Inter-class Correlation-based EEG Channel Selection for ADHD Classification

Vandana Joshi, Nirali Nanavati

***Abstract*—Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental condition that affects millions of children. In this manuscript, we propose methods to classify two groups of children, viz. healthy children, and children with ADHD, providing supplementary information to the doctors for analysis and prediction. An electroencephalogram (EEG) correlation and EEG channel selection are used to discriminate between ADHD EEG and healthy children's EEG. The primary objective of the channel selection method is to reduce the dimensions of the data, reduce computational complexity, improve model performance, and provide faster processing. We use EEG signals recorded in three different scenarios. We propose two models for channel selection using Pearson's correlation coefficient and Hoeffding's D correlation coefficient. Various nonlinear features are extracted from selected channels and used with multiple classifiers. The performance of different classifiers has been tested by calculating the accuracy, precision, recall, and ROC curves among different datasets. The proposed algorithms achieved similar or better accuracy above 90 %with significantly fewer features and channels.**

***Index Terms*—EEG Signal Processing, ADHD, Channel Selection, Correlation, Nonlinear features, EEG classification**

I. Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a prevalent neurodevelopmental disorder affecting approximately 5%–8% of children. It typically begins in childhood and can persist through adolescence and into adulthood. Children with ADHD often struggle with maintaining attention, listening attentively, following instructions, and remaining still. The classification of ADHD aids in understanding the diverse symptom profiles, guiding appropriate treatment strategies, facilitating research on its etiology and management, and promoting awareness and support for individuals affected by the disorder [1]. Neural functioning and human activity are strongly related. Electroencephalography (EEG) records the electrical activity of the brain, and it shows a highly complex electrical activity with distinct nonlinear and dynamic features. Brain activity is measured by placing electrodes on the subject's scalp. The 10–20 International Electrode Positioning System is used for electrode placement. Due to its high temporal resolution and ease of data acquisition compared to other brain imaging techniques, many researchers adopted EEG to assess and evaluate ADHD disorder in their research [1].

Various studies have shown that EEG can differentiate brain function in normal subjects with ADHD. In [2], researchers reviewed several combinations of diagnosis, evaluation, and analysis treatment methods for ADHD using EEG. EEG features like statistical features [3], linear/nonlinear features [4][5], entropy[6], and spectral features [7] [8] are used to characterize the EEG channels, and afterward, they are used for ADHD classification with different machine learning [9][10][11] and deep learning methods [12][13][14]. The acquired EEG signals are generally multichannel. Although a large number of EEG channels provide more information about cerebral activity, it also increases redundancy due to noise, resulting in high-dimensional data.

The setup procedure with several channels in EEG signal processing is time-consuming and inconvenient for the subjects. Additionally, it adds to the system's computational complexity, which needs to be maintained to a minimum in some applications. The above reasons lead to the need for dimension-reduction methods. Dimension reduction can be achieved in EEG signal processing by limiting the number of EEG channels.

This paper aims to classify the ADHD subjects and control subjects with high accuracy with reduced dimensions. We proposed models, Model 1 and Model 2, for EEG channel selection to reduce the dimension. Our models are based on Pearson's and Hoeffding's correlation coefficient. After applying the proposed models, six nonlinear features were extracted, namely, Approximate entropy, Singular Value Decomposition entropy (SVD), Spectral entropy, Petrosian FD, Katz FD, and Higuchi Fractal dimension (HFD) for each selected channel. These features are used with classifiers such as K-nearest neighbor (KNN), Random Forest, Naive Bayes (NB), Support vector machine (SVM), Decision tree (DT), and Multi-layer perceptron (MLP). The performance of classifiers was compared with reduced dimensions for three different datasets. We are able to improve the performance by up to 10% with less dimensions. The classifier's performance gives good results using Hoeffding's correlation coefficient compared to Pearson's coefficient for Model 1. The same highest accuracy for Dataset 1 is achieved by using proposed Model 2 at 98.45%, for Dataset 2, 73%, and for Dataset 3 it is 68%.

The details about existing EEG channel selection methods and applications are given in section 2. Section 3 discusses the proposed methods for channel selection and feature extraction, followed by results analysis and discussion in Sections 4 and 5, respectively.

II. Related work

Electroencephalography (EEG) signals have been extensively utilized in diverse applications, such as motor imagery classification, mental task classification, emotion recognition, brain-computer interfaces (BCI), seizure detection, and sleep stage classification. In numerous studies, EEG channels are treated as features, and feature engineering in EEG

Manuscript received Decembar 11, 2023; revised June 26, 2025.

Vandana Joshi is a Research Scholar of Gujarat Technological University, Ahmedabad, Gujarat, India (Email: vandana.joshi@scet.ac.in).

Dr. Nirali Nanavati is an Associate Professor at Sarvajanik College of Engineering and Technology, Surat, Gujarat, India (Email:nirali.nanavati@scet.ac.in).

signal analysis involves selecting the optimal measuring electrodes. However, the choice of specific electrodes may vary depending on the objectives of the application [15]. The EEG channel selection task falls under the features selection category. The popular feature selection methods are classified as filter methods, wrapper methods, Embedded techniques, and Hybrid techniques [15]. Apart from these, optimization bases such as particle swarm optimization and genetic algorithm optimization [16][17] are also used.

A novel approach for optimizing EEG channel selection using Relief and its extended Relief function was introduced by [18]. These functions aim to assign weights to features based on their ability to distinguish samples in proximity to each other. Various classifiers were employed to evaluate different combinations of EEG channels for distinguishing between real and imagined movements of hands and feet. The study began with two channels (C3-C4, CP3-CP4) located on the contralateral and ipsilateral sensorimotor cortex, defining the Region of Interest (ROI) centered on C3 or Cp3 and C4 or Cp4. Subsequently, a data-driven automatic channel selection method was tested to identify the optimal channel combination that enhances classification accuracy [19]. Furthermore, a novel approach using Granger causality (GC) analysis was proposed by the researcher to classify motor imagery (MI) tasks based on left-hand and right-hand MI [20].

A modified grey wolf optimizer (MGWO), which is a swarm-primarily based meta-heuristic approach as integration with two modifications proposed to achieve the balance between exploration and exploitation [16], has evaluated its performance on different BCI EEG datasets.

Likewise, many researchers have studied the relationship between emotional states and brain regions in selecting emotion-based EEG channels. In [21], the authors referred to the method of selecting channels based on brain region as channels from frontal and adjacent frontal and temporal channels of frontal, parietal, temporal, and occipital EEG signals. The mRMR-FS and mRMR-CS methods show the best classification performance for emotion classification in each channel combination [21].

Extracting useful information from EEG signals efficiently is a complex and demanding endeavor. Hence, many researchers advocate for automated feature learning, which serves as a driving force behind the adoption of deep learning techniques. In [22], CNN was utilized alongside the Reversed Correlation Algorithm (RCA) to identify optimal combinations of electrodes and their associated frequency bands. Furthermore, researchers in [23] introduced a multi-objective optimization approach for EEG channel selection using the non-dominated sorting genetic algorithm (NSGA). This method was applied to classify epileptic seizures using SVM, KNN, and Random Forest machine learning techniques.

In the same way, correlation-based EEG channel selection is a prominent technique many studies utilize. In [24], the authors implemented channel reduction through subject-specific channel selection. The analysis aids in selecting highly correlated EEG channels based on correlation, taking C3, C4, and Cz as a separate reference channel for each subject. A new correlation-based channel selection (CCS) method is proposed by [25], followed by a novel regularized common spatial pattern (RCSP). CCS was used to optimize the motor imagery features to improve classification accuracy.

Similarly, the Filter Bank Common Spatial Pattern (FBCSP) method was used to examine the classification performance of a selection of EEG channels that had been chosen based on the correlation coefficient of spectral entropy. This channel selection strategy improved classification accuracy from 1.25% to 8.22%, according to the results in [26]. Correlation analysis is valuable for assessing the relationship between two variables. In [27], researchers demonstrated that correlation analysis provides significant insights into the brain's functional organization during cognitive tasks and effectively discriminates between individuals based on how inter-channel correlation values vary across subjects.

Recent studies have investigated the correlation between EEG signals using various methods in both the frequency and time domains; in the frequency domain, methods such as correlation coefficients, coherence analysis, auto-correlation, wavelet coefficients, and cross-correlation [28]. Time domain metrics, including Kendall rank order correlation, mutual information, Pearson correlation, and Spearman rank order correlation, have been employed to assess channel correlations in EEG recordings from healthy individuals under different behavioral conditions, such as open and closed eyes [29]. Among these metrics, cross-correlation is particularly well-suited for analyzing EEG signals in the time domain due to its ability to evaluate signal similarity across all potential time delays [28].

Another study [30] has used features such as the autoregressive (AR) parameters extracted from EEG attention activity of ADHD and non-ADHD subjects. Two different classification methods, KNN and GMM-UBM, were used iteratively to find the best combination of fewer channels. However, some researchers implemented channel reduction through subject-specific channel selection or considered channels according to brain region. Most channel selection methods are based on linear correlation or monotonic relation. These approaches may not always work with EEG data. This paper aids in selecting channels according to the nonlinear and non-monotonic behavior of EEG with the help of Hoeffding's correlation coefficient. We compared our results using linear and monotonic correlation (Pearson's) and nonlinear and non-monotonic (Hoeffding's). The results using Hoeffding's correlation coefficient are better than Pearson's.

III. MATERIALS AND METHODS

To analyze the effect of dimension reduction methods, we processed EEG through different steps, namely data acquisition, pre-processing, channel selection, feature extraction, and classification, as shown in Figure 1. Different preprocessing methods were applied to each dataset. For Dataset 1 and Dataset 2, preprocessing involved removing slow drift below 1 Hz and high-frequency noise above 20 Hz, which included line noise. Additionally, a high-pass filter at 0.05 Hz, a low-pass filter at 70 Hz, and a notch filter at 50 Hz were applied on Dataset 3. Afterward, the proposed correlation-based channel selection Model 1 and Model 2 are applied to pre-processed data. Nonlinear features (as stated in section 3.4) are extracted from selected channels set to form feature vectors and processed for classification. In subsequent sections, we describe the methodologies employed

for data preparation and processing to ensure suitability for classification.

A. *Dataset*

To check the robustness of the proposed model, we used three datasets. In our experiment, three different data scenarios have been considered. Dataset 1 belongs to cognitive tasks; Dataset 2 is based on working memory, and Dataset 3 refers to auditory tasks. The details of each of the datasets are mentioned below.

Dataset 1: This dataset is utilized in the Iran National Brain Mapping Laboratory's project and is available to the public [13]. The dataset comprises thirty-one children in the ADHD group (twenty-two boys and nine girls, ages 7 to 10 years) and thirty children in the normal group (twenty-five boys and five girls, ages 8 to 11 years). During the EEG recording, children viewed various images on a monitor, such as animal figures or cartoon characters, and were instructed to count them. The goal was to keep the child engaged in a continuous mental task. EEG recordings were made using the international 10-20 system, which included reference electrodes on the earlobes and nineteen channels (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, and Pz). Figure 2 illustrates the nineteen EEG channel positions and their corresponding labels used in Dataset 1.

All trial data were sampled at intervals of 30 seconds. A total of 328 samples were obtained, of which 196 comprised the group with ADHD and 132 of the control group.

Dataset 2: This dataset [31], meticulously designed, consists of EEG data of 59 participants (aged 9 to 16). Among them, 34 were diagnosed with ADHD (21 with combined ADHD, and one each with inattentive and hyperactive ADHD), and 25 were healthy controls. EEG recordings were made using a SynAmps amplifier (Neuroscan, Sterling, VA, United States) across 21 channels. Healthy participants completed the single n-back, go/no-go, and combined n-back/no-go tasks in one session, while ADHD participants performed only the combined n-back/no-go task. Participants were asked to respond quickly and accurately to tasks displayed on a flat screen. The study's paradigms can be accessed at https://doi.org/10.6084/m9.figshare.c.493332. In our experiment, we focused on the combined n-back/no-go task. In this task, a list of capital letters (A, D, E, H, I, N, R, S, T, U) was presented, and participants determined whether the current stimulus matched the one presented two trials earlier (n = 2). They were also asked to withhold their response when the letter X appeared, with no n-back target trials occurring immediately after a no-go trial. Figure 3 illustrates the tasks.

Dataset 3: In this dataset [32], a novel recording of 20 neurotypical and 25 ADHD young adults was considered. EEG data was collected using 64 channels while performing auditory tasks. Human speech streams with syllable sequences of /ba/, /da/, and /ga/ were used for the stimuli. Each trial featured a three-syllable: Target, Distractor, and Interrupter. The target stream heard from the center, the Distractor of the five-syllable stream spatialized to the right, and the Interrupter stream spatialized to the left, beginning either 1 second (Early Interrupter) or 1.5 seconds (Late Interrupter) after the Target. The experimental design balanced the number of trials among No Interrupter, Early Interrupter, and Late Interrupter conditions. Each trial started with a visual cue indicating the required attentional state. Participants were asked to keep their eyes open and focused on a central fixation dot. During FOCAL attention trials, participants focused on the Target and reported the order of the /ba/, /da/, and /ga/ syllables. In BROAD attention trials, participants monitored the Target and were prepared to shift attention to the Interrupter if it appeared while always ignoring the right-lateralized Distractor.

B. *Correlation Coefficient*

The concept of channel selection based on correlation is derived from the manner in which signals propagate between neurons. When a task or activity is performed, a specific set of neurons is activated, leading to high correlations between signals from these particular neurons. This observation suggests that correlations between signals have strong discriminating properties, which can be leveraged to differentiate between various conditions. In this study, to distinguish between ADHD and normal EEG and analyze the functional connectivity of brain regions, we compute an inter-class correlation coefficient matrix for the ADHD class (C1) and the normal class (C2) using Pearson's correlation [33] and Hoeffding's correlation [34]. The EEG signal associated with channel k is represented as a vector (Eq. 1).

$$\mathbf{X}x(k)D = [x(k)1, x(k)2, ...x(k)N] \tag{1}$$

In this context, k (k=1,2...K) represents the channel number, and N denotes the number of temporal samples per channel. Our study involves 19 channels with 15,630 sample values for each channel. The average correlation coefficient of class C is represented by (Eq.2,3)

$$P_C^{(k,P)}(C = 1, 2) \tag{2}$$

$$P_C^{(k,P)} = \frac{1}{|I_c|} \sum_{i \in Ic} P_i^{(k,P)} \tag{3}$$

$|I_c|$ represent the number of samples for Class c.

We consider a training dataset consisting of *I* EEG signal samples, where each sample has N sample points, indexed by $x^{(k)}i \in R^{NX1}$, where $i = 1, \ldots, I$.

a) Pearson's correlation coefficient: Pearson's correlation coefficient of the i[th] sample EEG channel pair k and p is expressed as (Eq.4)

$$P_i^{(k,P)} = \frac{Cx_i^{(k,P)}}{\sqrt{(p(x_i^{(k)})}\sqrt{(p(x_i^{(p)})}} k, p = 1, 2, \ldots K \tag{4}$$

$Cx_i^{(k,P)}$is the sample Covariance and $p(x_i^{(k)})$ is the sample variance

$$Cx_i^{(k,P)} = \sum_{n=1}^{N} (x_i^{(k)}(n) - x_i^{(-k)})(x_i^{(p)}(n) - x_i^{(-p)}) \tag{5}$$

$$p(x_i^{(k)}) = \frac{1}{N} \sum_{n=1}^{N} (x_i^{(k)}(n) - x_i^{(-k)})^2 \tag{6}$$

$x_i^{(k)} = \frac{1}{n} \sum_{(n=1)}^{N} x_i^{(k)}(n)$denotes the sample mean of$x_i^{(k)}$.

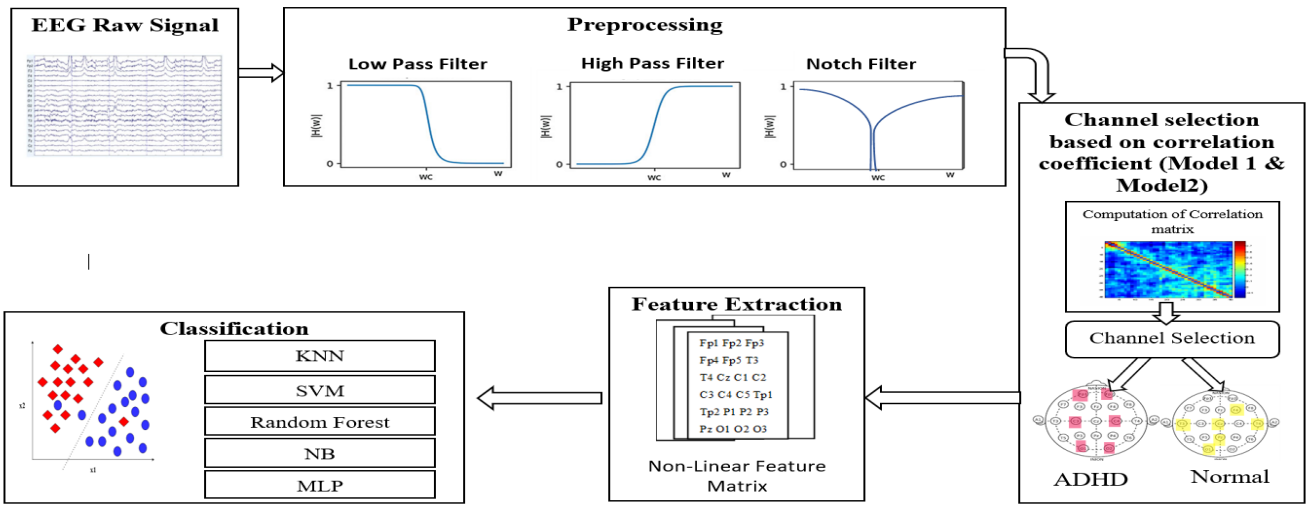


Fig. 1. Block diagram of EEG processing steps using the proposed model

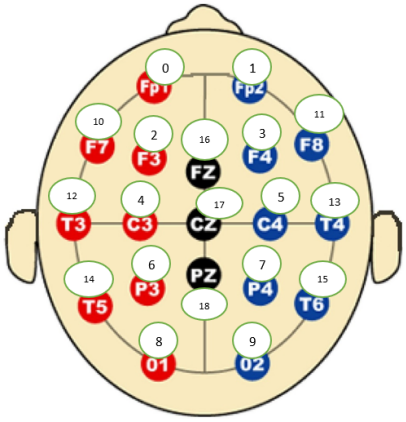


Fig. 2. 19 EEG Channel position with Number labels of Dataset1

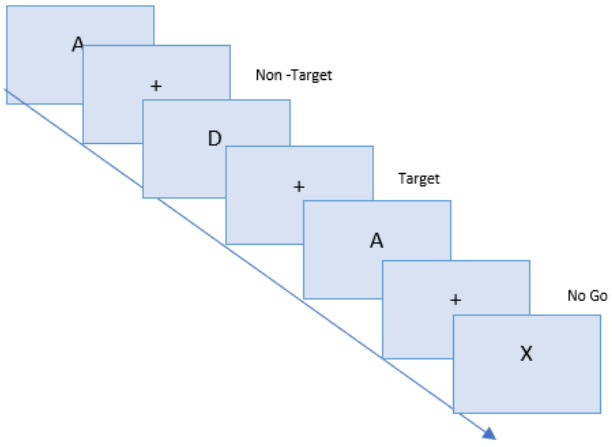


Fig. 3. Schematic illustration of the combined n-back/nogo task (redrawing based on [31])

b) Hoeffding's D Correlation Coefficient: Hoeffding's D, introduced in the 1940s, is a rank-based nonparametric test of independence. Unlike the Pearson correlation, which assesses linear relationships, Hoeffding's D statistic evaluates nonlinear or non-monotonic relationships. A Hoeffding's D statistic greater than 0 indicates dependence between variables.

It measures the distance D between F(x,y) and FG(x)H(y), where F(x,y) represents the joint cumulative distribution function (CDF) of X and Y, and G and H are marginal CDFs.

$$D = \int (F - GH)dF \tag{7}$$

Equation 7 quantifies the difference between the joint ranks of (X, Y) and the product of their marginal ranks. A higher value of D indicates a stronger dependence between X and Y across various types of dependencies.

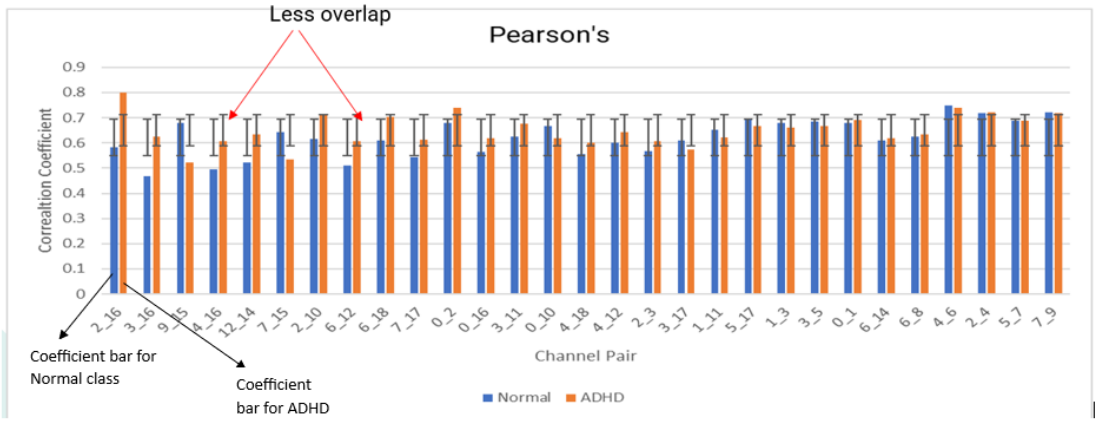
After calculating Pearson's and Hoeffding's D coefficient matrix for each class, we analyzed channels paired for the ADHD and normal classes. We used standard deviation error bars to determine whether a correlation difference is significant in studying discrimination between ADHD and normal classes. The difference may be significant if the standard deviation error bars do not overlap, and it is most likely not statistically significant if there is even less overlap.

For Pearson's, we got less overlapped bars for many channel pairs, while for Hoeffding's coefficient, we got almost all non-overlap error bars. Consequently, we can state that Hoeffding's correlation shows more discrimination between ADHD and Normal EEG channel pair correlation compared to Pearson's. A visual analysis of these is given in Figure 4 to support our findings.

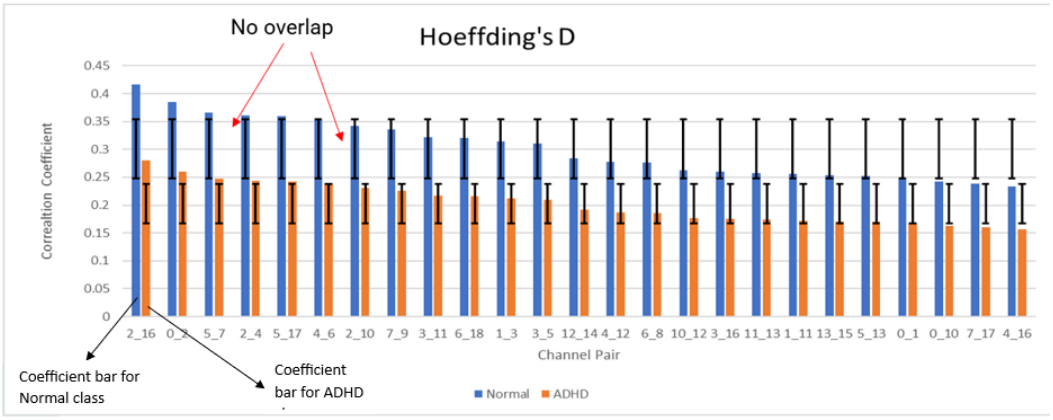
Likewise, we applied a t-statistic method to identify discriminative EEG channel pairs. The t-score for correlation coefficients represents the normalized difference between the average correlation coefficients of the two classes. This test yields a t-value and the associated probability for the null hypothesis. A high t-value, and consequently a low P-value, signifies effective discrimination. In this context, the null hypothesis hypothesizes that there is no difference in the correlation coefficients between EEG channel pairs for the ADHD and normal groups.

C. *EEG Channel Selection*

The correlation coefficient between EEG channel pairs indicates brain connectivity and has been used to evaluate the



(a)



(b)

Fig. 4. Standard Deviation error bar analysis to study discrimination between ADHD and normal class. The X-axis denotes the channel pair, and the Y-axis denotes the correlation coefficient value. (a): Standard Deviation error bar for Pearson's Correlation(b) Standard Deviation error bar for Hoeffding's Correlation

relationships between EEG signals across different channels. In this step, our goal is to reduce the dataset's dimensionality by excluding EEG channels that show relatively low or negligible correlations with one another.

We have proposed two models: Model 1 and Model 2. To build these models, we assume that the behavior of channels of the ADHD or Normal classes should exhibit common information across all samples where participants engage in the same tasks. Following this hypothesis, we evaluate intra-class similarity by employing a correlation coefficient. Here, we assess the similarity between channels rather than the directional influence among them.

In Algorithm 1 and Algorithm 2, we discussed a comprehensive exploration of the algorithmic steps employed in the proposed models.

For dataset 1, we used 132 samples (I) of 30 seconds for each class. One sample contains 15600 sample values of 19 channels (N x K). Then, we calculated Pearson's correlation coefficient matrix and Hoeffding's correlation coefficient matrix for each sample for each class. After taking the mean of the correlation matrix of each class for each correlation, we sorted the coefficient vector and selected the top 200 correlation coefficient values with their channel pairs from each class.

The top 200 average correlation coefficients from the ADHD and Normal classes were compared using a t-test. The associated probability of the null hypothesis (p-value) was considered. The null hypothesis hypothesizes that there are no correlations between channels. Selected channel pairs from both classes whose correlation coefficient rejects the Null hypothesis. The standard deviation error bar was used to get a sense of whether a correlation difference is significant in studying discrimination between ADHD and normal class. From the top 200 channel pairs, we made a two-class channel set named sets s1 and s2 with channel pairs whose coefficient values are more significant than the threshold value. For Pearson's coefficient, we kept 0.60 as the threshold for coefficient value; for Hoeffding's, we selected the top 25 coefficient values. Channel selection is done by selecting channels after union and intersection operation between the two-class channel sets. So, we got Pearson's intersection set with channel number as {0,1,2,3,4,5,6,7} channel (0-7) and Pearson's union set: {0,1,2,3,4,5,6,7,9,12} channel (0-7) (9,12). Similarly, In Model 2 for Hoeffding's correlation, we

got intersection/union: {0,1,2,3,4,5,6,7, 10,11,12,13} channel (0-7) (10-13). For dataset2 we got 8(C3, C4, CZ, FZ, FC5, FC6, F3, F4) channels out of 21 after using Model2. Moreover 14 (C2, C4, CP1, CP2, CP3, CP4, CP5, CP6, CPZ, P1, P2, P3, P4, PZ) channels get selected out of 64 for dataset3.

Algorithm 1 Proposed Model 1

1: Select sample of the fixed time window and form the EEG data of dimension N x K x I (*N =number of sample values, K= number of channels, I= number of samples*)
2: **for** $C = 1, 2$ **do**
3: **for** $i = 1, I$ **do**
4: Compute correlation coefficients and get the matrix:
RP=Correlation.p(Ic_i)
RH=Correlation.h (Ic_i)
5: **end for**
6: **end for**
7: Get the mean of the correlation matrix of each class for each correlation *(Pearson's and Hoeffding's)*
8: Sort the values in descending order
9: Make set 'S' of channel pair for each class *(s1 - ADHD, s2- Normal)* choosing coefficient value greater than threshold values
10: Apply $(s1 \cup s2)$ and$(s1 \cap s2)$ to select the channel

**RP= Pearson's correlation coefficient matrix, RH= Hoeffding's correlation coefficient matrix, C=Class 1-ADHD; Class 2- Normal*

Algorithm 2 Proposed Model 2

1: Select sample of the fixed time window and form the EEG data of dimension N x K x I*(N =number of sample values, K= number of channels, I= number of samples)*
2: **for** $C = 1, 2$ **do**
3: **for** $i = 1, I$ **do**
4: Compute correlation coefficients and get the matrix:
RP=Correlation.p(Ic_i)
RH=Correlation.h (Ic_i)
5: **end for**
6: **end for**
7: Get the mean of the correlation matrix of each class for each correlation *(Pearson's and Hoeffding's)*
8: Convert matrix to unweighted graph basis of the threshold value of correlation coefficients
9: Select the node with the top 3 degrees and make set S of the channel of each degree for each class
10: Apply $(s1 \cup s2)$ to select the channel

**RP= Pearson's correlation coefficient matrix, RH= Hoeffding's correlation coefficient matrix, C=Class 1-ADHD; Class 2- Normal*

D. *Feature Extraction*

The brain's highly nonlinear dynamic network reveals dynamic behavior during activity. Nonlinear dynamics and chaos theory are used to model brain dynamics and identify abnormalities in EEG recordings. Various nonlinear features are reported in the literature for EEG analysis, viz. Lempel-Ziv complexity, fuzzy entropy, spectral entropy [35], and correlation dimension (CD) [6]. Incorporating nonlinear mathematics minimizes the reliance on subjective methods for ADHD diagnosis. To achieve this, we have employed six nonlinear features to construct a feature vector. Three of these features are derived from Fractal Dimension (FD), while the remaining three are based on entropy. Fractal dimension (FD) is computed using methods such as Higuchi, Katz, and Petrosian [36]. It is a metric used to characterize a signal's complexity or irregularity. Additional features include Approximate entropy [37], Singular Value Decomposition entropy (SVDEn), and Spectral entropy. Entropy metrics are used to quantify the uncertainty in the EEG, which is essentially equivalent to the predictability of probable configurations.

The details of each of the features are mentioned below.

*1) **Fractal Dimension**:*

a) Katz Fractal Dimension: The Katz method calculates the fractal dimension of a signal trial as follows (Eq.8).

$$FD = \frac{lnln(N-1)}{lnln(N-1) - lnln(d/L))} \tag{8}$$

Here, L represents the sum of distances between consecutive points, N denotes the length of the data sequence, and d signifies the diameter of the data sequence

b) Higuchi Fractal Dimension: In Higuchi, using a time series $x_{(1)}, x_{(2)}, \ldots, x_{(N)}$ as input, a new time series is generated as follows (Eq. 9).

$$x_m^k = \{x(m), x(m+k), x(m+2k), .., x(m + \left|\frac{N-m}{K}\right|k\} \tag{9}$$

For m= 1, 2,3...K. where m indicates the initial time and k indicates the discrete interval time. The average lengthL_m^k is computed for each curve x_m^k, as (Eq.10).

$$L_m^k = \frac{\sum_{i=1}^{(N-m)/k} \left|x(m+ik) - x(m+(i-1)k\right|}{\lfloor (N-m)/K \rfloor k}(N-1) \tag{10}$$

Where N represents the total length of the data, $\frac{(N-1)}{(N-m)/k}$ serves as a normalization factor. Then, the average length L(k) is computed across the all-time series as (Eq.11).

$$L(k) = 1/k \sum_{m=1}^{k} L_m^k \tag{11}$$

c) Petrosian Fractal Dimension: Another method utilized to calculate the Fractal Dimension (FD) of a signal is the *Petrosian* method, computed as (Eq.12)

$$D = \frac{log_1 0n}{log_1 0n + log_1 0(n/(n+0.4N\Delta))} \tag{12}$$

Here, n represents the number of samples, and $N\Delta$ signifies the number of sign changes in the binary sequence.

*2) **Entropy**:*

a) Approximate entropy (ApEn): By using one-time point augmentation, approximate entropy (ApEn) quantifies the logarithmic value of the frequency at which the neighborhoods of temporal patterns of the same duration within a specific distance in phase space remain close for the patterns [37].

The state space of N-dimensional EEG signal x = $\{x_1, x_2..., x_N\}$ is computed given the time lag J and the embedding dimension m. ApEn is derived from the correlation integral $C_i^m(r)$, which denotes the number of points within distance r from the i[th] point of the EEG time series when embedded in phase space with embedding dimension m. The following equation is used to get the correlation integral (Eq.13)

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{J=1}^{N-m+1} \theta(r - || X_i - X_j ||) \quad (13)$$

In the given equation (Eq.13), θ represents the Heaviside function, while X_i and X_j denote vectors in the state space, and r is a threshold value for distance. Finally, the ApEn is defined as follows (Eq. 14):

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (14)$$

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} logC_i^m(r) \quad (15)$$

We have considered the value of embedding dimension as 2.

b) Spectral Entropy: Spectral entropy is a comprehensive measure of signal disorganization, and its mathematical expression is as follows (Eq.16):

$$H(x) \sum_{i=1}^{N} p(x_i) log_1 0 p(x_i) \quad (16)$$

where $x = (x_1, x_2, x_3, \ldots x_N)$ is the signal in the time domain.$p(x_i)$ is the probability of x_i

c) Singular Value Decomposition Entropy (SVDEn):: Singular Value Decomposition Entropy (SVDEn), representing the dimensions of the data, reveals the number of eigenvectors required for a comprehensive understanding of the dataset. The calculation process is as follows (Eq.17).

$$SvdE = -\sum_{i=1}^{M} \sigma_i(\sigma_i) \quad (17)$$

$$Y = [y_1, y_2, ..y_{(N-(r-1)\tau)}]^T \quad (18)$$

$$y_i = [x_i, x_{i+\tau}, ..x_{i+(r-1)\tau}]^T \quad (19)$$

Where M represents the number of singular values of the embedded matrix Y, obtainable through Eq. 18. $\sigma_1, \sigma_2, \ldots, \sigma_M$ denote the normalized singular values of Y. r indicates the order of permutation entropy, and τ represents the time delay, which was set to 3 and 1, respectively, in this study.

By considering selected channels from the proposed Model 1 and Model 2, feature extraction for each channel is done for the classification of Feature vectors with nonlinear features, as shown in Figure 5.

$$FV(SV_N, Class) = \begin{bmatrix} SV_1 & \cdots & Class\ label \\ \vdots & \ddots & \vdots \\ SV_N & \cdots & Class\ label \end{bmatrix}$$

(a)

$$SV\ (k, m) = \begin{bmatrix} F1_{Ch0} & \cdots & Fm_{Ch0} \\ \vdots & \ddots & \vdots \\ F1_{Chk} & \cdots & Fm_{Chk} \end{bmatrix}$$

(b)

Fig. 5. Feature vector with nonlinear features (a) Feature vector for classification (b) sample vector
where FV: Feature vector, SV: Sample vector, k: Number of selected channels, m: Number of features, N: number of samples

The total dimension number for classification is calculated as follows: If the selected channel numbers are 8 and are used to extract 6 nonlinear features, then the total number of dimensions will be 8 X 6= 48.

IV. Classification

Deep learning algorithms have been successful in image processing and other domains. However, they have not consistently demonstrated improvements over the most advanced approaches available to date when using EEG [38]. Moreover, its success depends on using a large number of instances, which is rare when using EEG data. Indeed, DNN's computational complexity is generally high, both for training and testing [13]. To assess the effectiveness of the proposed channel selection model, which is based on non-linear features, in distinguishing between control and ADHD groups, the recorded EEG signals must undergo classification. To achieve this, we opt to use some effective and less complex classifiers with little data to train.

The performance of multiple classifiers, such as Decision tree (DT), Random Forest [20], Naive Bayes (NB) [23], K—nearest neighbor (KNN) [39], Support vector machine (SVM) [40], and multi-layer perceptron (MLP) [36][41], was investigated. We have used all these classifiers for Dataset 1, Dataset 2, and Dataset 3. KNN (K-Nearest Neighbors) is a simple and effective classifier, whereas SVM (Support Vector Machine) exhibits strong generalization abilities, reducing the risk of overfitting [19]. Ensemble classifiers enhance accuracy by integrating multiple less accurate models, often outperforming single classifiers [9]. Decision Tree (DT) classifiers can adeptly learn various representations and demonstrate resilience to noisy data. KNN, SVM, Random Forest, DT, and Multilayer Perceptron (MLP) are among the most commonly employed techniques for achieving successful classification and are consequently examined in our study.

For SVM, two kernels, including Linear and RBF, were used. In the KNN classifier, we selected the optimal k value, which gives maximum accuracy. The four-layered feed-forward Multi-Layer Perceptron (MLP) with 256,128,64,32 neurons in hidden layers, respectively, was used for fitness computation. The random forest (RF) was also tested using the optimal number of trees (n_estimator =50). Furthermore, the Decision Tree and Naive Bayes (NB) classifiers were also evaluated to analyze their performance in classification.

V. Result Analysis

In this section, we present the experimental results for different classifiers and datasets for ADHD EEG classification with different numbers of dimensions. Classification results are measured through accuracy, precision, and recall with 10-fold cross-validation. We analyzed the performance of classifiers for datasets 1, 2, and 3. In dataset 1, we analyzed the performance of classifiers with all 19 channels using single features such as Higuchi and six non-linear features. Table 1 shows the results.

Then, we select different channels by applying our proposed models and compare their results as given in Table 2. For Model 1, we analyzed the result for Pearson's and Hoeffding's correlation coefficient; however, for Model 2, we only considered Pearson's correlation coefficient. The classifier's performance gives good results using Hoeffding's correlation coefficient compared to Pearson's. For instance, Random Forest for the model using Hoeffding's given 93.97 % with 72 dimensions (Table 3), whereas 90.51% and 90.07 % were given using Pearson's coefficient with 48-dimensions and 60-dimension size, respectively. Likewise, the highest accuracy for dataset1 is achieved by using model2 as 98.45% with MLP classifier. In the previous studies [13] on the same dataset1, researchers were able to achieve accuracies of 98.48% but with high dimensions. Our models achieved nearly the same accuracy (98.45%) with less dimension. In Table 4. the state-of-the-art and our proposed model are compared.

Model 2, which demonstrated the highest accuracy for Dataset 1, was further tested for its robustness by applying it to Dataset 2 and Dataset 3. The results were encouraging, as model 2 continued to perform well with these additional datasets (see Tables 5 and 6). For Dataset 2, the Random Forest method yielded the highest accuracy (78 %) with a total of 126 and 48 dimensions. However, Model 2 showcased its strength by improving the performance of MLP with just 48 dimensions, compared to the 126 dimensions required by the Random Forest method. This further underscores the robustness and efficiency of Model 2 in achieving high accuracy with fewer dimensions.
Likewise, in Dataset 3, Model 2 gives the highest accuracy (68 %) by the Random Forest method with 78 dimensions compared to 384 dimensions. It also improves accuracy by 5 % with MLP. NB classifier also shows performance improvement (61 %) using Model2 for Dataset 3.

Furthermore, we analyzed our proposed models using the mean Area Under the Curve (AUC) metric. The receiver operating characteristic (ROC) curve plots the true positive rate (TPR) against the false positive rate (FPR) at different threshold settings. The AUC of the ROC curve indicates the likelihood that a classifier will rank a randomly selected positive instance higher than a randomly selected negative instance.

The ROC curve and AUC calculation for all classifiers for all datasets with proposed model 1 and model 2 are shown in Figures 6 to 8. The result indicates that performance with a smaller number of channels shows a better AUC mean with a greater number of channels. For instance, as in Figure 6, the AUC mean of Random Forest shows similar values for 19 channels and 12 channels with model 2. Moreover, MLP and GaussianNB had higher AUC values of 0.75 and 0.71 than 64 channels, as shown in Figure 8.

VI. Conclusion

Our objective was to determine the most effective strategy for addressing the curse of dimensionality inherent in multichannel EEG recording data. To achieve this, we proposed two channel selection models, namely Model 1 and Model 2, aimed at identifying the optimal subset of channels from the available set. These models are based on Pearson's and Hoeffding's correlation coefficients, which give information about the most correlated channels. Correlation also helps to reduce redundancy.

We analyzed Pearson's and Hoeffding's correlation coefficient matrix for ADHD and normal classes using Standard Deviation and t-statistic methods. Hoeffding's correlation shows more discrimination between ADHD and Normal EEG channel pair correlation compared to Pearson's. Moreover, after applying the proposed models, we also found that most of the channels that belong to the central lobe had the maximum correlation with other regions of the brain. Similarly, majorly selected channels belong to the central lobe and parietal lobe or frontal lobe, irrespective of the recording scenario.

To check the robustness of the proposed channel selection method, we used three different EEG recordings with different working scenarios. As one dataset was recorded during cognitive activity, the second recording was considered while performing a memory task, and the third with an auditory task. With the correlation, we have used Fractal dimension (FD) (which is calculated based on Higuchi, Katz, and Petrosian methods) non-linear features extraction to get complexity or roughness of signals as well as Approximate entropy, Singular Value Decomposition entropy (SVD), and Spectral entropy measures to quantify the uncertainty in the EEG. The performance of different classifiers such as KNN, SVM, random forest, DT, and MLP are evaluated using accuracy, precision, recall, and AUC mean.

For Dataset 1, Model 1 gives good results using Hoeffding's correlation coefficient compared to Pearson's coefficient. We also compared our results with the state-of-the-art; both models show a good result with fewer channels (12 out of 19). Model 2 gives the highest accuracy at 98.45 % using MLP for Dataset 1. Thus, we apply it to Datasets 2 and 3 to test the robustness of the model.

In Dataset 2, 8 channels are selected out of 21 by Model 2, and it gives the highest 78% accuracy with random forest and 73% accuracy with MLP. Similarly, Model 2 selected 13 channels out of 64 from Dataset 3. It improves the 5-9% performance with MLP and random forest, respectively, with less dimension. Model 2 with the NB classifier also shows performance improvement (61%) for Dataset 3. After analyzing different classifiers and proposed models, we found that irrespective of the EEG recording scenario, our proposed models yielded good results with a lesser number of channels.

TABLE I
RESULTS FOR 19 CHANNELS WITH A SINGLE FEATURE AND 6 NON-LINEAR FEATURES

Dataset 1						
	19 channels With 6 nonlinear features Dimension: 114			**19 channels with Single feature (Higuchi) Dimension: 19**		
Classifier	Accuracy	Precision	Recall	Accuracy	Precision	Recall
KNN	90.91	0.9294	0.88868	93.14	0.9348	0.9401
SVM (RBF)	75.42	0.7517	0.7824	90.89	0.9102	0.917
Linear SVC	90.12	0.9056	0.9093	86.75	0.8672	0.8873
Random Forest	93.6	0.9536	0.9395	96.22	0.9652	0.9703
Logistic	71.23	0.8481	0.9098	96.6	0.8219	0.8357
DT	88.98	0.9038	0.8703	89.78	0.9432	0.9087
NB	71.23	0.7201	0.7137	71.25	0.7176	0.7137
MLP	**98.10**	0.9802	0.9696	**99.23**	0.9928	0.9923

TABLE II
RESULTS OF VARIOUS METRICS FOR DATASET 1 OF PROPOSED MODEL1

Dataset 1									
Proposed Model 1									
	8 Channel with 6 nonlinear Features			**10 Channel with 6 nonlinear Features**			**12 Channel with 6 nonlinear Features**		
	(Pearson's-Intersection)			**(Pearson's -Union)**			**(Hoeffding's D)**		
	Dimension: 48			**Dimension: 60**			**Dimension: 72**		
Classifier	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
KNN	87.54	0.9606	0.7879	90.58	0.9391	0.8714	87.17	0.9718	0.7653
SVM (RBF)	74.64	0.74644	0.7732	73.05	0.7321	0.7818	75.39	0.7545	0.7878
Linear SVC	79.99	0.7862	0.8401	79.14	0.7718	0.8478	84.48	0.8362	0.8708
Random Forest	90.51	0.9187	0.9318	92.07	0.9025	0.9093	93.97	0.9174	0.9247
Logistic	71.23	0.71944	0.706	78.07	0.7657	0.8483	81.08	0.7918	0.8631
DT	87.09	0.8375	0.8934	87.49	0.8575	0.8791	84.52	0.8549	0.8175
NB	71.39	0.71944	0.706	70.09	0.7062	0.706	71.62	0.734	0.706
MLP	93.6	0.9507	0.9318	95.82	0.9642	0.9538	**97.35**	0.979	0.9692

TABLE III
RESULTS OF VARIOUS METRICS FOR DATASET 1 OF PROPOSED MODEL2

Dataset 1			
Proposed Model 2			
	12 Channel with 6 nonlinear Features		
	Dimension: 72		
Classifier	**Accuracy**	**Precision**	**Recall**
KNN	91.69	0.9594	0.8719
SVM (RBF)	77.33	0.765	0.8054
Linear SVC	88.27	0.8707	0.9098
Random Forest	93.2	0.9269	0.9283
Logistic	93.21	0.9152	0.9395
DT	88.2	0.9104	0.8774
NB	73.91	0.757	0.7208
MLP	**98.45**	0.979	0.962

TABLE IV
COMPARISON OF MODEL ACCURACY WITH STATE-OF-THE-ART

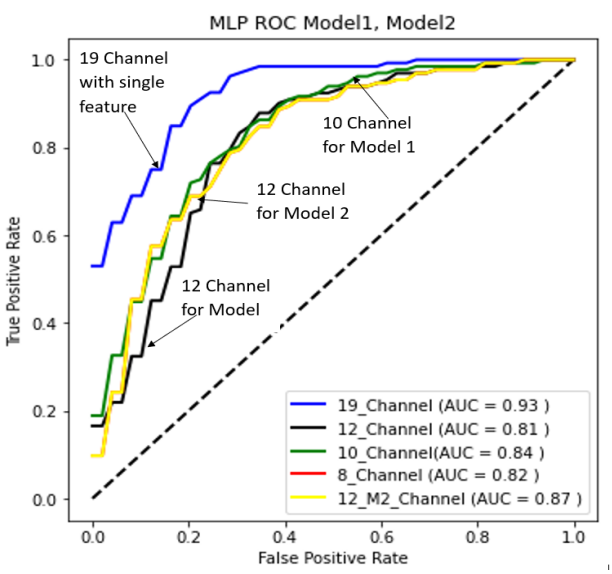
Comparison of Model accuracy with state-of-the-art				
Dataset 1				
	Method and Feature Extraction	**Dimension**	**Classifier**	**Accuracy**
M. Moghaddari [13]	Frequency band separation making RGB images	(19,512,3)	CNN	98.48%
Our Proposed Models	Higuchi Fractal dimension (HFD)	19	MLP	99.23%
	Approximate entropy, Singular Value Decomposition entropy (SVD), Spectral entropy, Petrosian FD, Katz FD	114	MLP	98.10%
		72 (Model 2)	MLP	98.45%
		72 (Model 1)	MLP	97.35%

TABLE V
RESULTS OF VARIOUS METRICS FOR DATASET 2 OF PROPOSED MODEL2

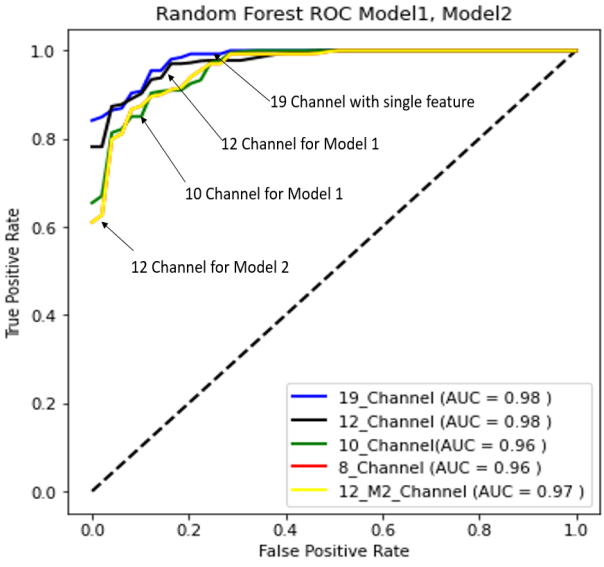
Dataset 2						
Proposed Model 2						
	21 channels with 6 nonlinear Features			**8 Channel with 6 nonlinear Features**		
	Dimension: 126			**Dimension: 48**		
Classifier	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
KNN	71	0.85	0.66	71.66	0.73	0.7
SVM (RBF)	59.33	0.59	1	59.33	0.59	0.93
Linear SVC	73	0.78	0.74	71.66	0.77	0.68
Random forests	78.33	0.77	0.74	78.33	0.72	0.8
Logistic	75	0.77	0.77	70	0.72	0.74
DT	67.99	0.72	0.72	61	0.62	0.59
NB	69.33	0.72	0.59	69.66	0.75	0.59
MLP	**72**	0.76	0.71	**73**	0.81	0.725

TABLE VI
RESULTS OF VARIOUS METRICS FOR DATASET 3 FOR PROPOSED MODEL2

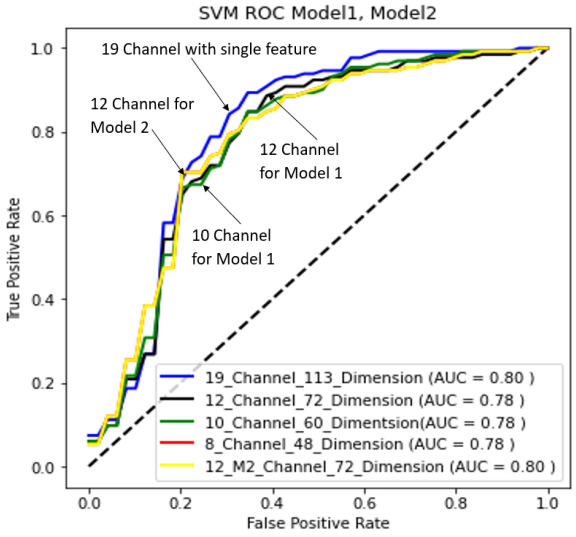
Dataset 3						
Proposed Model 2						
	64 channels with 6 nonlinear Features			**13 Channels with 6 nonlinear Features**		
	Dimension:384			**Dimension:78**		
Classifier	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
KNN	53.49	0.7	0.5	51	0.59	0.78
SVM (RBF)	61	0.61	1	61	0.61	1
Linear SVC	60.5	0.61	0.766	61	0.63	0.9
Random forests	**57.77**	0.568	0.6	**68.88**	0.56	0.78
Logistic	55.99	0.601	0.833	61	0.61	0.1
DT	59	0.666	0.733	58.5	0.37	0.58
NB	**41.5**	0.416	0.466	**61**	0.61	0.68
MLP	**49.5**	0.508	0.65	**55.4**	0.625	0.666



(a)

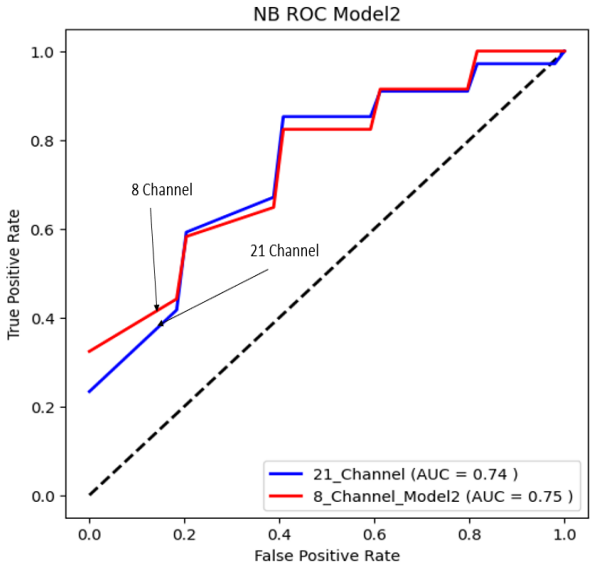


(b)

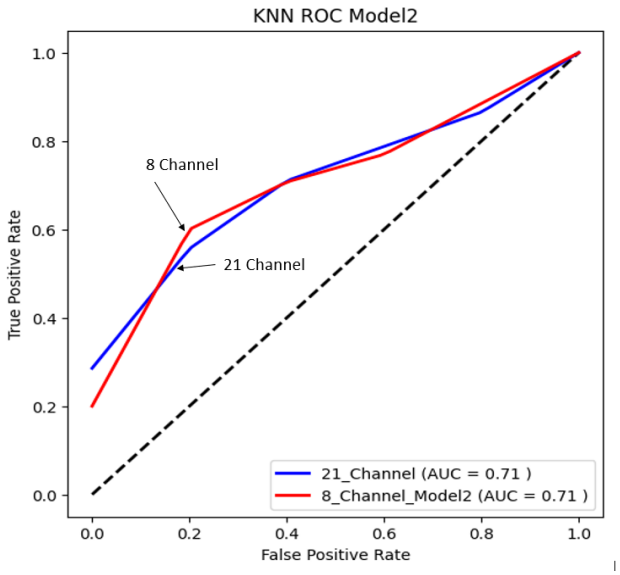


(c)

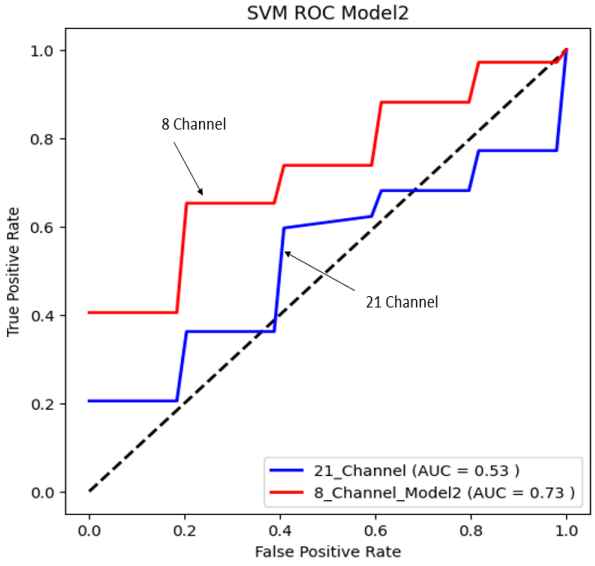
Fig. 6. ROC Curve and AUC mean for different classifiers with different dimensions for Dataset1 with comparison of Proposed Model1 and Model2 (a) Multi-Layer Percepton Classifier ROC curve of Dataset 1 (b) Random forest classifier ROC curve of Dataset 1 (c) Support Vector Machine classifier ROC curve of Dataset 1



(a)

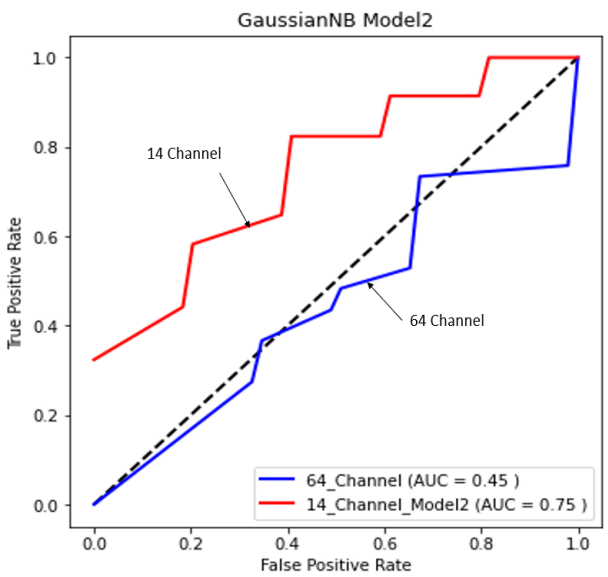


(b)

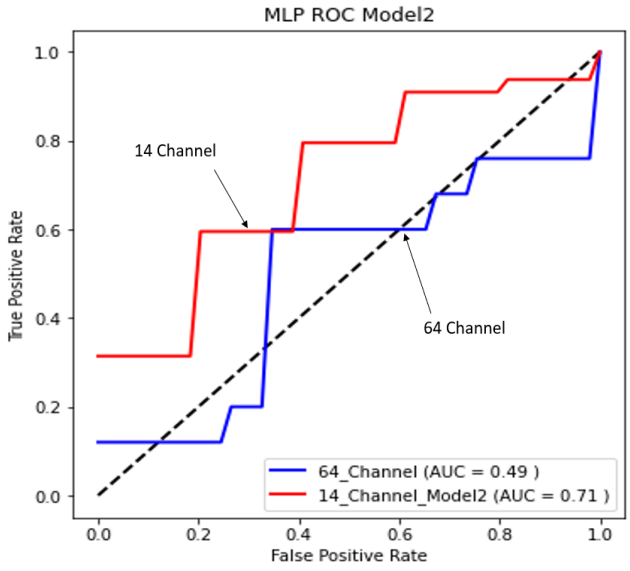


(c)

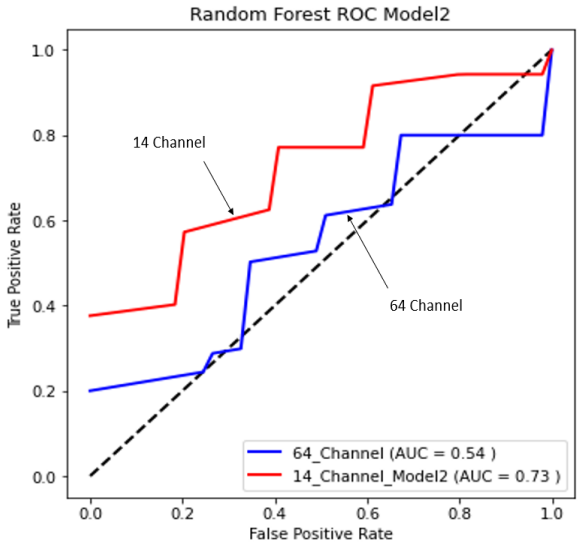
Fig. 7. ROC Curve and AUC mean for different classifiers with different dimensions for Dataset2 with the comparison of Proposed Model2 (a) GaussianNB ROC curve of Dataset 2 (b) KNN ROC curve of Dataset 2 (c) SVM ROC curve of Dataset 2



(a)



(b)



(c)

Fig. 8. ROC Curve and AUC mean for different classifiers with different dimensions for Dataset3 with the comparison of Proposed Model2 (a) GaussianNB ROC curve of Dataset 3 (b) MLP ROC curve of Dataset 3 (c) Random forest ROC curve of Dataset 3

AVAILABILITY OF DATASET

The data which was used in this research is publicly available on the following website:
Dataset1:https://nbml.ir/fa/scientific-tournament/First-Iranian-EEG-competition
Dataset2:https://doi.org/10.6084/m9.figshare.c.4933326
Dataset 3: https://osf.io/4285y/

REFERENCES

[1] V. Joshi and N. Nanavati, "A review of EEG signal analysis for diagnosis of neurological disorders using machine learning," Journal of Biomedical Photonics and Engineering, vol.7, no.4.2021, doi:10.18287/JBPE21.07.040201.

[2] S. M, N. G Cholli, and S. Nayak, "Classification of Attention Deficit Hyperactivity Disorder (ADHD) Considering Diagnosis and Treatment," Int. J. Mod. Educ. Comput. Sci., vol. 11, no. 6, pp. 26–42, 2019, doi:10.5815/ijmecs.2019.06.04.

[3] M. Y. Chang et al., "A New Method of Diagnosing Attention-Deficit Hyperactivity Disorder in Male Patients by Quantitative EEG Analysis," Clin. EEG Neurosci., vol. 50, no. 5, pp. 339–347, 2019, doi:10.1177/1550059419859164.

[4] F. Ghassemi, M. Hassan, M. Tehrani-Doost, and V. Abootalebi, "Using non-linear features of EEG for ADHD/normal participants' classification," Procedia - Soc. Behav. Sci., vol. 32, pp. 148–152, 2012, doi:10.1016/j.sbspro.2012.01.024.

[5] M. Rezaeezadeh, S. Shamekhi, and M. Shamsi, "Attention Deficit Hyperactivity Disorder Diagnosis using non-linear univariate and multivariate EEG measurements: a preliminary study," Physical and Engineering Sciences in Medicine, vol. 43, no. 2. pp. 577–592, 2020, doi: 10.1007/s13246-020-00858-3.

[6] Y. K. Boroujeni, A. A. Rastegari, and H. Khodadadi, "Diagnosis of attention deficit hyperactivity disorder using non-linear analysis of the EEG signal," IET Syst.Biol., vol. 13, no. 5, pp. 260–266,2019, doi:10.1049/iet-syb.2018.5130.

[7] R. Gabriel, M. M. Spendola, A. Mesquita, and A. Z. Neto, "Identification of ADHD cognitive pattern disturbances using EEG and wavelets analysis," Proc. - 2017 IEEE 17th Int. Conf. Bioinforma. Bioeng. BIBE 2017, vol. 2018-Janua, pp. 157–162, 2017, doi: 10.1109/BIBE.2017.00-62.

[8] A. Kamida et al.,"EEG power spectrum analysis in children with ADHD," Yonago Acta Med.,vol.59,no.2, pp.169–173,2016.

[9] P. T. Viet Huong, N. A. Tu, and T. A. Vu, "Ensemble learning in detecting ADHD children by utilizing the non-linear features of EEG signal," in CEUR Workshop Proceedings, 2021, vol. 3026, pp. 129–141, [Online]. Available: http://ceur-ws.org.

[10] H.Öztoprak,M.Toycan, Y. K. Alp, O. Arıkan, E. Doğutepe, and S. Karakaş, "Machine-based classification of ADHD and non ADHD participants using time/frequency features of the event-related neuroelectric activity," Clin. Neurophysiol., vol. 128, no. 12, pp. 2400–2410, 2017, doi: 10.1016/j.clinph.2017.09.105.

[11] A. Mueller, G. Candrian, J. D. Kropotov, V. A. Ponomarev, and G. M. Baschera, "Classification of ADHD patients on the basis of independent ERP components using a machine learning system," Nonlinear Biomed. Phys., vol. 4, no. SUPPL. 1, pp. 1–12, 2010, doi: 10.1186/1753-4631-4-1.

[12] H. Chen, Y. Song, and X. Li, "A deep learning framework for identifying children with ADHD using an EEG-based brain network," Neurocomputing, vol. 356, pp. 83–96, 2019, doi: 10.1016/j.neucom.2019.04.058.

[13] M. Moghaddari, M. Z. Lighvan, and S. Danishvar, "Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG," Comput. Methods Programs Biomed., vol. 197, 2020, doi: 10.1016/j.cmpb.2020.105738.

[14] L. Dubreuil-Vall, G. Ruffini, and J. Camprodon, "A deep learning approach with event-related spectral EEG data in attentional deficit hyperactivity disorder," pp. 1–21, 2019, doi: 10.1101/19005611.

[15] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," EURASIP J. Adv. Signal Process., vol. 2015, no. 1, Dec. 2015, doi:10.1186/s13634-015-0251-9.

[16] H. R. Hussien, E.-S. M. El-Kenawy, and A. I. El-Desouky, "EEG Channel Selection Using A Modified Grey Wolf Optimizer," Eur. J. Electr. Eng. Comput. Sci., vol. 5, no. 1, pp. 17–24, Jan. 2021, doi:10.24018/ejece.2021.5.1.265.

[17] J. Yang et al., "Channel selection and classification of electroencephalogram signals: An artificial neural network and genetic algorithm-based approach," Artif. Intell. Med., vol. 55, no. 2, pp. 117–126, Jun. 2012, doi: 10.1016/j.artmed.2012.02.001.

[18] H. Shan, H. Xu, S. Zhu, and B. He, "A novel channel selection method for optimal classification in different motor imagery BCI paradigms," Biomed. Eng. Online, vol. 14, no. 1, p. 1, Oct. 2015, doi:10.1186/s12938-015-0087-4.

[19] F. Ferracuti, S. Iarlori, Z. Mansour, A. Monteriù, and C. Porcaro, "Comparing between Different Sets of Preprocessing, Classifiers, and Channels Selection Techniques to Optimise Motor Imagery Pattern Classification System from EEG Pattern Recognition," Brain Sci., vol. 12, no. 1, Jan. 2022, doi:10.3390/brainsci12010057.

[20] H. Varsehi and S. M. P. Firoozabadi, "An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality," Neural Networks, vol. 133, pp. 193–206, Jan. 2021, doi:10.1016/j.neunet.2020.11.002.

[21] H. Xu, X. Wang, W. Li, H. Wang, and Q. Bi, "Research on EEG channel selection method for emotion recognition," in IEEE International Conference on Robotics and Biomimetics, ROBIO 2019, 2019, pp. 2528–2535, doi:10.1109/ROBIO49542.2019.8961740.

[22] A. Dura and A. Wosiak, "EEG channel selection strategy for deep learning in emotion recognition," in Procedia Computer Science, 2021, vol. 192, pp. 2789–2796, doi:10.1016/j.procs.2021.09.049.

[23] L. A. Moctezuma and M. Molinas, "EEG Channel-Selection Method for Epileptic-Seizure Classification Based on Multi-Objective Optimization," Front. Neurosci., vol. 14, Jun. 2020, doi:10.3389/fnins.2020.00593.

[24] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "An automatic subject-specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation," Biomed. Signal Process. Control, vol. 68, Jul. 2021, doi: 10.1016/j.bspc.2021.102574.

[25] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," Neural Networks, vol. 118, pp. 262–270, Oct. 2019, doi: 10.1016/j.neunet.2019.07.008.

[26] T. Yang et al., "EEG Channel Selection Based on Correlation Coefficient for Motor Imagery Classification: A Study on Healthy Subjects and ALS Patient," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2018, vol. 2018-July, pp. 1996–1999, doi:10.1109/EMBC.2018.8512701.

[27] D. M. Hernán, F. M. Córdova, L. Cañeteb, F. Palominosc, F. Cifuentes, and G. Rivas, "Inter-channel correlation in the EEG activity during a cognitive problem solving task with an increasing difficulty questions progression.," in Procedia Computer Science, 2015, vol. 55, pp. 1420–1425, doi:10.1016/j.procs.2015.07.136.

[28] R. Bhavsar, Y. Sun, N. Helian, N. Davey, D. Mayor, and T. Steffert, "The correlation between EEG signals as measured in different positions on scalp varying with distance," in Procedia Computer Science, 2018, vol. 123, pp. 92–97, doi:10.1016/j.procs.2018.01.015.

[29] J. D. Bonita et al., "Time domain measures of inter-channel EEG correlations: A comparison of linear, nonparametric and nonlinear measures," Cognitive Neurodynamics, vol. 8, no. 1. pp. 1–15, Feb. 2014, doi: 10.1007/s11571-013-9267-8.

[30] J. L. L. Marcano, M. A. Bell, and A. A. L. Beex, "EEG channel selection for AR model based ADHD classification," 2017, doi: 10.1109/SPMB.2016.7846851.

[31] C. Breitling-Ziegler, J. Tegelbeckers, H. H. Flechtner, and K. Krauel, "Economical Assessment of Working Memory and Response Inhibition in ADHD Using a Combined n-back/Nogo Paradigm: An ERP Study," Front. Hum. Neurosci., vol. 14, no. August, pp. 1–15, 2020, doi:10.3389/fnhum.2020.00322.

[32] J. Kwasa, A. Noyce, and B. Shinn-Cunningham, "Top-down attention modulates neural responses in neurotypical, but not ADHD, young adults," J. Acoust. Soc. Am., vol. 150, no. 4, pp. A64–A64, 2021, doi:10.1121/10.0007633.

[33] Y. Park and W. Chung, "Optimal Channel Selection Using Correlation Coefficient for CSP Based EEG Classification," IEEE Access, vol. 8, pp. 111514–111521, 2020, doi:10.1109/ACCESS.2020.3003056.

[34] Y. Gao et al., "Automatic epileptic seizure classification in multichannel EEG time series with linear discriminant analysis," Technol. Heal. Care, vol. 28, no. 1, pp. 23–33, 2020, doi:10.3233/THC-181548.

[35] S. J. Ruiz-Gómez et al., "Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and mild cognitive impairment," Entropy, vol. 20, no. 1, pp. 1–15, 2018, doi:10.3390/e20010035.

[36] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," Biomed. Eng. Lett., vol. 6, no. 2, pp. 66–73, 2016, doi:10.1007/s13534-016-0218-2.

[37] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy maturity in premature infants Physiological time-series analysis using approximate entropy and sample entropy," Am. J. Physiol. Hear. Circ. Physiol., vol. 278, pp. H2039–H2049, 2000.
[38] A. M. S. M. M. Otrokov, I. I. Klimovskikh, F. Calleja, J. H. D. O. Vilkov, A. G. Rybkin, D. Estyunin, S. Mu, H. O. A. L. Vázquez de Parga, R. Miranda, and A. A. F. Guinea, J. I. Cerdá, E. V. Chulkov, "A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update," no. 111, pp. 0–13, 2018.
[39] H. T. Tor et al., "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," Comput. Methods Programs Biomed., vol. 200, p. 105941, 2021, doi:10.1016/j.cmpb.2021.105941.
[40] S. Diwaker, S. K. Gupta, and N. Gupta, "Classification of EEG Signal using Correlation Coefficient among Channels as Features Extraction Method," Indian J. Sci. Technol., vol. 9, no. 32, Aug. 2016, doi:10.17485/ijst/2016/v9i32/100742.
[41] A. Einizade, M. Mozafari, M. Rezaei-Dastjerdehei, E. Aghdaei, A. M. Mijani, and S. Hajipour Sardouie, "Detecting ADHD children based on EEG signals using Graph Signal Processing techniques," in 27th National and 5th International Iranian Conference of Biomedical Engineering, ICBME 2020, 2020, pp. 264–270, doi:10.1109/ICBME51989.2020.9319456.

Vandana Joshi Vandana Joshi has completed her M. Tech from Dr. Babasaheb Ambedkar University at Aurangabad, Maharashtra, India, and is currently pursuing a PhD at Gujarat Technological University, Ahmedabad, Gujarat, India. Since 2013, she has been serving as an Assistant Professor at SCET, Surat. Prior to this, she accumulated five years of teaching experience at SVNIT, Surat. Vandana Joshi serves as the advisor of the IEEE Computer Society chapter in the Gujarat section and is also a member of IAENG. She has authored numerous research papers and patents. Her research interests lie primarily in the fields of Machine Learning, Deep Learning, and Biomedical Signal Processing.

Nirali Nanavati Nirali Nanavati obtained her PhD from the Computer Engineering Department at SVNIT, Surat, India, and has been serving as an Associate Professor at SCET, Surat, since 2016. Prior to this, she accumulated five years of teaching experience at SVNIT, Surat. She holds an M.S. degree in Computer Science from the New Jersey Institute of Technology, Newark, NJ, USA. Before joining SVNIT, she worked as a Technical Consultant at IBM, France, and as a Programmer Analyst at Infosys Technologies Ltd. She has authored numerous research papers, book chapters, and patents. Her research interests primarily focus on efficient algorithms for machine learning and artificial intelligence.

1) Date of modification – June 26, 2025.
2) Brief description of the changes: The affiliation of Vandana Joshi has been updated.