Received: 5 September 2025 | Revised: 17 November 2025 | Accepted: 15 January 2026 | Published online: 20 March 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 4(2) 237–246
DOI: 10.47852/bonviewJDSIS62027587

BON VIEW PUBLISHING

Modeling Markers for Detection of Psychiatric Disorders Using EEG Signals

Steven Chris[1], Debasish Mishra[1,*], Lakshman Varanasi[2] and Varun Viswanathan[3]

[1] *Data Science and Information Systems, Krea University, India*

[2] *Biological Sciences, Krea University, India*

[3] *Psychology, Krea University, India*

Abstract: The diagnosis of mental (psychiatric) disorders is challenging, and there is a lack of consensus on objective diagnostic criteria that are based on definitive signs that accompany the disorder. There is a need, therefore, to develop objective tools for the examination of these disorders. We present here a novel machine learning (ML) approach that accurately identifies disorders. The approach uses electroencephalography (EEG) signals for diagnosis, which are processed to extract novel region based markers that are found to contain key information about the types of disorders. Subsequently, a support vector machine (SVM) classifier is modeled, integrated with sequential feature (marker) selection (SFS), which identifies optimal and compact marker subsets for disorder detection. The proposed system has been validated using a publicly available dataset. The developed model was benchmarked against existing models and was shown to perform superior to the models it was extensively compared with; it demonstrated a 98.33% accuracy in detecting obsessive-compulsive disorder (OCD). Our findings indicate that an accurate psychiatric diagnosis system can be achieved using EEG signals with significantly fewer, and more interpretable markers. This simpler and transparent approach improves the practicality and trustworthiness of AI/ML-driven diagnostic tools, making them more suitable for real-world clinical integration and understanding by medical professionals.

Keywords: disorder markers, medical diagnosis, mental health, psychiatry, well-being

1. Introduction

Health is a spectrum of overlapping states from perfect to semi-functional to defunct, and the “compass needle” of one’s physiological or psychological condition can continuously transition between these and other intermediate states [1]. The quality of life is measured along multiple dimensions, such as those in the WHOQoL questionnaire: physical health, psychological health, level of independence, social relationships, environment, and spirituality, religion, or personal beliefs. Of these, the hardest to assess is, arguably, psychological health (Figure 1 [2]).

Much effort has focused on identifying those who suffer from psychiatric disorders, and to then grade them in order of the intensity of the affliction or according to its nature. This is not trivial, as the mental or psychiatric disorder is often more complicated, for various reasons, than a physical ailment. To manage this complexity, psychiatry attempts to establish clear definitions and classifications, which involves naming the phenomena and then systematically grouping diverse characteristics or cases (classification) so that individual cases can be accurately assigned to a specific group (diagnosis) for effective scientific study and communication. Despite the foundational work aimed at standardizing psychiatric diagnoses, various schools of thought within psychiatric medicine continue to disagree radically on what qualifies as a mental illness, what its causes are, and how it can be best treated [3, 4].

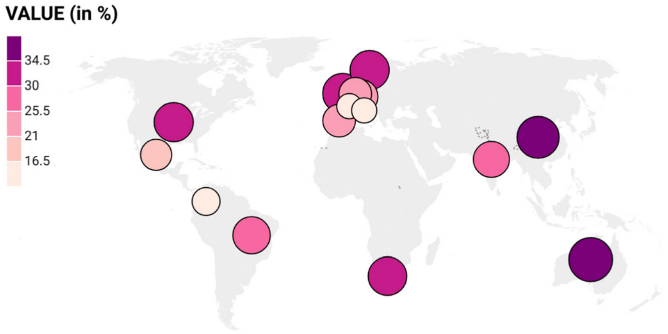
While early observations, such as Emil Kraepelin’s identification of dementia praecox in geographically and racially diverse populations (including Native Americans, African Americans, and Latin Americans) provided compelling evidence that some psychiatric conditions may be identifiable across racial and geographical boundaries, active debate continues to persist even a century later [5]. This ongoing discussion centers on the crucial role of culture in mental disorders and the extent to which biomedical psychiatric diagnoses are cross-culturally applicable [6].

With these complexities and disagreements, the criteria for diagnosing and classifying psychiatric illnesses and consequently, clinical decision-making, prognostication, and prediction are contested and constantly evolving. The current classification system is therefore provisional; although it is helpful for treatment and communication, it is not universal or definitive in the way classification systems for physical diseases often are. This issue is further compounded by the lack of clear, objective biological markers for psychiatric disorders (such as those found in blood tests or brain scans), unlike physiological conditions like cardiovascular disease and diabetes, which have established biomarkers [7]. It is in this context that the need for definitive and reliable biological indicators for the said purposes becomes significant.

Recent research efforts on this front have focused on studying electroencephalography (EEG) signals, which record the brain’s electrical activity through sensors placed on the scalp [8–10]. They are valuable because they capture unobtrusively and in real time brain activity related to thoughts, emotions, and mental states. Such signals offer a unique window into brain function, which physicians can explore to detect unusual patterns in brain waves. Further, the non-invasive characteristic of EEG signals allows for use in many clinical settings. Although EEG signals are effective in capturing brain-related data, the extraction of biomedically meaningful information from them is made difficult by concomitant signal artifacts or noise.

*Corresponding author: Debasish Mishra, Data Science and Information Systems, Krea University, India. Email: debasish.mishra@krea.edu.in

Figure 1
A recent study capturing adults in select countries experiencing mental health issues: Australia: 39%, China: 36%, United Kingdom: 32%, United States: 31%, Sweden: 31%, South Africa: 30%, Brazil: 28%, India: 26%, Germany: 22%, Spain: 21%, Netherlands: 21%, Mexico: 19%, Colombia: 15%, France: 12%, and Italy: 12%



Artificial intelligence (AI), including advanced machine learning (ML) and deep learning (DL) algorithms have been employed to understand complex patterns hidden in the EEG signals. Patterns so observed have utility, through AI/ML models, in the detection of various mental states. We observed that these complex ML/DL models were developed using raw (as-received) EEG signals, (or slightly processed signals) to identify features on their own. Although these complex ML/DL models have the capability for finding patterns in data, they suffer from their own drawbacks and challenges.

Our study is premised on the belief that for AI/ML approaches to truly help mental health diagnosis, they must not only be accurate, but also simple (minimally complex), transparent and interpretable (easy to understand), and empirically reproducible (robust). In this study, we present a decision-making model, accompanied by a ML model, that uses novel markers extracted from EEG signals to assist clinicians in the diagnosis and classification of various psychiatric conditions. We argue that applying these markers within a ML framework, can help clinicians detect and distinguish between various mental states, including anxiety, trauma/stress, addiction, obsessive-compulsive disorder (OCD), schizophrenia, and mood disorders. Our results demonstrate that by using relatively simple AI/ML systems and a minimal amount of brain wave data, we can achieve high diagnostic accuracy. This not only simplifies the process but it also reduces our reliance on complex systems of decision making.

Previous research has extensively explored the application of AI, from using ML and DL models, to analyze EEG signals to identify mental and psychiatric disorders [10–12]. For instance, several ML/DL algorithms, such as artificial neural networks (ANN), long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), convolutional neural networks (CNN), and a combination of CNN-LSTM architecture, have been employed to detect mental disorders from EEG data [13]. Researchers have developed a recurrent neural network (RNN) model to classify schizophrenia [14]. Other studies have modeled depression using XGboost, random forest, and CNN, where CNN model was found to be superior to the other models [15–17]. CNN models have also been explored to detect OCD [17]. Multi-layer ANN models have been used to detect general neurological disorders and suicidal tendencies from EEG signals [18, 19]. Unsupervised methods, such as the *k*-means algorithm, have also been applied successfully, in detecting major depressive disorder using EEG signals [20]. These reviews highlight the importance of EEG signals in the detection and diagnosis of psychiatric disorders [21, 22].

Literature reveals the use of complex DL models, such as LSTM, CNN, and RNN, to detect psychiatric conditions. It is also observed that these models typically rely on raw (as-received) or minimally processed EEG signals to identify clinical features. These complex DL models have the capability for finding features on their own, but they face significant challenges that limit their clinical applicability. The first challenge is the reliance of DL architectures on a large amount of data (features). This severely limits their clinical application. The second challenge is the inherent complexity of DL models. The internal logic and algorithm of these models are often opaque to researchers. This black box nature prevents reliable understanding and replication, making them difficult to trust or deploy them in a biomedical/clinical setting. Even with high accuracy, lack of transparency to know which clinical features or brain regions were instrumental in the diagnosis (known as lack of explainability) can limit their adoption as standard-of-care in clinical settings where clinicians need trustworthy mechanisms to explain and communicate diagnostic decisions. The third challenge is the computational cost of these models. Training these DL models requires significant computing power and time, which makes them expensive to develop.

The prevailing use of complex DL models (like CNN, LSTM, and RNN) for analyzing EEG signals faces major clinical barriers. To address these challenges of complex models, we present the following contributions in this study. To deal with the first challenge of reliance on massive data, we studied EEG signals and identified region based markers that are highly informative of mental states. These markers reduce the reliance on large datasets while capturing essential information required to assess mental state. To address the second challenge of lack of trust (explainability) in these models, we developed a support vector machine (SVM) model using the said markers. The SVM model demonstrated superior performance to the models it was benchmarked against; which ensures that the decision-making process is more interpretable and transparent—a critical factor for clinicians to develop trust and can pave the way for this to be adopted into clinical/biomedical settings as a standard-of-care. In order to address the third challenge of data burden, we conducted a sequential feature (marker) selection analysis, where we identified the most relevant set of markers that efficiently captures the essential information required to assess the clinical condition. This further minimizes the data input and makes the resulting ML system cheaper, faster, and more practical for clinical deployment.

2. Signal Acquisition

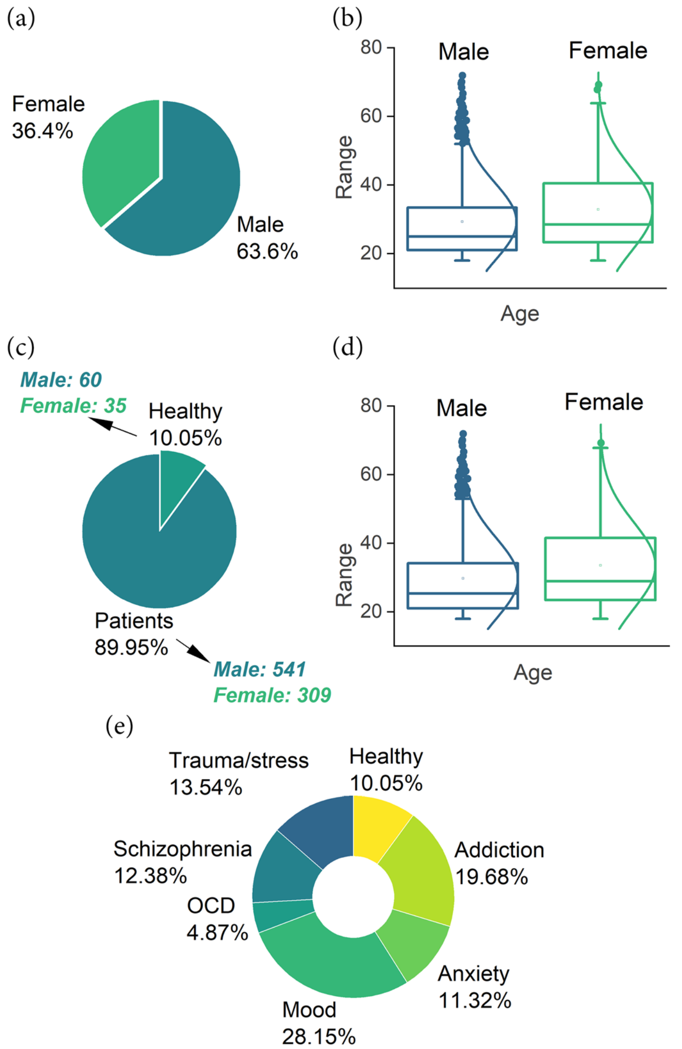
In this study, we used a public dataset that consists of EEG signals collected from 945 individuals, including 95 healthy individuals and 850 patients diagnosed with various mental (psychiatric) conditions [23]. All participants included in the study were between 18 and 70 years of age. Table 1 represents the central age of participants in this study. In addition, it was also ensured that none of the participants have a neurological disorder or injury in brain, neurodevelopmental disorder, intellectual disability, or borderline intellectual functioning, attention deficit hyperactivity disorder, or any neurocognitive disorder. Note the intellectual quotient listed in Table 1, which was utilized as a measure of any sort of neurological/neurodevelopmental disorder within the individuals. Individuals with intellectual quotient < 70 were excluded from the study. Moreover, the assessment of psychiatric disorders in the individuals was conducted comprehensively using their medical records, psychological assessments, and EEG signals. Next, the assessment of disorder was made by a psychiatrist using established standards (DSM-IV/DSM-5) and using the Mini-International Neuropsychiatric Interview. Figure 2 is a description of the cohort.

This study focuses on individuals diagnosed with six distinct mental health conditions: schizophrenia, mood disorders, anxiety, OCD,

Table 1
Participants information

Condition	Sample size	Age		Intellectual quotient	
		Mean	Standard deviation	Mean	Standard deviation
Healthy	95	25.72	4.55	116.24	10.94
Addiction	186	29.63	10.89	103.88	16.19
Anxiety	107	29.01	10.56	98.31	16.31
Mood	266	30.87	12.10	101.58	15.70
Obsessive-compulsive	46	28.48	9.83	107.8	15.24
Schizophrenia	117	31.73	12.10	89.62	17.51
Trauma/stress	128	36.09	13.82	98.89	15.86

Figure 2
Participants of this study: (a) percentage of male and female participants, (b) the range of age of the participants, (c) the proportion of healthy individuals and patients diagnosed with mental health issues, (d) the range of age of both male and female patients, and (e) the percentage of healthy individuals and patients diagnosed with various psychiatric disorders



addiction, and trauma/stress. Although varied in nature, these disorders fundamentally involve common struggles that impact the quality of daily life, such as difficulty in regulating emotions, disruptions in sleeping and eating habits, and disturbances to social, occupational and relational life. For instance, individuals diagnosed with schizophrenia experience delusions, which distorts the experience of reality and affects the ability to recognize what is real and what is not [24]. Mood disorders manifest as intense prolonged bouts of sadness, anger or elated mood, which can adversely affect one's functioning. In the case of anxiety disorders, people suffer from excessive fear, worries, or uncertainty in certain situations, which can trigger debilitating anxiety. Besides these conditions, some individuals may do things repeatedly, feeling compelled to engage in actions unnecessarily. This condition, where people are obsessed and feel compelled to repeat a behavior, is known as OCD. Individuals diagnosed with OCD suffer from repeated excessive unwanted thoughts known as obsessions. Often in response to the obsession, they feel compelled to engage and repeat a behavior. This cycle severely disrupts their everyday functioning. Addiction is another significant condition, where individuals are seen as having an uncontrollable urge to seek out substances or non-substances. In this dataset, addiction disorder classification includes individuals diagnosed with substance (alcohol) use. Finally, trauma/stress as a category includes individuals who have experienced traumatic past events that have left a significant negative impact on their lives, disturbing their lifestyle. In all these scenarios, it is important to identify and diagnose these conditions as early as possible to help these individuals restore their sense of well-being. The number of individuals diagnosed with each mental disorder (psychiatric) condition is also presented in Figure 2.

To collect data related to these mental states, EEG signals were recorded during a resting state for 5 minutes with the participants' eyes closed. The recordings were acquired using a Neuroscan system (Scan 4.5, NeuroScan), at an acquisition rate ranging from 500 to 1000 Hz and a filter set between 0.1 and 100 Hz. Nineteen (19) standard EEG channels were used (Figure 3). The instrument included a mastoid reference electrode and a ground channel positioned between the Fz and FPz electrodes, and the electrode impedances were maintained below 5 kΩ.

The raw EEG signals were downsampled to a frequency of 128 Hz. The Fast Fourier Transform (FFT) was then applied to convert the time-domain EEG signals into the frequency domain. The dataset includes two parameters for each of the 19 channels across 6 standard frequency bands (Table 2). These parameters are the power spectral density (PSD) and functional connectivity (FC). The PSD quantifies the power distribution of EEG signals, which means it measures the strength of brain activity within specific frequency bands. The FC measures the synchronization of EEG signals, which indicates the functional relationships between the electrical activities of two distinct regions of the brain. A high FC value indicates that two areas of brain are working together in a coordinated manner at a particular frequency. This experimentation process resulted in a dataset of the dimension: 1×1140 per subject. This includes PSD of 19 channels and 171 FC pairs, calculated across the 6 frequency bands (Table 2).

Figure 3
EEG electrode positions: EEG signals were acquired through 19 channels, which are FP1, FP2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2

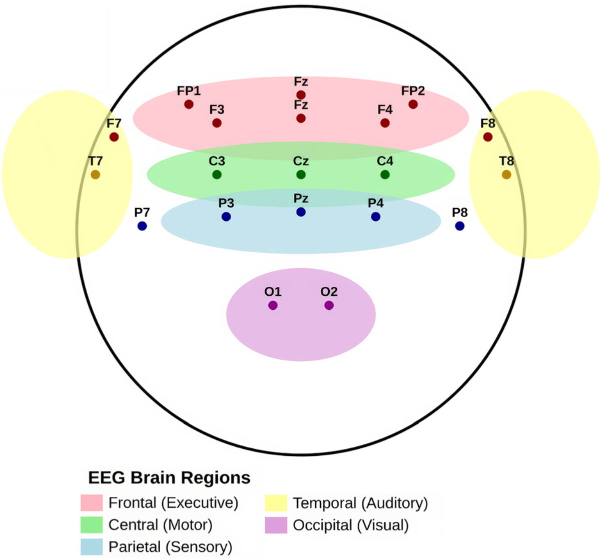


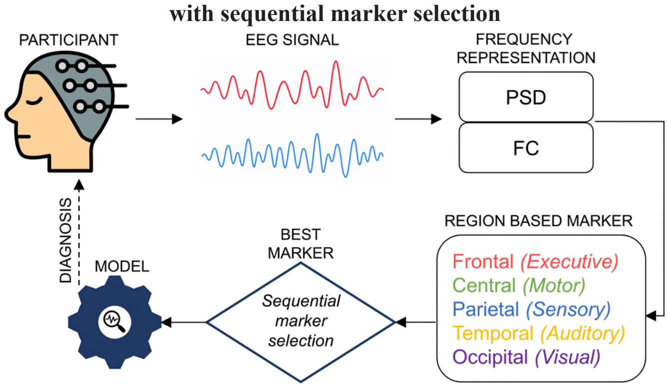
Table 2
EEG frequency bands

Channel	Frequency range
Delta	0.5–4 Hz
Theta	4–8 Hz
Alpha	8–13 Hz
Beta	13–22 Hz
High beta	22–30 Hz
Gamma	30–50 Hz

3. Methodology

Figure 4 schematically presents our overall approach designed to assess the mental state of individuals. Each step of this approach is discussed in this section.

Figure 4
Schematic diagram of modeling approach designed for mental state assessment. EEG signals are processed to calculate markers, which are subsequently evaluated in an optimized SVM model with sequential marker selection



3.1. Extraction of markers

As mentioned, it is important to derive information from the EEG signals that are relevant to mental states. To extract this comprehensive information from the signals, first we grouped them into the following regions of the brain: (a) frontal (comprising of FP1, FP2, F3, F4, F7, F8, and Fz channels); (b) central (comprising of C3, C4, and Cz channels); (c) temporal (comprising of T3, T4, T5, and T6 channels); (d) parietal (comprising of P3, P4, and Pz channels); and (e) occipital (comprising of O1 and O2 channels). Second, we computed the statistical markers from these selected regions and across the frequency bands for each individual. The statistical markers are mean, variance, skewness, and kurtosis. The mathematical expressions of these statistical markers are presented in Equations (1)–(4), respectively.

Consider each region of the brain represented by (r), frequency band of the EEG signal by (b), and the number of electrodes represented as ($N_{r,b}$), then the statistical measure, mean, can be expressed as:

$$\overline{F}_{r,b} = \frac{1}{N_{r,b}}\sum_{i=1}^{N_{r,b}} F_{r,b,i} \tag{1}$$

where, ($\overline{F}_{r,b}$) denotes the mean of the EEG signal for region (r) and frequency band (b), ($N_{r,b}$) is the number of electrodes used to measure the EEG signals for region (r) and frequency band (b), and ($F_{r,b,i}$) represent the PSD value of the EEG signal from the (i)th electrode in region (r) and frequency band (b). The next statistical measure calculated is the variance ($S^2_{r,b}$), which is expressed in Equation (2).

$$S^2_{r,b} = \frac{1}{N_{r,b}}\sum_{i=1}^{N_{r,b}}\left(F_{r,b,i} - \overline{F}_{r,b}\right)^2 \tag{2}$$

Finally, the statistical measures kurtosis ($K_{r,b}$) and skewness ($SK_{r,b}$) were calculated, which are presented in Equation (3) and Equation (4), respectively. Note that ($S_{r,b}$) in the skewness and kurtosis mathematical expressions denote the standard deviation, which is determined from the variance, as shown in Equation (5).

$$SK_{r,b} = \frac{\frac{1}{N_{r,b}}\sum_{i=1}^{N_{r,b}}\left(F_{r,b,i} - \overline{F}_{r,b}\right)^3}{\left(S_{r,b}\right)^3} \tag{3}$$

$$K_{r,b} = \frac{\frac{1}{N_{r,b}}\sum_{i=1}^{N_{r,b}}\left(F_{r,b,i} - \overline{F}_{r,b}\right)^4}{\left(S_{r,b}\right)^4} - 3 \tag{4}$$

$$S_{r,b} = \sqrt{S^2_{r,b}} \tag{5}$$

These statistical measures provide information about the distribution characteristics of brain power, which represents how concentrated or spread out the power is with respect to mental states.

Next, we computed the mean FC across all electrode pairs (spanning two regions) for every pair of defined brain regions. Let ($F_{r_1,r_2,j}$) denote the FC value for electrodes in regions (r_1) and (r_2), and (N_{r_1,r_2}) represents the number of such pairs. Then, the mean can be expressed as:

$$\overline{FC}_{r_1,r_2} = \frac{1}{N_{r_1,r_2}}\sum_{j=1}^{N_{r_1,r_2}} F_{r_1,r_2,j} \tag{6}$$

Finally, we computed the ratio (R) of the mean values determined for various frequency bands within each region. An expression is shown below, as an example, for calculating the ratio related to alpha and theta bands:

$$R_{r,\,alpha/theta} = \frac{\overline{FC}_{r,alpha}}{\overline{FC}_{r,theta}} \tag{7}$$

3.2. Selection of best markers

In this step, we determine the best set of markers required to assess a mental state. To achieve this, we used a sequential feature selection process. The sequential feature selector (SFS) algorithm selects an optimal subset of features from a range of features that maximize the performance of a model. In this study, we used the forward sequential feature selection approach, which iteratively adds the most meaningful marker to the current subset until a desired number of markers is reached, resulting in a compact yet discriminative set.

Let (S) be the set of available markers, and $P(X, y)$ denote the performance metric of the model trained on markers (X) and target (y). The objective of SFS is to find a subset of markers ($X_k^* \subset S$) with cardinality ($\left|X_k^*\right| = k$) such that $P\left(X_k^*, y\right)$ is maximized. The selection process begins with an empty set of selected markers, which can be represented as: ($X_0 = \emptyset$). Next, at each step (j), from the set of currently unselected markers ($S \backslash X_{j-1}$), the algorithm identifies the marker (x_{add}) that when added to (X_{j-1}), yields the best performance ($P(X_{j-1} \cup \{x_{add}\}, y)$:

$$x_{add} = ar\,g \max_{x\,\in S\backslash X_{j-1}} P(X_{j-1} \cup \{x\}, y) \tag{8}$$

The set of selected markers is then updated as:

$$X_j = X_{j-1} \cup \{x_{add}\} \tag{9}$$

This process iteratively continues until the desired number of markers, (k), is selected. In this study, (k) was varied from 20 to 30. In addition, the performance of the model at each step was evaluated using a 3-fold cross-validation approach.

3.3. Modeling

In this step, the best set of markers was explored in a ML algorithm to model the mental state. We used the SVM algorithm, which constructs an optimal hyperplane that separates data points into different classes [25]. Mathematically, for a given data sample (set of markers, (x)), the model (SVM) assigns it to a class, which is given by:

$$f(x)\mathrm{sgn}\left(\sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b\right) \tag{10}$$

where, $f(x)$ is the label predicted by the model for the given input (x), (α_i) denotes the weight learned by the model during training, (y_i) is the class label of support vectors, (K) is the kernel function, and (b) is the bias. Prior to modeling, any imbalance in the dataset was addressed. Imbalance is often found in real-world problems, and medical diagnosis is one such data-scarce area, where the number of patients with a specific disorder (mental condition) might be significantly lower than healthy individuals. To deal with this problem, the synthetic minority over-sampling technique was employed. This method generates synthetic samples for the minority class, resulting in balancing the class distribution and preventing the model from becoming biased towards the majority class [26, 27]. For a minority class sample (x_i), the method generates a synthetic sample (x_{new}) along the line segment joining (x_i) and one of its (k) nearest neighbors (x_{zn}):

$$x_{new} = x_i + \mathrm{rand}(0,1) \cdot (x_{zn} - x_i) \tag{11}$$

where rand(0,1) is a random number between 0 and 1. Note that the method was only applied to the training portion.

It should be noted that the PSD and FC parameters of EEG signals have different numerical ranges. Therefore, a few markers derived from them could dominate the learning process of the model. To avoid this, the dataset was normalized by transforming it to have a mean of 0 and a standard deviation of 1. This process weighs all markers equally to the model, preventing those with larger absolute values from disproportionately influencing the model's decision-making process. For a marker (x_i), its normalized value (x_i') is calculated as:

$$x_i' = \frac{x_i - \overline{x}}{S_x} \tag{12}$$

where ($\overline{x}$) is the mean and (S_x) is the standard deviation of marker (x). Finally, this normalized dataset was used for modeling, where 80% of the data was used for training the model, and the remaining 20% was reserved for testing. Note that the testing data was never a part of the training, which ensured model generalization. A binary classification strategy, where a distinct model was developed for each mental disorder against the set of healthy individuals, was created.

3.4. Model optimization

In this step, we optimized the model hyperparameters to enhance its performance, reduce dimensionality, and improve the interpretability of the results. The hyperparameters within a model can be thought of as configuration settings whose values are set before the learning process of the model begins. As a result, they are critical for ensuring the learning and performance of the model. To identify the optimal combination of these hyperparameters in this study, we performed an exhaustive search, where a predefined set of values for each hyperparameter was evaluated, and the model performance was assessed for each set. In this study, the hyperparameters studied are the regularization parameter (c) and the kernel (K). The regularization parameter ensures a smooth decision boundary required to classify the markers according to their respective mental state. This hyperparameter was varied from 0.1 to 10 with a step size of 1. Further, we evaluated two kernels, linear and radial basis functions, to apply the kernel transform to the data.

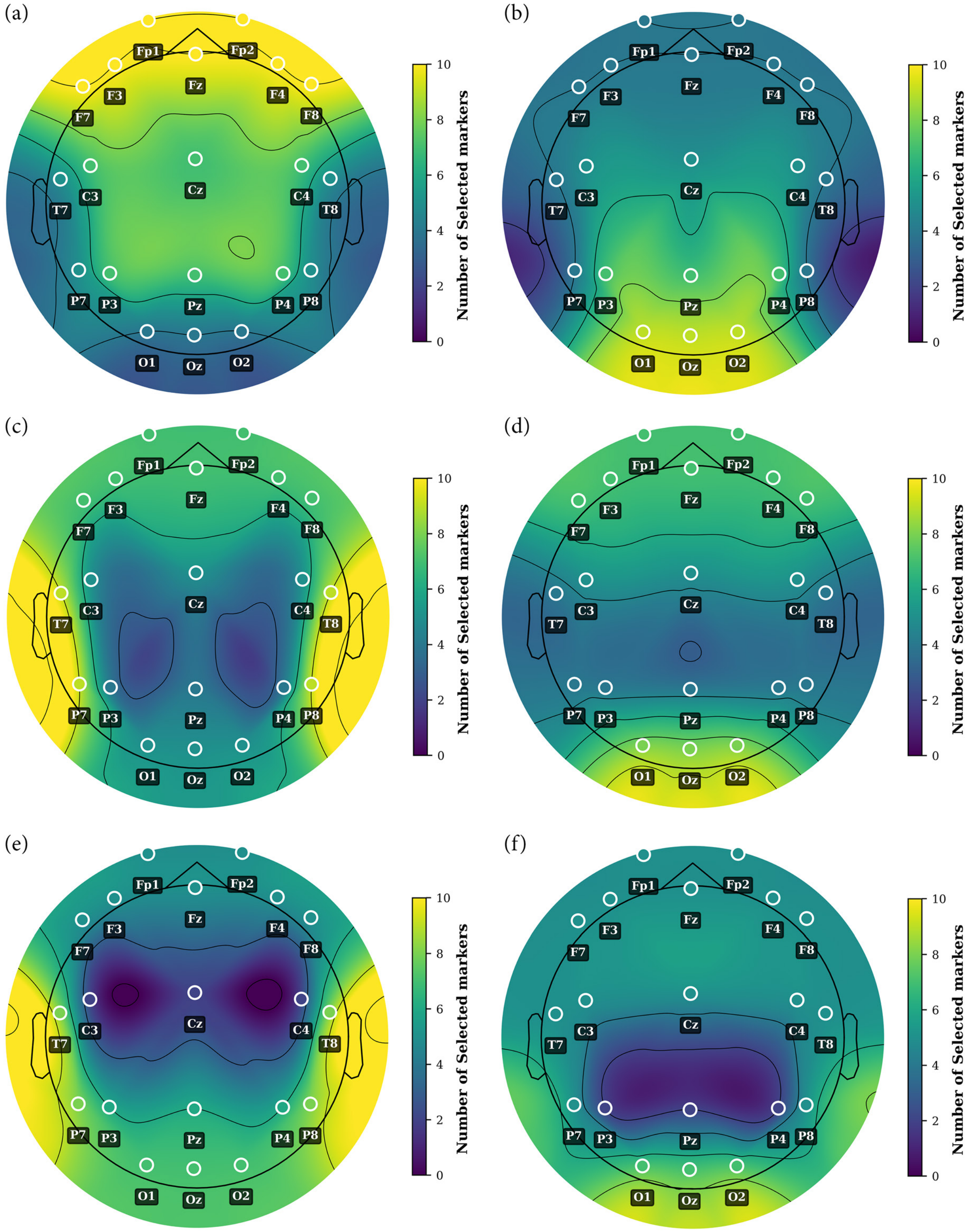
4. Results

In this section, we present the results obtained by applying the discussed methods.

4.1. Markers of mental states

Figure 5 presents a granular understanding of the spatial distribution of the computed markers for each mental disorder. These maps show the areas within the scalp where EEG alterations are most prominent and relevant to classify a specific mental condition. The color intensity of both the interpolated field and, critically, the individual electrode markers (represented by dots), indicates the prominence of markers from that area, with darker shades corresponding to a higher count of relevant markers. Furthermore, the labels for each electrode position (e.g., Fp1, Cz) are explicitly shown, allowing precise localization of these significant markers. This visual representation facilitates the identification of disorder-specific spatial signatures. For example, a strong concentration of selected markers over frontal electrodes might be observed for mood disorders, while a more pronounced pattern could appear in temporal or parietal regions for conditions like schizophrenia (Figure 5). These unique topographical patterns are crucial for hypothesizing localized neurophysiological

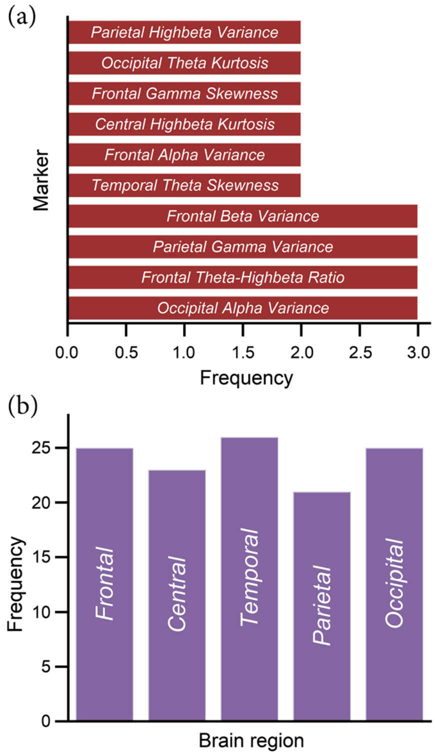
Figure 5
Distribution of selected EEG markers across brain regions for different mental (psychiatric) disorders. The color intensity represents the number of markers identified within each region for: (a) addiction, (b) trauma/stress, (c) mood, (d) obsessive-compulsive, (e) schizophrenia, and (f) anxiety. The electrode position labels provide precise anatomical localization



dysfunction, serving as compelling visual biomarkers that can help in clinical assessment and provide deeper insight into the underlying pathophysiology of each condition.

The analysis of discriminative EEG markers provides unique information on the neurophysiological underpinnings of mental (psychiatric) disorders. The 10 optimal markers most consistently selected, across all disorders, are presented in Figure 6(a), which highlights the most robust and generally relevant markers. Furthermore, the regional distribution from which these optimal markers are predominantly selected can be observed in Figure 6(b). This indicates the areas of the scalp that contribute significantly to diagnostic differentiation.

Figure 6
(a) The top 10 most frequently selected markers across all disorders and (b) the frequency of brain regions where the informative markers, presented in (a), are found relevant to disorders



4.2. Mental state assessment model

The derived markers were used to develop the SVM model to diagnose mental state. Figure 7(a) presents the impact of the selected markers (*k*) on the accuracy of the SVM model. Our initial analysis began with the examination of all markers calculated from the EEG signals. In this examination, it was identified that the model converged and accuracy saturated with approximately 30 markers; note the total of 30 markers presented in Figure 7(a). This analysis further revealed that higher precision was consistently achieved within the range of 20 to 30 markers in most mental (psychiatric) disorders. It is also noteworthy that the detection/diagnosis of mental state is less accurate with less than 20 markers.

Therefore, to further refine these findings and pinpoint optimal performance, a more granular SFS was subsequently performed, which specifically focused on the (*k*) markers from 21 to 30 with increments of 1. Results of this analysis are presented in Figure 7(b). It presents the optimal number of (*k*) markers and the corresponding precision of the model for each mental (psychiatric) disorder. Further, the optimal settings (markers and maximum accuracy achieved) for each mental (psychiatric) disorder are summarized in Table 3. The radial basis function was found to be the most optimum kernel for all disorders, with a (*c*) value of 10.

As presented in Table 3, the developed SVM model demonstrated strong classification performance in various disorders. The highest precision is achieved for OCD (0.9833), followed by trauma/stress disorder (0.9205) and anxiety disorder (0.9107). The mood disorder presented a more challenging classification task; nonetheless, this accuracy is still clinically significant, especially given the reduced feature set. These results indicate the significance of the derived markers in diagnosing mental disorders.

Figure 7
Model sensitivity with markers

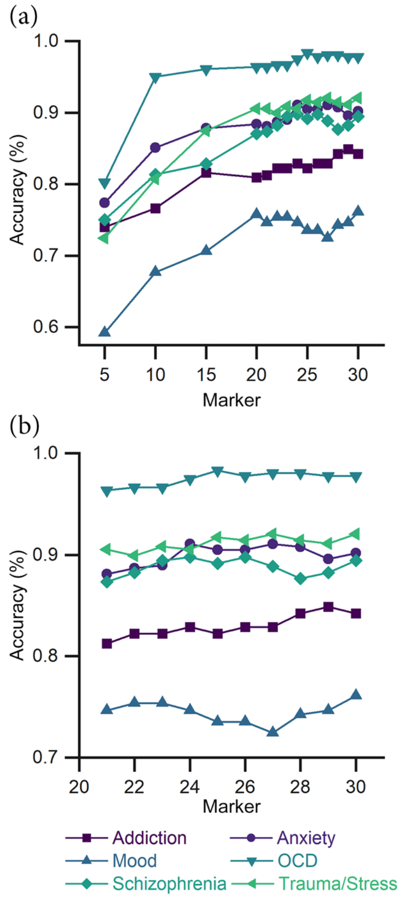


Table 3
Modeling results

Disorder	Optimal (*k*) markers	Accuracy
Addiction	29	0.8487
Anxiety	24	0.9107
Mood	30	0.7610
Obsessive-compulsive	25	0.9833
Schizophrenia	24	0.8976
Trauma/stress	27	0.9205

4.3. Model benchmarking

Clearly, the developed SVM model achieved superior accuracy with the derived markers that are significantly a small size of data. This developed diagnosis system makes the decision-making process simpler and efficient compared to the complex ML/DL models, developed in the previous research works. It is also important to recall that the models developed in prior research often relied on a significantly larger number of EEG signals (1140 features), while our approach uses only 20–30 markers that are meaningful to the mental disorders. Table 4 benchmarks our approach with prior research reports. It can be observed that the model developed in this study consistently achieved higher accuracy in diagnosing disorders using a highly reduced set of informative markers, when compared to the complex DL models developed in previous studies.

Table 4
Benchmarking

Disorder	Model	Features used	Accuracy	Ref.
Addiction	ANN	Entire band	0.7830	[13]
	Bi-LSTM	Entire band	0.8150	[13]
	CNN-LSTM	Entire band	0.8200	[13]
	KNN	Entire band	0.7940	[13]
	LSTM	Entire band	0.7780	[13]
	Elastic Net	Theta band	0.8566	[23]
	This study	29 markers	0.8487	-
Anxiety	ANN	Entire band	0.8730	[13]
	Bi-LSTM	Entire band	0.8680	[13]
	CNN-LSTM	Entire band	0.8840	[13]
	KNN	Entire band	0.8890	[13]
	LSTM	Entire band	0.9150	[13]
	Elastic Net	PSD	0.9103	[23]
	This study	24 markers	0.9107	-
Mood	ANN	Entire band	0.6510	[13]
	Bi-LSTM	Entire band	0.7300	[13]
	CNN-LSTM	Entire band	0.7040	[13]
	KNN	Entire band	0.6880	[13]
	LSTM	Entire band	0.7720	[13]
	Elastic Net	Theta FC	0.8926	[23]
	CNN	Entire band	0.7000	[28]
	Bi-LSTM	Entire band	0.7600	[28]
	This study	30 markers	0.7610	-
Obsessive-compulsive	ANN	Entire band	0.9680	[13]
	Bi-LSTM	Entire band	0.9310	[13]
	CNN-LSTM	Entire band	0.9210	[13]
	KNN	Entire band	0.9470	[13]
	LSTM	Entire band	0.9360	[13]
	Elastic Net	Gamma FC	0.7452	[23]
	CNN	Entire band	0.7100	[28]
	Bi-LSTM	Entire band	0.8200	[28]
	This study	25 markers	0.9833	-
Schizophrenia	ANN	Entire band	0.8360	[13]
	Bi-LSTM	Entire band	0.8840	[13]
	CNN-LSTM	Entire band	0.8620	[13]
	KNN	Entire band	0.8410	[13]
	LSTM	Entire band	0.8470	[13]
	Elastic Net	Alpha PSD	0.9383	[23]
	CNN	Entire band	0.7500	[28]
	Bi-LSTM	Entire band	0.6600	[28]
	This study	24 markers	0.8976	-
Trauma/stress	ANN	Entire band	0.8780	[13]
	Bi-LSTM	Entire band	0.8620	[13]
	CNN-LSTM	Entire band	0.8680	[13]
	KNN	Entire band	0.8620	[13]
	LSTM	Entire band	0.8680	[13]
	Elastic Net	Beta FC	0.9121	[23]
	This study	27 markers	0.9205	-

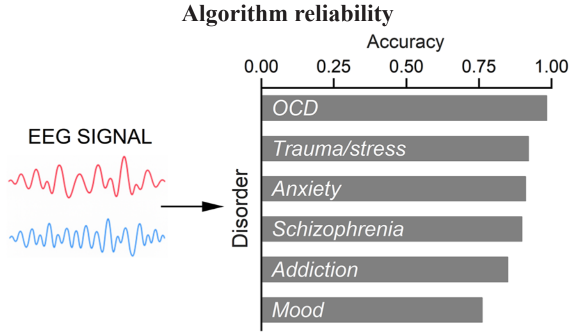
For addiction, obsessive-compulsive, schizophrenia, and trauma/stress disorders, the model significantly outperformed all DL and KNN/ANN models developed in literature. In the case of anxiety disorder, the developed model achieved an accuracy that is equivalent to the LSTM model. However, while the LSTM model operates on a huge amount of data, our model achieved it with a small set of markers. In the case of mood disorder, although our model achieved a slightly lower accuracy than the LSTM model, the difference in accuracy is minimal. In addition, it is noteworthy that our model uses far fewer data (relevant markers) in comparison and is more interpretable.

Apart from these studies, a recent study reported experiments conducted with 20 participants, where 10 participants were diagnosed with OCD and the remaining 10 were healthy participants [17]. Resting-state EEG signals were collected from these participants to distinguish between healthy controls and patients diagnosed with the disorder. The CNN algorithm was modeled using time-frequency representations (images) of the EEG signals derived using wavelets. The accuracy for this study was reported as 0.8200. In our study (with a different dataset), we have achieved a greater accuracy of 0.9833, just by using fewer (25) interpretable markers.

5. Discussion

The high accuracy achieved by the developed algorithm highlights its reliability to be used as a decision support system for real-world clinical screening. As presented in Figure 8, most disorders are detected by the algorithm with exceptional accuracy. This indicates the generalizability of the model to detect various mental (psychiatric) disorders.

Figure 8
Algorithm reliability



Mood disorder exhibits relatively lower reliability, even after the extraction of meaningful information, including the selection of markers, and optimization of model. While we acknowledge the different school of thoughts on this issue, we present our point of view. First, we are intrigued by our observation about the severity of mood disorder. Individuals with this condition (particularly, mood disorder) may exhibit overlapping symptoms, such as anger or sadness, or might experience emotional instability (depression). These are probable consequences that co-exist with other disorders, such as anxiety, fear, or stress. Therefore, the severity (and deviation) of health of an individual with mood disorder may not differ significantly from the healthy controls. The overlapping conditions may make it more difficult to distinguish mood states through an algorithmic approach. Second, we also wonder whether resting-state EEG signals alone could be the best indicator of mood disorder, given the involvement of dynamics.

To address these questions, future work will focus on expanding the dataset to include a larger and a more diverse population, as well as additional physiological/behavioral predictors to detect the severity (magnitude) of mood disorder.

Overall, the algorithm uses a subset of the data-markers that its peers do, to achieve comparable or superior outcomes. It achieves an economy of computation and provides a more objective set of criteria for diagnosis than has hitherto been possible; the implications are of value and bode well for psychiatric care. There is a caveat—the electroencephalograph required for EEG measurements is expensive and precludes the adoption of the device and algorithm as the current standard of care. This may change in the future though as the technology matures and becomes more affordable. Whether or not the same markers can be used for prediction or prognostication will be evident only when the appropriate dataset is available.

6. Conclusion

Mental health and physical health are intertwined states—two sides of the same coin—of any healthy individual. Over the years, they have been perceived as separate states; however, they move in tandem. Scholars advocate that the quality of life of an individual can be measured by assessing how well the mental health status of that person is. This being true, any average individual today primarily opts for blood tests (or the like) to monitor physical health as they transition through their life stages, (which in our study are the following stages, i.e., perfect, semi-functional, to non-functional). It is worth noting, from past research, that problems in physical health lead to issues in mental health and vice versa. The number of people globally with mental health issues has burgeoned, and the need for measuring tools to assess mental health statuses has never been more urgent. This identification is important for physicians and clinical researchers, including an average person, as it can assist in the development of targeted interventions, diagnoses, and awareness.

To this end, this study presents a novel interpretable ML approach for mental (psychiatric) disorder diagnosis using EEG signals. We derived region based informative markers from EEG signals and implemented sequential marker selection to achieve superior diagnostic performance using significantly fewer data. Such markers could potentially expedite a communal consensus. The key contribution of our work lies in demonstrating that high diagnostic accuracy can be achieved through interpretable ML with a reduced set of markers. This approach addresses the critical challenges in mental (psychiatric) diagnosis, which are: (1) lack of objectivity in traditional diagnosis methods, (2) computational complexity and lack of interpretability in DL models, and (3) the reliance on excessive amounts of potentially irrelevant data.

In the future, we aim to implement this approach in prospective clinical trials in real-world healthcare settings. In addition, we plan to extend this work to include multi-modal data, such as heart rate variability, which could enhance diagnostic accuracy and provide more insights.

Funding Support

This study was supported by the Office of Research Support at Krea University, bearing grant number: 25-26-120175.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in OSF repository at https://osf.io/8bsvr/.

Author Contribution Statement

Steven Chris: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Debasish Mishra:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision. **Lakshman Varanasi:** Methodology, Resources, Writing – review & editing, Supervision, Funding acquisition. **Varun Viswanathan:** Methodology, Writing – review & editing, Supervision.

References

[1] White, R. L., Vella, S., Biddle, S., Sutcliffe, J., Guagliano, J. M., Uddin, R., …, & Teychenne, M. (2024). Physical activity and mental health: A systematic review and best-evidence synthesis of mediation and moderation studies. *International Journal of Behavioral Nutrition and Physical Activity*, *21*(1), 134. https://doi.org/10.1186/s12966-024-01676-6

[2] Statista. (2022) *Share of adults in select countries worldwide currently experiencing mental health conditions (e.g. depression, anxiety) as of 2022*. https://www.statista.com/statistics/1428794/adults-with-select-countries-mental-health-issues-worldwide-by-country/?srsltid=AfmBOopOYrr-WmgvyFZy53OUDejimpfKQinjfEESTD51midYWKvdnDjn

[3] Doan, T., Ha, V., Strazdins, L., & Chateau, D. (2023). Healthy minds live in healthy bodies – Effect of physical health on mental health: Evidence from Australian longitudinal data. *Current Psychology*, *42*(22), 18702–18713. https://doi.org/10.1007/s12144-022-03053-7

[4] Mahindru, A., Patil, P., & Agrawal, V. (2023). Role of physical activity on mental health and well-being: A review. *Cureus*, *15*(1), e33475. https://doi.org/10.7759/cureus.33475

[5] Jilek, W. G. (1995). Emil Kraepelin and comparative sociocultural psychiatry. *European Archives of Psychiatry and Clinical Neuroscience*, *245*(4), 231–238. https://doi.org/10.1007/BF02191802

[6] Summerfield, D. (2008). How scientifically valid is the knowledge base of global mental health? *BMJ*, *336*, 992. https://doi.org/10.1136/bmj.39513.441030.AD

[7] Yun, S. (2024). Advances, challenges, and prospects of electroencephalography-based biomarkers for psychiatric disorders: A narrative review. *Journal of Yeungnam Medical Science*, *41*(4), 261–268. https://doi.org/10.12701/jyms.2024.00668

[8] Tashakori, M., Rusanen, M., Karhu, T., Huttunen, R., Leppänen, T., & Nikkonen, S. (2025). Optimal electroencephalogram and electrooculogram signal combination for deep learning-based

sleep staging. *IEEE Journal of Biomedical and Health Informatics*, *29*(7), 4741–4747. https://doi.org/10.1109/JBHI.2025.3541453

[9] Boubchir, L., Daachi, B., & Pangracious, V. (2017). A review of feature extraction for EEG epileptic seizure detection and classification. In *2017 40th International Conference on Telecommunications and Signal Processing*, 456–460. https://doi.org/10.1109/TSP.2017.8076027

[10] de Bardeci, M., Ip, C. T., & Olbrich, S. (2021). Deep learning applied to electroencephalogram data in mental disorders: A systematic review. *Biological Psychology*, *162*, 108117. https://doi.org/10.1016/j.biopsycho.2021.108117

[11] Uyanik, H., Sengur, A., Salvi, M., Tan, R.-S., Tan, J. H., & Acharya, U. R. (2025). Automated detection of neurological and mental health disorders using EEG signals and artificial intelligence: A systematic review. *WIREs Data Mining and Knowledge Discovery*, *15*(1), e70002. https://doi.org/10.1002/widm.70002

[12] Sun, J., Dong, Q.-X., Wang, S.-W., Zheng, Y.-B., Liu, X.-X., Lu, T.-S., …, & Han, Y. (2023). Artificial intelligence in psychiatry research, diagnosis, and therapy. *Asian Journal of Psychiatry*, *87*, 103705. https://doi.org/10.1016/j.ajp.2023.103705

[13] Ahmed, Z., Wali, A., Shahid, S., Zikria, S., Rasheed, J., & Asuroglu, T. (2024). Psychiatric disorders from EEG signals through deep learning models. *IBRO Neuroscience Reports*, *17*, 300–310. https://doi.org/10.1016/j.ibneur.2024.09.003

[14] Supakar, R., Satvaya, P., & Chakrabarti, P. (2022). A deep learning based model using RNN-LSTM for the detection of schizophrenia from EEG data. *Computers in Biology and Medicine*, *151*, 106225. https://doi.org/10.1016/j.compbiomed.2022.106225

[15] Ksibi, A., Zakariah, M., Menzli, L. J., Saidani, O., Almuqren, L., & Hanafieh, R. A. M. (2023). Electroencephalography-based depression detection using multiple machine learning techniques. *Diagnostics*, *13*(10), 1779. https://doi.org/10.3390/diagnostics13101779

[16] Ahire, N. (2025). EEG-derived brainwave patterns for depression diagnosis via hybrid machine learning and deep learning frameworks. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2025.2457999

[17] Zaboski, B., Fineberg, S., Skosnik, P., Kichuk, S., Fitzpatrick, M., & Pittenger, C. (2026). Classifying obsessivecompulsive disorder from resting-state EEG using convolutional neural networks: A pilot study. *Computational Psychiatry*, *10*(1), 1–17. https://doi.org/10.5334/cpsy.149

[18] Zhao, Y., Guo, Y., Wu, D., Zhang, J., Zheng, S., Xie, W., …, & Tian, Y. (2025). Alpha oscillatory dysregulation: Mapping EEG oscillatory in suicidal depression. *Frontiers in Human Neuroscience*, *19*, 1582330. https://doi.org/10.3389/fnhum.2025.1582330

[19] Dursun, İ., Akın, M., Aluçlu, M. U., & Uyar, B. (2025). Multilayer neural-network-based EEG analysis for the detection of epilepsy, migraine, and schizophrenia. *Applied Sciences*, *15*(16), 8983. https://doi.org/10.3390/app15168983

[20] Li, Q., Detke, M., Paul, S. M., Potter, W. Z., Zhang, F., Breier, A., …, & Wang, K. (2025). *Machine learning-enabled EEG biomarkers predict divergent antidepressant and placebo response in a clinical trial of major depression*. medRxiv. https://doi.org/10.1101/2025.05.29.25328167

[21] Ghasemi, F., Shalbaf, A., & Esteki, A. (2025). Application of artificial intelligence in obsessive-compulsive disorder detection and response to treatment: A systematic review. *IEEE Access*, *13*, 62930–62960. https://doi.org/10.1109/ACCESS.2025.3558845

[22] García-Ponsoda, S., García-Carrasco, J., Teruel, M. A., Maté, A., & Trujillo, J. (2023). Feature engineering of EEG applied to mental disorders: A systematic mapping study. *Applied Intelligence*, *53*(20), 23203–23243. https://doi.org/10.1007/s10489-023-04702-5

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

[24] GBD 2019 Mental Disorders Collaborators. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

[25] Mypati, O., Mukherjee, A., Mishra, D., Pal, S. K., Chakrabarti, P. P., & Pal, A. (2023). A critical review on applications of artificial intelligence in manufacturing. *Artificial Intelligence Review*, *56*(1), 661–768. https://doi.org/10.1007/s10462-023-10535-y

[26] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

[27] Nnamoko, N., & Korkontzelos, I. (2020). Efficient treatment of outliers and class imbalance for diabetes prediction. *Artificial Intelligence in Medicine*, *104*, 101815. https://doi.org/10.1016/j.artmed.2020.101815

[28] Alawieh, H., Rasheed, M., Abboud, R., Hamawy, L., Hage-Diab, A., & Sbeity, F. (2024). Mental illnesses detection employing features extracted from EEG signals using deep learning. In *2024 Second Jordanian International Biomedical Engineering Conference*, 119–122. https://doi.org/10.1109/JIBEC63210.2024.10932013

How to Cite: Chris, S., Mishra, D., Varanasi, L., & Viswanathan, V. (2026). Modeling Markers for Detection of Psychiatric Disorders Using EEG Signals. *Journal of Data Science and Intelligent Systems*, *4*(2), 237–246. https://doi.org/10.47852/bonviewJDSIS62027587