BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Centre, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2026, Volume 17, Issue 3, pages: 275-291
Submitted: May 29th, 2026 | *Accepted for publication:* August 23rd, 2026

Multiclass EEG-Based Classification of Psychiatric Disorders Using Interpretable Machine Learning on Spectral and Connectivity Features

Aangi Shah
The Riverside School, Sardarnagar, Ahmedabad Cantonment, Ahmedabad, Gujarat 380004, India.
aangishah1020@gmail.com ,
https://orcid.org/0009000210876535

Ajay Goyal*
Institute of Design, Nirma University, Ahmedabad, Gujarat 382481 India.
ajay.goyal@nirmauni.ac.in,
https://orcid.org/0000-0001-6305-7432

* *Corresponding author*

Abstract: *Psychiatric diagnosis relies heavily on subjective clinical evaluation, limiting objective differentiation across co-occurring disorders. This study presents an interpretable multiclass electroencephalography (EEG)-based framework for —Addictive Disorder, Anxiety Disorder, Mood Disorder, Obsessive-Compulsive Disorder (OCD), Schizophrenia, Trauma and Stress-Related Disorder—and Healthy Controls. A dataset of 945 subjects with 1,140 EEG-derived spectral and coherence features was analysed, with variance-based selection reducing the feature space to 300 descriptors. Classification was performed using Extreme Learning Machine (ELM), Naïve Bayes (NB), and Support Vector Machine (SVM) within a one-vs-all architecture. NB demonstrated comparatively limited discriminative capability (F1: 35.33%–72.76%). ELM achieved stable performance with F1-scores ranging from 66.03% to 89.31% and consistently higher testing accuracies across all classes, including 95.12% for OCD and 90.34% for Healthy Controls. SVM yielded the highest precision–recall balance across most classes, achieving F1-scores of 94.00% for OCD, 84.08% for Anxiety Disorder, and 79.84% for Schizophrenia. Mood Disorder remained the most challenging class (maximum F1: 66.03%). Bandwise analysis identified theta-band features as the most discriminative within the present dataset , achieving 91.57% standalone accuracy. These results suggest that EEG-derived spectral and coherence features, when combined with lightweight machine learning classifiers, can support multiclass psychiatric classification with variable class-wise performance. However, the lower performance observed for Mood Disorder and the reliance on a single public dataset indicate that further external validation, statistical comparison, and multimodal feature integration are required before clinical translation.*

Keywords: electroencephalography; multiclass psychiatric classification; support vector machine; power spectral density; machine learning.

How to cite: Shah, A., & Goyal, A. (2026). Multiclass EEG-based classification of psychiatric disorders using interpretable machine learning on spectral and connectivity features. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 17*(3), 275-291. https://doi.org/10.70594/brain/17.3/17

1. Introduction

Psychiatric diagnosis continues to rely largely on subjective clinical assessment, highlighting the need for objective, quantifiable biomarkers. Electroencephalography (EEG) is a non-invasive neurophysiological assessment technique that provides millisecond-level temporal resolution for monitoring brain activity (Weisz & Obleser, 2014). Its clinical application in psychiatry is limited due to its complexity and vulnerability to artefacts from ocular, muscular, and environmental sources (Jiang et al., 2019). Nonetheless, EEG's capacity to reflect functional brain alterations has been enhanced with the advent of machine learning (ML) techniques (Sarisik et al., 2025). This enables the discovery of diagnostic-relevant patterns across spectral, temporal, and connectivity domains (Rasool et al., 2025). Resting-state EEG studies have shown that psychiatric disorders are associated with altered spectral activity and functional connectivity patterns, although these abnormalities often overlap across diagnostic categories, supporting the need for multiclass and connectivity-aware machine learning approaches (Newson & Thiagarajan, 2019; Park et al., 2021a; Wu et al., 2021; Miljevic et al., 2023).

ML-based EEG diagnostic classification has demonstrated promising results across a range of neuropsychiatric disorders. Supervised models have successfully classified patients with conditions such as schizophrenia, obsessive-compulsive disorder (OCD), major depressive disorder (MDD), and anxiety relative to healthy controls (Mahato & Paul, 2020; Park et al., 2021b; Vázquez et al., 2021; Ranjan et al., 2024). However, these studies were limited to binary classification, lacked generalisation testing, and used controlled task-related EEG. Rashid et al. (2019) reported an efficient machine-learning classification model for cognitive task EEG using power spectral density (PSD) features applied to k-Nearest Neighbours (k-NN) and Support Vector Machines (SVM). Zhang (2019) employed event-related potentials (ERPs) and random forests to classify schizophrenia.

Multiclass ML frameworks offer broader diagnostic utility but present scalability challenges. Park et al. (2021b) classified six disorders using a dataset comprising 945 subjects and achieved >89% accuracy using Power Spectral Density (PSD) and functional connectivity (FC) features. Yet, performance declined markedly for Obsessive-Compulsive-Disorder (OCD) (74.52%) and required IQ-adjusted inputs, thus limiting clinical robustness. Similarly, Mumtaz et al. (2018) leveraged synchronisation likelihood on the PhysioNet database to classify Major Depressive Disorder (MDD) with 98% accuracy, but within a binary setup. Ieracitano et al. (2019) reached 95.76% accuracy for an Alzheimer's EEG dataset using time-frequency features, underscoring the domain's potential, albeit outside psychiatry.

Deep learning (DL) has accelerated EEG research, but often at the expense of interpretability and clinical transferability (Bouazizi & Ltifi, 2024; Sharma & Meena, 2024). Xu et al. (2023) used a hybrid multi-receptive-field convolutional neural network–long short-term memory model for MDD classification, reporting 98.48% accuracy from 10-minute frontal EEG segments.

Rafiei et al. (2022) applied InceptionTime to a 19-channel dataset, achieving 91.67% accuracy and 100% sensitivity.

Many prior studies rely on narrow datasets and lack disorder diversity. Sharma et al. (2021) used the CAP (Cyclic Alternating Pattern) database to classify six sleep disorders (91.3% accuracy). AlSharabi et al. (2022) applied discrete wavelet transforms to Alzheimer's EEG data, attaining 99.98% accuracy with k-NN. However, most of these studies address non-psychiatric or binary endpoints, use small sample sizes, or depend on feature-specific tuning.

Model generalisation across subjects and datasets remains a major challenge. Abdulbaqi et al. (2022) achieved 81.23% accuracy and 98% specificity on the CHB-MIT (Children's Hospital Boston – Massachusetts Institute of Technology) dataset using a Support Vector Machine (SVM), but performance deteriorated on unseen participants. Bouallegue et al. (2020) and Khalfallah et al. (2025) achieved strong results on autism and epilepsy datasets via consumer-grade EEG headsets, yet lacked external validation. Recent reviews (Gautam & Sharma, 2020; Rivera et al., 2022; Zhou et al., 2022) highlight these concerns, calling for models that prioritise transparency, clinical interpretability, and robustness across heterogeneous populations.

To address these limitations, this study proposes a clinically oriented, interpretable, and reproducible EEG classification framework for differentiating seven diagnostic classes comprising six psychiatric disorders groups-Addictive Disorder, Anxiety Disorder, Mood Disorder, Obsessive–Compulsive Disorder, Schizophrenia, and Trauma and Stress-Related Disorder-and one healthy-control group within a unified modelling strategy.

Unlike most prior studies that predominantly focus on binary classification or single-disorder detection, the proposed approach reflects real-world diagnostic complexity by enabling simultaneous multiclass differentiation. The framework integrates Power Spectral Density (PSD) and Coherence (COH) features to capture both spectral dynamics and functional brain connectivity, thereby preserving physiological interpretability while maintaining discriminative capability. Furthermore, the use of lightweight and transparent machine learning models (Extreme Learning Machine (ELM), Naïve Bayes (NB), and Support Vector Machine (SVM)) within a one-vs-all (OvA) architecture supports scalability and interpretability, addressing key limitations associated with black-box deep learning approaches.

In addition to multiclass classification, a complementary binary classification setting is evaluated to emulate potential clinical screening scenarios, where distinguishing pathological conditions from healthy states is critical. This dual evaluation strategy allows for a more comprehensive assessment of model behaviour across diagnostic contexts. By systematically combining interpretable features with computationally efficient classifiers on a relatively large and heterogeneous dataset, this study contributes towards advancing reproducible machine-learning approaches for multiclass EEG-based psychiatric classification. To the best of our knowledge, this work is among the few studies to jointly investigate multiclass psychiatric classification and model interpretability within a unified framework. Although the proposed approach demonstrates promising classification performance, its potential application to clinical decision support, early screening, differential diagnosis, and objective psychiatric assessment requires further validation using independent datasets and prospective clinical studies.

2. Methodology

2.1. Dataset Description

This study conducted a secondary analysis of the publicly available feature-level quantitative electroencephalography dataset originally reported by Park et al. (2021).

The feature table analysed in the present study was accessed through the EEG Psychiatric Disorders Dataset hosted on Kaggle:

https://www.kaggle.com/datasets/shashwatwork/eeg-psychiatric-disorders-dataset .

However, the original publication and its associated Open Science Framework repository (https://osf.io/8bsvr/) were treated as the primary sources of information regarding participant recruitment, diagnostic assessment, EEG acquisition, feature generation, and ethical approval. The original cohort was retrospectively assembled from medical records, psychological assessment batteries, and resting-state quantitative EEG assessments conducted between January 2011 and December 2018 at the Seoul Metropolitan Government–Seoul National University Boramae Medical Center, Seoul, Republic of Korea. The cohort comprised 945 participants, including 850 patients and 95 healthy controls. The original study was approved by the Institutional Review Board of the SMG-SNU Boramae Medical Center under approval number 20-2019-16. The requirement for written informed consent was waived because of the retrospective study design. The present study involved no new participant recruitment, intervention, or EEG data collection.

In the original study, the initial clinical diagnosis was established by a psychiatrist according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM) , Fourth or Fifth Edition. Diagnoses were additionally assessed using the Mini-International Neuropsychiatric

Interview during psychological evaluation. Final confirmation of the primary diagnosis was performed by two psychiatrists and two psychologists through retrospective review of the diagnoses recorded in the electronic medical records and psychological assessments completed within one month before or after the EEG assessment. The eligibility criteria included an age of 18–70 years, a primary diagnosis within one of the specified psychiatric categories, and no difficulty reading, listening to, writing, or understanding Korean. Participants with neurological disorders or brain injury, neurodevelopmental disorders, or neurocognitive disorders were excluded. Healthy controls were selected from studies conducted at the same medical centre; however, the source publication did not provide detailed information regarding their recruitment or the diagnostic screening procedure used to confirm the absence of psychiatric disorders.

Participants were assigned to seven diagnostic classes comprising six psychiatric disorder groups and one healthy-control group: Addictive Disorder (n = 186), Anxiety Disorder (n = 107), Mood Disorder (n = 266), Obsessive–Compulsive Disorder (n = 46), Schizophrenia (n = 117), Trauma and Stress-Related Disorder (n = 128), and Healthy Controls (n = 95). These broad diagnostic categories were inherited from the source dataset and included more specific diagnoses such as depressive disorder, bipolar disorder, panic disorder, social anxiety disorder, alcohol use disorder, behavioural addiction, post-traumatic stress disorder, acute stress disorder, and adjustment disorder.

In the original study, five-minute eyes-closed resting-state EEG recordings were acquired from either 19 or 64 channels at sampling rates of 500–1,000 Hz using Neuroscan Scan 4.5, with online filtering between 0.1 and 100 Hz. Electrode impedances were maintained below 5 kΩ. The recordings were subsequently downsampled to 128 Hz, and a common set of 19 electrodes corresponding to the international 10–20 system was retained using a mastoid reference. Frequency-domain processing and EEG feature generation were performed using the NeuroGuide system, NG Deluxe version 3.0.5. Fast Fourier transformation was applied using two-second epochs, 256 digital time points, a frequency range of 0.5–40 Hz, a frequency resolution of 0.5 Hz, and a cosine taper window. At least 60 seconds of edited EEG data were used for the spectral calculations. The original publication referred to its supplementary material and an earlier study for further details of signal preprocessing and artefact rejection.

The released dataset contained 1,140 EEG-derived features for each participant. For each frequency band, the feature set comprised 19 channel-level absolute-power features, identified by the prefix AB in the released table, and 171 pairwise coherence features, identified by the prefix COH. These corresponded to power spectral density and coherence-based functional connectivity, respectively. The resulting 190 features per frequency band were calculated across six bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–25 Hz), high beta (25–30 Hz), and gamma (30–40 Hz), producing a total of 1,140 EEG features. Demographic and psychological variables were provided separately and were not included in the count of 1,140 EEG-derived features.

The demographic variables included age, sex, years of education, and intelligence quotient. In the released feature table used for the present analysis, age ranged from 18 to 71.88 years, with a mean of 30.60 ± 11.78 years; education ranged from 0 to 20 years, with a mean of 13.44 ± 2.55 years; and intelligence quotient ranged from 49 to 145, with a mean of 101.60 ± 17.02. The sample included 601 males (63.6%) and 344 females (36.4%). The source publication did not clarify whether a value of zero years of education represented an absence of formal schooling, another coding convention, or an undocumented data-entry value. Because the original eligibility criteria required participants to have no difficulty reading, listening to, writing, or understanding Korean, a value of zero years of education was reported as recorded but was not interpreted as evidence of illiteracy or automatically treated as missing data.

The maximum age of 71.88 years and the minimum intelligence quotient of 49 observed in the released feature table were not fully consistent with the age and neurodevelopmental exclusion criteria described in the original publication. These discrepancies could not be resolved from the publicly available documentation and were therefore retained as reported and considered limitations of the secondary dataset.

Because the present study used a feature-level dataset rather than raw or epoched EEG recordings, channel quality, artefact rejection, independent component analysis, epoch rejection, and the amount of clean EEG retained for each participant could not be independently inspected or reproduced. The EEG descriptors were generated from recordings preprocessed by the original investigators. Accordingly, the processing undertaken in the present study was restricted to feature-level cleaning, normalisation, and feature selection, as described in Section 2.2. Residual artefacts, differences in the original acquisition configuration, medication exposure, comorbidities, symptom severity, and the single-centre Korean composition of the cohort may have influenced the published spectral and coherence features. The original study also identified medication, comorbidity, diagnostic timing, single-centre recruitment, and the absence of external validation as limitations.

2.2. Data Preprocessing and Feature Selection

EEG features containing undefined or zero variance across all subjects were excluded. The exclusion of delta- and gamma-band features was an a priori methodological decision rather than a post-hoc decision based on classification performance. The analysis was restricted to theta, alpha, beta, and high-beta activity to reduce the dimensionality of the feature space and focus on mid-frequency resting-state rhythms frequently associated with psychiatric EEG alterations. Because only feature-level data were available and raw-signal preprocessing could not be independently repeated, the restriction also reduced reliance on delta and gamma activity, which may be more sensitive to residual slow-frequency drift and ocular artefacts, and high-frequency muscular contamination, respectively. We have clarified this rationale in Section 2.2 and acknowledged that excluding these bands may have omitted potentially informative features. Following the removal of undefined and zero-variance descriptors, the analysis was restricted a priori to theta (4–8 Hz), alpha (8–12 Hz), beta (12–25 Hz), and high-beta (25–30 Hz) features. Beta and high-beta features were considered collectively as the broader beta-range activity for descriptive purposes. Channel-level absolute-power descriptors, identified by the prefix AB, and pairwise coherence descriptors, identified by the prefix COH, were included. This restriction was specified before classifier evaluation and was not based on class labels, exploratory classification performance, or holdout-test results. The selected bands were retained to reduce the dimensionality of the feature space and focus on mid-frequency rhythms frequently implicated in resting-state psychiatric EEG findings. In addition, because the present study used derived features and could not independently repeat raw-signal artefact correction, delta and gamma features were excluded to reduce dependence on bands that may be particularly sensitive to residual slow-frequency drift or ocular activity and high-frequency myogenic contamination, respectively. Nevertheless, this restriction may have excluded diagnostically informative features and is acknowledged as a limitation. The original feature table contained 1,140 EEG-derived descriptors across six frequency bands. Each band contributed 19 channel-level absolute-power features and 171 pairwise coherence features, giving 190 descriptors per band. Restricting the analysis to theta, alpha, beta, and high-beta features resulted in 760 candidate descriptors before the removal of undefined and zero-variance features. The remaining eligible descriptors were ranked according to variance, and the 300 highest-variance features were retained for classification. To reduce dimensionality and enhance discriminative learning, features were ranked by variance and the top 300 were selected for modelling. The variance of each feature x_j was computed as:

$$Var\left(x_j\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \underline{x}_j\right)^2 \tag{1}$$

where x_{ij} is the j^{th} feature of the i^{th} subject, and $\underline{x}_j$ is the mean across all subjects. These selected features were normalised using z-score scaling to ensure numerical consistency:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{2}$$

where μ_j and σ_j are the mean and standard deviation of feature j, respectively. Notably, the 10 most variable features were exclusively from the alpha band, with the top feature AB_C_alpha_o_Pz having the highest variance of 1686.31.

In the implemented analysis, feature filtering, variance calculation, top-300 feature selection, and z-score normalisation were performed on the complete feature matrix before the stratified train-test split. These preprocessing operations were independent of the diagnostic labels and did not use classifier performance or test-set outcomes. Accordingly, the procedure did not introduce label-driven feature-selection leakage. However, because the feature variances, means, and standard deviations were estimated using all participants, a degree of distributional information leakage between the subsequently defined training and holdout subsets cannot be excluded. This limitation may have resulted in optimistic performance estimates and should be considered when interpreting the findings.

Variance-based feature selection was used as an unsupervised dimensionality-reduction procedure because it does not directly use diagnostic labels. It therefore avoids label-driven selection of features; however, unsupervised preprocessing does not fully eliminate information leakage when feature statistics are estimated before the train-test split. Given the high feature-to-sample ratio of the dataset, this approach provided a computationally simple and reproducible method for removing low-information descriptors before classifier training. However, variance does not necessarily capture class-discriminative relevance. Therefore, supervised or wrapper-based methods such as recursive feature elimination, minimum redundancy maximum relevance selection, sequential backward selection, or embedded regularization-based selection should be evaluated in future work.

2.3. Classification Models

The classification problem was framed as a multiclass, one-vs-all task using the main disorder field as the label. For each class c, a binary target was constructed by encoding:

$$y_i^{(c)} = \{1, \quad if\ subject\ i\ belongs\ to\ class\ c\ 0, \quad otherwise \tag{3}$$

Following feature-level preprocessing and selection, the processed dataset was divided into 80% training and 20% holdout testing subsets using stratified sampling to preserve the relative distribution of the seven diagnostic classes. ELM, NB, and SVM classifiers were trained on this setup. These models are explained below:

ELM: ELM has been reported to provide good generalisation in high-dimensional data. The ELM model is a single-hidden-layer feedforward neural network (SLFN) with randomly initialised input weights and biases. The hidden layer output matrix H is computed as:

$$H = g(XW + b) \tag{4}$$

where $X \in R^{n\times d}$ the input matrix of n samples by d features, $W \in R^{n\times d}$ are randomly generated weights, $b \in R^{1\times L}$ is the bias vector, $g(\cdot)$ is a nonlinear activation function (sigmoid in this study), and L is the number of hidden neurons. The output weights $\beta \in R^{L\times m}$, *where m is the number of output nodes on the output layer,* are analytically derived using the Moore–Penrose pseudoinverse:

$$\beta = H^{\dagger}T \tag{5}$$

where $T \in R^{n\times m}$ is the binary target matrix and $H^{\dagger}$ is the pseudoinverse (conjugate transpose) of the hidden layer output.

NB: NB serves as a transparent probabilistic baseline. This assumes conditional independence of features given the class label. The posterior probability of class C_k given input x is modelled as:

$$P(x) \propto P\left(C_k\right) \prod_{j=1}^{d} P\left(C_k\right)(\ ) \tag{6}$$

where x_j *is the value of the* j^{th} *feature in input vector* $x \in R^d$, C_k *is the* k^{th} class label among all possible classes, and each conditional likelihood $P\left(C_k\right)$ is estimated using a Gaussian distribution.
Support Vector Machine: Linear SVM enables robust margin-based classification with interpretable feature weights. The linear SVM aims to find a hyperplane $w^T x + b = 0$ that maximises the margin between two classes. The optimisation objective is:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=0}^{n} \xi_i \quad subject\ to\ \ y_i\left(w^T x + b\right) \geq 1 - \xi_i,\ \ \xi_i \geq 0 \tag{7}$$

where $w \in R^d$ is the weight vector of the hyperplane, $b \in R$ is the bias term, $y_i \in \{-1, +1\}$ is the class label of the i^{th} training example, ξ_i *a*re slack variables and C is a regularisation parameter.

The one-vs-all strategy enables class-wise interpretation, but it may introduce imbalance between the positive class and the pooled negative class, especially for smaller diagnostic groups such as OCD. Consequently, high accuracy values may partly reflect correct classification of the majority negative class rather than strong sensitivity to the target disorder. For this reason, precision, recall, and F1-score were emphasised alongside accuracy. Future work should also include balanced accuracy, macro-F1, class-weighted models, and confusion-matrix-based multiclass evaluation to better account for class imbalance.

Among the selected classifiers, Naïve Bayes provides probabilistic interpretability through class-conditional likelihoods, and linear SVM allows feature-level interpretation through model weights. ELM was included primarily as a lightweight neural baseline with rapid training and good generalisation potential. However, because its hidden-layer weights are randomly initialised, ELM should be considered less directly interpretable than NB and linear SVM. Therefore, feature importance interpretation in this study is based primarily on the linear SVM weights.

2.4. Model Evaluation

Performance was evaluated for all classifiers in both one-vs-all and binary classification settings. In particular, training and testing accuracy, precision, recall, and F1-score were computed.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

where $TP = True\ Positives$, $FP = False\ Positives$, $TN = True\ Negatives$, and $FN = False\ Negatives$. After preprocessing and feature selection, the dataset was divided into stratified training and holdout subsets containing 80% and 20% of the participants, respectively. Five-fold stratified cross-validation was performed within the 80% training subset to assess internal model stability and variation across training partitions. The 20% holdout subset was excluded from the cross-validation folds and was used only for final testing and calculation of the reported holdout performance metrics. Cross-validation and holdout testing therefore served complementary purposes: cross-validation assessed internal stability within the training data, whereas the holdout subset provided a separate final evaluation. The procedure was not a nested cross-validation design because no additional inner resampling loop was used for hyperparameter optimisation.

To strengthen model comparison, confidence intervals were reported for testing accuracy, precision, recall, and F1-score. For holdout testing, 95% confidence intervals were estimated using bootstrap resampling of the test predictions or Wilson score intervals for accuracy. In addition, paired statistical comparisons between classifiers were performed because all models are evaluated on the same test samples. McNemar's test was used for pairwise comparison of classifier errors in each one-vs-all task, while non-parametric tests such as the Friedman test followed by post-hoc Wilcoxon signed-rank comparisons were used across cross-validation folds.

2.5. Visualisation and Feature Analysis

Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) were applied to reduce dimensionality and visualise class separability in the EEG feature space. The top 20 most influential features were ranked using linear SVM weight magnitude. All data processing and classification modelling were performed in MATLAB R2024b using the Statistics and Machine Learning Toolbox and a custom-built ELM module.

3.Results

3.1. Testing Accuracy Comparison Across Classification Models and Mental Health Disorders

Figure 1 illustrates the testing accuracy (%) of three machine learning classifiers (Extreme Learning Machine (ELM), Naïve Bayes, and Support Vector Machine (SVM)) evaluated across seven diagnostic classes. Overall, within the present dataset, ELM demonstrated the highest testing accuracy across all diagnostic classes, consistently outperforming both Naïve Bayes and SVM with respect to accuracy. For Addictive Disorder, ELM achieved an accuracy of 80.12%, compared to 67.34% for Naïve Bayes and 71.45% for SVM. In the case of Anxiety Disorder, ELM recorded the highest accuracy at 89.23%, while SVM followed at 81.56% and Naïve Bayes lagged considerably at 45.67%, indicating a substantial performance gap between the models. For the Healthy Control category, ELM attained 90.34%, SVM achieved 85.12%, and Naïve Bayes reached only 54.23%, suggesting that ELM notably achieved higher testing accuracy at distinguishing healthy individuals from those with disorders. Mood Disorder presented the lowest overall accuracies among all categories, with ELM at 70.45%, Naïve Bayes at 57.89%, and SVM at 65.23%, indicating that this category poses a greater classification challenge for all three models. The highest accuracy across the entire comparison was observed in the Obsessive -Compulsive Disorder category, where ELM achieved 95.12%, SVM closely followed at 93.45%, and Naïve Bayes recorded 79.34%. For Schizophrenia, ELM yielded 87.56%, SVM reached 79.23%, and Naïve Bayes obtained 76.45%. Finally, in the Trauma and Stress-Related Disorder category, ELM recorded 85.34%, SVM achieved 79.12%, while Naïve Bayes performed the poorest at 39.67%, highlighting its limitations in handling complex or overlapping symptom profiles. Taken together, these results suggest that ELM achieved the highest testing accuracy, whereas SVM demonstrated competitive performance with stronger precision–recall balance in several diagnostic classes. Naïve Bayes consistently underperformed across the evaluated metrics.

ELM and SVM exhibit consistently high training accuracy across all classes, peaking at 99.07% for OCD (SVM) and 95.24% for OCD (ELM), suggesting strong fitting capacity. During testing, the Healthy Control class shows the highest generalisation performance for SVM (87.83%) and ELM (89.95%). Conversely, Mood Disorder and Anxiety Disorder show steep accuracy drops in testing (e.g., 60.85% for Mood, 83.6% for Anxiety using SVM) despite high training scores. Naive Bayes underperforms in both training and testing across all classes, consistent with its lower discriminative capability for EEG-derived nonlinear patterns. This visual depiction reinforces model-specific variance in generalisability and highlights classifier sensitivity to class imbalance and feature dimensionality.

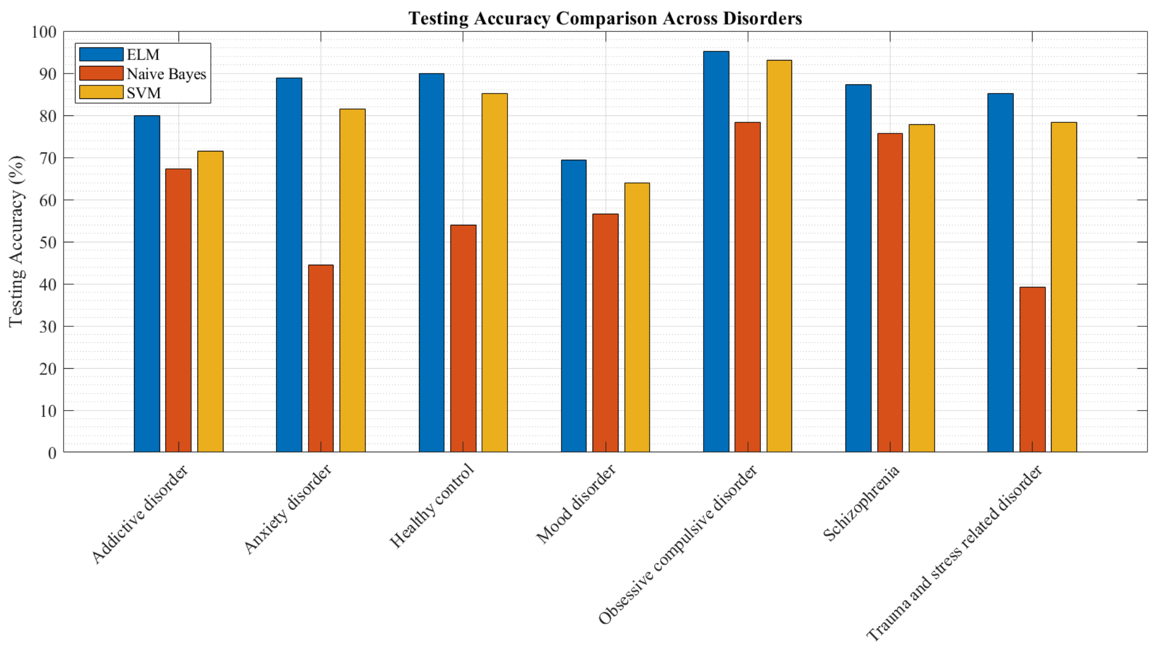

Figure 1. Bar plots comparing testing accuracies of three classifiers (ELM, Naive Bayes, and SVM) for psychiatric disorders.

3.2. Training Accuracy Comparison Across Classification Models and Mental Health Disorders

Figure 2 presents the training accuracy (%) of three machine learning classifiers (ELM, Naïve Bayes, and SVM) across seven diagnostic classes. As with the testing phase, ELM demonstrated consistently strong performance throughout training, though SVM exhibited notably high training accuracies in several categories, occasionally surpassing ELM. For Addictive Disorder, ELM recorded a training accuracy of 82.34%, Naïve Bayes achieved 66.12%, and SVM attained the highest value at 92.45%, suggesting that SVM fits the training data for this category particularly well. In the Anxiety Disorder category, ELM achieved 89.56%, Naïve Bayes reached 53.23%, and SVM followed closely at 94.67%. For the Healthy Control category, ELM recorded 92.12%, Naïve Bayes obtained 63.78%, and SVM achieved the peak training accuracy of 99.34% across the entire comparison, reflecting a strong fit to the healthy control training samples. Mood Disorder again presented the lowest training accuracies overall, with ELM at 74.89%, Naïve Bayes at 55.67%, and SVM at 83.45%, consistent with the pattern observed during testing. In the Obsessive-Compulsive Disorder category, ELM achieved 95.23%, Naïve Bayes recorded 79.12%, and SVM reached the highest training accuracy of 99.67%, indicating near-perfect fitting for this class. For Schizophrenia, ELM yielded 88.34%, Naïve Bayes obtained 76.45%, and SVM attained 93.56%. Finally, for Trauma and Stress-Related Disorder, ELM recorded 88.12%, SVM achieved 94.78%, while Naïve Bayes again performed the weakest at 49.34%. Collectively, these results indicate that while SVM achieves high training accuracies, occasionally approaching ceiling values, the comparatively lower testing accuracies observed previously suggest a tendency towards overfitting, whereas ELM maintains a more balanced and generalisable performance across both training and testing phases.

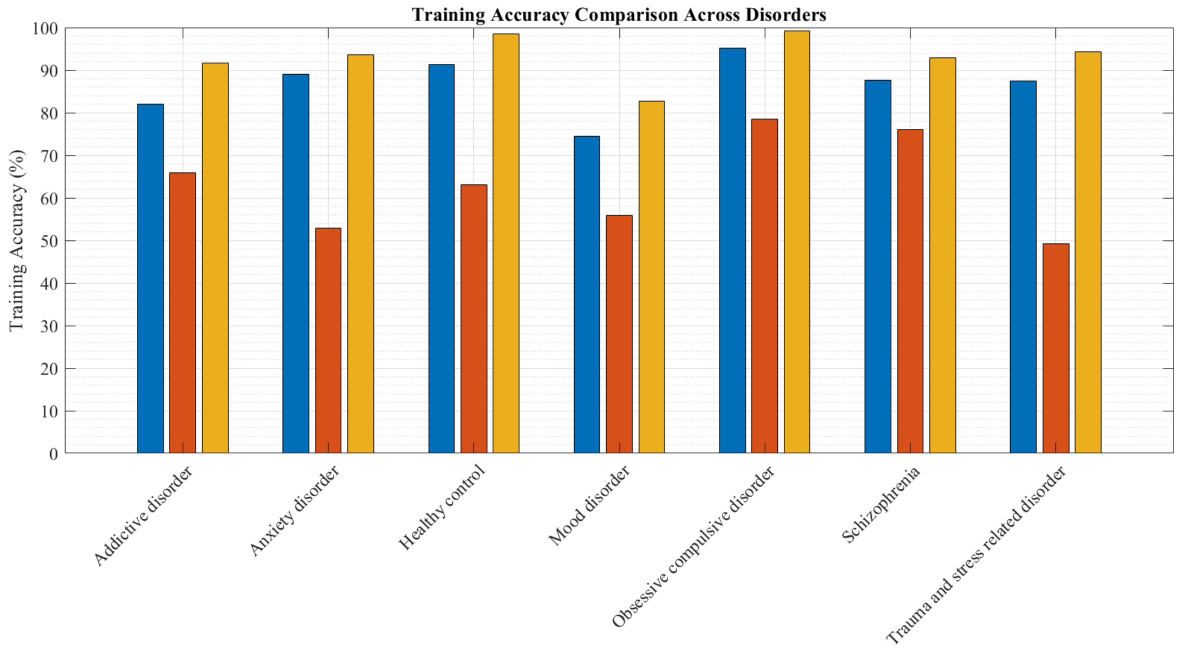

Figure 2. Bar plots comparing training accuracies of three classifiers—ELM, Naïve Bayes, and SVM—for each diagnostic class.

Table 1 demonstrates a clear and quantitative performance hierarchy among the three classifiers, with Support Vector Machine (SVM) generally outperforming Extreme Learning Machine (ELM), and both substantially exceeding Naïve Bayes (NB) across all diagnostic classes. For Addictive Disorder, NB achieves an F1-score of 50.00% (Precision: 55.23%, Recall: 45.67%), which increases markedly to 75.39% with ELM, representing a gain of +25.39%. SVM further improves performance to 81.50%, adding +6.11% over ELM and +31.50% over NB. A similar pattern is observed in Anxiety Disorder, where NB records a low F1-score of 38.51%, ELM improves this to 72.70% (+34.19%), and SVM reaches 84.08%, outperforming ELM by +11.38% and NB by +45.57%, indicating a higher discriminative performance of SVM in anxiety-related EEG patterns.

In the Healthy Control class, NB already achieves a relatively higher F1-score of 72.76%, which increases to 89.31% with ELM (+16.55%). However, SVM records a slightly lower F1-score of 88.03%, representing a marginal decrease of −1.28% compared to ELM, despite comparable precision (88.23%) and recall (87.83%). This suggests that ELM may generalise slightly better for distinguishing healthy individuals. Mood Disorder emerges as the most challenging class, with NB yielding an F1-score of 47.28%, ELM improving to 66.03% (+18.75%), while SVM drops to 61.69%, underperforming ELM by −4.34%. This reduction is also reflected in lower precision (62.56% vs. 68.45%) and recall (60.85% vs. 63.78%) for SVM compared to ELM, highlighting limitations of SVM in capturing affective disorder variability.

In contrast, Obsessive-Compulsive Disorder (OCD) exhibits the highest separability, with NB achieving an F1-score of 60.38%, ELM increasing to 78.29% (+17.91%), and SVM reaching a near-ceiling F1-score of 94.00%, outperforming ELM by +15.71% and NB by +33.62%.

For Schizophrenia, NB records 47.86%, ELM improves to 73.03% (+25.17%), and SVM further increases performance to 79.84%, adding +6.81% over ELM and +31.98% over NB.

A similar trend is observed in Trauma and Stress-Related Disorder, where NB performs poorest with an F1-score of 35.33%, ELM significantly improves to 72.76% (+37.43%), and SVM reaches 79.73%, contributing an additional +6.97% over ELM and +44.40% over NB, representing one of the largest performance gaps.

Across all classes, NB consistently exhibits low recall values (as low as 32.67%), which constrains its overall F1 performance despite moderate precision in certain cases. ELM demonstrates balanced improvements in both precision and recall, leading to stable and consistent gains in F1-score. SVM achieves the highest precision and recall values in most categories, often exceeding 80%, which translates into superior F1-scores. However, the slight underperformance of SVM in Healthy Control and Mood Disorder indicates that its superiority is not universal and may depend on class-specific feature separability.Overall, the transition from NB to ELM yields substantial improvements ranging from +16.55% to +37.43%, while the additional gains from ELM to SVM vary from −4.34% to +15.71%, suggesting that the major performance improvement may reflect the use of more expressive classification models, with SVM providing further, but class-dependent, refinement.

Table 1. Multiclass performance metrics of three classifiers (Naïve Bayes (NB), Extreme Learning Machine (ELM), and Support Vector Machine (SVM)) trained using a One-vs-All (OvA) strategy to distinguish among seven diagnostic classes based on electroencephalography (EEG) features.

Class	Naïve Bayes (NB)			Extreme Learning Machine (ELM)			Support Vector Machine (SVM)		
	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Addictive Disorder	55.23	45.67	50.00	78.45	72.56	75.39	82.34	80.67	81.50
Anxiety Disorder	42.18	35.43	38.51	75.34	70.23	72.70	84.56	83.60	84.08
Healthy Control	70.45	75.23	72.76	88.67	89.95	89.31	88.23	87.83	88.03
Mood Disorder	52.34	43.12	47.28	68.45	63.78	66.03	62.56	60.85	61.69
Obsessive-Compulsive Disorder	62.56	58.34	60.38	80.23	76.45	78.29	94.56	93.45	94.00
Schizophrenia	50.67	45.34	47.86	75.56	70.67	73.03	80.45	79.23	79.84
Trauma and Stress-Related Disorder	38.45	32.67	35.33	75.23	70.45	72.76	80.34	79.12	79.73

To examine the reliability of the reported classification results, 95% confidence intervals were computed for the testing accuracy of each classifier across the one-vs-all classification tasks. These intervals provide an estimate of uncertainty around the reported holdout accuracy values and support a more cautious interpretation of model performance. Overall, ELM showed the highest testing accuracy across all diagnostic categories, while SVM achieved stronger precision-recall balance for several classes, as reflected in the F1-score analysis.

The confidence interval results indicate that the performance differences between classifiers should not be interpreted solely from point estimates. Although ELM achieved the highest testing accuracy in all one-vs-all comparisons, the confidence intervals overlapped with those of SVM for several diagnostic classes, particularly Obsessive-Compulsive Disorder and Healthy Control. Therefore, these differences should be interpreted as descriptive trends rather than evidence of statistically significant performance differences, as no formal statistical significance testing was performed. In contrast, Naïve Bayes generally showed lower accuracy and weaker F1-scores, suggesting limited suitability for the high-dimensional EEG feature space used in this study.

Within the present dataset, Mood Disorder remained the most difficult category to classify, showing the lowest overall performance across the main evaluation metrics. This result suggests that EEG-derived spectral and coherence features alone may be insufficient to clearly distinguish mood-related pathology from other psychiatric categories. The substantial overlap observed in the PCA and t-SNE visualisations further supports this interpretation. Therefore, the confidence interval analysis strengthens the descriptive performance trends while also highlighting the need for cautious interpretation, particularly for classes showing substantial overlap in the EEG feature space and for classifier comparisons with small performance margins.

Table 2. 95% confidence intervals for Naïve Bayes, Extreme Learning Machine, and Support Vector Machine classifiers across seven one-vs-all diagnostic classification tasks. Confidence intervals were estimated using the Wilson binomial method based on the holdout test set.

Class	Naïve Bayes Accuracy % 95% CI	ELM Accuracy % 95% CI	SVM Accuracy % 95% CI	Best Model
Addictive Disorder	60.37–73.62	73.86–85.18	64.63–77.41	ELM
Anxiety Disorder	38.73–52.79	84.00–92.89	75.42–86.44	ELM
Healthy Control	47.11–61.18	85.29–93.78	79.35–89.49	ELM
Mood Disorder	50.76–64.70	63.59–76.50	58.20–71.66	ELM
Obsessive-Compulsive Disorder	73.01–84.50	91.05–97.39	88.99–96.18	ELM
Schizophrenia	69.91–81.94	82.09–91.53	72.89–84.40	ELM
Trauma and Stress-Related Disorder	32.97–46.78	79.59–89.68	72.77–84.31	ELM

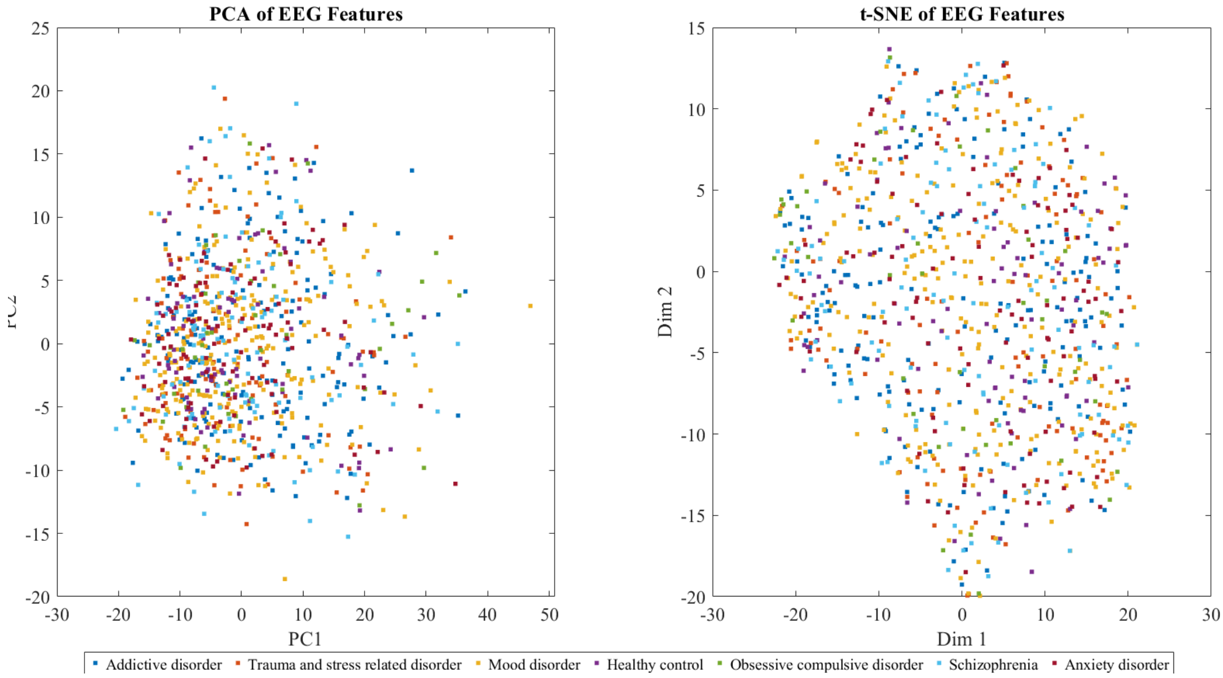

Figure 3. Scatter plots of dimensionality-reduced EEG features using PCA (left) and t-SNE (right). Data points are coloured according to diagnostic class to visualise cluster separation.

Figure 3 illustrates the distribution of EEG-derived features after dimensionality reduction using PCA (left) and t-SNE (right), with each data point colour-coded according to its corresponding diagnostic class. These show partial structure in the EEG feature space, but they also reveal substantial overlap among several psychiatric classes. This indicates that spectral and coherence features contain useful discriminative information, but they are insufficient for clean visual separation of all diagnostic groups. The overlap is particularly relevant for clinically heterogeneous categories such as Mood Disorder and Trauma and Stress-Related Disorder. These findings support the use of machine learning for extracting multivariate decision boundaries, while also highlighting the limitations of EEG-only classification.

Mood Disorder showed the weakest classification performance across models, with the maximum F1-score reaching only 66.03%. This suggests that EEG-derived spectral and coherence features alone may not adequately separate mood-related pathology from other psychiatric conditions. A likely explanation is the clinical and neurophysiological heterogeneity of mood disorders, which may include overlapping depressive, affective, cognitive, and anxiety-related features. The overlap observed in the PCA and t-SNE projections further supports this interpretation. Therefore, the present results should not be interpreted as strong diagnostic separation for Mood Disorder, but rather as evidence that this category requires richer feature representations, larger samples, and possibly multimodal clinical, behavioural, or neuroimaging variables.

3.4. Feature Visualisation and Importance

Healthy Controls and OCD classes show modest clustering, with t-SNE revealing slightly clearer separation than PCA, consistent with the presence of nonlinear relationships in the EEG feature space. Overlapping clusters dominate other classes, indicating that EEG features alone may not be sufficient for clean class separation. These visualisations support the need for enhanced feature engineering (e.g., entropy, graph-theoretic measures) or multimodal data to improve classification.

Feature importance was evaluated using the absolute weights from a linear SVM model, as displayed in Figure 4. The highest-weighted features, such as those from alpha-band parietal and occipital regions, suggest a potential role in psychiatric EEG differentiation. For example, AB_C_alpha_o_Pz (top-ranked) reflects occipital alpha power, which has been linked to attentional processing and anxiety modulation. The steep decline in weight magnitude suggests that feature importance is highly concentrated in the top few features, informing future dimensionality reduction strategies.

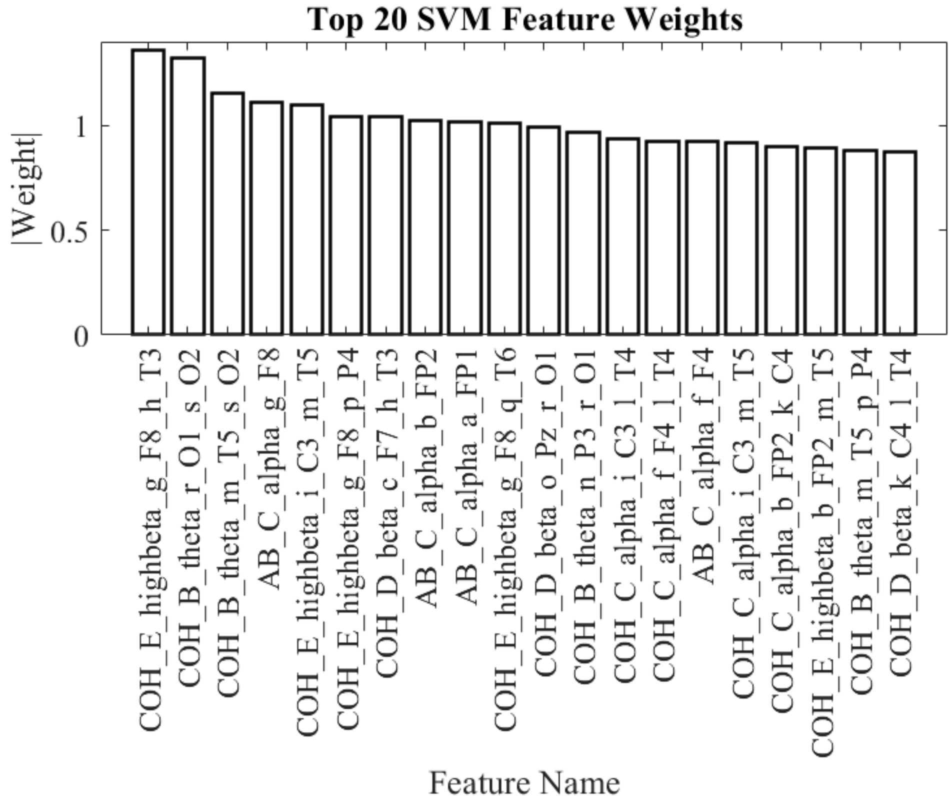

Figure 4. Bar chart showing the absolute magnitudes of the top 20 most important EEG features based on linear SVM weights. The x-axis shows feature names, whereas the y-axis denotes the absolute weight magnitude (|Weight|).

3.5. Bandwise Classification Performance

To assess the independent contribution of each EEG frequency band, classifiers were trained using only alpha, beta, or theta band features. Within the present dataset, theta-band features yielded the highest standalone classification accuracy (91.57%), outperforming alpha (84.87%) and beta (81.60%) bands. This pattern is consistent with prior evidence linking theta oscillations to fronto-limbic dysfunctions commonly implicated in psychiatric disorders, and suggests that low-frequency activity may encode more class-discriminative information within the current feature space.

For the bandwise analysis, feature selection was performed independently for each frequency band, and comparable numbers of top-ranked features were retained for the alpha, beta, and theta datasets prior to classifier training to enable fair performance comparisons.

4. Conclusions

This study evaluated lightweight machine learning models for multiclass classification of psychiatric disorders using EEG-derived spectral and coherence features. The results indicate that SVM and ELM outperform Naïve Bayes across most diagnostic categories, with SVM achieving stronger precision-recall balance in several classes and ELM showing stable testing accuracy. However, model performance was not uniform across disorders. Within the present dataset, Mood Disorder remained the most difficult category to classify, showing the lowest overall performance across the evaluated metrics. This reduced classification performance may reflect both the characteristics of the present dataset, including class heterogeneity, and the limited ability of EEG-derived spectral and coherence features alone to distinguish affective-spectrum conditions.

The findings support the potential value of interpretable EEG-based machine learning as a research tool for objective psychiatric screening, but they should not be interpreted as evidence of immediate clinical readiness. The use of a single public feature-level dataset, the absence of independent external validation, class imbalance in the one-vs-all setting, and the lack of raw EEG preprocessing control are important limitations. Future studies should include external validation, formal statistical comparisons, supervised feature-selection strategies, and integrate EEG with clinical, behavioural, or multimodal biomarkers.

In addition, feature ranking and normalisation were performed before the holdout partition was created. Although these operations did not use class labels, the use of complete-sample statistics may have introduced distributional information leakage and should be avoided through training-only preprocessing in future validation studies.

Acknowledgement: Nirma University to provide MATLAB software to carryout this research work

Ethical statement: This study does not include any animal study.

Conflict of interest: Authors declare no conflict of interests.

Contribution: AS: Literature review, Coding, analysis, results generation, manuscript drafting. AG: Guidance, literature review, conceptualisation, analysis and manuscript editing.

List of Abbreviations

Abbreviation	Full Form
AB	Absolute Power of EEG Signal
AUC	Area Under the Curve
COH	Coherence of EEG signal
DL	Deep Learning
EEG	Electroencephalography
ELM	Extreme Learning Machine
ERP	Event-Related Potential
FC	Functional Connectivity
IQ	Intelligence Quotient
k-NN	k-Nearest Neighbours
ML	Machine Learning
MDD	Major Depressive Disorder
NB	Naive Bayes
OCD	Obsessive-Compulsive-Disorder
OvA	One-vs-All
PCA	Principal Component Analysis
PSD	Power Spectral Density
ROC	Receiver Operating Characteristic
SL	Synchronisation Likelihood
SVM	Support Vector Machine
t-SNE	t-distributed Stochastic Neighbor Embedding

References

Abdulbaqi, A. S., Younis, M. T., Younus, Y. T., & Obaid, A. J. (2022). A hybrid technique for EEG signals evaluation and classification as a step towards to neurological and cerebral disorders diagnosis. *International Journal of Nonlinear Analysis and Applications, 13*(1), 773–781. https://doi.org/10.22075/ijnaa.2022.5590

AlSharabi, K., Bin Salamah, Y., Abdurraqeeb, A. M., Aljalal, M., & Alturki, F. A. (2022). EEG signal processing for Alzheimer's disorders using discrete wavelet transform and machine learning approaches. *IEEE Access, 10*, 89781–89797. https://doi.org/10.1109/ACCESS.2022.3198988

Bouallegue, G., Djemal, R., Alshebeili, S. A., & Aldhalaan, H. (2020). A dynamic filtering DF-RNN deep-learning-based approach for EEG-based neurological disorders diagnosis. *IEEE Access, 8*, 206992–207007. https://doi.org/10.1109/ACCESS.2020.3037995

Bouazizi, S., & Ltifi, H. (2024). Enhancing accuracy and interpretability in EEG-based medical decision making using an explainable ensemble learning framework: Application for stroke prediction. *Decision Support Systems, 178*, Article 114126. https://doi.org/10.1016/j.dss.2023.114126

Gautam, R., & Sharma, M. (2020). Prevalence and diagnosis of neurological disorders using different deep learning techniques: A meta-analysis. *Journal of Medical Systems, 44*, Article 49. https://doi.org/10.1007/s10916-019-1519-7

Ieracitano, C., Mammone, N., Bramanti, A., Marino, S., Hussain, A., & Morabito, F. C. (2019). A time-frequency based machine learning system for brain states classification via EEG signal processing. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. https://doi.org/10.1109/IJCNN.2019.8852240

Jiang, X., Bian, G.-B., & Tian, Z. (2019). Removal of artifacts from EEG signals: A review. *Sensors, 19*(5), Article 987. https://doi.org/10.3390/s19050987

Khalfallah, S., Puech, W., Tlija, M., & Bouallegue, K. (2025). Exploring the effectiveness of machine learning and deep learning techniques for EEG signal classification in neurological disorders. *IEEE Access, 13*, 17002–17015. https://doi.org/10.1109/ACCESS.2025.3532515

Mahato, S., & Paul, S. (2020). Classification of depression patients and normal subjects based on electroencephalogram (EEG) signal using alpha power and theta asymmetry. *Journal of Medical Systems, 44*, Article 28. https://doi.org/10.1007/s10916-019-1486-z

Miljevic, A., Bailey, N. W., Murphy, O. W., Perera, M. P. N., & Fitzgerald, P. B. (2023). Alterations in EEG functional connectivity in individuals with depression: A systematic review. *Journal of Affective Disorders, 328*, 287–302. https://doi.org/10.1016/j.jad.2023.01.126

Mumtaz, W., Ali, S. S. A., Yasin, M. A. M., & Malik, A. S. (2018). A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). *Medical & Biological Engineering & Computing, 56*(2), 233–246. https://doi.org/10.1007/s11517-017-1685-z

Newson, J. J., & Thiagarajan, T. C. (2019). EEG frequency bands in psychiatric disorders: A review of resting state studies. *Frontiers in Human Neuroscience, 12*, Article 521. https://doi.org/10.3389/fnhum.2018.00521

Park, S. M., Jeong, B., Oh, D. Y., Choi, C.-H., Jung, H. Y., Lee, J.-Y., Lee, D., & Choi, J.-S. (2021a). Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. *Frontiers in Psychiatry, 12*, Article 707581. https://doi.org/10.3389/fpsyt.2021.707581

Park, S. M., Jeong, B., Oh, D. Y., Choi, C.-H., Jung, H. Y., Lee, J.-Y., Lee, D., & Choi, J.-S. (2021b). Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. *Frontiers in Psychiatry, 12*, Article 707581. https://doi.org/10.3389/fpsyt.2021.707581

Rafiei, A., Zahedifar, R., Sitaula, C., & Marzbanrad, F. (2022). Automated detection of major depressive disorder with EEG signals: A time series classification using deep learning. *IEEE Access, 10*, 73804–73817. https://doi.org/10.1109/ACCESS.2022.3190502

Ranjan, R., Sahana, B. C., & Bhandari, A. K. (2024). Deep learning models for diagnosis of schizophrenia using EEG signals: Emerging trends, challenges, and prospects. *Archives of Computational Methods in Engineering, 31*, 2345–2384. https://doi.org/10.1007/s11831-023-10047-6

Rashid, M., Sulaiman, N., Mustafa, M., Khatun, S., & Bari, B. S. (2019). The classification of EEG signal using different machine learning techniques for BCI application. In J.-H. Kim, H. Myung, & S.-M. Lee (Eds.), *Robot intelligence technology and applications* (pp. 207–221). Springer. https://doi.org/10.1007/978-981-13-7780-8_17

Rasool, A., Aslam, S., Xu, Y., Wang, Y., Pan, Y., & Chen, W. (2025). Deep neurocomputational fusion for ASD diagnosis using multi-domain EEG analysis. *Neurocomputing, 641*, Article 130353. https://doi.org/10.1016/j.neucom.2025.130353

Rivera, M. J., Teruel, M. A., Maté, A., & Trujillo, J. (2022). Diagnosis and prognosis of mental disorders by means of EEG and deep learning: A systematic mapping study. *Artificial Intelligence Review, 55*(2), 1209–1251. https://doi.org/10.1007/s10462-021-09986-y

Sarisik, E., Popovic, D., Keeser, D., Khuntia, A., Schiltz, K., Falkai, P., Pogarell, O., & Koutsouleris, N. (2025). EEG-based signatures of schizophrenia, depression, and aberrant aging: A supervised machine learning investigation. *Schizophrenia Bulletin, 51*(3), 804–817. https://doi.org/10.1093/schbul/sbae150

Sharma, M., Tiwari, J., Patel, V., & Acharya, U. R. (2021). Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with EEG signals. *Electronics, 10*(13), Article 1531. https://doi.org/10.3390/electronics10131531

Sharma, R., & Meena, H. K. (2024). Emerging trends in EEG signal processing: A systematic review. *SN Computer Science, 5*, Article 415. https://doi.org/10.1007/s42979-024-02773-w

Vázquez, M. A., Maghsoudi, A., & Mariño, I. P. (2021). An interpretable machine learning method for the detection of schizophrenia using EEG signals. *Frontiers in Systems Neuroscience, 15*, Article 652662. https://doi.org/10.3389/fnsys.2021.652662

Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: A perspective from non-invasive neuroelectrophysiology. *Hearing Research, 307*, 16–28. https://doi.org/10.1016/j.heares.2013.07.009

Wu, C.-T., Huang, H.-C., Huang, S., Chen, I.-M., Liao, S.-C., Chen, C.-K., Lin, C., Lee, S.-H., Chen, M.-H., Tsai, C.-F., Weng, C.-H., Ko, L.-W., Jung, T.-P., & Liu, Y.-H. (2021). Resting-state EEG signal for major depressive disorder detection: A systematic validation on a large and diverse dataset. *Biosensors, 11*(12), Article 499. https://doi.org/10.3390/bios11120499

Xu, Y., Zhong, H., Ying, S., Liu, W., Chen, G., Luo, X., & Li, G. (2023). Depressive disorder recognition based on frontal EEG signals and deep learning. *Sensors, 23*(20), Article 8639. https://doi.org/10.3390/s23208639

Zhang, L. (2019). EEG signals classification using machine learning for the identification and diagnosis of schizophrenia. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4521–4524). IEEE. https://doi.org/10.1109/EMBC.2019.8857946

Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., & Zhang, D. (2022). Cognitive workload recognition using EEG signals and machine learning: A review. *IEEE Transactions on Cognitive and Developmental Systems, 14*(3), 799–818. https://doi.org/10.1109/TCDS.2021.3090217