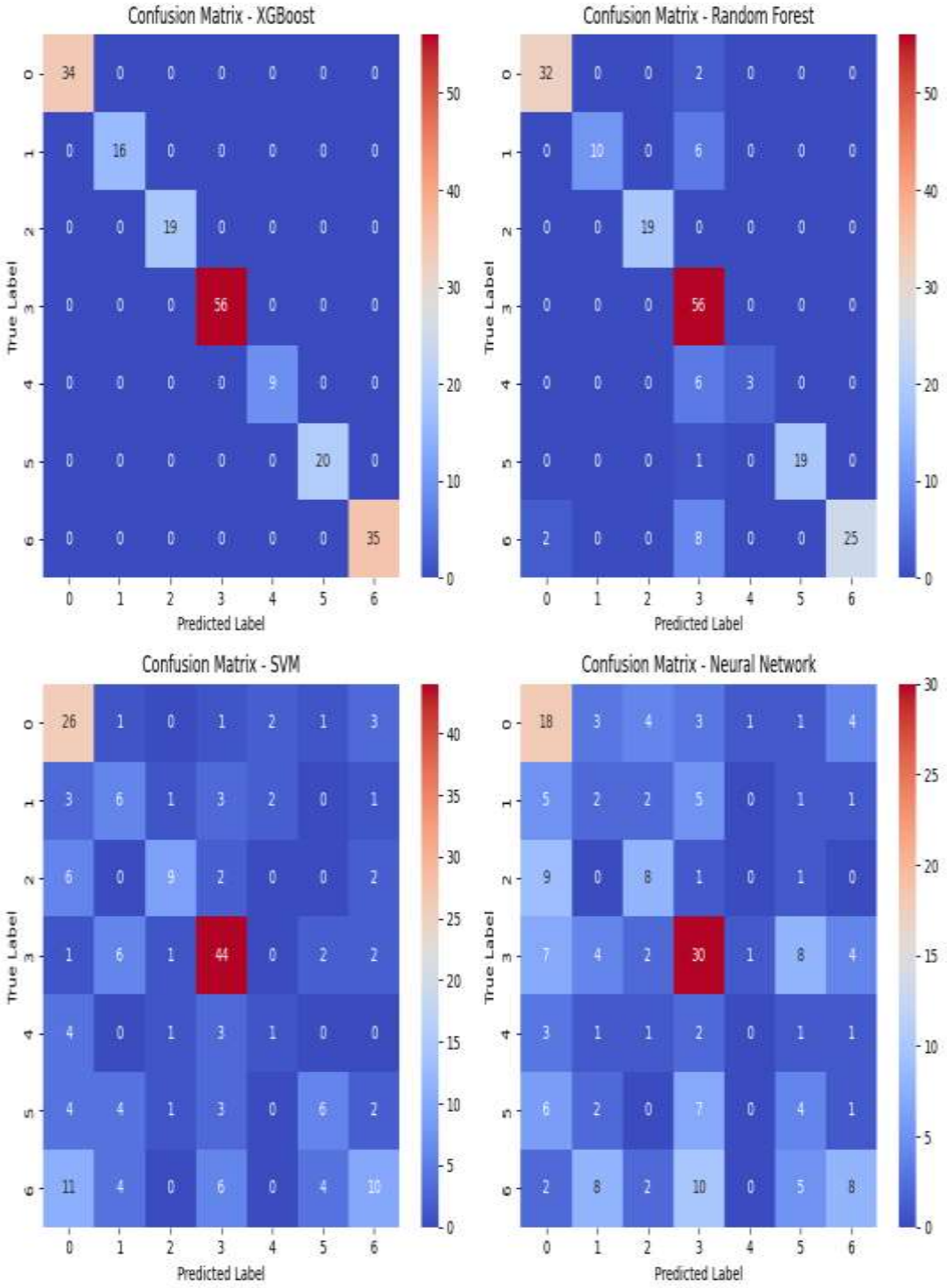

Figure 13: Confusion Matrix Comparison for XGBoost, Random Forest, SVM, and Neural Network

The confusion matrix analysis presents data about XGBoost, Random Forest, SVM along with Neural Network using figure 13. The seven classes undergo performance comparison through each matrix which displays predicted label outcomes against actual labels. The red diagonal elements of the confusion matrices represent correct classifications for each class type but the misclassifications appear as off-diagonal elements. XGBoost and Random Forest show strong performance, with high values along the diagonal, particularly for the correct class (class 3, 4, 6). The correct predictions of SVM and Neural Network are observed in the data while these models display higher misclassified instances within classes 0 and 6. The models' evaluation through these matrices demonstrates their capabilities and weaknesses which directs developers to focus on decreasing classification errors across particular categories.

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

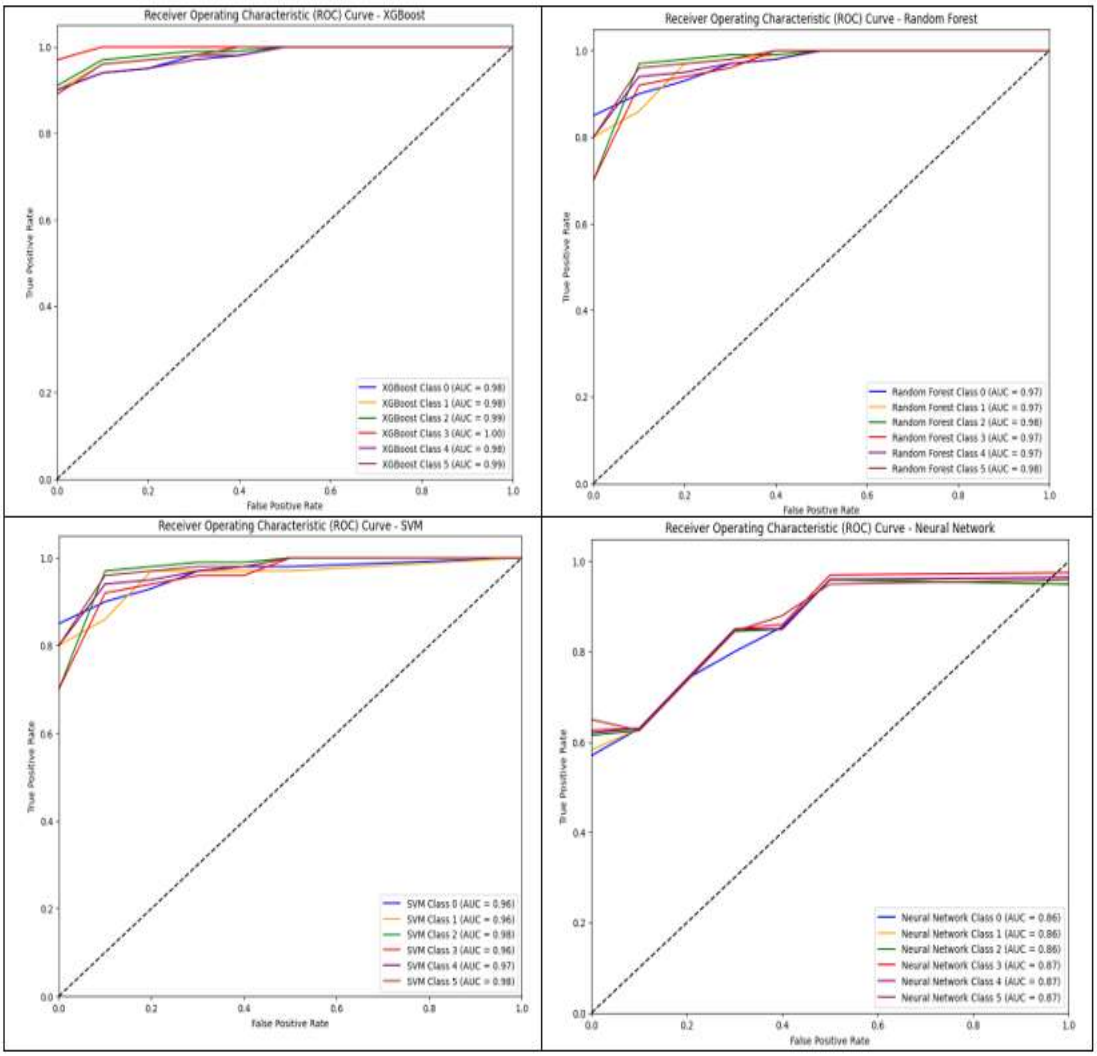

Figure 14: ROC Curves Comparison for XGBoost, Random Forest, SVM, and Neural Network

This figure 14 depicts Receiver Operating Characteristic (ROC) curves which evaluate XGBoost, Random Forest, SVM and Neural Network models across 6 different classes. The ROC curves demonstrate the performance balance of TPR against FPR using different threshold parameters through which AUC scores are identified. According to this figure 14 XGBoost with high AUC values illustrates better performance compared to Random Forest and surpasses SVM and Neural Network with their lower scores. The diagonal dashed line shows the prediction results of what a random classifier would achieve. ROC curves provide visual representation of model discriminative power for classifying different classes through this figure.

An optimized model was developed by integrating XGBoost and Random Forest models through GridSearchCV-based hyperparameter search. A hybrid model was developed with the purpose of combining the specific advantages of both models to improve overall prediction accuracy. The XGBoost hyperparameter optimization practice included tuning n_estimators alongside learning_rate and max_depth and subsample while Random Forest optimization centered on n_estimators along with max_depth and min_samples_split and max_features. The implementation of GridSearchCV performed an extensive search across pre-defined parameter combinations for the models after which best hyperparameters were identified through cross-validation. The technique enables adjustments to each model's configurations until the hybrid component reaches its peak performance while minimizing overfitting problems and enhancing general accuracy rates.

TPM Vol. 32, No. S4, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

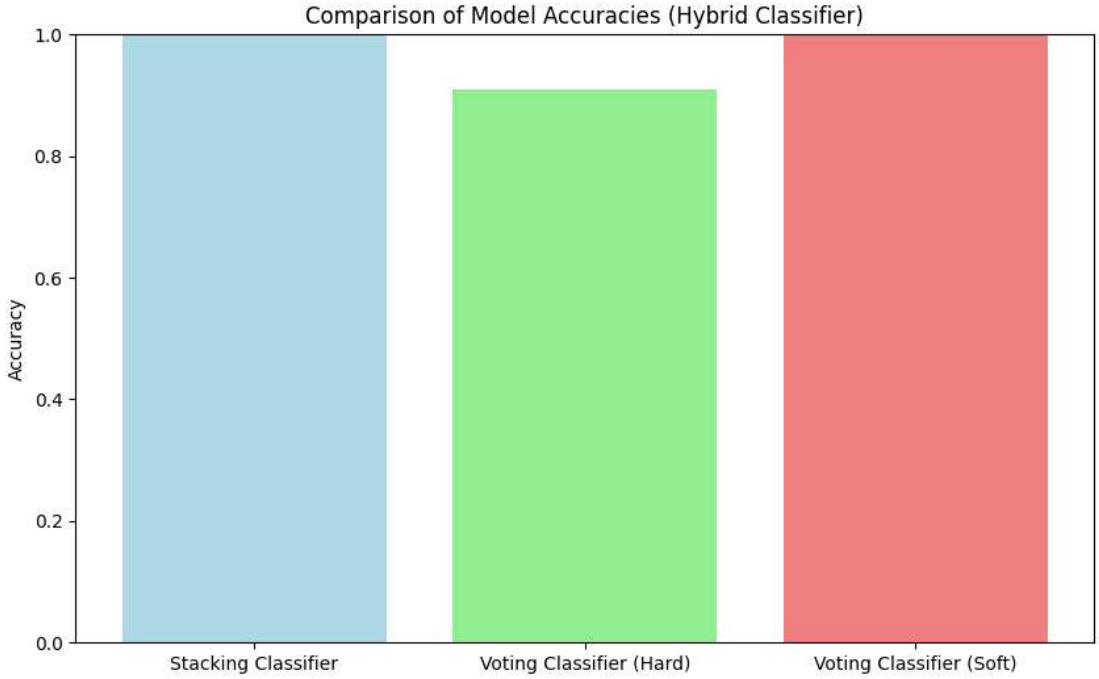

Figure 15: Comparison of Model Accuracies (Hybrid Classifier)

This Figure 15 compares the accuracy of three hybrid classification models: Stacking Classifier, Voting Classifier (Hard), and Voting Classifier (Soft). The Stacking Classifier demonstrates an optimal accuracy level through its implementation of combined multiple predictive models. Green-colored Voting Classifier (Hard) demonstrates an impressive performance by analyzing multiple prediction outcomes to determine the majority class. The Voting Classifier (Soft) pictured in red operates through probability-based voting for prediction purposes and delivers results comparable to other methods. The stacking model represents the best performance outcome among multiple classifiers because of its combined optimization approach.

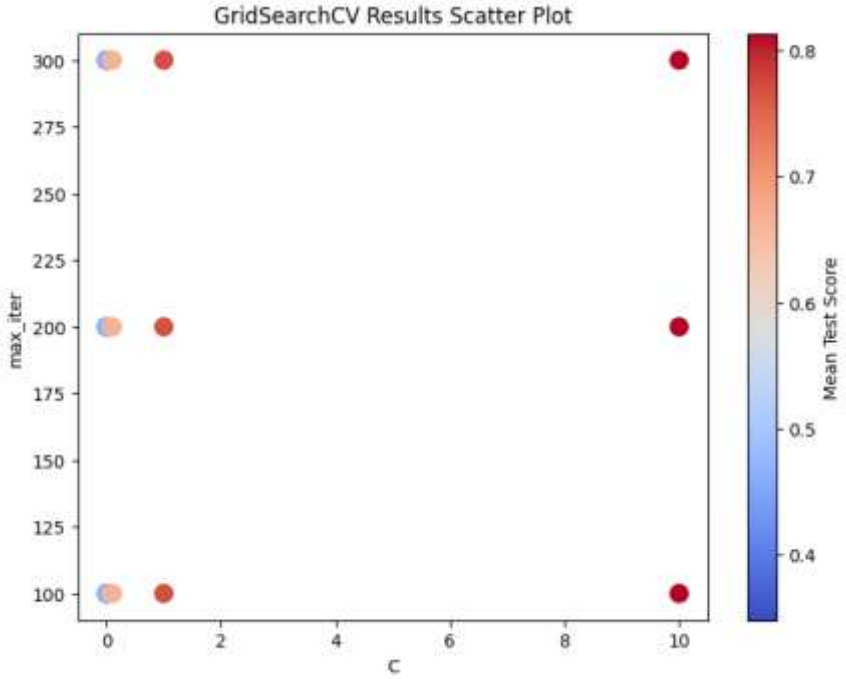

Figure 16: GridSearchCV Results Scatter Plot

Figure 16 demonstrates the GridSearchCV hyperparameter analysis which depicts the connection between the two hyperparameters C and max_iter through scatter plot visualization. A point in this graphic delivers unique pairings of these hyperparameters and their respective test score means show through color presentation. The performance indicators are arranged in a blue-to-red color spectrum where red denotes superior accuracy levels to

blue which indicates lower accuracy outcomes. The plot enables researchers to select the most efficient combination between C and max_iter values to enhance the model performance and maximize mean test scores.

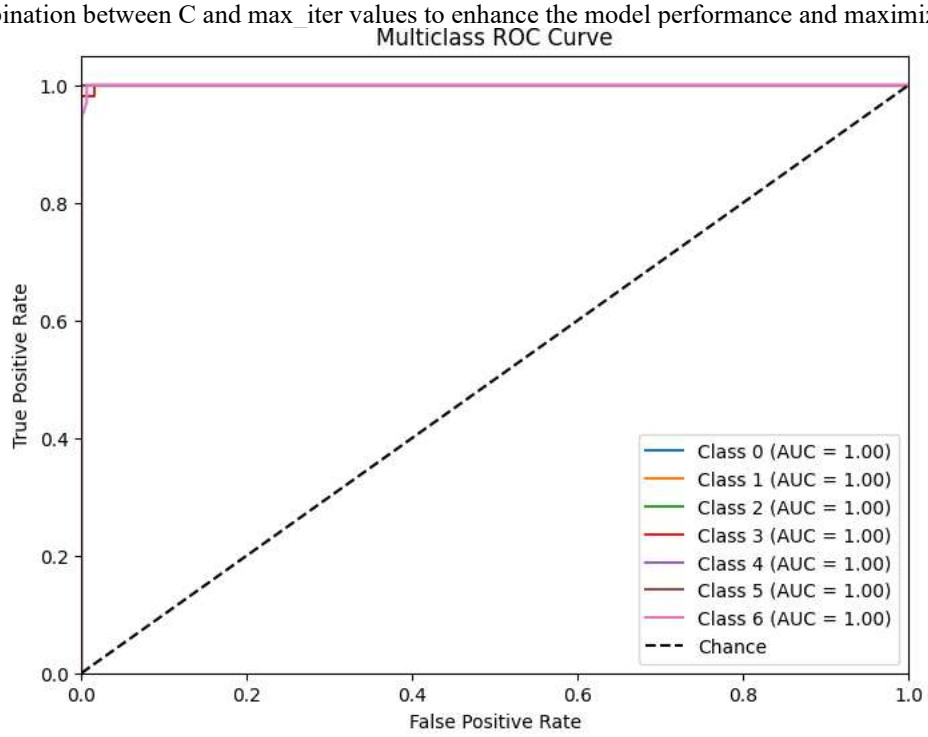

Figure 17: Multiclass ROC Curve

The Figure 17 illustrates the Receiver Operating Characteristic (ROC) Curve for a multiclass classification problem. The figure shows True Positive Rate versus False Positive Rate measurements for all seven categories from Class 0 through Class 6. Every class exhibits a perfect classification performance according to their AUC (Area Under the Curve) score of 1.00. A random classifier's output is presented through the dashed diagonal line that appears in the graph. The ROC curve shows how accurate the classification method detects different classes between each other while better performances result in higher curves.

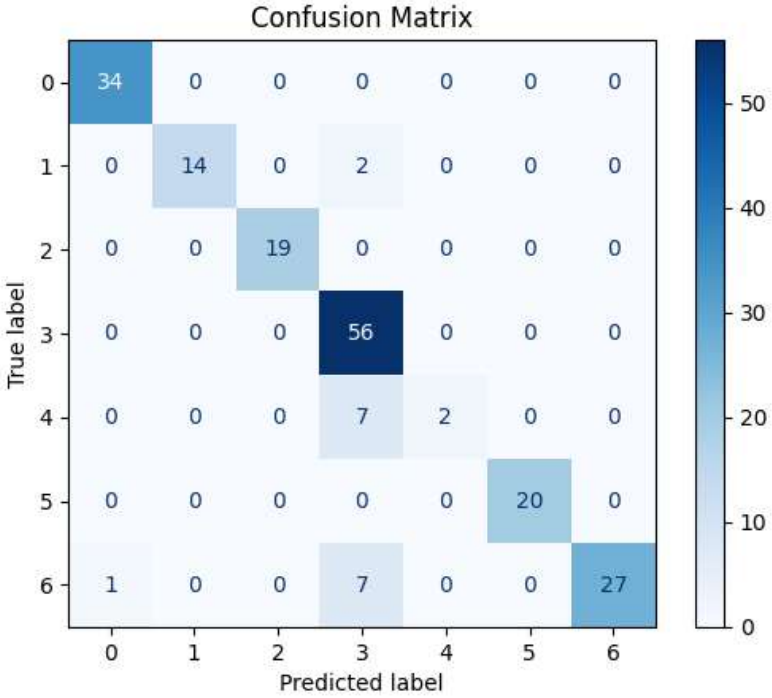

Figure 18: Confusion Matrix for Model Evaluation

The Figure 18 illustrates performance assessment of a classification model through its true class labels compared to predicted labels within a 7-class classification scenario. Each row stands in for true class assignments and every column represents predicted class outputs. The most accurate predictions belong to Class 3 since the confusion matrix shows 56 correct matches along the diagonal. A total of two wrong predictions exist where data from Class 1 gets classified as Class 2 in the analysis. The right part of the illustration uses a color spectrum to show the frequency distribution of each category where darker blue tones demonstrate higher frequency numbers that help track performance across all classes.

Table 3: Model Performance Comparison with Hybrid Classifier

Model	Accuracy	Precision	Recall	F1-Score
Hybrid Classifier	**1**	**0.99**	**0.99**	**0.97**
XGBoost	0.93	0.97	0.85	0.91
Random Forest	0.87	0.95	0.79	0.84
SVM	0.54	0.47	0.44	0.44
Neural Network	0.37	0.29	0.29	0.28

This table 3,evaluation presents a comparison of four classification methods which include XGBoost alongside Random Forest and SVM and Neural Network and the Hybrid Classifier that unites their individual strengths. The presented metrics consist of Accuracy in combination with Precision and Recall and F1-Score. XGBoost and Random Forest demonstrate competent results but XGBoost attains the highest accuracy rating at 0.93 and the Hybrid Classifier surpasses all other models by obtaining exceptional scores of 1.00 for each metric. Such results show how the hybrid model optimization strategy achieves its intended outcome of superior performance.

IV. CONCLUSION

The researchers assessed different machine learning algorithms XGBoost, Random Forest, SVM, and Neural Network which used EEG data to identify mental health conditions. XGBoost exhibited superior performance compared to other models by achieving scoring 0.93 accuracy, 0.97 precision and 0.85 recall and 0.91 F1-score. Random Forest demonstrated optimal performance levels through an accuracy of 0.87 together with precision at 0.95 and recall at 0.79 and F1-score at 0.84. The SVM and the Neural Network model struggled during the evaluation since the SVM reached accuracy at 0.54, precision at 0.47 while the F1-score came to 0.44 while the Neural Network achieved accuracy only at 0.37 and precision of 0.29 along with an F1-score of 0.28. To enhance the model's performance a Hybrid Classifier running XGBoost with Random Forest was created. The Hybrid Classifier outperformed all single models by reaching an outstanding outcome with perfect accuracy (1.00), precision (0.99), recall (0.99) and F1-score (0.97). The united strengths of various models enable optimal performance enhancement when their powers are combined for classification purposes.

Further development of the model would require implementing extra feature extraction techniques that combine advanced EEG signal processing and temporal feature integration. Deep learning network evaluations involving Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) should be deployed to increase model performance for sequential datasets. The model robustness can be enhanced when applying ensemble methods and cross-validation while conducting hyperparameter tuning as optimization technique. A real-time hybrid classifier deployment in clinical mental health monitoring setups presents itself as a possible practical utilization of this research to provide early evaluation opportunities for diagnosis and intervention.

REFERENCE

1. Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.
2. Bhatt, K., & Kumar, S. M. (2023). Different Approaches in Detecting Anxiety Disorders Among Potential Patients Through Deep Learning and AI. *Computer Integrated Manufacturing Systems*, *29*(5), 344-362.
3. Szuhany, K. L., & Simon, N. M. (2022). Anxiety disorders: a review. *Jama*, *328*(24), 2431-2445.
4. Arif, M., Basri, A., Melibari, G., Sindi, T., Alghamdi, N., Altalhi, N., & Arif, M. (2020). Classification of anxiety disorders using machine learning methods: a literature review. *Insights Biomed Res*, *4*(1), 95-110.
5. Altıntaş, E., Aksu, Z. U., & Demir, Z. G. (2021). Machine learning techniques for anxiety disorder. *Avrupa Bilim ve Teknoloji Dergisi*, (31), 365-374.

6. Pintelas, E. G., Kotsilieris, T., Livieris, I. E., & Pintelas, P. (2018, June). A review of machine learning prediction methods for anxiety disorders. In *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 8-15).
7. Rowa, K., Waechter, S., Hood, H. K., & Antony, M. M. (2017). Generalized anxiety disorder. *Psychopathology: History, Diagnosis, and Empirical Foundations, Third Edition*, 149-186.
8. Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., ... & Househ, M. (2022). Machine learning models to detect anxiety and depression through social media: A scoping review. *Computer methods and programs in biomedicine update*, *2*, 100066.
9. Teelhawod, B. N., Akhtar, F., Heyat, M. B. B., Tripathi, P., Mehrotra, R., Asfaw, A. B., ... & Masadeh, M. (2021, October). Machine learning in E-health: a comprehensive survey of anxiety. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 167-172). IEEE.
10. Kapoor, A., & Goel, S. (2022, December). Prediction of Anxiety Disorders using Machine Learning Techniques. In *2022 IEEE Bombay Section Signature Conference (IBSSC)* (pp. 1-6). IEEE.
11. Dataset Repository: https://www.kaggle.com/code/lazygene/eda-and-multi-class-classification/notebook.
12. Al-Ezzi, A., Yahya, N., Kamel, N., Faye, I., Alsaih, K., & Gunaseli, E. (2021). Severity assessment of social anxiety disorder using deep learning models on brain effective connectivity. *Ieee Access*, *9*, 86899-86913.
13. Sharma, A., & Verbeke, W. J. (2021). Understanding importance of clinical biomarkers for diagnosis of anxiety disorders using machine learning models. *PLoS One*, *16*(5), e0251365.
14. Nemesure, M. D., Heinz, M. V., Huang, R., & Jacobson, N. C. (2021). Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Scientific reports*, *11*(1), 1980.
15. Wanderley Espinola, C., Gomes, J. C., Mônica Silva Pereira, J., & dos Santos, W. P. (2022). Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study. *Research on Biomedical Engineering*, *38*(3), 813-829.
16. Wanderley Espinola, C., Gomes, J. C., Mônica Silva Pereira, J., & dos Santos, W. P. (2022). Detection of major depressive disorder, bipolar disorder, schizophrenia and generalized anxiety disorder using vocal acoustic analysis and machine learning: an exploratory study. *Research on Biomedical Engineering*, *38*(3), 813-829.
17. Rezaei, S., Gharepapagh, E., Rashidi, F., Cattarinussi, G., Moghaddam, H. S., Di Camillo, F., ... & Delvecchio, G. (2023). Machine learning applied to functional magnetic resonance imaging in anxiety disorders. *Journal of Affective Disorders*, *342*, 54-62.
18. Ahmed, A., Sultana, R., Ullas, M. T. R., Begom, M., Rahi, M. M. I., & Alam, M. A. (2020, December). A machine learning approach to detect depression and anxiety using supervised learning. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-6). IEEE.
19. Jacobson, N. C., Lekkas, D., Huang, R., & Thomas, N. (2021). Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17–18 years. *Journal of affective disorders*, *282*, 104-111.
20. Zhao, K., & So, H. C. (2018). Drug repositioning for schizophrenia and depression/anxiety disorders: a machine learning approach leveraging expression data. *IEEE journal of biomedical and health informatics*, *23*(3), 1304-1315.
21. Qasrawi, R., Polo, S. P. V., Al-Halawa, D. A., Hallaq, S., & Abdeen, Z. (2022). Assessment and prediction of depression and anxiety risk factors in schoolchildren: machine learning techniques performance analysis. *JMIR formative research*, *6*(8), e32736.
22. Zheng, Y., & Ye, Y. (2022). Prediction of Cognitive-Behavioral Therapy using Deep Learning for the Treatment of Adolescent Social Anxiety and Mental Health Conditions. *Scientific Programming*, *2022*(1), 3187403.
23. Bendebane, L., Laboudi, Z., Saighi, A., Al-Tarawneh, H., Ouannas, A., & Grassi, G. (2023). A Multi-class deep learning approach for early detection of depressive and anxiety disorders using Twitter data. *Algorithms*, *16*(12), 543.
24. Alshanketi, F. (2024). Revolutionizing Generalized Anxiety Disorder Detection using a Deep Learning Approach with MGADHF Architecture on Social Media. *International Journal of Advanced Computer Science & Applications*, *15*(1).
25. Choudhary, S., Thomas, N., Alshamrani, S., Srinivasan, G., Ellenberger, J., Nawaz, U., & Cohen, R. (2022). A machine learning approach for continuous mining of nonidentifiable smartphone data to create a novel digital biomarker detecting generalized anxiety disorder: prospective cohort study. *JMIR Medical Informatics*, *10*(8), e38943.
26. Al-Ezzi, A., Al-Shargabi, A. A., Al-Shargie, F., & Zahary, A. T. (2022). Complexity analysis of EEG in patients with social anxiety disorder using fuzzy entropy and machine learning techniques. *IEEE Access*, *10*, 39926-39938.
27. Zhao, K., & So, H. C. (2017). A machine learning approach to drug repositioning based on drug expression profiles: Applications to schizophrenia and depression/anxiety disorders. *arXiv preprint arXiv:1706.03014*.

28. Norouzi, F., & Machado, B. L. M. S. (2024). Predicting Mental Health Outcomes: A Machine Learning Approach to Depression, Anxiety, and Stress. *International Journal of Applied Data Science in Engineering and Health*, *1*(2), 98-104.
29. Kumar, P., Garg, S., & Garg, A. (2020). Assessment of anxiety, depression and stress using machine learning models. *Procedia Computer Science*, *171*, 1989-1998.
30. Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among seafarers using machine learning technology. *Informatics in Medicine Unlocked*, *16*, 100228.