Received: 18 October 2024 | Revised: 7 April 2025 | Accepted: 24 April 2025 | Published online: 19 May 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–10
DOI: 10.47852/bonviewAIA52024579

BON VIEW PUBLISHING

Leveraging Artificial Immune Systems for Mental Health Research: Anomaly Detection in EEG Data

Shivani Bhandari[1], Neil Buckley[1] and Emanuele Lindo Secco[1,*]

[1]*School of Computer Science and the Environment, Liverpool Hope University, UK*

Abstract: Mental Health is a physical, mental, and social state affecting 970 million people in the world. *Artificial Intelligence* and deep learning techniques classifying *ElectroEncephaloGraphy* (EEG) data have emerged as a promising technology for the detection of mental health disorders. In this context, one underexplored area is the application of *Artificial Immune Systems*, which is a technique inspired by the human immune system that has been useful in many computational tasks, including anomaly detection. This paper aims to bridge the gap by leveraging Artificial Immune Systems for Mental Health through anomaly detection in EEG Data: a novel *Negative Selection Clonal for Anomaly Detection* (NSCAD) algorithm is presented and applied on a dataset of 945 samples with individuals diagnosed with disorders and a control group of healthy participants. Efficacy of NSCAD on anomaly detection was assessed using precision, recall, *F*1-score, and accuracy metrics. Results are promising, with a precision of 0.92, a recall of 0.83, an *F*1-score of 0.88, and an accuracy of 0.78. A comparative analysis between the evaluation metrics and anomaly detection of NSCAD vs other methods is finally reported together with a critical analysis of the limitations.

Keywords: artificial intelligence, artificial immune systems, mental health, electroencephalography

1. Introduction

Mental health is important for well-being, and the World Health Organization (WHO) defines health as a state of physical, mental, and social well-being, rather than simply absence of disease or infirmity [1, 2]. There are about 970 million people affected worldwide [3], so it is important to have good diagnostic and treatment strategies. Diagnosing mental disorders traditionally depends on the WHO's International Classification of Diseases 11th Revision and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders Fifth Edition Text Revision (DSM-5-TR) [4]. These aim to group mental health symptoms based on the assumption they reflect underlying dysfunctions in biological, psychological, or developmental processes [5]. However, these methods face significant challenges as they rely on subjective assessments and do not currently use biomarkers for defining disorders, which is a limitation in this area.

At the same time, psychiatry is exploring approaches to understand neuropsychological factors underlying psychiatric symptoms. ElectroEncephaloGraphy (EEG) has emerged as a promising technology, offering insights into pathophysiological aspects of mental health disorders [6]. Studies have suggested the potential significance of EEG signals, such as resting-state power spectral and functional connectivity (FC) analysis, in diagnosing such disorders [7]. Despite advancements, the integration of these into clinical practice is still in its infancy [8].

Workload in EEG analysis is cumbersome, with neurologists dedicating lots of time interpreting raw data. To mitigate this, there is ongoing research into automating aspects of EEG interpretation. Machine learning (ML), including deep learning techniques, has shown promise in classifying signal data for detecting mental health disorders, with recent work demonstrating their efficacy [9–12]. However, anomaly detection, which is a critical step in identifying pathological patterns, remains underexplored in EEG research.

In general, we intend for anomaly some data or patterns of data which may not follow expected behaviors within the dataset: focusing on the EEG signals, the anomaly usually refers to signs who may underly some neurological disorders. In this context, artificial immune systems (AIS) have been used for the detection of these anomalies [13–15].

This work presents an attempt to cover current gaps between the detection of the anomalies in the EEG signals with AIS according to the following objectives:

1) To investigate the possibility of using AIS in the detection of anomalies of the EEG signals vs mental health conditions.
2) To present a novel algorithm which we called the Negative Selection Clonal for Anomaly Detection (NSCAD) algorithm, combining the Negative Selection Algorithm (NSA) with Clonal Selection Algorithm (CSA).
3) To estimate the performances of this novel NSCAD algorithm through a cross-validation and performance analysis on the dataset of the test data.
4) To discuss the implications of these findings for mental health research and the potential of the NSCAD model in diagnostics.

*Corresponding author: Emanuele Lindo Secco, School of Computer Science and the Environment, Liverpool Hope University, UK. Email: seccoe@hope.ac.uk

© The Author(s) 2025. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

The structure of the paper is made of 4 more Sections: in the 2nd Section, we explore previous examples of AIS, while the 3rd Section presents the dataset and how we process these data and determine the main features of the data. In this Section, we also present the NSCAD algorithm. Section 4 reports the main implications of this novel approach providing a comparative analysis vs other methodologies. Finally, Section 5 concludes the study, addresses its limitations, and suggests potential applications of our approach in other medical and research fields.

2. Background

AIS have been used to address challenges in real-world and engineering use cases, demonstrating their adaptability and potential in anomaly detection [16]. Notably, hybrid algorithms combining NSA and CSA have been effective in industrial settings and network intrusion detection [16, 17]. For example, AIS has been used in order to reconstruct signal and for the development of security systems [18–27]. Even though these advancements, the application of AIS in medical diagnosis, and in particular, in the EEG-based diagnosis, is quite underexplored. This gap persists despite a trend of using ML and deep learning techniques for the detection of anomalies in the EEG signals [6, 15].

A few studies have used EEG datasets for pathology detection. Arslan et al. [19] combined artificial neural networks (ANN) with AIS for diagnosing epilepsy, employing CSA to update weights and achieving high accuracy. Ba-Karait et al. [20] proposed adaptive particle swarm negative selection for epileptic seizure detection, showing high effectiveness. Another study combined the artificial immune recognition system (AIRS) with principal component analysis (PCA) and fast Fourier transform (FFT) for epilepsy diagnosis, achieving near-perfect classification [21]. Rashid et al. [22] applied AIS to classifying EEG signals related to limb movements, achieving mean accuracy of 86.39% across subjects. Ramalingam et al. [18] used CSA for prosthetic limb movement classification using EEG signals, enhancing accuracy through a fitness-based antibody selection process. For example, EEG time series dataset have been analyzed for the diagnosis of epilepsy with different classifiers [19–21], showcasing the effectiveness of data-driven methodologies in specific medical contexts [21–25]. Similarly, industrial applications of anomaly detection combining NSA and CSA algorithms have shown proper accuracies in the classification of datasets about vacuum valves and bearings [16, 26].

Its ability to manage imbalanced datasets further underlines its clinical applicability. The exploration of specialized techniques like synthetic minority oversampling could enhance its anomaly detection capabilities, increasing its usefulness in clinical settings. Table 1 provides a complete comparative analysis between prior methods. In this context, it is also important to stress that the proposed table is not exhaustive and to underline how a set of different technologies could also be embedded into our research method in order to (1) improve the overall estimation and (2) provide recovery tools (e.g., [28–31]).

3. Proposed Methodology

3.1. Feature in EEG-based anomaly detection

The usual source of the EEG signals is a set of electrodes which are positioned on the scalp of the subject. These signals are inherently high-dimensional data which, in turn, reflect the activity of the brain within the scalp. Because of the nature of these signals, it is mandatory to determine and define a proper set of features which can then condition the success and performance of the algorithm. Previous studies on the diagnosis of epilepsy have adopted a different set of these features: for example, Arslan et al. [19] pre-process the raw EEG signals into vectors containing the minimum, maximum, average, and standard deviation values and then they classify these parameters with ANN and AIS. This effectively differentiated between epileptic patients and healthy individuals. In [20] discrete wavelet transform was used on the same EEG dataset, in order to capture the transient features in the time and frequency domains, which are significant vs the detection of seizure. In another study based on the use of Brain-Computer Interface IV-Graz dataset, Mel-Frequency Cepstral Coefficients were selected, and then, a dimensionality reduction was achieved by means of a two-layer stacked auto-encoder: here the feature selection was performed with a wrapper method, presenting subsets of feature to a support vector machine (SVM) [22]. In another study on the application of CSA for prosthetic limb movements, the mean, median, and standard deviation of the EEG signals were simply used. A J48 decision tree classifier was also used to select the most relevant features [18]. Additionally, the Welch method was used for power spectrum estimation on EEG data, with PCA reducing the dataset to a lower-dimensional representation, retaining the most informative components [21]. The dataset used in this study also uses feature selection combining quantitative EEG (qEEG) parameters, such as power spectrum density and function connectivity, across different frequency bands. Different feature combinations were tested, and models achieved high classification accuracies for a range of psychiatric disorders, adjusting for covariates like age, sex, education, and IQ [15].

3.2. Data pre-processing

The dataset of this study was sourced from a database created by Park et al. in 2020 [15]. It was originally collected at a Medical Centre in Seoul, South Korea, and spans from January 2011 to December 2018. It includes a variety of medical records, psychological assessments, and qEEG data, particularly focusing on resting-state assessments.

The main disorder in the dataset was identified as the target variable. Although there was also a “specific disorder” column, this was excluded to simplify the model, as were other clearly irrelevant columns. The dataset details are summarized in Table 2.

Mode imputation was used to fill the missing values in the IQ and education columns, and categorical columns, such as sex, were processed using an ordinal encoder. For anomaly detection, a binary classification approach was used, where all main disorder types, such as Addictive disorder, Anxiety disorder, Mood disorder, Obsessive-compulsive disorder, Schizophrenia, and Trauma and stress-related disorder, were labeled as “1.” In contrast, the Healthy Control category was labeled as “0,” establishing a clear distinction between healthy individuals and those with disorders. Finally, the dataset was standardized with min-max normalization and then z-score, in order to enhance accuracy and interpretability of the results.

3.3. Feature selection

We used univariate feature selection using statistical tests, particularly ANOVA-F, to identify the most relevant features. This test carries out a univariate analysis for each feature against the target feature to determine if there is a statistically significant relationship. While ANOVA-F is a widely known method,

Table 1
Evaluation metrics and corresponding equations

Study	Dataset	Medical/ Industrial Context	Method	Metric(s)	Performance
Current Study	EEG dataset collected from SMG-SNU Boramae Medical Center (Publicly available EEG data described by Park et al. [15])	Psychiatric Disorders	NSCAD	• Precision • Recall • *F*1-Score • Accuracy	• Precision: 92.2% • Recall: 82.6% • *F*1-Score: 87.2% • Accuracy: 77.8%
Park et al. [15]	EEG signal dataset collected from SMG-SNU Boramae Medical Center	Psychiatric Disorders	• SVM • EN • RF	AUC	Mean AUC for all disorders adjusted for IQ: • EN: 87.59 ± 7.92% • SVM: 86.02 ± 8.89% • RF = 87.18 ± 8.08%
Sahu [25]	EEG dataset collected from Institute of Psychiatry and Neurology Warsaw, Poland	Schizophrenia Diagnosis	Various classifiers (10)	• Precision • Recall • Accuracy	Best performance by XGBoost Classifier with Central, Parietal-Occipital, and Frontal-Prefrontal electrodes: • Accuracy: 78.75 % • Precision: 72.97% • Recall: 79.41%
Rashid et al. [22]	BCI IV-Graz dataset 2a-EEG dataset	Human Limb Movement	GA+NSA	Accuracy	Mean Classification Accuracy: 86.39%
Ramalingam et al. [18]	EEG signals from healthy volunteers performing limb movements	Four different limb movement	CSCA	• Accuracy • Sensitivity • Specificity	• Total classification accuracy: 79.62% • Average Sensitivity: 79.01% • Average Specificity: 91.77%
Ba-Karait et al. [20]	The Bonn EEG time series dataset (Publicly available EEG data described by Andrzejak et al. [27])	Epilepsy Diagnosis	APSNS	• Accuracy • Sensitivity • Specificity	Performance with different training-test partitions: • 50–50%: Accuracy-99.44%, Sensitivity-99.69%, Specificity-99.19% • 60–40%: Accuracy-99.60%, Sensitivity:99.73%, Specificity-99.47% • 10-fold cross-validation: Accuracy-99.66%, Sensitivity: 99.63%, Specificity-99.69%
Kim et al. [16]	Vacuum valve dataset and IMS bearing dataset	Industrial Application	NSA+ CSA	Accuracy	2-class classification: • Vacuum valve dataset: 99.43% • IMS bearing data: 98.66%
Xinmin et al. [26]	Bearings dataset for fault detection	Industrial Application	NSCA	Accuracy	Achieved accuracy of 95.368% with a distance threshold 0.826
Kim and Bentley [17]	Wisconsin breast cancer dataset, vote dataset, and Iris dataset	Breast cancer diagnosis, vote Iris data	NSA+ CSA	• TPR • FPR	• For all three datasets: Average TPR: >93%. Average FPR: <10% • For the iris dataset, TPR achieved 100%
Arslan and Işik [19]	The Bonn EEG time series dataset (Publicly available EEG data described by Andrzejak et al. [27])	Epilepsy Diagnosis	ANN+AIS	Accuracy	• Accuracy rates: 93%, 96% with 5 and 10 antibodies respectively • Average success rate: 95%
Polat and Güneş [21]	The Bonn EEG time series dataset (Publicly available EEG data described by Andrzejak et al. [27])	Epilepsy Diagnosis	AIRS+ PCA +FFT	• Accuracy • Sensitivity • Specificity	Performance with different trained and tested partitions: • 50–50%: Accuracy-99.81%, Sensitivity: 99.62%, Specificity: 100% • 70–30%: Accuracy-100%, Sensitivity: 100%, Specificity: 100% • 80–20%: Accuracy-100%, Sensitivity: 100%, Specificity:100%

Table 2
List of features in the dataset

Feature category	Features	Description
Patient's Information	no.	Patient's Unique ID
	sex	Male (M), Female (F)
	age	18 to 70 years (continuous)
	eeg.date	Date of EEG
	education	Years of Education
	IQ	IQ level (discrete)
Diagnosis	main.disorder	Addictive disorder, Trauma and stress-related disorder, Mood disorder, Healthy control, Obsessive-compulsive disorder, Schizophrenia, Anxiety disorder
	specific.disorder	Alcohol use disorder, Acute stress disorder, Depressive disorder, Healthy control, Behavioral addiction disorder, Obsessive-compulsive disorder, Schizophrenia, Panic disorder, Social anxiety disorder, Posttraumatic stress disorder, Adjustment disorder, Bipolar disorder
qEEG parameters	Power Spectrum Density (PSD)	114 variables (continuous) Absolute power values for specific frequency bands
	Functional Connectivity (FC)	1026 variables(continuous) Coherence for each frequency band and pair of channels.
Blank Column	Unnamed: 122	Empty column with no data

we specifically adapted it to identify the significant features for EEG anomaly detection related to mental health disorders. This was done by fine-tuning the number of selected features to optimize the performance of the NSCAD algorithm. This step-by-step feature selection process ensured that only the most informative features are retained, reducing dimensionality while preserving discriminative power for accurate anomaly detection.

ANOVA-F works by calculating the ratio of variance between groups to the variance within groups, per Equation (1).

$$F = \frac{var_b}{var_w} \tag{1}$$

Here, var_b is the variance between groups and var_w is the variance within groups. The test follows a filter-based approach, ranking features based on their statistical significance in relation to the target. Specifically, it selects the top k features with the highest scores for further analysis, where k is a parameter representing the number of features to select. This is calculated as shown in Equations (2) and (3).

$$var_b = \frac{1}{k-1}\sum_{i=1}^{j} j_i(\overline{x}_i - \overline{\overline{x}})^2 \tag{2}$$

$$var_w = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{p=1}^{j_i}\left(x_{ip} - \overline{x}_i\right)^2 \tag{3}$$

Here, ji is the number of observations in the ith group, xi is the mean of the ith group, xip is the pth observation in the ith group, and $\overline{\overline{x}}$ is the overall mean across all groups.

To address the challenge of high-dimensional data, the process of feature selection in our study involved fine-tuning k to determine its optimal value. This entailed systematically experimenting with different values of k and selecting one that maximized performance of the proposed NSCAD algorithm. By focusing on the most significant features, the aim was to have results as accurate and generalizable as possible.

3.4. The NSCAD algorithm

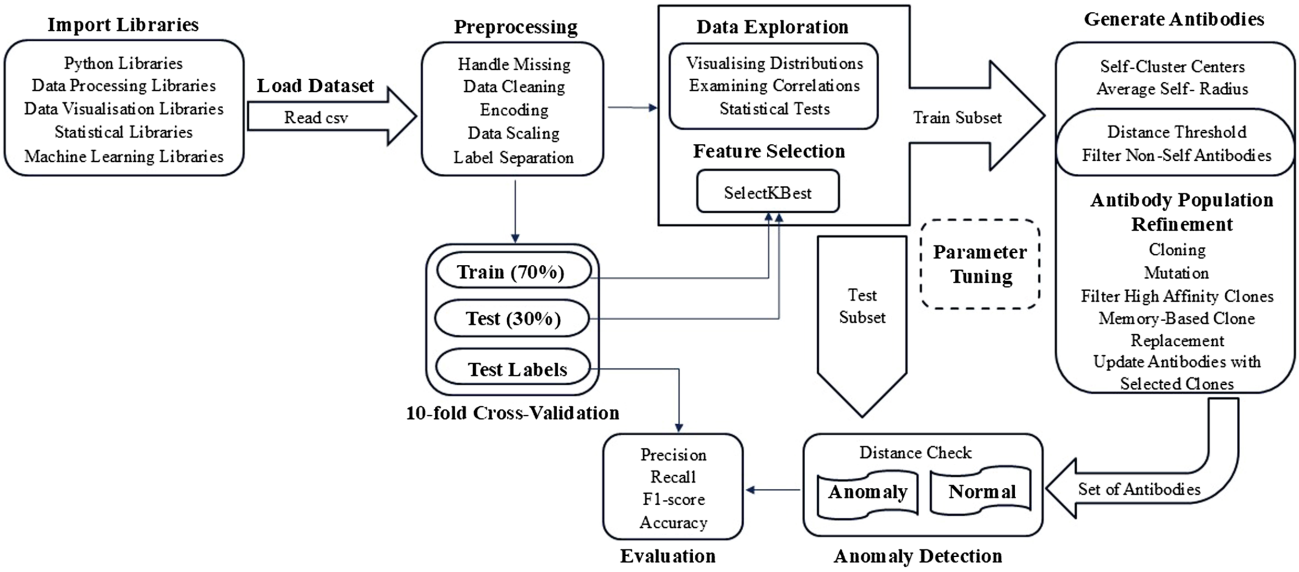
This study proposes the NSCAD algorithm, which is a hybrid of the NSA and the CSA, mimicking essential functions of the immune system. NSCAD integrates these to perform anomaly detection in EEG data, inspired on the immune system's ability to distinguish between self from non-self cells.

NSCAD functions by self-clustering and generating antibodies based on a distance threshold, aiming to effectively identify anomalies. For self-clustering, Agglomerative Clustering is used to group normal samples based on feature similarity. The number of clusters (*n_clusters*) was chosen through experimentation to best represent the diversity in the normal training data. Cluster centers are calculated as the average of samples in each cluster, and the recognition radius is determined using the mean pairwise distances within clusters. Antibodies are generated from abnormal samples by computing the Euclidean distance to cluster centers. NSCAD uses the recognition radius and a threshold (tuned using a percentile of the distance distribution) to select abnormal samples for antibody creation.

NSCAD leverages the generalization and development processes of the CSA to refine antibody population. A small subset of antibodies with the highest affinity underwent controlled mutations to enhance diversity and adaptability. Clones with improved affinity are selected, and weaker antibodies are replaced in each generation, ensuring robust anomaly detection. This refinement is important to adapt to evolving data patterns and maintain diversity within the antibody population. The primary objective is a robust mechanism that can adapt to new and unseen data patterns while identifying anomalous data points.

The algorithm generates antibodies to distinguish between normal and anomalous samples. Test samples are classified by comparing their distance to the generated antibodies. Samples within a tuned distance threshold are marked as anomalies, with the threshold optimized to balance false positives and negatives. This adaptive nature of NSCAD, summarized in Figure 1, is critical in handling the complexity and variability inherent in EEG data.

Figure 1
Blueprint of the NSCAD algorithm



The rational of this procedure is further summarized in Algorithm 1.

```
Algorithm 1. The NSCAD Algorithm
Inputs: training and testing data, X_train, Y_train and X_test
Outputs:  label
Procedure:

Select k best features based on X_train and Y_train
//OBTAIN CLUSTERS
normal_samples := samples from X_train where Y_train is 0
n_clusters := number of desired clusters
clusters := result of hierarchical clustering on normal_samples using n_clusters
//OBTAIN CLUSTER CENTROIDS
centroids := empty list
For each cluster in clusters, do,
   centroid := mean of points in cluster
   add centroid to centroids list
End For
//CALCULATE RECOGNITION RADIUS
distances[] := list of pairwise distances between all centroids
radius := mean(distances)
//CREATE ANTIBODIES
abnormal_samples := samples from X_train where y_train is 1
initial_antibodies := empty list
threshold := a multiplier to adjust recognition radius
For each abnormal_sample in abnormal_samples, do,
   If distance to closest cluster centroid <= radius * threshold, then,
      add abnormal_sample to initial_antibodies
   End If
End For
```

```
    //OPTIMISE ANTIBODIES
    num_generations := number of generations for optimisation
    For i = 1 TO num_generations, do,
       //CREATE CLONES
       top_antibodies := list of highest affinity antibodies from initial_antibodies
       clones := empty list
       For each antibody in top_antibodies, do,
          mutate antibody and append to clones
       End For
        //CALCULATE CLONE AFFINITY
       For each clone in clones, do,
          affinity := 1 - distance to closest cluster
       End For
        //SELECT IMPROVED CLONES
       improved_clones := clones with affinity > threshold
       replace lowest affinity antibodies in initial_antibodies with improved_clones
    End For
    //CALCULATE DISTANCE THRESHOLD FOR CLASSIFICATION
    antibody_distances := list of distances from each abnormal sample to its closest
    antibody in initial_antibodies
    q := percentile rank
    threshold := percentile of antibody_distances using rank q
    //PERFORM FINAL CLASSIFICATION ON TEST DATA
    For each sample in X_test, do,
       distance := distance to closest antibody from initial_antibodies
       If distance <= threshold, then,
          label := 1 (abnormal)
       Else,
          label := 0 (normal)
       End If
End For
```

3.5. Evaluation metrics

The efficacy of the proposed method was assessed using the following metrics: precision, recall, *F*1-score, and accuracy. This multi-metric approach provided a holistic understanding of the algorithm's effectiveness in classifying EEG samples as normal or anomalous [23]. The computation of these metrics is summarized in Table 3.

TP (True Positive) represents correctly classified positive samples, *FP* (False Positive) represents incorrectly classified positive samples, *TN* (True Negative) represents correctly classified negative samples, and *FN* (False Negative) represents incorrectly classified negative samples. The evaluation metric scores, ranging from 0 to 1, were computed using the ground truth and predicted labels [24].

3.6. Validation techniques

Validation involved cross-validation and parameter tuning. First, the study used *k-fold* cross-validation, a technique that divides the dataset into k subsets. Each subset is used once as a validation set, while the remaining subsets form the training set. This process is repeated k times, with each of the k subsets used exactly once as the validation data. This technique helps to assess the algorithm's effectiveness across subsets of the data.

In addition, parameter tuning was used to optimize performance. This involved systematically varying the algorithm's parameters and selecting the set that maximized its effectiveness.

To evaluate NSCAD's performance, the *OneClassSVM* algorithm was chosen as a benchmark, following its effective use in prior studies alongside NSA methods. *OneClassSVM*, an extension of the conventional SVM, specializes in novelty detection by classifying new data based on its similarity to the training set. For this study, *OneClassSVM* was applied to the test dataset using the same feature selection as NSCAD. The model, trained on normal samples with an RBF kernel and a "nu" parameter set at 0.05 to control outlier proportion, was used to predict anomalies in the test data. The performance of *OneClassSVM* was evaluated using precision, recall, *F*1 score, and accuracy metrics to provide a detailed comparison with NSCAD in anomaly detection.

Table 3
Evaluation metrics and corresponding equations

Metrics	Equation
Accuracy	(TP + TN)/(TP + FP + TN + FN)
Recall	TP/(TP + FN)
Precision	TP/(TP + FP)
F1-score	2 * TP/(2 * TP + FP + FN)

4. Results and Discussion

4.1. Dataset

According to Section 3, the dataset for this study was sourced from Park et al. in 2020 [15]. The set has 945 samples, with 850 for individuals diagnosed with various psychiatric disorders and 95 from a control group of healthy participants. The psychiatric diagnoses were determined based on DSM-5 criteria and confirmed by an experienced team of psychiatrists and psychologists. The subjects ranged in age from 18 to 70 and included a broad range of

conditions, covering seven major diagnostic categories and twelve specific disorders. Categories include Schizophrenia, Mood Disorders (including depressive and bipolar disorders), Anxiety Disorders, Obsessive-Compulsive Disorder, Addictive Disorders, and Trauma and Stress-Related Disorders, in addition to the Control Group. Subjects with Neurological, Neurodevelopmental, or Neurocognitive disorders were excluded from the dataset to maintain focus on psychiatric conditions.

The dataset is characterized by high dimensionality, consisting of 1148 features, including the target variable. For signal processing, EEG data were transformed into the frequency domain utilizing FFT and analyzed using a sliding window technique which minimized noise. *Power Spectral Density* (PSD) values were calculated for each frequency band, including 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). FC values were computed using coherence across channels to capture synchronization between brain regions. In addition, the dataset includes patient demographic and clinical details such as unique ID, sex, age, date of EEG recording, education level, IQ, and diagnosis (both main and specific disorders). A detailed description, including the EEG pre-processing steps, is available in the original study, which includes down-sampling, Butterworth filtering, artifact rejection, and FFT for frequency domain analysis [15].

4.2. Evaluation of NSCAD

4.2.1. Cross-validation findings

The NSCAD model underwent 10-fold cross-validation to ensure its reliability and generalizability, as shown in Table 4. This process involved tuning the model's hyperparameters to optimize the chosen metrics. It demonstrated a high precision score of 0.889, indicating its ability to accurately identify anomalies with minimal false positives. Additionally, it achieved recall, *F*1-score, and accuracy scores of 0.834, 0.860, and 0.758, respectively, reflecting its sensitivity and overall effectiveness in anomaly detection.

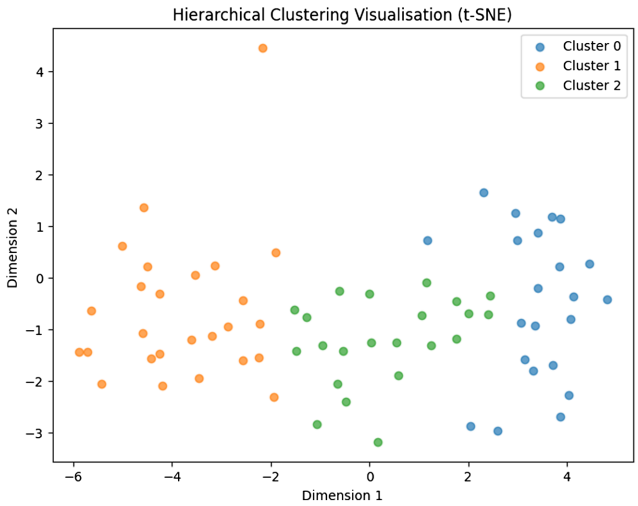
4.2.2. Testing

Performance was further validated on unseen test data using the optimal parameters identified with cross-validation. A 2D visualization using t-SNE was plotted for the hierarchical clustering results, applied to normal samples. This was to provide deeper insight and is shown in Figure 2. In this plot, Dimension 1 and Dimension 2 represented the coordinates in the 2D space generated by *t-SNE*. These dimensions captured the structure and clustering patterns of the data in a lower-dimensional space. This visualization provided insights into the distribution and separation of normal samples within the reduced feature space.

Table 4
Cross-validation results

Metrics	Precision	Recall	*F*1 score	Accuracy
Score	0.889	0.834	0.860	0.758
K	30	10	10	10
$n_{cluster}$	7	3	3	3
$num_{generations}$	30	10	10	10
$threshold_{percentile}$	60	50	50	50
$memory_{percentile}$	60	50	50	50

Figure 2
Visualization of clusters (t-SNE)



Furthermore, the model had exceptional precision (0.922), recall (0.826), *F*1-score (0.872), and accuracy (0.778) on the test dataset.

4.3. Comparative analysis: NSCAD model vs. one-class SVM

A comparative analysis with One-class SVM, shown in Table 5, highlighted the NSCAD model's strengths and areas for improvements.

While the One-class SVM slightly outperformed NSCAD in precision, NSCAD was better when it came to recall, *F*1-score, and accuracy. The higher precision of One-class SVM may be due to its semi-supervised learning approach, which is effective in identifying clear-cut anomalies. However, this might not be as sensitive to subtler variations in EEG data associated with mental health disorders, which could explain its lower recall and overall accuracy.

4.4. Statistical analysis

4.4.1. Variance homogeneity via Levene's test

Levene's test was used to assess the homogeneity of variance for *Mean* PSD and *Mean* FC among different mental health disorder groups and the control group. This test is calculated as in Equation (4).

$$W = \frac{(N-k)\sum_{i=1}^{k} N_i(\overline{Z_i}-\overline{Z})^2}{(k-1)\sum_{i=1}^{N}(Z_{ij}-\overline{Z_i})^2} \quad (4)$$

where N is the total sample size, k is the number of groups, Ni is the sample size for the ith group, $\overline{Z}$ is the mean of the ith group, and Zij is the jth observation in the ith group.

Table 5
Comparative performance analysis between NSCAD model and one-class SVM

Metric	NSCAD	OC SVM
Precision	0.922	0.925
Recall	0.826	0.479
*F*1 Score	0.872	0.631
Accuracy	0.778	0.489

The majority of comparisons resulted in non-significant *p*-values (greater than 0.05), suggesting consistent variance among groups. However, for individuals with anxiety disorders, significant *p*-values (less than 0.05) were observed, indicating potential differences in variance for both Mean PSD and Mean FC when compared to the control group. This suggests that while most mental health disorders may not exhibit significant variance differences compared to the control group, individuals with anxiety disorders might display distinct brain connectivity patterns.

4.4.2. T-test findings

The independent t-tests provided further insights into potential mean differences in Mean PSD and Mean FC between the groups. We calculate the t-tests as in Equation (5).

$$t = \frac{\overline{X_1} - \overline{X_2}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{5}$$

where X_1 and X_2 are the sample means, s_p is the pooled standard deviation, and n_1 and n_2 are the sample sizes. Remarkably, comparisons involving trauma and stress-related disorder and schizophrenia against the control group demonstrated significant differences in mean PSD. Similarly, significant differences in mean FC values were observed between the trauma and stress-related disorder, as well as anxiety disorder groups, and the healthy control group. Results indicate distinct neural mechanisms potentially associated with these disorders, offering valuable insights for targeted interventions and understanding specific neural patterns.

4.5. Discussion

4.5.1. Implications for mental health research

Compared to traditional methods such as SVM, RF, and EN models, NSCAD provides a novel approach to identifying subtle and diverse abnormal patterns characteristic of mental health disorders. This is important in the context of diagnosis, where conventional classification may not fully capture the nuances in EEG data. The study from which the dataset originated achieved a mean AUC of 87.59% using EN models [15], highlighting the efficacy of ML in psychiatric disorder diagnosis. Another ML study addressed the classification of healthy individuals and those with Schizophrenia, using different classifiers that achieved varying accuracies: 85.71% for the Ada-Boost Classifier with Central electrodes, 80.00% for the Gradient-Boosting Classifier with Parietal-Occipital electrodes, 76.67% for the Decision-Tree Classifier & Ada-Boost Classifier with Frontal-Prefrontal electrodes, and 78.75% for the XGBoost Classifier with the combination of Central, Parietal-Occipital, and Frontal-Prefrontal electrodes [25]. Importantly, the NSCAD model's anomaly detection capabilities offer an additional layer of insight for identifying more complex patterns.

When comparing the NSCAD model with other studies based on medical diagnosis, Ba-Karait's work [20] achieved remarkable accuracies in epilepsy diagnosis using a hybrid combination of PSO and NSA. Using training-test datasets partitions of 50–50%, 60–40%, and 10-fold cross-validation, the model achieved accuracies of 99.44%, 99.60%, and 99.66%, respectively. Similarly, another study [21] employed the AIRS algorithm with a fuzzy resource allocation classifier on the same dataset and achieved classification accuracies of 99.81%, 100%, and 100% on different trained and tested partitions. Furthermore, in a study [19] that used the same publicly available dataset, ANN and AIS were employed, achieving an accuracy of nearly 100% with 100 iterations and 0.05 mutation rate. Additionally, its average success rate was 95% for all test scores. It is important to recognize that these approaches targeted different medical contexts, whereas the NSCAD model in the current study focused on anomaly detection in EEG data for mental health disorders, providing a complementary perspective.

Bidgoli et al.'s generative adversarial networks (GAN)-based approach for EEG anomaly detection achieved an AUC of 72% using GANomaly on the Temple University Hospital's abnormal EEG corpus dataset, outperforming One-Class SVM, Isolation Forest, and One-Class CNN. Unlike GAN-based models, which reconstruct normal data distributions, NSCAD uses a clustering-based, immune-inspired approach for anomaly detection, refining patterns through optimization to improve sensitivity to psychiatric conditions [28].

Compared to the NSCAD model, which relies on handcrafted distance thresholds and immune-inspired optimization, *SincVAE* leverages deep learning for adaptive feature extraction and anomaly detection. While NSCAD's clustering-based approach offers interpretability, its reliance on predefined recognition radii may limit generalization. In contrast, *SincVAE*'s probabilistic latent space enables greater flexibility in handling unseen anomalies. Nevertheless, NSCAD's clustering-based approach and immune-inspired optimization provide a level of transparency and interpretability, allowing for better understanding of how decisions are made during anomaly detection. This is crucial in clinical settings where transparency is needed for validation and trust in the model's outputs [29].

The NSCAD model uses a hybrid approach inspired by the immune system, refining antibodies through mutation and selection for robust anomaly detection, but lacks the flexible feature extraction capabilities of deep learning models like GATs-LSTM [30].

4.5.2. Limitations of this study

While NSCAD demonstrates promise, one limitation pertains to overfitting. Rigorous validation including k-fold cross-validation and parameter tuning were used in our study to mitigate this, but more efforts are needed to ensure its reliability on unseen data.

In addition, we did not specifically address the issue of dataset imbalance in this study. The dataset used was heavily imbalanced, with 95 healthy control samples and 850 mental illness samples, which may have affected the generalizability of the results. Future work should explore strategies to handle such imbalances during training, such as oversampling or cost-sensitive learning, to improve performance on underrepresented classes.

Lastly, the study employed a specific public dataset, which could introduce bias or limitations. To ensure the model's performance across diverse scenarios and its generalizability, diverse clinical datasets with varying characteristics should be explored.

5. Conclusion

This study introduced the NSCAD model for detecting anomalies in EEG data associated with mental health disorders. It performed well, compared to traditional classification-focused models. We showed generally better performance over the One-class SVM, not only in terms of accuracy but in its maintaining balanced anomaly detection, even with imbalanced datasets.

However, further work is needed to address overfitting. This may be done by incorporating additional methods such as L1 and L2 regularization to penalize excessive model complexity and enhance generalization. Furthermore, in order to optimize the precision-recall trade-off, the use of more diverse datasets will likely yield better results.

Future research will focus on improving the interpretability of the NSCAD model, giving medical professionals more trust and confidence in validating its outputs. While the model shows potential for identifying anomalies in EEG data, the current study did not include specific examples or tests demonstrating how these outputs could be directly interpreted in clinical practice. Future work should include developing interpretability tools, such as visualizations of the detected anomalies or integrating medical decision-support systems to ensure the outputs are understandable and actionable by healthcare providers [31–33].

Acknowledgement

This work was presented in Dissertation form in fulfillment of the requirements for the MSc Data Science for the student Shivani Bhandari from the School of Computer Science and the Environment, Liverpool Hope University.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

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

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Shivani Bhandari: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Neil Buckley:** Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Visualization, Supervision, Project administration. **Emanuele Lindo Secco:** Resources, Writing – review & editing, Visualization.

References

[1] World Health Organization. (2022). *Health and well-being*. Retrieved from: https://www.who.int/data/gho/data/major-themes/health-and-well-being

[2] Manderscheid, R. W., Ryff, C. D., Freeman, E. J., McKnight-Eily, L. R., Dhingra, S., & Strine, T. W. (2009). Evolving definitions of mental illness and wellness. *Preventing Chronic Disease*, *7*(1), A19.

[3] World Health Organization. (2022). *Mental disorders*. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/mental-disorders

[4] First, M. B., Gaebel, W., Maj, M., Stein, D. J., Kogan, C. S., Saunders, J. B., . . . , & Reed, G. M. (2021). An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*, *20*(1), 34–51. https://doi.org/10.1002/wps.20825

[5] Raskin, J. D., Maynard, D., & Gayle, M. C. (2022). Psychologist attitudes toward DSM-5 and its alternatives. *Professional Psychology: Research and Practice*, *53*(6), 553. https://doi.org/10.1037/pro0000480

[6] Yasin, S., Hussain, S. A., Aslan, S., Raza, I., Muzammel, M., & Othmani, A. (2021). EEG based major depressive disorder and bipolar disorder detection using neural networks: A review. *Computer Methods and Programs in Biomedicine*, *202*, 106007. https://doi.org/10.1016/j.cmpb.2021.106007

[7] Tran, B. X., McIntyre, R. S., Latkin, C. A., Phan, H. T., Vu, G. T., Nguyen, H. L. T., . . . , & Ho, R. C. (2019). The current research landscape on the artificial intelligence application in the management of depressive disorders: A bibliometric analysis. *International Journal of Environmental Research and Public Health*, *16*(12), 2150. https://doi.org/10.3390/ijerph16122150

[8] Shor, O., Yaniv-Rosenfeld, A., Valevski, A., Weizman, A., Khrennikov, A., & Benninger, F. (2023). EEG-based spatio-temporal relation signatures for the diagnosis of depression and schizophrenia. *Scientific Reports*, *13*(1), 776. https://doi.org/10.1038/s41598-023-28009-0

[9] Bajpai, R., Yuvaraj, R., & Prince, A. A. (2021). Automated EEG pathology detection based on different convolutional neural network models: Deep learning approach. *Computers in Biology and Medicine*, *133*, 104434. https://doi.org/10.1016/j.compbiomed.2021.104434

[10] Luján, M. Á., Mateo Sotos, J., Torres, A., Santos, J. L., Quevedo, O., & Borja, A. L. (2022). Mental disorder diagnosis from EEG signals employing automated leaning procedures based on radial basis functions. *Journal of Medical and Biological Engineering*, *42*(6), 853–859. https://doi.org/10.1007/s40846-022-00758-9

[11] 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*, 4521–4524. https://doi.org/10.1109/EMBC.2019.8857946

[12] Ranjan, R., Neeti, & Sahana, B. C. (2022). Automatic detection of mental health status using alpha subband of EEG data. In *2022 IEEE International Symposium on Medical Measurements and Applications*, 1–6. https://doi.org/10.1109/MeMeA54994.2022.9856586

[13] Ishida, Y. (1990). Fully distributed diagnosis by PDP learning algorithm: Towards immune network PDP model. In *1990 IJCNN International Joint Conference on Neural Networks,* 777–782. https://doi.org/10.1109/IJCNN.1990.137663

[14] Forrest, S., Perelson, A. S., Allen, L., & Cherukuri, R. (1994). Self-nonself discrimination in a computer. In *Proceedings of 1994 IEEE Computer Society Symposium on Research in Security and Privacy*, 202–212. https://doi.org/10.1109/RISP.1994.296580

[15] 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

[16] Kim, Y. J., Nam, W., & Lee, J. (2022). Multiclass anomaly detection for unsupervised and semi-supervised data based on a combination of negative selection and clonal selection algorithms. *Applied Soft Computing*, *122*, 108838. https://doi.org/10.1016/j.asoc.2022.108838

[17] Kim, J., & Bentley, P. J. (2001). Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator. In *Proceedings of the 2001 Congress on Evolutionary Computation, 2,* 1244–1252. https://doi.org/10.1109/CEC.2001.934333

[18] Ramalingam, V. V., Mohan, S., & Sugumaran, V. (2016). Prosthetic arm control using clonal selection classification algorithm (CSCA)-A statistical learning approach. *Indian Journal of Science and Technology*, *9*(16), 1–8. https://doi.org/10.17485/ijst/2016/v9i16/86816

[19] Arslan, S., & Işik, H. (2014). The hybrid classification model thanks to artificial neural network and artificial immune systems for diagnosis of epilepsy from electroencephalography. *Journal of Advances in Computer Networks*, *2*(1), 31–34 https://doi.org/10.7763/JACN.2014.V2.77

[20] Ba-Karait, N. O., Shamsuddin, S. M., & Sudirman, R. (2012). EEG signals classification using a hybrid method based on negative selection and particle swarm optimization. In *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, 7376*, 427–438. https://doi.org/10.1007/978-3-642-31537-4_34

[21] Polat, K., & Güneş, S. (2008). Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals. *Expert Systems with Applications*, *34*(3), 2039–2048. https://doi.org/10.1016/j.eswa.2007.02.009

[22] Rashid, N., Iqbal, J., Mahmood, F., Abid, A., Khan, U. S., & Tiwana, M. I. (2018). Artificial immune system–negative selection classification algorithm (NSCA) for four class electroencephalogram (EEG) signals. *Frontiers in Human Neuroscience*, *12*, 439. https://doi.org/10.3389/fnhum.2018.00439

[23] Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, *12*(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

[24] Rupapara, V., Rustam, F., Shahzad, H. F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model. *IEEE Access*, *9*, 78621–78634. https://doi.org/10.1109/ACCESS.2021.3083638

[25] Sahu, P. K. (2023). Artificial intelligence system for verification of schizophrenia via theta-EEG rhythm. *Biomedical Signal Processing and Control*, *81*, 104485. https://doi.org/10.1016/j.bspc.2022.104485

[26] Xinmin, T., Baoxiang, D., & Yong, X. (2007). One-class bearing fault detection using negative clone selection algorithm. In *IECON 2007-33rd Annual Conference of the IEEE Industrial Electronics Society*, 2672–2677. https://doi.org/10.1109/IECON.2007.4460003

[27] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*(6), 061907. https://doi.org/10.1103/PhysRevE.64.061907

[28] Bidgoli, N. S., Kheradpisheh, S. R., & Farahani, H. (2024). EEG anomaly detection using generative adversarial networks (GANs). In *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing*, 1–6. https://doi.org/10.1109/AISP61396.2024.10475243

[29] Pollastro, A., Isgrò, F., & Prevete, R. (2024). SincVAE: A new approach to improve anomaly detection on EEG data using SincNet and variational autoencoder. *arXiv Preprint: 2406.17537*.

[30] Duan, J., Xie, F., Huang, N., Luo, N., Guan, Z., Zhao, W., & Gao, G. (2024). An EEG abnormality detection algorithm based on graphic attention network. *Multimedia Tools and Applications*, *83*(6), 17941–17960. https://doi.org/10.1007/s11042-023-16280-2

[31] Manolescu, D., Buckley, N., & Secco, E. L. (2024). Machine learning models for probability classification in spectrographic EEG seizures dataset. *WSEAS Transactions on Biology and Biomedicine*, *21,* 260–271. https://doi.org/10.37394/23208.2024.21.27

[32] Manolescu, D., Buckley, N., & Secco, E. L. (2023). Convolutional neural network applied to X-ray medical imagery for pneumonia identification. In *International Conference on Communication and Intelligent Systems,* 183–197. https://doi.org/10.1007/978-981-97-2053-8_14

[33] Innes, M., & Secco, E. L. (2023). An understanding of how technology can assist in the epidemic of medicine nonadherence with the development of a medicine dispenser. *European Journal of Applied Sciences*, *11*(3), 522–550. https://doi.org/10.14738/aivp.113.14878

How to Cite: Bhandari, S., Buckley, N., & Secco, E. L. (2025). Leveraging Artificial Immune Systems for Mental Health Research: Anomaly Detection in EEG Data. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA52024579