GMR

The Original

EXPLAINABLE EEG-BASED SCHIZOPHRENIA DETECTION USING ANOVA FEATURE SELECTION AND OPTIMIZED EXTRA TREES WITH STRATIFIED CROSS-VALIDATION

Rabi Narayan Panda*[1] , Anubhav Kumar[2]

[1*]Research Scholar, HRIT University, Ghaziabad, UP, India, rabipanda@rediffmail.com
[2]Professor, HRIT University, Ghaziabad, UP, India, ds.dranubhavkumar@gmail.com

ABSTRACT
Schizophrenia is a chronic psychiatric disorder where the perception, cognition and behaviors are disturbed, which requires a proper and timely diagnosis to give an effective clinical intervention. EEG has been a recently viable non-invasive technique for identifying neurophysiological markers related to schizophrenia. EEG-derived features, however, are usually very high dimensional and include redundant features and reduce classification performance. We propose a novel explainable machine learning approach for automated schizophrenia detection using EEG features. A publicly available EEG Psychiatric Disorders dataset was used to extract 1,144 features from both the demographic information and the EEG. To minimize feature redundancy, Analysis of Variance (ANOVA) based SelectKBest feature-selection was used to select 50 most discriminative-features. The hyperparameters of the Extra Trees classifier were then optimized by grid search and the selected features were classified using this optimized Extra Trees classifier. The model performance was assessed by an independent hold-out test set and Stratified-5-Fold Cross-Validation. The optimized model attained 95.35%, 95.29%, 95.83%, and 96.27% for hold-out test accuracy, balanced accuracy, F1-score and ROC-AUC, respectively. Moreover, the Stratified-5-Fold Cross-Validation provided a mean accuracy of 85.37 ± 5.01% and a mean ROC-AUC of 93.14 ± 2.51%, showing good generalization across the different data folds. Additionally, SHapley Additive exPlanations (SHAP) analysis revealed the most important demographic and EEG biomarkers for classification, improving the model interpretability. The results show that integrating ANOVA-based feature-selection with an optimized Extra Trees classifier, significantly improves classification performance while providing clinically interpretable predictions.

KEYWORDS: Schizophrenia, Electroencephalography, EEG, Explainable Artificial Intelligence, Feature-selection, ANOVA, Extra Trees, SHAP, Machine Learning.

1. INTRODUCTION

Schizophrenia is a chronic psychiatric illness that involves the alteration of perception, cognition, emotion and behavior, which has adverse impact on the quality of the life of a person. The early diagnosis is crucial for early intervention, effective treatment planning and better clinical outcomes in the long-term. The diagnosis is essentially subjective and dependent on the clinical expertise of the specialist and is carried out mainly during psychiatric interviews, assessment of behavior and application of a psychiatric diagnostic criterion. This has led to the development of objective computer-aided diagnostic systems that can aid clinical decision-making by performing quantitative analysis on neurophysiological signals, which has been an increasing interest. Electroencephalography (EEG) is cheap to acquire, portable, has high temporal resolution and measures neuronal activity directly. The attributes have been exploited in numerous ways for the diagnosis of neurological diseases and for the evaluation of cognitive state in humans as well as for the use of brain-computer interfaces (BCIs) [1] and [2]. The availability of public repositories like PhysioNet [5] and open-source analysis tools like MNE [3] have also facilitated research by ensuring that data is available in a standardized format and signal processing tools are shared to enable consistent analysis of EEG. Furthermore, advanced pre-processing and spatial filtering approaches have shown good results in improving the quality of EEG signals and discriminating neural information [4]. Recent developments in machine learning have greatly enhanced the automatic analysis of EEG signals. The machine learning classifiers that have demonstrated impressive results in many EEG classification problems [6, 13]. More recently, deep learning has been introduced to revolutionize EEG analysis, allowing to automatically extract hierarchical representations directly from EEG recordings. Craik et al. [7] and Roy et al. [8] showed that convolutional neural networks (CNNs) consistently beat traditional handcrafted feature-based methods for many EEG applications. To efficiently capture spatial-temporal EEG characteristics in a relatively low computational complexity, representative architectures like Deep ConvNet [9] and EEGNet [10] have been generally accepted. Moreover, the Transformer-based architectures have recently gained significant interest for the modelling of long-range temporal dependencies in EEG data, achieving promising results in different EEG decoding tasks including Xie et al. [15] and Song et al. [16] for the reconstruction of long-term memory, Cui et al. [24] for the seizure detection task, and Altaheri et al. [25] for the task of hand motion prediction from brain activity. The features derived from EEG data are usually hundreds or thousands of parameters comprising inter alia spectral power, coherence and connectivity, as well as demographic variables. These high dimensional feature spaces often have redundant and weakly informative variables, which result in high computation cost, low generalization ability and may cause overfitting for the classifiers. Hence, proper Feature-Selection is a key challenge to discover most informative EEG biomarkers to enhance classification performance and computational efficiency [6, 13]. Furthermore, intra-subject variability is still a big challenge for reliable EEG classification. To achieve better generalization across various recording sessions and participants, transfer learning, domain adaptation, and subject independent learning strategies have been proposed [11], [12], [21]-[23]. Although complex classifiers frequently give a very accurate prediction, clinicians need to explain how the decision is made to learn about the logic or rationale behind the automated decision-making and find clinically meaningful biomarkers. In medical decision-support systems, the use of Explainable Artificial Intelligence (XAI) techniques is growing in significance. More recently, it has been shown that interpretable deep learning can be applied to EEG analysis [19] and that the need for XAI is increasing in the context of trustworthy AI systems for healthcare applications [17, 18, 20]. In an attempt to address these challenges, this study aims to present an interpretable machine learning solution for EEG classification of schizophrenia, focusing on feature selection, model optimization and robust validation methodology. At first, 1,144 features are taken into account, consisting of demographic and EEG-derived features. The dimensionality of EEG features is then further reduced by using an ANOVA-based SelectKBest approach to extract the discriminative features, which are the most informative features among the candidate features, to select the Top-50 features. Next, the selected features are classified by an optimized Extra Trees classifier with optimized hyperparameters by systematic Grid Search optimization. The proposed framework is assessed on an independent hold-out test set as well as Stratified-5-Fold Cross-Validation, to guarantee consistent performance estimation.

The findings of this manuscript are summarized as follows:

- A feature-selection strategy based on the ANOVA is presented to achieve even though the number of the original EEG-derived and demographic features is large (1,144), the Top most fifty discriminative-features are selected to achieve better computational efficiency with minimal compromise on feature dimensionality.
- To enhance the classification performance of EEGs, an optimized Extra Trees (ET) classifier is developed by optimizing its hyper-parameters in a systematic way.
- Explainability is built into SHAP to uncover EEG biomarkers that influence the classifier's predictions and to give interpretable insights into the classifier's predictions.

The proposed framework is thoroughly tested on an independent holdout test set and Stratified-5-Fold Cross-Validation (5FCV), generating a strong evaluation of the classification performance and the model generalization.
The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the proposed methodology. Section 4 describes the experimental results, and Section 5 concludes the paper.

2. RELATED WORK

Electroencephalography (EEG) is popular neurophysiological methods used to explore brain activity, due to its high temporal-resolution, low acquisition cost-and-real-time capture of brain dynamics. The properties of these have made EEG a useful component in brain-computer interface (BCI) systems, diagnosis of neurological disorders, monitoring of cognitive state, and healthcare applications. The advent of robust machine learning models, effective feature-selection strategies, and explainable artificial intelligence (XAI) models is gaining more significance as EEG data is inherently high dimensional, noisy, and nonlinear. In this section, recent advances in these research areas are reviewed.

2.1 EEG-Based Machine Learning

The initial research work on EEG analysis was dedicated to the creation of effective BCIs that can convert brain signals into control commands. The basic principles of BCI technology were laid down by Wolpaw et al. [1] who showed that EEG signals can be used for communication and device control. Later, Nicolas-Alonso and Gomez-Gil [2] gave a full review of BCI systems, covering the acquisition and preprocessing of the signals, feature extraction, classification methods and practical applications. The advent of standardised EEG processing tools and public datasets has greatly propelled the research of machine learning based EEG. PhysioNet created a freely available archive of physiological signals and analysis tools, which allows for reproducible biomedical research [5]. Likewise, Gramfort et al. [3] designed the MNE software package, which is one of the most popular open-source software packages for EEG and MEG signal pre-processing, visualization, and analysis. Additionally, Blankertz et al. [4] showed that suitable spatial filtering methods could significantly boost the accuracy of EEG classification by boosting the signal to noise ratio and by highlighting discriminative neural patterns. Classification problems of EEG signals have been extensively handled by conventional machine learning algorithms with the growing computational power. Lotte et al. [6] gave an extensive review of the EEG classification methods, advantages and disadvantages of linear classifiers, support vector machines, ensemble methods and neural networks. More recently, Shoorangiz et al. [13] reviewed theoretical background and application of machine learning algorithms in the field of EEG analysis, highlighting the necessity of good feature engineering, dimensionality reduction and the selection of classifiers for reliable biomedical decision support.

2.2 Deep-Learning for EEG-Classification

With the development of deep-learning, compared to traditional handcrafted features, EEG analysis of signals has radically changed because the deep learning techniques enable automatic feature learning directly from the raw or least-processed EEG signals. Craik et al. [7] found that convolutional neural networks (CNNs) consistently outperform traditional handcrafted method through various applications of EEG. Likewise, Roy et al. [8] performed a systematic review of the deep learning architectures for EEG-analysis and found that deep learning models significantly outperformed traditional surface level models with respect to classification accuracy and feature dependence. In EEG decoding, there are some CNN architectures specially developed. In contrast, deep convolutional neural networks (CNNs) have been introduced by Schirrmeister et al. [9] that are able to learn spatial and temporal EEG representations and also to visualize learned neural patterns. Lawhern et al. [10] , which preserved the competitive performance of CNN on several BCI paradigms while simultaneously reducing the computational complexity. The studies showed that compact models of deep learning can effectively extract the EEG properties without much feature engineering. Recent years, however, have seen the rise of Transformer-based models, which offer the potential to capture long-range temporal information, as a promising alternative for EEG analysis. Xie et al. [15] designed a spatial-temporal fusion Transformer network for a raw EEG classification task, while Song et al. [16] proposed a novel EEG Conformer model comprising of convolutional layers and Transformer encoders to enhance the performance of EEG decoding. The recent trend of Transformer models has spurred the creation of brain-inspired foundation models for EEG representation learning, such as those proposed by Cui et al. [24] and the success of temporal convolutional Transformers for EEG decoding problems by Altaheri et al. [25].

2.3 Generalization and Transfer Learning in EEG Analysis

The variability of subjects, recordings, and protocols are major hitches in EEG-based machine-learning, and sometimes restrict the generalizability of the models. To tackle this problem, some of the researchers have explored transfer learning and domain adaptation techniques. Jayaram et al. [21] surveyed transfer learning techniques for BCI systems and showed that knowledge transfer between subjects can significantly reduce the amount of calibration required. In addition to that, Wu et al. [22] conducted a comprehensive overview of the progress in transfer learning for EEG-based BCI, and identified domain adaptation, feature alignment, and distribution matching techniques as effective approaches to enhance cross-subject performance. Many efforts have been undertaken for deep domain adaptation approaches have also been studied in depth. To address this issue, Hang et al. [11] designed a deep domain adaptation network (DDAN) for cross-subject EEG classification, which achieved significant inter-subject variability reduction. Likewise, Kwon et al. [23] created subject independent CNN models that would perform well on a variety of participants regardless of the specific subject with minimal subject dependent calibration. All of these studies are evidence of the ongoing research effort in making models more robust and generalizable for EEG, which encourages the use of solid evaluation methods like Stratified-5-Fold Cross-Validation [12].

2.4 Explainable Artificial Intelligence for EEG-Based Decision Support

Deep learning and ensemble learning models have proven to have a very high predictive performance, but the black-box nature of these models makes them less popular for healthcare applications. As a result, XAI has become a crucial element in trustworthy and reliable medical decision-making support systems. The Local-Interpretable-Model-Agnostic-Explanations (LIME) framework was introduced by Ribeiro et al. [18] to give local explanations that are independent of the model used for prediction. Later, Lundberg and Lee [17] came up with a unified game-theoretic framework, called SHAP, that offers both feature importance and local feature importance and is, at the same time, theoretically sound. In the field of EEG analysis, Sturm et al. [19] showed that an interpretable deep network can uncover physiologically relevant EEG patterns without compromising classification accuracy. Recently, Ali et al. [20] gave an extensive overview on XAI methods, highlighting their significance in enhancing transparency, clinicians trust, fairness, and accountability in healthcare-oriented AI systems. The findings from these studies suggest that explainability is becoming ever more critical for clinical deployment of machine learning models.

2.5 Research Gap

Many EEG classification studies are performed in high dimensional feature spaces, which add complexity to the problem, and frequently contain redundant or weakly-informative features. Secondly, while powerful deep learning architectures have demonstrated remarkable results, they often necessitate extensive training sets and high computing power, which may make them impractical in healthcare environments where resources are limited. Third, only a few studies systematically study the joint effect of statistical feature-selection, optimization of ensemble learning, explainability and robust validation in a holistic approach. In order to overcome these limitations, this paper introduces an explainable framework for EEG classification by combining the process of feature-selection using the ANOVA method to reduce the number of the initial 1144 EEG-derived features to the Top-50 discriminative features with an optimized Extra Trees classifier. Additionally, the contribution of the individual EEG biomarkers is interpreted using SHAP, and the robustness and generalization capability is assessed by Stratified-5-Fold Cross-Validation.

3. PROPOSED METHODOLOGY

The illustrated framework is made up of six stages: (i) data acquisition and preprocessing, (ii) feature normalization, (iii) feature-selection through ANOVA, (iv) optimized Extra Trees classification, (v) explainability analysis via SHAP, and (vi) performance evaluation through hold-out testing and Stratified-5-Fold Cross-Validation. The proposed framework includes the complete workflow. (Figure 1).

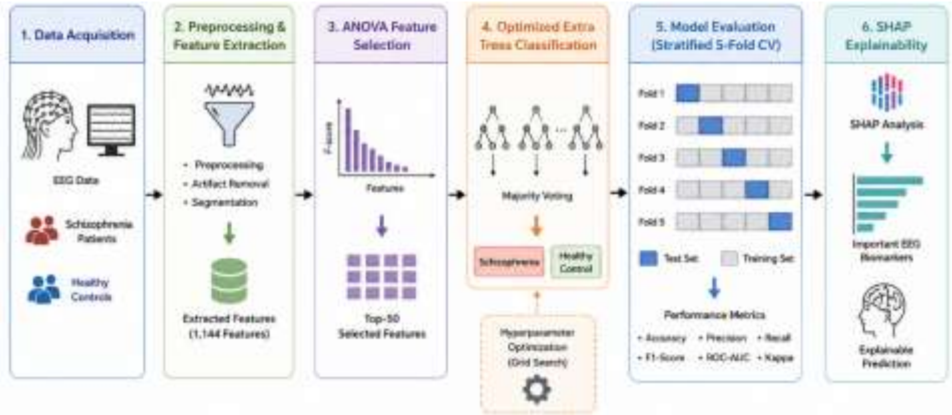

Figure 1. Proposed explainability in schizophrenia detection using ANOVA feature-selection, optimized ETC and SHAP explainability and Stratified-5-Fold Cross-Validation.

The overall process of the proposed explainable EEG classification framework is shown in Figure 1. The first stage is the collection of demographic and neuro physiological features from EEG, which then undergo data preprocessing involving data cleaning, missing value imputation, label encoding and feature standardization. Then, ANOVA-based SelectKBest is used to rank the extracted-features and select the Top-50 most discriminative features out of the original 1,144 predictor variables that reduce the dimensionality of the features and remove redundant information. These selected features are then classified with an optimized Extra Trees classifier, with hyper parameters optimized with Grid Search. The robustness and generalization capability of the proposed framework is evaluated using Stratified-5-Fold Cross-Validation and keeping the class distribution same in each fold.

3.1 Dataset Description

Experimental data was employed from the publicly available EEG Psychiatric Disorders Dataset, and included quantitative EEG features derived from resting-state EEG recordings of patients with psychiatric disorders and healthy controls. The original data set includes demographic data and quantitative biomarkers derived from the EEG data across multiple frequency bands and in terms of functional connectivity. In the current study we only considered the Healthy Controls and Schizophrenia subjects, and therefore a binary classification problem. The processed data set consists of 212 subjects: 95 healthy controls and 117 schizophrenia patients.

The predictor variables consist of 1,144 features, including:

- demographic characteristics,
- spectral power measurements,
- coherence features,
- EEG band-power descriptors.

Table 1. Dataset characteristics used in this study.

Attribute	Value
Total Subjects	212
Healthy Controls	95
Schizophrenia	117
Predictor Features	1,144
Target Classes	2

3.2 Data Preprocessing

Before developing a model, several pre-processing steps were taken to enhance data quality and ensure the data was suitable for machine learning algorithms. Initially, irrelevant variables that were not required for schizophrenia classification were removed. Only demographic parameters and EEG-derived quantitative parameters were kept. There were missing values on two numerical variables education and IQ. Median imputation was employed to fill in the missing values, maintaining the median property of the data and minimizing the effects of outliers. Label encoding was used to convert categorical features like the sex attribute and diagnostic labels into numeric values. Then all predictor variables were normalized in order to get rid of the differences in the scale of the numerical values of EEG features.

Feature standardization was performed using z-score normalization, defined as:

$$z = \frac{x - \mu}{\sigma}$$

where $x$denotes the original feature value, μrepresents the mean of the feature, σdenotes the standard deviation of the feature, and $z$is the standardized feature value. After preprocessing, the dataset contained standardized-numerical-features suitable for feature-selection and classifier training.

3.3 ANOVA-Based Feature-selection

The EEG data file has many quantitative descriptors. Most of these features are informative but redundant or poorly correlated with the schizophrenia diagnosis. The curse of dimensionality may result in a decrease in model generalization and increase the computational complexity of the feature space if the dimensionality is increased. The SelectKBest algorithm in conjunction with Analysis of Variance (ANOVA) was used to identify the most discriminative EEG biomarkers. For each feature, F-statistic was computed to evaluate its discriminative ability by measuring the fraction of between-group variance to within-group variance:

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}}$$

where MS_{Between} denotes the mean square between diagnostic groups, reflecting the variability among group means, and MS_{Within} represents the mean-square within groups, indicating the variability-of-samples within each diagnostic category. Features with higher F-statistics exhibit stronger discriminative capability. All 1,144 predictor variables were ranked according to their ANOVA scores, and only the Top-50 features were retained for classifier training. Consequently, the feature dimensionality was reduced by approximately 95.6%, substantially decreasing computational complexity while preserving the most informative EEG biomarkers.

Figure 2. ANOVA-based feature-selection process reducing the original feature space from 1,144 features to the Top-50 discriminative EEG features.

The entire set of 1,144 demographic and EEG-derived features were initially used as the input feature space. An ANOVA F-test was then performed to assess the discriminative power of each feature on each class. All the features were ranked in order of statistical significance using the computed F-scores. Next, the top-50 features with most significant scores were chosen from the original features space using SelectKBest algorithm which resulted in 95.63% reduction in feature space and preserved the best and most informative features. The resulting reduced-feature-set was then used for training the Extra Trees classifier that was optimized. This substantial reduction in feature dimensionality decreased computational complexity while improving classification performance by eliminating redundant and irrelevant features.

Table 2. Top-20 EEG Features Selected by ANOVA SelectKBest

Rank	Feature	Feature Category
1	IQ	Demographic
2	sex	Demographic
3	age	Demographic
4	AB.A.delta.r.O1	Absolute Band Power (Delta)
5	AB.D.beta.g.F8	Absolute Band Power (Beta)
6	AB.E.highbeta.b.FP2	Absolute Band Power (High Beta)
7	AB.A.delta.p.P4	Absolute Band Power (Delta)
8	AB.D.beta.q.T6	Absolute Band Power (Beta)
9	COH.D.beta.a.FP1.e.Fz	Coherence (Beta)
10	AB.A.delta.q.T6	Absolute Band Power (Delta)
11	AB.A.delta.k.C4	Absolute Band Power (Delta)
12	AB.D.beta.m.T5	Absolute Band Power (Beta)
13	COH.A.delta.b.FP2.e.Fz	Coherence (Delta)
14	AB.B.theta.e.Fz	Absolute Band Power (Theta)
15	AB.A.delta.c.F7	Absolute Band Power (Delta)
16	COH.C.alpha.m.T5.n.P3	Coherence (Alpha)
17	COH.D.beta.m.T5.r.O1	Coherence (Beta)
18	AB.A.delta.s.O2	Absolute Band Power (Delta)
19	AB.B.theta.n.P3	Absolute Band Power (Theta)
20	COH.A.delta.a.FP1.f.F4	Coherence (Delta)

The ANOVA feature ranking shows that the EEG-derived spectral and coherence features and both demographic attributes (IQ, sex and age) are significant features for schizophrenia classification. The majority of the selected EEG biomarkers were delta and beta band power features, followed by delta, alpha, and beta band coherence measures. The original feature space was reduced from 1,144 features to 50 features (95.63% reduction) by using ANOVA-based feature-selection to reduce the computational complexity without losing the most discriminative information for further classification.

3.4 Optimized Extra Trees Classifier

It is ensemble classifier was used to classify the reduced feature set. In Extra Trees, the splitting threshold is also randomly chosen for each node, which adds more randomness than conventional Random Forests and helps to lower variance and boost generalization. The final prediction of the ensemble classifier is determined using the majority voting strategy:

$$\hat{y} = \text{Mode}(T_1, T_2, \dots, T_n)$$

where $\hat{y}$ denotes the final-predicted-class-label, T_i represents the prediction generated by the i^{th} decision-tree, and n is the total-number-of-trees in the ensemble. The majority voting mechanism selects the class receiving the highest number of votes across all decision trees, thereby improving prediction robustness and reducing model variance. To maximize classification performance, the classifier parameters were optimized using Grid Search Cross-Validation. The optimal parameter configuration is presented in Table 3.

Table 3. Optimal hyperparameters of the Extra Trees classifier.

Hyperparameter	Value
Number_of_Trees	300
Maximum_Depth	10
Maximum_Features	None
Minimum_Samples_Split	2
Minimum_Samples_Leaf	1

3.5 SHAP-Based Explainability

Ensemble learning algorithms are known to have high predictive power, but are of the type that are hard to understand. For the purpose of increasing the transparency of the model, a method called SHAP was used to measure the contribution-of-each feature that was chosen to the final-prediction.

For a prediction function $f(x)$, the SHAP framework decomposes the model prediction into the sum of a baseline prediction and the individual contributions of each feature:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

where $f(x)$ denotes the predicted output for input sample x, ϕ_0 baseline value, $\phi_i$denotes the Shapley value corresponding to the contribution of the i^{th} feature, and M is the total number of input features. SHAP provides both global feature importance and local prediction explanations as it can determine the most important demographic and EEG biomarkers linked to schizophrenia.

3.6 Experimental Design

To comprehensively evaluate the proposed framework, two evaluation protocols were adopted.

3.6.1 Hold-Out Evaluation:Here 80% training and 20% testing was used, maintaining the class distribution. For the classifiers, the training set was employed for feature-selection and classifiers training whereas test set was not used for training and was only used for final evaluation.

3.6.2 Stratified Five-Fold Cross-Validation: The framework was validated using Stratified-5-Fold Cross-Validation to test its robustness and generalization ability. For each fold, the preprocessing, feature standardization, ANOVA feature-selection and optimized Extra Trees training were performed on the training partition, and the validation partition was only evaluated after the training of the model. This protocol prevents information leakage and gives a more reliable estimate of the model performance.

4. RESULTS

4.1 Baseline Performance

As a baseline, the Extra Trees classifier was first trained with the 1144 EEG extracted features without any feature-selection. The model discriminated between healthy controls and schizophrenia subjects with limited accuracy (62.79%), balanced accuracy (60.64%), F1 score (70.37%) and ROC/AUC (68.53%). The relatively low Cohen's Kappa (0.2200) and MCC (0.2300) further indicate that many redundant and non-discriminative features caused the classifier to perform poorly. These baseline results encouraged the use of feature-selection for finding EEG biomarkers with the highest information content.

4.2 Effect of ANOVA Feature-selection

To decrease the dimension of the EEG feature space, dimension reduction method ANOVA SelectKBest was utilized. The Top-50 features were selected for further classification. The introduction of ANOVA feature-selection produced a substantial improvement in classification performance. The classification accuracy increased from 62.79% to 81.40%, while the balanced accuracy improved from 60.64% to 80.59%. Similarly, the F1-score increased from 70.37% to 84.00%, and the ROC-AUC improved from 68.53% to 88.27%. Moreover, Cohen's Kappa increased from 0.2200 to 0.6186, and

MCC improved from 0.2300 to 0.6215, indicating a considerable enhancement in classification consistency. This result shows that redundancy removal of EEG features and only keeping the most discriminative biomarkers significantly increases the prediction ability of the Extra Trees classifier.

Table 4. Performance Comparison Before and After ANOVA Feature-selection

Performance Metric	Extra Trees (All Features)	Extra Trees + ANOVA SelectKBest (Top-50 Features)	Improvement
Number of Features	1,144	50	−95.63%
Accuracy	0.627907	0.813953	+18.60%
Balanced Accuracy	0.606360	0.805921	+19.96%
Precision	0.633333	0.807692	+17.44%
Recall	0.791667	0.875000	+8.33%
F1 Score	0.703704	0.840000	+13.63%
ROC-AUC	0.685307	0.882675	+19.74%
Cohen's Kappa	0.219955	0.618625	+0.398670
Matthews Correlation Coefficient (MCC)	0.230015	0.621456	+0.391441
Training Time (s)	0.804330	0.747708	−7.04%
Prediction Time (s)	0.231214	0.209831	−9.25%

The model has reduced the original feature space from 1,144 to 50 features, and has helped to improve the results in all the evaluation metrics. Specifically, the accuracy of classification rose from 62.79% to 81.40%, and the balanced accuracy rose from 60.64% to 80.59% (Table 4). In addition, feature-selection showed that it reduced train and prediction times, thereby proving that feature-selection not only resulted in better predictive performance but also increased computational efficiency. The results demonstrate that feature-selection using ANOVA can efficiently select the most informative EEG feature biomarkers, and yield a short feature representation for classification.

4.3 Performance of the Optimized Extra Trees Classifier

Optimization of the Extra Trees classifier was done after feature-selection with grid-search hyperparameter tuning. The best hyperparameters were 300 Decision Trees, max tree depth of 10, max_features = None, minimum samples split of 2, and minimum samples leaf of 1. The best configuration generated an accuracy of 95.35% for the hold-out test set with a balanced accuracy of 95.29%, a precision of 95.83%, a recall of 95.83%, an F1-score of 95.83%, and an ROC-AUC of 96.27%. Cohen's-Kappa value rose to 0.9057 along with MCC value, showing excellent agreement between the class labels predicted and the actual class labels. The optimized model also needed less computational time than the baseline model (0.466 s for training and 0.135 s for prediction).

Table 5. Performance of the Optimized Extra Trees Classifier Using the Selected Top-50 EEG Features

Performance Metric	Value
Number of Selected Features	50
Classifier	Optimized Extra Trees
Hyperparameter Optimization	Grid Search
Cohen's Kappa	0.905702
Matthews Correlation Coefficient (MCC)	0.905702
Training Time (s)	0.465874
Prediction Time (s)	0.134677

Performance of proposed Optimized Extra Trees classifier in Table 5, it exhibited reasonable performance on the Top-50 features selected from the ANOVA method. The optimized classifier achieved good discrimination between the two classes with ROC-AUC 96.27%. Moreover, the Cohen's Kappa and Matthews Correlation Coefficient values are very high (0.9057) which shows good agreement between the predicted class labels and the true class labels. The optimized model was cost-effective, and it took 0.466 s and 0.135 s to be trained and tested, respectively. The results have shown that the combination of the feature-selection based on the ANOVA method and systematic hyperparameter optimization can greatly improve the classification accuracy while preserving the computational efficiency. Table 5 shows the average performance achieved using Stratified-5-Fold Cross-Validation, but it does not show the variation in performance of the proposed classifier across the different folds (Figure 3).

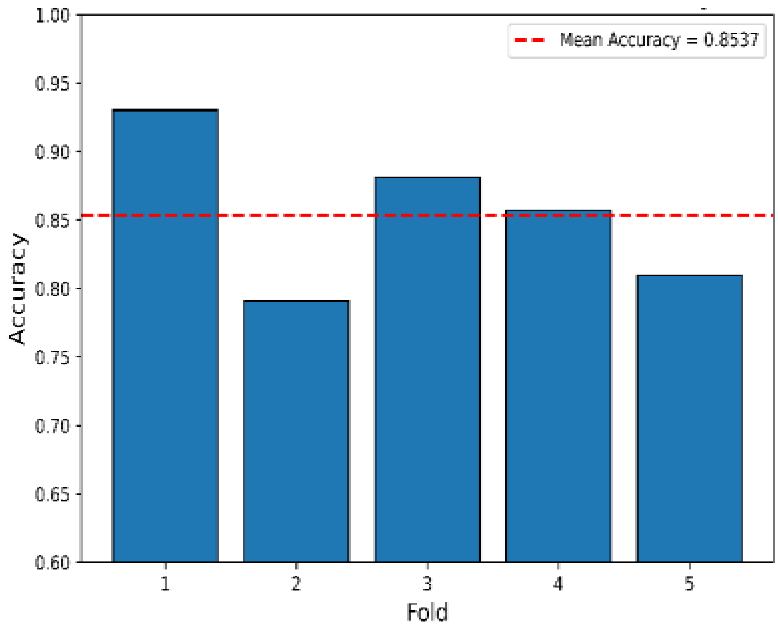

Figure 3. Proposed optimized Extra-Trees-classifier on the folds of Stratified-5-Fold Cross-Validation. Dashed red line is the average classification accuracy of the 5 folds.

The classification accuracy results in each fold during Stratified-5-Fold Cross-Validation are shown in Figure 3. The results showed that the proposed optimized Extra Trees classifier consistently gave good classification results with an average accuracy of 85.37% across all the validation subsets, although the results were found to be slightly different from one-fold to another. The difference in the folds is relatively low, proving the strength and stability of the proposed framework and thus generalisation of the classifier to unseen data. The results further confirm the reliability of the feature-selection method such as ANOVA and optimized Extra Trees classifier for EEG-based schizophrenia classification.

Table 6. Optimal Hyperparameters of the Proposed Extra Trees Classifier

Hyperparameter	**Optimal Value**
Number_of_Trees (n_estimators)	300
Maximum_Tree_Depth (max_depth)	10
Maximum_Features (max_features)	None
Minimum_Samples_Split (min_samples_split)	2
Minimum_Samples_Leaf (min_samples_leaf)	1
Best_Cross-Validation_Accuracy	0.8579

The optimal hyperparameter configuration is summarized in Table 6, and is the result of optimizing the hyperparameters using Grid Search. The optimized model used 300 trees of a max depth of 10 and used all features to split nodes (max_features=None). The default values of minimum support for splitting and leaf nodes were kept, so that the effect of further increase of diversity in the ensemble and control of tree depth was more significant in improving the performance than the other tree regularization. Before final model evaluation, the optimized configuration yielded the best CV accuracy of 85.79%, showcasing the strength of the hyperparameters chosen. The optimization process played a significant role in the enhanced prediction accuracy achieved within the proposed framework.

4.4 Stratified Five-Fold Cross-Validation

The Stratified-5-Fold Cross-Validation was implemented on the entire processing pipeline, which involved the application of preprocessing, ANOVA feature-selection and optimized Extra Trees (ET) Classifier in each fold. Across the five folds, the proposed framework achieved a mean-accuracy of 85.37 ± 5.01%, balanced accuracy of 85.17 ± 5.33%, precision of 86.70 ± 6.13%, recall of 87.17 ± 2.65%, and F1-score of 86.89 ± 4.23%. Moreover, the average ROC-AUC was 93.14 ± 2.51%, which suggests good discriminative performance on various partitions. The mean Cohen's Kappa (0.7034 ± 0.1029) and MCC (0.7045 ± 0.1026) also show the classification performance is stable. The obtained cross validation results demonstrate that the proposed framework is consistent across various train-test splits and is not relying on the random partition.

Table 7. Stratified-5-Fold_Cross-Validation Results of the Proposed Framework

Performance Metric	**Mean**	**Standard Deviation**
Accuracy	0.853710	0.050065
Balanced Accuracy	0.851659	0.053262
Precision	0.867036	0.061270

Recall	0.871739	0.026486
F1 Score	0.868857	0.042318
ROC-AUC	0.931369	0.025087
Cohen's Kappa	0.703428	0.102909
Matthews Correlation Coefficient (MCC)	0.704501	0.102619

The Stratified-5-Fold Cross-Validation results of the proposed framework are concise in Table 7. The model attained the classification accuracy of 85.37% (± 5.01%) and balanced accuracy of 85.17% (±5.33%) in all data partition, showing the consistent performance of the model. The average precision, recall and F1-score were 86.70%, 87.17% and 86.89%, respectively, showing the precision was balanced for both classes. Moreover, the suggested framework obtained an average ROC-AUC of 93.14% (± 2.51%), ensuring the framework's high discriminative power. The Cohen's Kappa coefficient (0.7034 ± 0.1029) and Matthew's correlation coefficient (0.7045 ± 0.1026) shows that there is a substantial agreement between the true class labels and those predicted for all folds. The relatively small standard deviations of most evaluation metrics also illustrate the stability and reliability of the proposed framework, confirming the robustness and generalization ability of the proposed framework over a single train-test split. The confusion matrix of the proposed optimized Extra Trees classifier is displayed in Figure 4 to further assess the classification performance of the proposed classifier. The confusion matrix provides a detailed analysis of correctly classified/incorrectly classified data instances for two classes Healthy and Schizophrenia.\

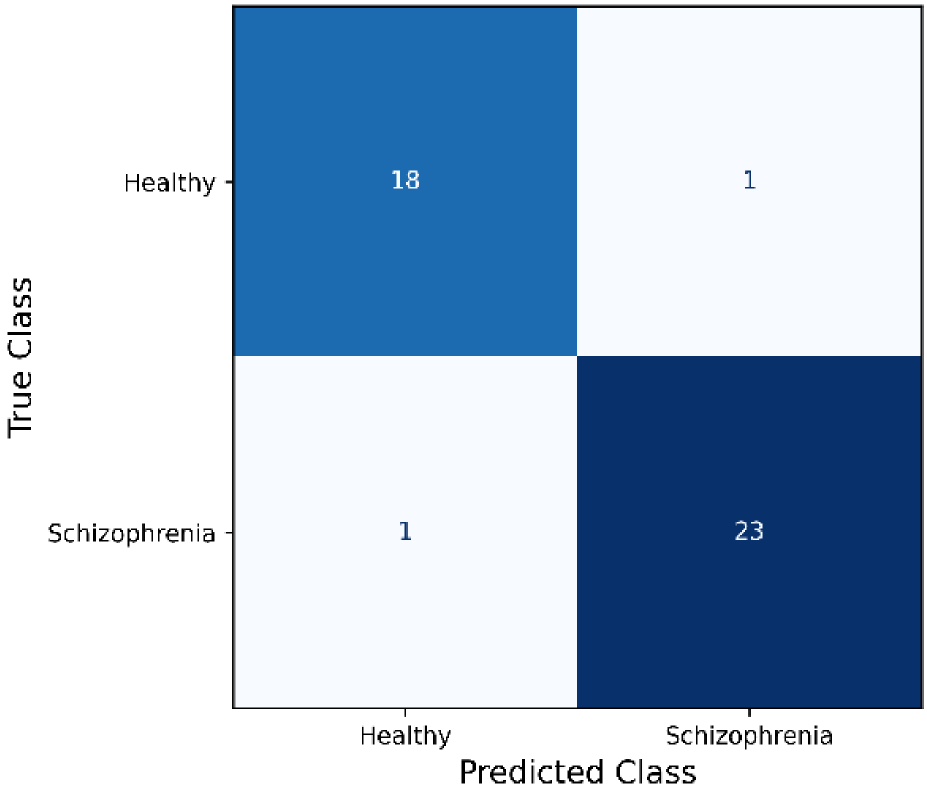

Figure 4. Confusion matrix of the proposed optimized classifier.

The proposed classifier correctly classified 18 out of 19 Healthy subjects and 23 out of 24 Schizophrenia subjects, and made just two misclassifications on the entire test set as shown in Figure 4. One Healthy subject was misclassified as Schizophrenia and one Schizophrenia subject was misclassified as Healthy. The overall accuracy obtained is 95.35%, the balanced accuracy is 95.29%, and the MCC is 0.9057, which indicate an excellent discrimination between the two classes. The almost balanced distribution of the errors indicates that the classifier is not biased toward any class and that the selected Top-50 features using ANOVA and the optimized Extra Trees classifier is effective. The confusion matrix is useful to analyze the classification results; however, it does not completely describe the discriminative power of the proposed classifier for all decision thresholds.

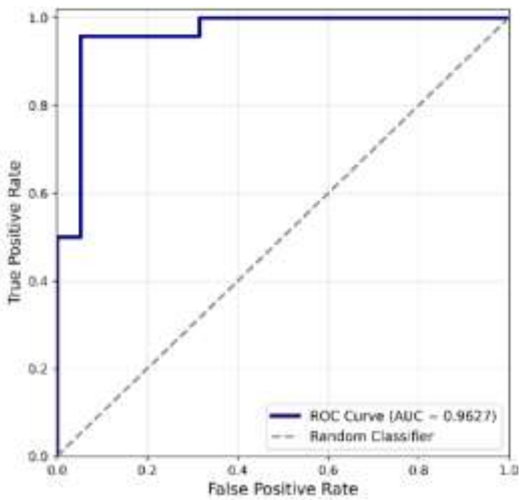

Figure 5. ROC Curve of the Proposed Classifier

The ROC curve of the proposed optimized Extra Trees classifier is shown in figure 5. The ROC curve is very tight, close to the upper left corner, meaning that, at various thresholds, the true positive rate is high, while also keeping the false positive rate low. In addition to the confusion matrix, the results validate the good prediction ability and reliability of the proposed framework for EEG based classification. The confusion matrix and ROC curve show the predictive accuracy of the proposed classifier, but do not give an indication of the respective contribution of the single features. Thus, feature importance scores from the optimized extra trees classifier were examined to understand the most important features affecting the classification.

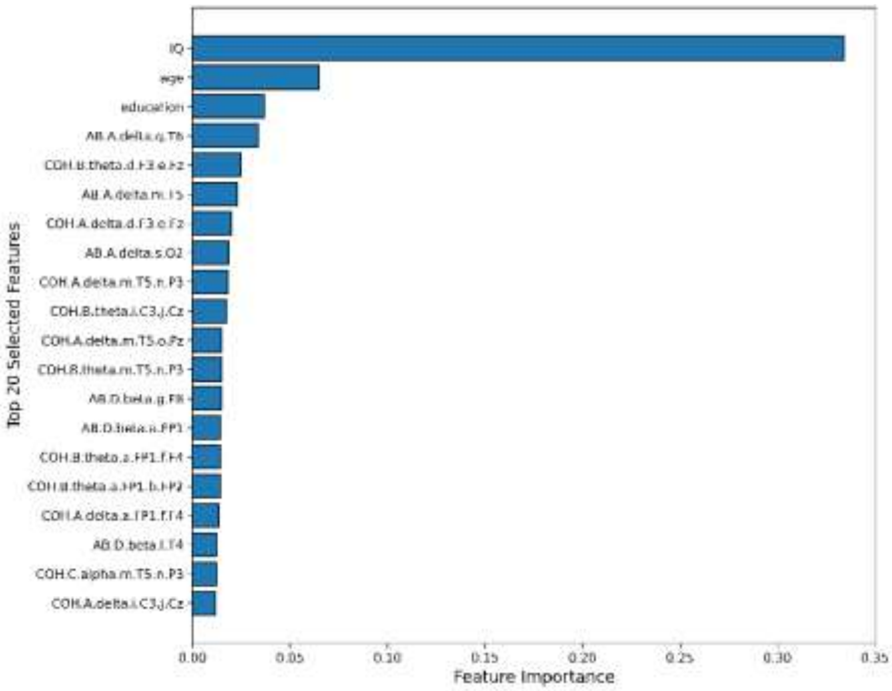

Figure 6. Top-20 feature importances obtained from the proposed classifier.

The relative importance of the Top-20 features learned by the optimized Extra Trees classifier is plotted in Figure 6. Of the variables selected, IQ, age, and education were the most influential predictors, followed by a number of features derived from the EEG obtained by using spectral and coherence analysis. Demographic and neurophysiological variables were among the highest ranked predictors, suggesting that the proposed structure is an effective incorporation of complementary sources of information for classification. Moreover, multiple features from the EEG remained after the feature-selection, indicating that the ANOVA-based feature-selection has preserved the informative ones while keeping a considerably small feature space from 1,144 to 50 features.

4.5 SHAP-Based Explainability Analysis

SHAP analysis was carried out on the optimized classifier to explore the contribution of individual EEG biomarkers. The SHAP summary plot showed that the sex, IQ and age, plus a number of spectral power and coherence EEG features, were the most important factors for the classification of schizophrenia. The features related to the delta, beta and gamma frequency bands consistently showed high SHAP values and are considered to be important for differentiation between schizophrenia and healthy controls among the spectral biomarkers. The results are interpretable and useful to support the clinical significance of the selected EEG biomarkers. The SHAP visualization also shows that the proposed framework can not only be used to train a model that achieves high classification performance but also explain their prediction results transparently.

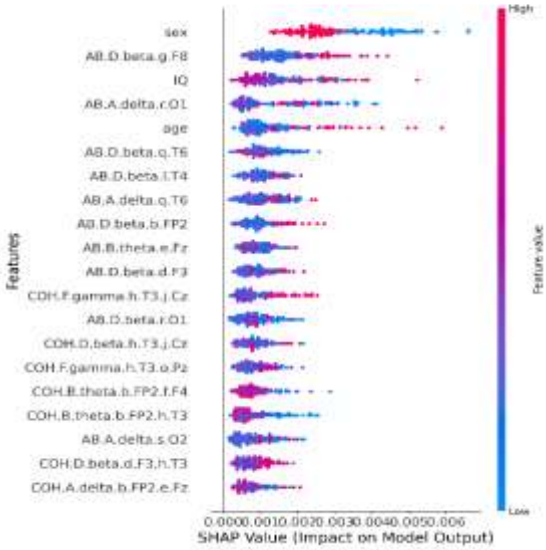

Figure 7. Global SHAP summary plot of the Top-20 predictive features selected by the proposed framework.

From the Figure 7, the global SHAP summary plot of the optimized Extra Trees classifier with the Top-50 features selected by ANOVA is shown in Figure 7. The features are ranked based on their mean absolute SHAP values, which help to identify the most influential features associated with the classification process. Satisfied, sex, IQ and age turned out to be the most influential demographic features, followed by a number of EEG-derived spectral and coherence features. The beta-band and delta-band power features and coherence-based measures had particularly high contributions to the model's predictions. From the distribution of SHAP values, it is also evident that the inclusion of both the demographic information and EEG biomarkers contribute to the classification result. Table 8 shows a summary of the performance of the experiments carried out at every step of the proposed framework in order to have a full picture of the experimental progression. ANOVA feature-selection, classifier optimization and Stratified-5-Fold Cross-Validation contribute in improving the classification performance and robustness of the model, as shown in the Table 8.

Table 8. Overall Performance Comparison of the Proposed Framework

Experiment	**Model**	**Features**	**Accuracy**	**Balanced Accuracy**	**Precision**	**Recall**	**F1 Score**	**ROC-AUC**	**Cohen's Kappa**	**MCC**
Baseline	Extra Trees	1,144 (All Features)	0.627907	0.606360	0.633333	0.791667	0.703704	0.685307	0.219955	0.230015
Feature-selection	Extra Trees + ANOVA SelectKBest	50	0.813953	0.805921	0.807692	0.875000	0.840000	0.882675	0.618625	0.621456
Proposed Framework	Optimized Extra Trees + ANOVA SelectKBest	50	0.953488	0.952851	0.958333	0.958333	0.958333	0.962719	0.905702	0.905702
Robustness Evaluation	Stratified 5-Fold CV (Optimized Extra Trees + Top-50)	50	0.853710 ± 0.050065	0.851659 ± 0.053262	0.867036 ± 0.061270	0.871739 ± 0.026486	0.868857 ± 0.042318	0.931369 ± 0.025087	0.703428 ± 0.102909	0.704501 ± 0.102619

As shown in Table 8, the proposed framework showed progressive improvements along the experimental pipeline. The feature-selection using ANOVA technique reduced the feature space from 1144 to 50 features and resulted in significant improvement in the classification performance and reduced computational complexity. Hold out test accuracy of 95.35%

was achieved for the Extra Trees classifier with a ROC-AUC of 96.27% after a few more hyperparameter optimizations. We can see that the proposed framework was demonstrated to be robust and capable of generalization using Stratified-5-Fold Cross-Validation with mean accuracy of 85.37% and mean ROC-AUC of 93.14%, respectively, across five separate folds. These results prove that the feature-selection, the model optimization and rigorous validation generate an effective and reliable EEG classification system.

5. CONCLUSION

This study proposed an explainable machine learning model to detect schizophrenia automatically based on EEG features. The proposed framework combined a feature-selection technique based on analysis of variance (ANOVA), an optimized Extra Trees (ET) classifier, and an explainability technique using SHAP. In total, 1,144 demographic and EEG-derived features were taken into account, initially. In order to reduce the dimensionality of the features, yet maintain features with diagnostically relevant information, the features were reduced to the 50 most discriminative features with ANOVA SelectKBest. The experimental results showed that the selection of features had significantly better classification results than the baseline Extra Trees classifier. Moreover, the predictive ability of the proposed framework was also improved through hyperparameter optimization with a hold-out test accuracy of 95.35%, balanced accuracy of 95.29%, an F1-score of 95.83%, and an ROC-AUC of 96.27%. Stratified-5-Fold Cross-Validation was used to evaluate the robustness of the model, which achieved a mean accuracy of 85.37 ± 5.01% and a mean ROC-AUC of 93.14 ± 2.51%, demonstrating its good generalization performance over the different folds. The Top three biomarkers that contributed to schizophrenia classification were demographic variables in addition to EEG spectral and coherence variables, identified by SHAP based explainability. The clinically meaningful insights obtained from the explainability analysis make the proposed framework more transparent and clinically meaningful. In general, the proposed framework shows that the feature-selection method based on Analysis of Variance (ANOVA), coupled with the optimized Extra Trees (ET) classifier is an effective and interpretable method for EEG-based schizophrenia detection.

REFERENCES

1. j. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clinical Neurophysiology, vol. 113, no. 6, pp. 767-791, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.
2. L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain Computer Interfaces, a Review,” Sensors, vol. 12, no. 2, pp. 1211-1279, Jan. 2012, doi: 10.3390/s120201211.

A. Gramfort et al., “MNE software for processing MEG and EEG data,” NeuroImage, vol. 86, pp. 446-460, Feb. 2014, doi: 10.1016/j.neuroimage.2013.10.027.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. Muller, “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis,” IEEE Signal Processing Magazine, vol. 25, no. 1, pp. 41-56, 2008, doi: 10.1109/MSP.2008.4408441.

A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” Circulation, vol. 101, no. 23, 2000, doi: 10.1161/01.CIR.101.23.e215.

3. F. Lotte et al., “A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update,” Journal of Neural Engineering, vol. 15, no. 3, p. 031005, Jun. 2018, doi: 10.1088/1741-2552/aab2f2.

A. Craik, Y. He, and J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” Journal of Neural Engineering, vol. 16, no. 3, p. 031001, Jun. 2019, doi: 10.1088/1741-2552/ab0ab5.

4. Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: a systematic review,” Journal of Neural Engineering, vol. 16, no. 5, p. 051001, Oct. 2019, doi: 10.1088/1741-2552/ab260c.
5. R. T. Schirrmeister et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” Human Brain Mapping, vol. 38, no. 11, pp. 5391-5420, Nov. 2017, doi: 10.1002/hbm.23730.
6. V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces,” Journal of Neural Engineering, vol. 15, no. 5, p. 056013, Oct. 2018, doi: 10.1088/1741-2552/aace8c.
7. W. Hang et al., “Cross-Subject EEG Signal Recognition Using Deep Domain Adaptation Network,” IEEE Access, vol. 7, pp. 128273-128282, 2019, doi: 10.1109/ACCESS.2019.2939288.
8. S. Liang et al., “Domain-Generalized EEG Classification With Category-Oriented Feature Decorrelation and Cross-View Consistency Learning,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 3285-3296, 2023, doi: 10.1109/TNSRE.2023.3300961.
9. R. Shoorangiz, S. J. Weddell, and R. D. Jones, “EEG-Based Machine Learning: Theory and Applications,” in Handbook of Neuroengineering, Singapore: Springer, 2023, pp. 2463-2501, doi: 10.1007/978-981-16-5540-1_70.
10. R. Lyu, “Deep learning approaches for EEG-based healthcare applications: a comprehensive review,” Frontiers in Human Neuroscience, vol. 19, p. 1689073, 2026, doi: 10.3389/fnhum.2025.1689073.
11. J. Xie et al., “A Transformer-Based Approach Combining Deep Learning Network and Spatial-Temporal Information for Raw EEG Classification,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2126-2136, 2022, doi: 10.1109/TNSRE.2022.3194600.
12. Y. Song, Q. Zheng, B. Liu, and X. Gao, “EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 710-719, 2023, doi: 10.1109/TNSRE.2022.3230250.
13. S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” 2017, arXiv. doi: 10.48550/ARXIV.1705.07874.

14. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 1135-1144, doi: 10.1145/2939672.2939778.
I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial EEG classification," Journal of Neuroscience Methods, vol. 274, pp. 141-145, Dec. 2016, doi: 10.1016/j.jneumeth.2016.10.008.
15. S. Ali et al., "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," Information Fusion, vol. 99, p. 101805, Nov. 2023, doi: 10.1016/j.inffus.2023.101805.
16. V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer Learning in Brain-Computer Interfaces," IEEE Computational Intelligence Magazine, vol. 11, no. 1, pp. 20-31, Feb. 2016, doi: 10.1109/MCI.2015.2501545.
17. D. Wu, Y. Xu, and B.-L. Lu, "Transfer Learning for EEG-Based Brain-Computer Interfaces: A Review of Progress Made Since 2016," 2020, doi: 10.48550/ARXIV.2004.06286.
18. O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-Independent Brain-Computer Interfaces Based on Deep Convolutional Neural Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 3839-3852, Oct. 2020, doi: 10.1109/TNNLS.2019.2946869.
19. S. Cui, D. Lee, and D. Wen, "Toward brain-inspired foundation model for EEG signal processing: our opinion," Frontiers in Neuroscience, vol. 18, p. 1507654, Dec. 2024, doi: 10.3389/fnins.2024.1507654.
20. H. Altaheri, F. Karray, and A.-H. Karimi, "Temporal convolutional transformer for EEG based motor imagery decoding," Scientific Reports, vol. 15, no. 1, p. 32959, Sep. 2025, doi: 10.1038/s41598-025-16219-7.