IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 March 2023, accepted 29 March 2023, date of publication 3 April 2023, date of current version 6 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3264266

RESEARCH ARTICLE

Optimal Channels and Features Selection Based ADHD Detection From EEG Signal Using Statistical and Machine Learning Techniques

MD. MANIRUZZAMAN[1], MD. AL MEHEDI HASAN[2], NOBUYOSHI ASAI[1], AND JUNGPIL SHIN[1], (Senior Member, IEEE)

[1]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu 965-8580, Japan
[2]Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh

Corresponding author: Jungpil Shin (jpshin@u-aizu.ac.jp)

This work was supported by the Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (KAKENHI), Japan, under Grant JP21H00891.

ABSTRACT Attention deficit hyperactivity disorder (ADHD) is one of the major psychiatric and neurodevelopment disorders worldwide. Electroencephalography (EEG) signal-based approach is very important for the early detection and classification of children with ADHD. However, diagnosing children with ADHD using full EEG channels with all features may lead to computational complexity and overfitting problems. To solve these problems, machine learning (ML)-based ADHD detection was designed by identifying optimal channels and its significant features. In this work, support vector machine and t-test based, two separate approaches were devised to select optimal channels individually and then proposed a hybrid channel selection approach by combining these two channel selection methods in order to select the optimal channels. After that, LASSO logistic regression-based model was used to select the important features from the selected channels. Finally, six ML-based classifiers, like Gaussian process classification (GPC), random forest, k-nearest neighbors, multilayer perceptron, decision tree, and logistic regression were applied for the detection of children with ADHD and evaluated their performances using accuracy and area under the curve (AUC). This study utilized a total of one hundred twenty-one children, with sixty-one children with ADHD, aged 7-12 years, and had nineteen channels. Ten different channels were selected by SVM based and an independent t-test-based approach separately and six overlapping channels were identified from both channel selection methods. Then, we selected twenty-eight features from selected six channels using LASSO. Using only six channels and twenty-eight features, GP-based classifier achieved an accuracy rate of 97.53% and AUC of 0.999. This is an improvement of 3% over previously developed techniques published in the literature. This study illustrated that LASSO with GP-based system performed outstanding performance in distinguishing children with ADHD from healthy children. This proposed system will be helpful to doctors and physicians in order to detect children with ADHD at an early stage and take the necessary steps for the patients to access appropriate healthcare services, receive effective treatment, and be more conscious of maintaining their lives.

INDEX TERMS Attention deficit hyperactivity disorder, electroencephalography, channel selection, t-test, feature selection, machine learning.

I. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is one of the major psychiatric and neurodevelopment disorders worldwide that affects 5% of children worldwide [1]. About 11% of U.S children aged 4-17 years are affected by ADHD [2]. Children with ADHD have various problems like inattention, impulsivity, and hyperactivity [3]. ADHD is usually diagnosed in children between the ages of 6 and 12 years and can last until adulthood [4], with serious implications, including suicide [5]. Over 40% of children and younger who suffer from ADHD develop behavioral problems until

The associate editor coordinating the review of this manuscript and approving it for publication was Utku Kose.

adults [6], [7], leading to serious problems [8]. ADHD is also significantly associated with comorbidities like asthma, depression, anxiety, and learning difficulties [9], [10]. Males are more likely to have ADHD than females and their behavior differs [11].

Nowadays, the efficient diagnosis of children with ADHD is still a major problem. Various research works have been carried out to propose an automated system for early diagnosis of children with ADHD [10], [12], [13], [14]. There is still a scope to propose an automated system for the early detection and classification of children with ADHD. Early detection of ADHD will be helpful for the patients to access appropriate healthcare services, receive effective treatment, and be more conscious of maintaining their lives. There are various neuroimaging and neurophysiological methods, like, electromyography (EMG), event-related potentials (ERP), and electroencephalography (EEG) which are widely used to investigate the effect of ADHD on brain signals.

In this study, we mainly focus on EEG analysis since it is a more popular, low-cost, non-invasive, and portable technique. Since EEG signals are captured directly from the human brain, they may be more accurate and valid in reflecting the brain's inner physiological conditions. Various types of works have used EEG signals to draw different types of conclusions in different fields. For example, detection or classification of seizure using EEG [15], [16], automatic stage of sleep-scoring using EEG [17], develop a portable wireless closed-loop seizure controller using EEG [18]. Moreover, EEG-based analysis is widely used for the detection of various neurological disorders like ADHD [12], [14], Parkinson's disease [19], etc. EEG data is a type of multivariate time series data. It typically consists of a collection of time-ordered data points linked to various time-dependent features containing local and global patterns. Local patterns describe the discriminative features of the dataset, whereas, global patterns show the overall trend of the dataset. Both local and global patterns of EEG data may be captured by any EEG classification system. Nowadays, effective feature extraction and feature selection from EEG signals is still a major challenge [20], [21], [22]. In order to address these challenges, various feature extraction methods like statistical features [23] and deep learning-based features [23] have been widely used to extract features for analyzing EEG data. Unlike other domains, ADHD can be also diagnosed from EEG signals [12], [14] and is necessary to extract features from EEG signals. Nowadays, there are various types of linear and non-linear features widely used for the diagnosis of children with ADHD. For example, Khaleghi et al. [24] extracted various morphological, time domain, frequency, and non-linear features from EEG signals for diagnosing children with ADHD. Alt1nkaynak et al. [4] also extracted morphological, non-linear, and wavelet features to diagnose children with ADHD. In our current study, we have also extracted time domain, morphological, and non-linear features on the basis of previous studies [4], [24], [25].

Generally, the recorded EEG datasets have several channels. While we analyze EEG signals with all channels, some irrelevant channels may reduce robustness and also may reduce the model performances. So, channel selection is very important in ADHD like other domains [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40]. Because, it can provide mainly three types of benefits such as (i) reduce the computational complexity of any processing task performed on EEG signals by selecting the relevant, (ii) helps to improve the performance, and (iii) reduce the setup time in some applications [26]. Moreover, the channel selection approach was used as an effective tool by many researchers in different fields, such as EEG emotion [26], [27], [28], [29], [30], personal identification [31], [32], user identification [33], seizure detection [34], [41], intruder detection [35], screening of alcoholism [36], depression detection [39], [40], detecting drowsiness [42], auditory attention detection [37], [38], brain-computer interfaces [43], [44] and so on. It was noted that several studies proposed effective predictive-based approaches for the detection of children with ADHD without selecting potential channels from EEG signals [4], [12], [13], [14], [24], [25], [45], [46], [47], [48]. Although only a few studies gave more attention on the selection of effective channels [49], [50].

From these surveys, we got the motivation to do this work which will be beneficial in the ADHD research domain as well as the clinical domain. So, there is still a scope to detect children with ADHD by selecting effective channels. In this study, we mainly focused on selecting relevant channels for ADHD detection from EEG signals. In literature, there were various methods like variance [34], entropy [34], normalized mutual information (NMI) [26], k-nearest neighbors (k-NN) [50], Gaussian mixture model (GMM) [50], and so on were widely used to select the relevant channels from EEG signals. In this work, we selected channels from two different views. One was from a discriminative view and the other was from a parametric statistical test view. Then, we proposed a hybrid channel selection approach by combining these two channel selection methods in order to select the optimal channels. The combined hybrid channels selection method provides more evidence for a channel to be selected as a more relevant channel. This hybrid channel selection approach is a new approach for channel selection.

Efficient feature selection-based approaches pay more attention to determining the potential features of ADHD. Feature selection is very important because it removes redundant features, improves the model's performance, and reduces overfitting problems. In the literature, there were various feature selection techniques such as principal component analysis (PCA) [49], [51], minimum redundancy maximum relevance (mRMR) [52], MI [53], t-test [4], [25], support vector machine recursive elimination (SVM-RFE) [52], least absolute shrinkage and selection operator (LASSO) [25], logistic regression (LR) [10], and so on, which were widely used to select the potential features in any domain like

ADHD. For example, Tenev et al. [54] used the forward selection method to select the significant features for ADHD. Khoshnoud et al. [51] implemented PCA for dimension reduction and extracted features which were strongly correlated to each other. Chen et al. [52] applied the mrMR-based approach to select the most discriminative features. Maniruzzaman et al. [25] implemented two effective feature selection methods as independent t-test and LASSO for the identification of potential biomarkers of children with ADHD. In this current study, we also applied LASSO logistic regression (LASSO-LR) based model for the selection of more effective features from the selected channels. These more effective features were fed into machine learning (ML) algorithms that can be easily classified children into two classes: ADHD and healthy controls.

Nowadays, ML-based approaches have been widely used in the field of medical imaging [46], time series [20], [21], [55], [56], and bioinformatics [57]. Like other domains, ML-based approaches were also widely used to discriminate children with ADHD from healthy control [4], [10], [13], [24], [25], [47], [49], [52], [58], [59], [60]. For example, Muller et al. [61] proposed SVM-based children with ADHD detection system. Tenev et al. [54] investigated the changes in the characteristics of mismatch negativity in adults with ADHD and healthy controls using ML-based algorithms and SVM obtained a classification accuracy of 82.3%. Kim et al. [62] performed SVM with linear kernel for classification and its performance was evaluated using classification accuracy. Khoshnoud et al. [51] recorded EEG data from 12 children with ADHD and 12 healthy controls during eyes-closed resting. SVM and neural networks (NN) were employed to discriminate children with ADHD from healthy controls and got a classification accuracy of 83.3%, which was achieved by SVM. Ahmadlou and Adeli [63] proposed an ADHD detection approach with the integration of non-linear features, wavelets, signal-based processing techniques, and neural networks and got a classification accuracy of 95.6%. Parashar et al. [13] also proposed an automated system for detecting children with ADHD. Helgadóttir et al. [60] proposed a classification approach for the detection of children with ADHD using EEG signals. They trained an SVM-based model with a 10-fold CV and obtained a classification accuracy of 76.0%.

In the present study, we adopted well-known, well-established, and widely used six ML-based classifiers as Gaussian process classification (GPC), random forest (RF), k-NN, multilayer perceptron (MLP), decision tree (DT), and LR for the discrimination of children either having ADHD or healthy controls. In a summary, the contributions of this study are as follows:

- We extracted different types of time domain, morphological and non-linear features from EEG signals.
- An relevant EEG channel selection method was proposed that used SVM and independent t-test to determine the optimal channels. Also, propose a hybrid channel selection approach by combining these two channel selection methods in order to select the most relevant channels.
- Moreover, an efficient feature selection method was also proposed that used LASSO to determine the more potential features for children with ADHD.
- Finally, we selected an ML-based classifier from the set of different classifiers: GPC, RF, k-NN, MLP, DT, and LR, respectively to classify children as ADHD and healthy controls with higher classification accuracy compared with other classifiers.
- Finally, we performed a performance comparison of our proposed system with channel selection and without channel selection.

The overall layout of this paper is as follows: Section II presents materials and methods are presented that include data acquisition, feature extraction, feature normalization, channel selection, feature selection techniques, and classification models. Experimental setup and performance metrics are presented in Section III. Results and discussion are presented in Section IV. Finally, the conclusion and future work direction is presented in Section V.

II. MATERIALS AND METHODS

The overview of the proposed ML-based approach for the prediction of children with ADHD from EEG signals is shown in Fig. 1. The first step is to divide the EEG-based dataset (that contains both children with ADHD and healthy subjects) into two parts: the training set and the testing set. These two parts are separated by a dotted line as follows: the training set or offline system (left) and the testing set or online system (right). The next step is the data prepossessing and data normalization and then, the extraction of different time domain, morphological, and non-linear features from the EEG signals. The fourth step is to select the potential EEG channels using two methods, such as SVM and independent t-test, and then, combined them to select the best combination of channels. Moreover, the top significant biomarkers or features are selected using LASSO. These significant biomarkers or features are trained on an ML-based framework for the classification of children with ADHD and healthy controls. The next step is to estimate the training parameters of the six classifiers (GPC, RF, k-NN, MLP, DT, and LR) and transform them into an online system to predict the children as ADHD and healthy controls.

A. DATA ACQUISITION

We utilized an ADHD-based EEG dataset for the experiment, which was extracted from a publicly available IEEE data port [64]. The sampling frequency of the ADHD-based EEG dataset was 128 Hz. The dataset was collected from sixty healthy children (males: 50 vs. females: 10) and sixty-one children with ADHD (males: 48 vs. females: 13). The average age of the one hundred twenty-one children was 7-12 years. Each child watched the images of cartoon characters and was asked to count the characters. There were different numbers

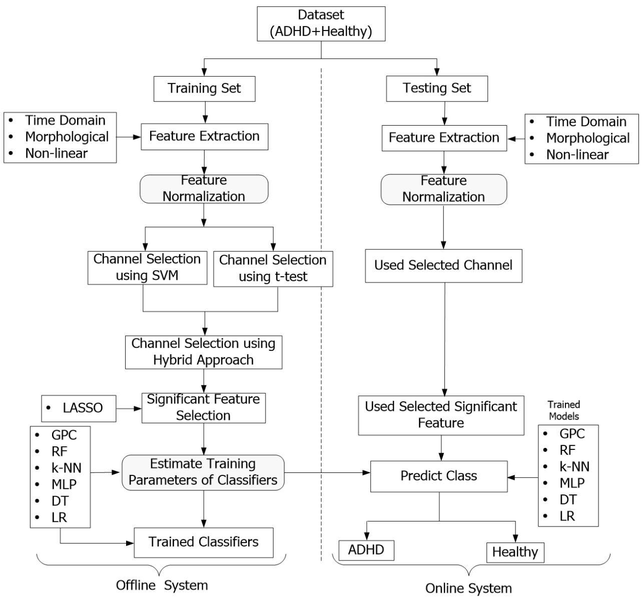


FIGURE 1. A proposed ML-based system for the prediction of children with ADHD from EEG signals.

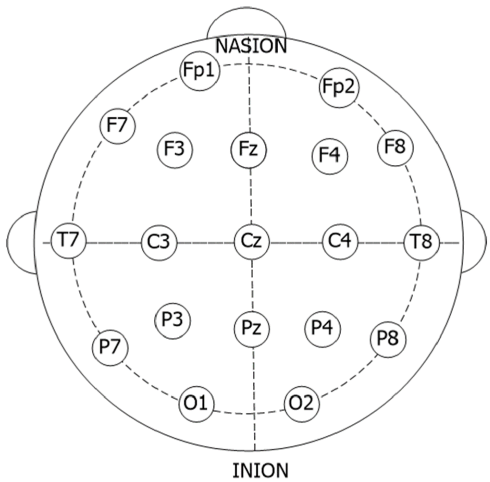


FIGURE 2. International 10-20 system-based electrode locations distribution.

in each image, from 5 to 16, and the size of the images was so large that the children were able to easily see and count. The dataset had 19 EEG channels such as Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz. The electrode distribution used in this work is depicted in Fig. 2. Since the performance of each child in the cognitive activities differed, the overall time period of EEG recordings varied across the dataset.

B. FEATURE EXTRACTION

Feature extraction is one of the most critical parts of classification. If the features are not selected properly, the classification's performance will be declined. Effective feature extraction from EEG data is more crucial that can be used to identify and forecast various brain illnesses [65]. There are available various feature extraction methods in the literature. In this paper, we used three statistical feature extraction methods such as (i) time domain, (ii) morphological, and (iii) non-linear features to extract various kinds of features from EEG signals. These feature extraction methods are briefly explained in detail as follows:

TABLE 1. List of extracted time domain features.

Features	Description	Calculation formula
Mean	Mean	$\frac{1}{n}\sum_{t=1}^{n} z_t$
Median	Median	Median (z_t)
Q1	1st quartile	Q1 (z_t)
Q3	3rd quartile	Q3 (z_t)
σ	Standard deviation	$\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(z_t-\mu)^2}$
CV	Coefficient of variation	$CV=\frac{\sigma}{\mu}$
β_1	Skewness	$E\left(z_t-\mu\right)^3/E\left(z_t-\mu\right)^2$
β_2	Kurtosis	$E(z_t-\mu)^4/E(z_t-\mu)^2$
ENG	Energy	$\sum_{t=1}^{n} z_t^2$
Power	Power	$\frac{1}{n}\sum_{t=1}^{n} z_t^2$
Activity	Hjorth Parameter Activity	$\frac{1}{n}\sum_{t=1}^{n}(z_t-\mu)^2$
Mobility	Hjorth Parameter Mobility	$\sqrt{(\sigma(dz/dt))^2/\sigma^2}$
Complexity	Hjorth Parameter Complexity	$\sqrt{\text{Mob.}(dz/dt)/\text{Mobi.}(z_t)}$

TABLE 2. List of extracted morphological features.

Features	Description	Calculation formula
AA	Absolute amplitude	$\max\lvert z_t\rvert$
PA	Positive area	$\sum_t(z_t+\lvert z_t\rvert)$
NA	Negative area	$\sum_t(z_t-\lvert z_t\rvert)$
TA	Total area	PA+NA
PP	Peak to peak	max $\lvert z_t\rvert$-min $\lvert z_t\rvert$

1) TIME DOMAIN FEATURES

We extracted 13 different time domain features from the EEG signals. These attributes were mean [24], [25], median [25], 1st and 3rd quartile [25], standard deviation [25], coefficient variations [25], skewness [24], [25], kurtosis [24], [25], energy [25], [66], power [25], activity [66], mobility, and complexity [66]. All of these 13-time domain features are computed from each of the 19 channels. The list of extracted feature names, their short description, and the calculation formula are presented in Table 1.

2) MORPHOLOGICAL FEATURES

Morphological features are one kind of statistical feature extraction method which are also computed from EEG signals. In previous studies, several available morphological features were computed from EEG signals [24], [67]. In the present study, we computed five morphological features which were also extracted from each of the 19 channels. Let z_t be the considered signal. The list of extracted morphological feature names, their short description, and calculation formula are shown in Table 2.

3) NON-LINEAR FEATURES

Non-linear analysis can provide crucial and helpful information about the electrical activity patterns of the brain. In the present study, we extracted four non-linear features as

TABLE 3. List of extracted non-linear features.

Features	Description	Calculation formula
PFD	Petrosian Fractal Dimension	$\frac{\log_{10}(N)}{\log_{10}(N)+\log_{10}\left(\frac{N}{N+0.4N_\delta}\right)}$
KFD	Katz Fractal Dimension	$\frac{\log_e(N-1)}{\log_e(N-1)+\log_e(d/L)}$
HFD	Higuchi Fractal Dimension	$\frac{1}{K}\sum_{M=1}^{k} L_m(k)$
DFA	Detrend Fluctuation Analysis	$L^H \times \sigma(z_t)$

N: Length of the time series; N_δ: No. of sign changes in the signal derivative; L: Sum of the spacing between sequential points; d is the distance between the first and farthest point of EEG signal, and H is Hurst parameter.

Petrosian, Katz, and Higuchi fractal dimensions, and detrend fluctuation analysis was computed from EEG signals [24]. In Table 3, we described the list of extracted non-linear features, their descriptions, and the calculation formula.

C. FEATURE NORMALIZATION

Feature normalization is a process to minimize redundancy and improve the efficiency of the data. It is also known as feature scaling. We normalized the features (see in Eq. 1) using the following formula:

$$z = \frac{X-\mu}{\sigma} \tag{1}$$

where X is the original feature vectors or input features; μ and σ are the mean and standard deviation of the respective features. z is the standardized value and its values lie between 0 to 1.

D. CHANNEL SELECTION METHODS

EEG-based processing has become one of the more attractive research fields in the last decades. The analysis of a large number of channels has several problems, such as overfitting, high computational cost, and time [34]. The optimal channel is needed in order to improve the performance of the model and save computational costs and time. In our current study, two-channel selection methods, such as SVM and independent t-test were adopted to determine the optimal channels.

1) SUPPORT VECTOR MACHINE-BASED CHANNEL SELECTION

SVM is a supervised ML algorithm that may be used for feature selection [68], classification [69], and regression [70]. The goal of SVM is to find a hyperplane in a high-dimensional space that can clearly classify children as ADHD or healthy controls, and it must be able to solve (see in Eq. 2 and Eq. 3) the following constraint problem:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j K(x_i,\ x_j) \tag{2}$$

where, α_i is value of the i^{th} coefficients.

Subject to

$$\sum_{i=1}^{n} {y_i}^T\alpha_i = 1, 0 \le \alpha_i \le C, i = 1, \ldots, n \ \&\ \forall\, i = 1, 2, 3, \ldots, n \tag{3}$$

The final discriminate function (see in Eq. 4) is written as follows:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i,\ x_j) + b \tag{4}$$

where b is the bias term.

In this research, we used radial basis kernel which was clearly defined in Eq. 5 as follows:

$$K(x_i,\ x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{5}$$

We tuned the hyperparameters as cost (C) and gamma (γ) on the basis of the grid search method. In the current study, we used SVM to select the most relevant channels from EEG signals. We selected the most relevant channels using the following algorithm:

Algorithm: SVM-based channel selection:

Step 1: Extract features (here, 22 features extracted) from each EEG channel.
Step 2: Take 80% of the dataset for training and 20% for the test set.
Step 3: SVM was trained on the training set.
Step 4: Compute the classification accuracy for each channel.
Step 5: Repeat **Step 2** to **Step 4** five times.
Step 6: Compute the average of the classification accuracy.
Step 7: Repeat **Step 1** to **Step 6** for all channels (here, the number of channels is 19).
Step 8: Sort the channels based on classification accuracy in descending order.
Step 9: Choose the channel that will produce a classification accuracy of more than 85.0%.

2) INDEPENDENT T-TEST BASED CHANNEL SELECTION

The independent t-test is a parametric test that is commonly used to compare the mean difference between two groups (ADHD vs. healthy control). It is also widely used for the identification of significant biomarkers for various diseases like cancers [57], ADHD [25], and diabetes [71] etc. In this study, we used an independent t-test for the selection of channels. The procedure of channel selection using an independent t-test is described in the following algorithm:

Algorithm: t-test based channel selection:

Step 1: Compute the mean and variance of each feature (here, 22 features are extracted from each channel) by ADHD vs. healthy control.
Step 2: Compute the t-test statistics under the null hypothesis as follows:

$$t = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{6}$$

where, μ_1 and μ_2 are the mean of the ADHD and healthy control; s_1^2 and s_2^2 are the variances of the ADHD and healthy control; $n = (n_1 + n_2)$ is the total of number respondents; n_1 and n_2 are total number of ADHD and healthy control. The above-mentioned statistic in Eq. 6 follows the t-distribution with $n_1 + n_2 - 2$ degrees of freedom (df).

Step 3: Compute the probability value (p-value) based on the value of the t-test statistic for each feature from t-distribution with $n_1 + n_2 - 2$ df and level of significance (α).

Step 4: Repeat **Step 1** to **Step 3** for extracted all features from each channel.

Step 5: Repeat **Step 1** to **Step 4** for all channels (Here, the number of channels is 19).

Step 6: Compute the mean of each channel (see in Eq. 7) overall features using the following formula:

$$\bar{X}_{i\cdot} = \frac{1}{f}\sum_{j=1}^{f} X_{ij} \tag{7}$$

where, i and j is the number of channels and features, respectively; f is the total number of features, X_{ij} is the value of the calculated p-value for the i^{th} channels and j^{th} features.

Step 7: Sort the channels on the basis of the p-value mean in ascending order.

Step 8: Choose the channel whose p-value mean is less than 0.05.

3) HYBRID APPROACH FOR CHANNEL SELECTION

To identify the most optimal channels and avoid missing the relevant channels, we identified the optimal channels using a hybrid approach by combing two-channel selection methods (SVM, and independent t-test). We identified the most relevant channels which were shown in Eq. 8:

$$\text{Hybrid approach} = \bigcap_{i=1}^{k} \text{Channel selection methods}_i \tag{8}$$

where, k is the no. of channel selection methods (here, k=2).

E. FEATURE SELECTION TECHNIQUE

Sometimes, the dataset has redundant, irrelevant, and inaccurate features. As a result, the performance of the predictive model is very low. Feature selection is needed to remove these redundant and irrelevant features and identify the most significant features as well as to improve the performance of the predictive model. In this work, the LASSO-LR model was applied to select more pertinent features, increase the performance of predictive models, and reduce computational time and cost [72], [73]. The calculation procedure of the LASSO-LR-based model is described in the following section.

1) LASSO

LASSO is one of the popular methods that is widely used to select the most significant features [74]. We adopted the LASSO-LR-based model [75], [76] because our response variable had two categories, like ADHD or healthy control. The aim of LASSO is to minimize the error sum of squares under the L_1 constraint on the regression coefficient. In order to estimate the regression coefficient, the log-likelihood function (see in Eq. 9) can be written as:

$$l(B) = \sum_{i=1}^{n}\left\{y_i X^T B - \log_e\left(1 + exp\left(X^T B\right)\right)\right\} \tag{9}$$

where $B = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ is regression coefficients; $X^T = (X_0, X_{i1}, X_{i2}, \ldots, X_{ik})$ are the input features with $X_0 = 1$, and y_i is the response variable that takes two values, "1" for ADHD and "0" for healthy control. Therefore, we have to minimize the following negative log-likelihood function with an L_1 penalty term which was explained in Eq. 10:

$$\sum_{i=1}^{n}\left\{\log_e\left(1 + exp\left(X^T B\right)\right) - y_i X^T B\right\} + \lambda\sum_{j=1}^{k}|B_j| \tag{10}$$

where λ is the amount of shrinkage. To obtain the optimal value of λ, we adopted a 5-fold cross-validation (CV) protocol. LASSO chooses only the features whose coefficients have non-zero coefficients. We implemented a LASSO-LR-based model using the *"glmnet"* package in R with version 4.1.2.

F. CLASSIFICATION MODEL

In the current study, we employed six ML-based algorithms such as GPC, DT, RF, k-NN, MLP, and LR to distinguish children with ADHD from healthy control. These six ML-based classifiers are briefly discussed in the following subsections:

1) GAUSSIAN PROCESS CLASSIFICATION

A GP is a generalization of the Gaussian probability distribution that can be utilized as a non-parametric supervised ML-based algorithm. It is a random function that is fully specified by a mean function and a covariance or kernel function [77]. The GP-based model attempts to take advantage of the best of two different schools of techniques: SVM, developed by Vapnik in the early nineties of the last century, and Bayesian methods. GP also has different types of kernel functions, like SVMs. GP can easily be capable of predicting class labels as well as class probabilities compared to SVM. Recently, GP-based classification has been widely used for the prediction of diabetes [78], cancer [57], etc. In this study, we have adopted three types of kernel functions, such as radial basis function (RBF), dot product, and rational quadratic kernel, which are defined in Table 4.

In addition to the three kernels, each has some additional parameters, like l, α, and σ^2, called hyperparameters (shown in TABLE 4). The explanation of these hyper-parameters

TABLE 4. Three types of GPC kernels names, calculation formula, and their descriptions.

SN	KT	Calculation formula
1	RBF	$K(x_i,\ x_j) = exp(-\frac{\|x_i - x_j\|^2}{2l^2}), l > 0$
2	Dot product	$K(x_i,\ x_j) = \begin{cases} \sigma_0^2 + x_i.x_j, & \text{if } \sigma_0^2 > 0 \\ x_i.x_j, & \sigma_0^2 = 0 \end{cases}$
3	Rational quadratic	$K(x_i,\ x_j) = (1 + \frac{\|x_i - x_j\|^2}{2l^2})^{-\alpha}$

KT: Kernel types

is more clearly explained in the existing paper [77]. These hyper-parameters are user-defined that needed to optimize in order to improve model performance. In our current study, we optimized these hyper-parameters using a grid search method. After optimizing parameters, we chose the better kernel on the basis of classification accuracy.

2) RANDOM FOREST

RF is an ensemble learning that was developed to solve the various problems, that occurred in DTs [79]. It is improved by averaging a set of DTs. It is easily enabled to solve the overfitting problem that is usually associated with DTs [80]. Moreover, it can be also used for both regression and classification problems. It is also a tree-based classifier that uses the input features of a set of DTs for making decisions over training sets. It is regarded to be more accurate and better performance compared to DT. In RF, there has also some parameters such as the number of trees (n_estimators), maximum depth (max_depth), minimum samples required to be a leaf node (min_samples_leaf), the minimum number of samples required to split an internal node (min_samples_split) and the maximum number of leaf nodes (max_leaf_nodes), and bootstrapping, which is more clearly explained in [81]. These parameters are used in order to build an RF-based model, which is also user-defined. In this work, we tuned these parameters using the grid search method in order to improve the performance of the RF-based model.

3) K-NEAREST NEIGHBORS

k-NN is a non-parametric method in statistics, first developed in 1951 by Fix and Hodges [82]. It is not only used for classification problems but also used for regression problems. In both problems, the input comprises of the k-nearest training samples in the dataset. If the output is continuous, then *k*-NN is used as regression, whereas *k*-NN is used as classification when the output is a categorical variable. Different metrics such as Euclidean distance, hamming distance, etc., can be used in the *k*-NN method. In the current study, we have used *k*-NN with Euclidean distance for the classification of children either with ADHD or healthy controls. In *k*-NN-based classification model, the value of *k* is the number of nearest neighbors to include in the majority of the voting process, in which *k* is user-defined and the optimum value of *k* is highly data-dependent [83]. We tuned the value of *k* using the grid search method and selected the optimum value of *k*, at which point the *k*-NN model provided the highest classification accuracy.

4) MULTILAYER PERCEPTRON

MLP is also a supervised ML-based technique that is a fully connected class of feed-forward neural networks. It can also be used for regression and classification [84]. MLPs are made up of neurons known as perceptions. MLP is comprised of three nodes of layers: input, hidden, and output layers. Input layers represent the input features while output layers represent the output or class label. A hidden layer is a layer between the input and output layers. Each node (except the input node) uses an activation function. There are various activation functions such as sigmoid, tanh, ReLU, etc. MLP has high processing power that can be used to solve both linear and non-linear problems. MLP has some additional parameters (activation functions, optimization functions, learning rate, etc.), called hyper-parameters. The choice of the optimum value of these hyper-parameters is user-defined and depends on the data structure [84]. In this work, we tuned these hyper-parameters using the grid search method to improve the model's performance.

5) DECISION TREE

DT is one of the supervised predictive models in statistics and ML. It may be used for both classification and regression problems; however, it is most often used to tackle classification problems [85]. It is a tree-structure-based classifier, with leaf nodes representing class labels, internal nodes representing the input features, and branches representing the decision rules [85]. The main goal of this model is to build a model on the training set and predict the class label on the test set on the basis of input features [85]. There are also some hyper-parameters (max_features, min_samples_leaf2, and min_samples_split) in the DT model, which are more clearly explained in [85], which are also user-defined. So, these parameters need to be defined during the trained model. In the present work, we also optimized these parameters using the grid search method.

6) LOGISTIC REGRESSION

LR is one of the simplest and most well-known supervised ML-based algorithms that is widely used to predict the probability of a response variable. The response variable must be dichotomous, which means it takes only two possible values: "1" for yes/success or "0" for no/failure. It has established a relationship between the response variable and a set of predictors. The logistic function can be clearly explained in Eq. 11:

$$\mathrm{P(z)} = \frac{1}{1 + \exp\text{ (-z)}} \tag{11}$$

where, $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +, \ldots, + \beta_k X_k$; $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ represents the sets of unknown parameters that must be estimated, and $X_1, X_2, \ldots, X_k$ represents the sets of given predictors. Now, we have trained the LR-based model by

TABLE 5. Hyperparameter optimization of classifiers using grid search method.

Classifiers	Search range of each parameter
GPC	kernel= ("RBF", "DotProducts", "RationalQuadratic"), length-scale= (1 to 5), alpha= (0.04, 0.05, 0.06), sigma= (0.01, 0.02, 0.03, 0.05, 0.06, 0.07, 0.08, 0.09)
RF	max_depth= (2,3,5, None), n_estimators= (15, 30,60,120), min_samples_split= (2, 3, 10), min_samples_leaf= (1, 3, 10), bootstrap= ("True", "False"), criterion= ("gini", "entropy")
k-NN	n_neighbors= (2 to 13), leaf_size= (4, 5, 6)
MLP	hidden_layer_sizes= [(120,120,50), (60,120,50), (60, 240, 100)), activation= ('relu','tanh','logistic'), alpha= (0.01, 0.05, 0.001), solver= ('adam'), learning_rate= ("constant", "adaptive"]
DT	max_features= ("auto", "sqrt", "log2"), min_samples_split= (2 to 15), min_samples_leaf= (1 to 11)
LR	None

estimating these unknown parameters using a maximum likelihood estimator over the training dataset. Using these estimated parameters, we have predicted the class label or response variable (here, ADHD and healthy controls) over the test dataset and also computed the probability of the response variable or class label.

III. EXPERIMENTAL SETUP AND PERFORMANCE METRICS

In this section, we have first discussed the experimental setup in order to conduct this study. Then, the performance metrics are also discussed in this section.

A. EXPERIMENTAL SETUP

The R-programming language with version 4.1.2 and Python with version 3.10 is used for this experiment. Windows 10 version 21H1 (build 19043.1151) 64-bit is used as the operating system. In terms of hardware, Intel(R) Core (TM) i5-10400 with 16 GB RAM setup is used. In this work, first, we selected the relevant channels from EEG signals. Then, we adopted 5-fold cross-validation for the feature selection and classification model. During training models, we optimized various hyper-parameters of utilized ML-based algorithms. We set a range of each hyper-parameter in each predictive model, which is shown in Table 5. After that, we performed another experiment to compare the performance of our proposed system with channel selection and without channel selection.

B. PERFORMANCE METRICS

Five performance metrics such as accuracy (ACC), recall (Rec), precision (Prec), F1-score, and area under the curve (AUC) computed from the ROC curve were used to evaluate the performance of predictive models. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are used to calculate the ACC, Rec, and Prec, which are mathematically defined in Eq. 12 to Eq. 15:

1) ACCURACY

ACC is the proportion of correctly classified cases (TP and TN) from the total number of cases and is mathematically defined as follows:

$$\mathrm{ACC}(\%) = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \times 100 \tag{12}$$

2) RECALL

Rec is the proportion of correctly positive classified cases (TP) from the cases which are actually positive and defined as follows:

$$\mathrm{Rec}\ (\%) = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100 \tag{13}$$

3) PRECISION

Prec is the proportion of correctly positive classified cases (TP) from the cases which are predicted as positive and defined as follows:

$$\mathrm{Prec}\ (\%) = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times 100 \tag{14}$$

4) F1-score

F1-score is the harmonic mean of recall and precision. It is mathematically defined as follows:

$$\text{F1-score}\ (\%) = 2\times\left(\frac{\mathrm{Rec}\ \times\ \mathrm{Prec}}{\mathrm{Rec}\ \times\ \mathrm{Prec}}\right)\times 100 \tag{15}$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We designed this section as follows: first, we discussed the result parts of channel selection with how the most relevant EEG channels were selected for ADHD using SVM and the independent t-test method. Secondly, we discussed the result part of feature selection with how the potential features were selected using the LASSO-LR-based model. Thirdly, we discussed the predictive performance evaluations of six classifiers. Finally, we discussed the comparison of our current study with existing studies in the literature.

A. EXPERIMENTAL RESULTS OF CHANNEL SELECTION

1) CHANNEL SELECTION USING SVM

The utilized dataset in this study had 19 channels. We extracted 22 features from each channel. Then, we employed SVM on the training set and optimized the hyperparameters as follows: cost (C): 120 and gamma (γ): 0.001. After optimizing the hyperparameters, we trained the models again and computed the classification accuracy of each channel for the test set, as shown in Fig. 3a. We chose the channels that provided a classification accuracy of more than 85.0%. As shown in Fig. 3a, we observed that ten channels provided more than 85.0% classification accuracy, and also their positions were presented in Fig. 3b. Therefore, we selected ten channels as Fz, F8, P7, P3, F3, Fp2, C4, Pz, C3, F7 for our next experiments.

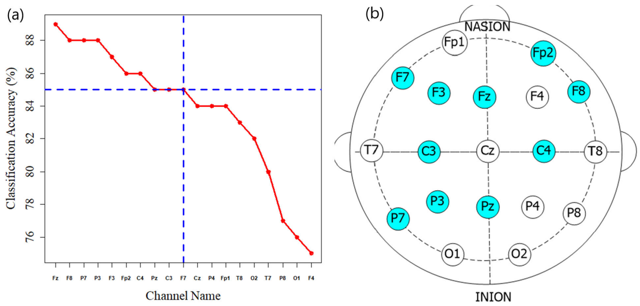


FIGURE 3. SVM-based Channel selection approach: (a) Classification accuracy of each channel (sorted in descending order); (b) Distribution of the electrode locations. Turquoise colors represent electrodes selected using SVM.

TABLE 6. Channels obtained from independent t-test.

Channel	p-value mean	Channel	p-value mean
F4	0.0162	Fp1	0.193
C3	0.0180	O2	0.245
F3	0.0180	P3	0.284
C4	0.0187	Fp2	0.294
T7	0.0237	P8	0.323
F8	0.0267	P4	0.350
T8	0.0398	Pz	0.372
Cz	0.0427	P7	0.433
F7	0.0435	O1	0.483
Fz	0.0451	–	–

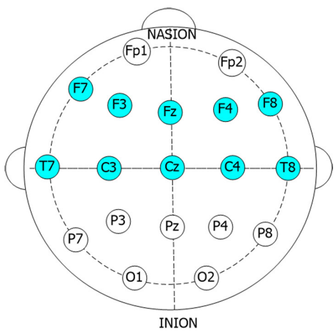


FIGURE 4. Distribution of the electrode locations. Turquoise colors represent electrodes selected using a t-test.

2) CHANNEL SELECTION USING INDEPENDENT T-TEST

To identify the region of the brain that was affected by ADHD and the more relevant EEG channels that can play an important role in discriminating the children with ADHD from healthy control children. We computed the p-values of the extracted 22 features for each channel using an independent t-test. Then, we calculated the mean of p-values of each channel over 22 features and ordered them from the smallest to largest, which are shown in Table 6. As shown in Table 6, we observed that the p-values of ten channels were less than 0.05. Therefore, we chose these ten channels (F4, C3, F3, C4, T7, F8, T8, Cz, F7, and Fz) as more relevant channels for children with ADHD, and their position was also shown in Fig. 4.

3) CHANNEL SELECTION USING HYBRID APPROACH

We observed that SVM and independent t-test identified ten different channels (See Fig. 3 and Table 6). The next experiment of this study was to choose the unique or overlapping channels from the two-channel selection methods. We identified six common or overlapping channels (Fz, F8, F3, C4, C3, and F7) among them, which are shown in Fig. 5a, and also their positions were illustrated in Fig. 5b. Finally, we extracted a total of 132 (6×22) features from selected six channels, which were used for feature selection in order to extract more potential features for children with ADHD.

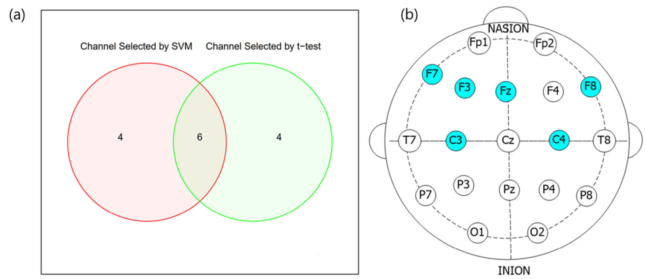


FIGURE 5. (a) Relevant channels selection from hybrid approach; (b) Distribution of the electrode locations. Turquoise colors represent electrodes selected using a hybrid channel selection approach.

B. EXPERIMENTAL RESULTS OF FEATURE SELECTION

In order to select the significant features, we adopted the LASSO-LR-based model and needed to determine the optimum value of λ. We optimized the value of λ using the minimum criteria for a 5-fold CV protocol. The identification process of significant features for children with ADHD using the LASSO-LR-based model is presented in Fig. 6. In order to determine the optimum value of λ, we generated a plot of binomial deviance versus $\log(\lambda)$ (See in Fig. 6a). The optimal values of the parameter (λ) are indicated by the dotted vertical lines, and a value λ of 0.0095 with $\log(\lambda) = -4.658$ was chosen (See in Fig. 6a). The optimum value (-4.658) obtained from Fig. 6a was used to select the significant features with non-zero coefficients in Fig. 6b. After removing the features with zero coefficients, twenty-eight features with non-zero coefficients were selected using LASSO-LR based model (See in Fig. 6b), and their (twenty-eight features) contributions are presented in Fig. 6c. The selected twenty-eight significant features were used in ML-based algorithms for the discrimination of children as ADHD and healthy control.

C. HYPERPARAMETER OPTIMIZATION OF CLASSIFIERS

In ML, each classifier or algorithm has some additional parameters called hyperparameters. These hyperparameters need to be tuned in order to improve the model's performances [86]. In this study, we set different hyperparameters of the classifiers and tuned these hyperparameters with a 5-fold CV on the training set using the grid search method. We chose the hyperparameters of the classifiers that provided the highest classification accuracy. The optimized values of the hyperparameters of different classifiers are shown in Table 7.

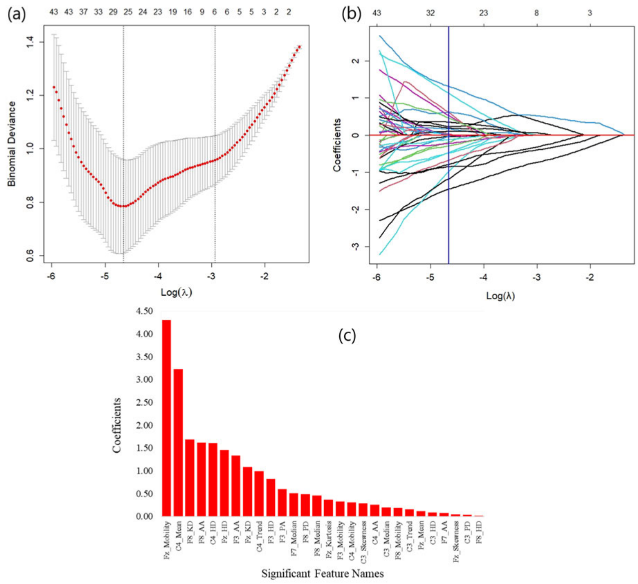


FIGURE 6. Significant features selection for children with ADHD using LASSO-LR based model: (a) Optimal parameter selection in LASSO-LR. (b) The binomial deviance was generated vs. log(λ). (c) Contribution of selected significant twenty-eight features for ADHD.

TABLE 7. Optimized value of the classifiers.

Classifiers	Optimized values of the hyperparameters
GPC	Kernel: RBF, Length-scale (l): 2
RF	leaf: 2,split: 2, n_estimators': 60.
k-NN	Leaf_size: 4, n_neighbors: 3
MLP	function: "tanh", α: 0.05, layer:(60, 120, 50)
DT	Max_features: "log2", leaf: 2, split:2
LR	None

D. PREDICTIVE PERFORMANCE OF SIX CLASSIFIERS

After selecting the more relevant six common channels along with twenty-eight features and optimized values of hyperparameters, we implemented six classifiers (GPC, RF, k-NN, MLP, DT, and LR) to discriminate the classification of children either having ADHD or healthy control. The different performance metrics of the proposed GP-based classifier with other classifiers are shown in Table 8. As shown in Table 8, it was noted that GP with RBF-based classifier gave comparatively better performance scores for the selected channels than other classifiers. The GP with RBF classifier provided 97.53% classification accuracy, 98.46% recall, 96.92% precision, and 0.999 AUC. Similarly, MLP and LR-based classifiers also performed better than RF, k-NN, and DT. The classification accuracy of MLP and LR were 95.03% and 94.23%, and the AUC values were 0.994 and 0.995. The normalized confusion matrix of our proposed system is presented in Figure 7.

E. PERFORMANCE COMPARISON WITH CHANNEL SELECTION VS WITHOUT CHANNEL SELECTION

In order to show the efficiency of our proposed system, we conducted another experiment to make a comparison between model performance with channel selection and without channel selection. We have computed the model performance using all channels in order to differentiate children with ADHD from healthy children. In both cases (channel selection and without channel selection), we have used the same feature extraction and classification methods. We observed that the GP-based classifier obtained a classification accuracy of 96.70% in the case of all channels approach, whereas our proposed system with the combination of hybrid-based channel selection-based classifier obtained a classification accuracy of 97.53%, which is a comparatively approximately 1% improvement over without a channel selection approach (See in Table 9).

TABLE 8. Performance scores of classifiers for six common channels.

Classifiers	ACC (%)	Rec (%)	Prec (%)	AUC
GPC	97.53	98.46	96.92	0.999
RF	75.17	85.51	75.34	0.898
k-NN	72.80	80.77	73.34	0.760
MLP	95.03	98.46	92.75	0.994
DT	68.60	69.10	70.44	0.707
LR	94.23	98.46	91.38	0.995

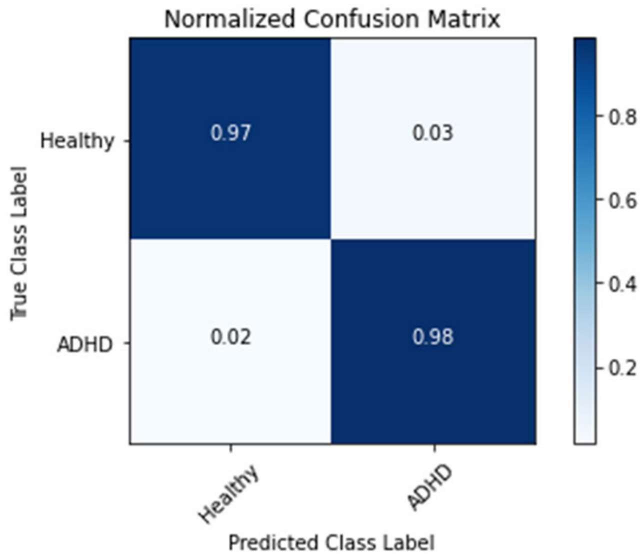


FIGURE 7. Normalized confusion matrix for the proposed system.

F. COMPARISON OF OUR PROPOSED WORK AGAINST PREVIOUS STUDIES

The comparison of different studies on the classification of ADHD is briefly illustrated in Table 10. Mohammadi et al. [47] extracted different types of fractal dimension-based non-linear features from EEG signals and identified more efficient features using DISR and mrMR methods, which were used to classify children as ADHD and healthy controls. They showed that 93.7% of classification accuracy was obtained by MLP. Moreover, different time domain features were also used to discriminate against children with ADHD and healthy controls. Yang et al. [48] adopted PCA for the dimension reduction, and k-NN produced a classification accuracy of 83.5%. Autoregressive (AR)-based features were extracted from EEG signals by Marcano et al. [50]. They adopted two classifiers, such as

TABLE 9. Comparison of performance between our proposed system with channel selection and without channel selection.

Proposed system	No. of CH	No. of EF	CST	SC	FST	SF	Classifiers	Performance
Without channel selection	19	22	×	19	LASSO	35	GPC, RF, k-NN, MLP DT, LR	ACC: 96.70% AUC: 0.999
With Channel Selection	19	22	SVM, t-test	6	LASSO	28	GPC, RF, k-NN, MLP DT, LR	ACC: 97.53% AUC: 0.999

EF: Extracted features; CST: Channel selection technique; FST: Feature selection technique; SC: Selected channel; SF: Selected features.

TABLE 10. Comparison with our works against existing published similar works in literature.

Authors	DS	CH No.	EF	CHS	FST	Classifiers	ACC	AUC
Mohammadi et al. [47]	60	19	NL	×	mRMR, DISR	MLP	93.7	×
Yang et al. [48]	30	×	TD	×	PCA	k-NN, SVM	3.5	×
Marcano et al. [50]	8	26	AR	Yes	×	k-NN, GMM UBM	90.0	0.980
Khoshnoud et al. [51]	24	×	NL		PCA	SVM, RBFNN	83.3	×
Khaleghi et al. [24]	60	×	Ml, TD, FrD, TF, NL, FD	×	×	k-NN	86.4	×
Chen et al. [52]	108	128	TD, FD	×	mRMR	SVM	84.6	0.916
Chen et al. [58]	107	128	×	×	×	SVM, MLP, CNN	94.7	×
Chow et al. [49]	60	32	Mobility, TBR	Yes	PCA	LR	79.2	0.885
Altınkaynak et al. [4]	46		Ml, NL, WV	×	t-test	MLP, NB, SVM, AB, k-NN, LR, RF	91.3	0.910
Parashar et al. [13]	120	19	×	×	×	AB, RF, SVM	84.0	×
Ekhlasi et al. [59]	121	19	×	x	GA	ANN	89.7	x
Maniruzzaman et al. [25]	121	19	Ml, TD	×	t-test, LASSO	SVM, k-NN, MLP, LR	94.2	0.964
Our proposed study	121	19	TD, Ml, NL	Yes	LASSO	GPC, RF, k-NN, MLP DT, LR	97.5	0.999

DS: Data size; CHS: Channel selection; NL: Non-linear; TD: Time domain; Ml: Morphological; FrD: Frequency domain; TF: Time frequency; WV: Wavelet; mRMR: Minimum redundancy and maximal relevance; DISR: Double input symmetrical relevance. ACC is presented in %.

k-NN and GMM to select the best combination of channels, and the highest accuracy rate of 90.0% and an AUC of 0.98 was obtained by k-NN.

Khoshnoud et al. [51] explored non-linear features from 19 EEG channels. They employed SVM and neural networks (NN) for the discrimination of children with ADHD after selecting efficient features using PCA. The classification accuracy rate of 83.3% was achieved by SVM. Khaleghi et al. [24] explored various features such as time domain, frequency domain, time-frequency, and non-linear features, respectively. Using these features, k-NN-based classifier was employed for classification and obtained an accuracy and recall rate of 86.4% and 91.8%, respectively. Chen et al. [52] employed the mRMR-based method for the identification of significant features from different time and frequency features. The results of their findings illustrated that SVM obtained an accuracy rate of 84.59% and 0.916 AUC. Chen et al. [58] also applied convolution neural networks for the prediction of ADHD children. They also illustrated that the accuracy rate of CNN was improved by almost 10% compared to their previous study [52].

Chow et al. [49] extracted two types of features: mobility and theta beta ratio (TBR), from 32 EEG-based channels. They adopted an independent t-test for the identification of optimum channels using mobility and TBR-based features. The optimum channels were identified on the basis of p-values ($p<0.05$). They identified 12 channels and adopted LR to classify children as ADHD and healthy controls. They showed that Hjorth mobility features were better predictors for ADHD than TBR. LR produced better accuracy, recall, and AUC of 79.2%, 79.6%, and 0.885, respectively. Alt1nkaynak et al. [4] discriminated children with ADHD and healthy controls using MLP, NB, SVM, k-NN, AdaBoost (AB), LR, and RF and obtained an accuracy rate of 91.3% by MLP. Parashar et al. [13] also studied sixty children with ADHD and sixty healthy control children. They adopted three classifiers (AB, RF, and SVM) for classification and obtained an 84.0% accuracy rate with AB. Ekhlasi et al. [59] proposed a system that can be easily classified into children with ADHD and healthy controls. They applied genetic algorithms for feature selection and NN for classification. They illustrated that NN produced the highest classification accuracy of 89.7%.

Our previous published papers [25], extracted different types of time domain and morphological features and adopted LASSO and t-test for feature selection. Moreover, four ML-based classifiers, such as SVM, k-NN, MLP, and LR were used for the classification of children with ADHD and healthy controls, and SVM produced 94.2% classification accuracy and 0.964 AUC. TABLE 10 confirmed that LASSO-LR based approach along with a GP-based classifier can identify and diagnose children with ADHD with 97.5% accuracy, which is comparatively higher than all existing methods of previous studies in the literature.

V. CONCLUSION AND FUTURE WORK DIRECTION

This study proposed an ML-based system for the diagnosis of children with ADHD on the basis of more relevant channel selection and feature extraction using EEG signals. The first step of this study was to extract twenty-two features from each EEG channel. The second step was to select the more relevant channels using two channel selection methods: SVM and independent t-test. Ten different channels out of nineteen were selected by SVM and an independent t-test. Then, six overlapping or unique channels were selected from both channel selection methods. Since each channel had twenty-two features, a total of one hundred thirty-two features were extracted from the selected six channels. The third step was to determine the more potential features using the LASSO-LR-based model. The fourth step was to classify the children as either having ADHD or healthy, where six classification methods were adopted and experimented. Our findings showed that GP-based classifier achieved the highest classification accuracy of 97.53% and AUC of 0.999. An improvement of almost 3% in classification accuracy and AUC was obtained over our previous published papers [25].

We will have the plan to extend this study by adding more children with ADHD subjects from EEG signals. We will also adopt our proposed method to mix EEG data of other similar psychiatric diseases. Furthermore, we will adopt a deep learning-based classifier and try to develop a web-based method for the automated prediction of children with ADHD that will help physicians to take the necessary steps for early diagnosis of ADHD.

ACKNOWLEDGMENT

The authors are grateful to Nasrabadi et al. [64] for providing this dataset in publicly available. They are also grateful to the editors and reviewers for their valuable suggestions in order to improve the manuscript.

REFERENCES

[1] G. Polanczyk, M. S. de Lima, B. L. Horta, J. Biederman, and L. A. Rohde, "The worldwide prevalence of ADHD: A systematic review and metaregression analysis," *Amer. J. Psychiatry*, vol. 164, no. 6, pp. 942–948, Jun. 2007.
[2] S. N. Visser, M. L. Danielson, R. H. Bitsko, J. R. Holbrook, M. D. Kogan, R. M. Ghandour, R. Perou, and S. J. Blumberg, "Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 53, no. 1, pp. 34–46, Jan. 2014.
[3] APA Diagnostic, *Diagnostic and Statistical Manual of Mental Health Disorders: DSM-5*, no. 5, 5th ed. Washington, DC, USA: American Psychiatric Publishing, 2013.
[4] M. Altinkaynak, N. Dolu, A. Guven, F. Pektas, S. Özmen, E. Demirci, and M. Izzetoglu, "Diagnosis of attention deficit hyperactivity disorder with combined time and frequency features," *Biocybernetics Biomed. Eng.*, vol. 40, no. 3, pp. 927–937, Jul. 2020.
[5] A. Stickley, A. Koyanagi, V. Ruchkin, and Y. Kamio, "Attention-deficit/hyperactivity disorder symptoms and suicide ideation and attempts: Findings from the adult psychiatric morbidity survey 2007," *J. Affect. Disorders*, vol. 189, pp. 321–328, Jan. 2016.
[6] J. Kooij, D. Bijlenga, L. Salerno, R. Jaeschke, I. Bitter, J. Balazs, J. Thome, G. Dom, S. Kasper, C. N. Filipe, and S. Stes, "Updated European consensus statement on diagnosis and treatment of adult ADHD," *Eur. Psychiatry*, vol. 56, no. 1, pp. 14–34, 2019.
[7] G. Castelpietra, A. K. S. Knudsen, E. E. Agardh, B. Armocida, M. Beghi, K. M. Iburg, G. Logroscino, R. Ma, F. Starace, N. Steel, and G. Addolorato, "The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990–2019: Findings from the global burden of disease study 2019," *Lancet Reg. Health Eur.*, vol. 16, pp. 100341–100359, May 2022.
[8] P. Preetha and R. Mallika, "Normalization and deep learning based attention deficit hyperactivity disorder classification," *J. Intell. Fuzzy Syst.*, vol. 40, no. 4, pp. 7613–7621, Apr. 2021.
[9] N. Tistarelli, C. Fagnani, M. Troianiello, M. A. Stazi, and W. Adriani, "The nature and nurture of ADHD and its comorbidities: A narrative review on twin studies," *Neurosci. Biobehavioral Rev.*, vol. 109, pp. 63–77, Feb. 2020.
[10] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with ADHD using behavioral activity: A machine learning analysis," *Appl. Sci.*, vol. 12, no. 5, pp. 2737–2749, 2022.
[11] E. M. Derks, J. J. Hudziak, and D. I. Boomsma, "Why more boys than girls with ADHD receive treatment: A study of Dutch twins," *Twin Res. Hum. Genet.*, vol. 10, no. 5, pp. 765–770, Oct. 2007.
[12] H. T. Tor, C. P. Ooi, N. S. Lim-Ashworth, J. K. E. Wei, V. Jahmunah, S. L. Oh, U. R. Acharya, and D. S. S. Fung, "Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals," *Comput. Methods Programs Biomed.*, vol. 200, pp. 105941–105954, Mar. 2021.
[13] A. Parashar, N. Kalra, J. Singh, and R. K. Goyal, "Machine learning based framework for classification of children with ADHD and healthy controls," *Intell. Autom. Soft Comput.*, vol. 28, no. 3, pp. 669–682, 2021.
[14] R. C. Joy, S. T. George, A. A. Rajan, and M. S. P. Subathra, "Detection of ADHD from EEG signals using different entropy measures and ANN," *Clin. EEG Neurosci.*, vol. 53, no. 1, pp. 12–23, Jan. 2022.
[15] S. Lahmiri and A. Shmuel, "Accurate classification of seizure and seizure-free intervals of intracranial EEG signals from epileptic patients," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 3, pp. 791–796, Mar. 2018.
[16] C. Lei, S. Zheng, X. Zhang, D. Wang, H. Wu, H. Peng, and B. Hu, "Epileptic seizure detection in EEG signals using discriminative Stein kernel-based sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–15, 2021.
[17] S.-F. Liang, Y.-H. Kuo, Y.-H. Hu, Y.-H. Pan, and Y.-H. Wang, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1649–1657, Jun. 2012.
[18] C.-P. Young, S.-F. Liang, D.-W. Chang, Y.-C. Liao, F.-Z. Shaw, and C.-H. Hsieh, "A portable wireless online closed-loop seizure controller in freely moving rats," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 513–521, Feb. 2011.
[19] S. K. Khare, V. Bajaj, and U. R. Acharya, "PDCNNet: An automatic framework for the detection of Parkinson's disease using EEG signals," *IEEE Sensors J.*, vol. 21, no. 15, pp. 17017–17024, Aug. 2021.
[20] Z. Xiao, X. Xu, H. Xing, S. Luo, P. Dai, and D. Zhan, "RTFN: A robust temporal feature network for time series classification," *Inf. Sci.*, vol. 571, pp. 65–86, Sep. 2021.
[21] Z. Xiao, H. Zhang, H. Tong, and X. Xu, "An efficient temporal network with dual self-distillation for electroencephalography signal classification," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2022, pp. 1759–1762.
[22] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Müller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, Mar. 2019.
[23] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 204–218, 2021.
[24] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," *Res. Biomed. Eng.*, vol. 36, no. 1, pp. 1–11, Mar. 2020.
[25] M. Maniruzzaman, J. Shin, M. A. M. Hasan, and A. Yasumura, "Efficient feature selection and machine learning based ADHD detection using EEG signal," *Comput., Mater. Continua*, vol. 72, no. 3, pp. 5179–5195, 2022.
[26] Z. Wang, S. Hu, and H. Song, "Channel selection method for EEG emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019.
[27] E. Yildirim, Y. Kaya, and F. Kilic, "A channel selection method for emotion recognition from EEG based on swarm-intelligence algorithms," *IEEE Access*, vol. 9, pp. 109889–109902, 2021.
[28] P. Sarma and S. Barma, "Emotion recognition by discriminating EEG segments with high affective content from automatically selected relevant channels," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
[29] J. W. Li, S. Barma, P. U. Mak, F. Chen, C. Li, M. T. Li, M. I. Vai, and S. H. Pun, "Single-channel selection for EEG-based emotion recognition using brain rhythm sequencing," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2493–2503, Jun. 2022.

[30] M. Javidan, M. Yazdchi, Z. Baharlouei, and A. Mahnam, "Feature and channel selection for designing a regression-based continuous-variable emotion recognition system with two EEG channels," *Biomed. Signal Process. Control*, vol. 70, pp. 102979–102991, Sep. 2021.
[31] Z. A. A. Alyasseri, O. A. Alomari, S. N. Makhadmeh, S. Mirjalili, M. A. Al-Betar, S. Abdullah, N. S. Ali, J. P. Papa, D. Rodrigues, and A. K. Abasi, "EEG channel selection for person identification using binary grey wolf optimizer," *IEEE Access*, vol. 10, pp. 10500–10513, 2022.
[32] Z. A. A. Alyasseri, O. A. Alomari, M. A. Al-Betar, M. A. Awadallah, K. H. Abdulkareem, M. A. Mohammed, S. Kadry, V. Rajinikanth, and S. Rho, "EEG channel selection using multiobjective cuckoo search for person identification as protection system in healthcare applications," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–18, Jan. 2022.
[33] Z. A. A. Alyasseri, O. A. Alomari, J. P. Papa, M. A. Al-Betar, K. H. Abdulkareem, M. A. Mohammed, S. Kadry, O. Thinnukool, and P. Khuwuthyakorn, "EEG channel selection based user identification via improved flower pollination algorithm," *Sensors*, vol. 22, no. 6, pp. 2092–2108, 2022.
[34] J. Duun-Henriksen, T. W. Kjaer, R. E. Madsen, L. S. Remvig, C. E. Thomsen, and H. B. D. Sorensen, "Channel selection for automatic seizure detection," *Clin. Neurophysiol.*, vol. 123, no. 1, pp. 84–92, Jan. 2012.
[35] L. A. Moctezuma and M. Molinas, "Multi-objective optimization for EEG channel selection and accurate intruder detection in an EEG-based subject identification system," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Apr. 2020.
[36] S. Bavkar, B. Iyer, and S. Deosarkar, "Rapid screening of alcoholism: An EEG based optimal channel selection approach," *IEEE Access*, vol. 7, pp. 99670–99682, 2019.
[37] A. M. Narayanan and A. Bertrand, "Analysis of miniaturization effects and channel selection strategies for EEG sensor networks with application to auditory attention detection," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 1, pp. 234–244, Jan. 2020.
[38] A. Gonzalez, I. Nambu, H. Hokari, and Y. Wada, "EEG channel selection using particle swarm optimization for the classification of auditory event-related potentials," *Sci. World J.*, vol. 2014, pp. 1–11, Mar. 2014.
[39] J. Shen, X. Zhang, X. Huang, M. Wu, J. Gao, D. Lu, Z. Ding, and B. Hu, "An optimal channel selection for EEG-based depression detection via kernel-target alignment," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2545–2556, Jul. 2021.
[40] Y. Zhang, K. Wang, Y. Wei, X. Guo, J. Wen, and Y. Luo, "Minimal EEG channel selection for depression detection with connectivity features during sleep," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105690.
[41] J. S. Ra, T. Li, and Y. Li, "A novel permutation entropy-based EEG channel selection for improving epileptic seizure prediction," *Sensors*, vol. 21, no. 23, pp. 7972–7983, 2021.
[42] V. P. Balam and S. Chinara, "Statistical channel selection method for detecting drowsiness through single-channel EEG-based BCI system," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.
[43] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.
[44] 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 Netw.*, vol. 133, pp. 193–206, Jan. 2021.
[45] A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin, and M. Elsharnouby, "FOCUS: Detecting ADHD patients by an EEG-based serious game," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1512–1520, Jul. 2018.
[46] B. Miao, L. L. Zhang, J. L. Guan, Q. F. Meng, and Y. L. Zhang, "Classification of ADHD individuals and neurotypicals using reliable RELIEF: A resting-state study," *IEEE Access*, vol. 7, pp. 62163–62171, 2019.
[47] 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.
[48] J. Yang, W. Li, S. Wang, J. Lu, and L. Zou, "Classification of children with attention deficit hyperactivity disorder using PCA and K-nearest neighbors during interference control task," in *Advances in Cognitive Neurodynamics (V)* (Advances in Cognitive Neurodynamics). Singapore: Springer, 2016, pp. 447–453.
[49] J. C. Chow, C.-S. Ouyang, C.-T. Chiang, R.-C. Yang, R.-C. Wu, H.-C. Wu, and L.-C. Lin, "Novel method using Hjorth mobility analysis for diagnosing attention-deficit hyperactivity disorder in girls," *Brain Develop.*, vol. 41, no. 4, pp. 334–340, Apr. 2019.
[50] J. L. L. Marcano, M. A. Bell, and A. A. L. Beex, "EEG channel selection for AR model based ADHD classification," in *Proc. IEEE Signal Process. Med. Biol. Symp. (SPMB)*, Dec. 2016, pp. 1–6.
[51] S. Khoshnoud, M. A. Nazari, and M. Shamsi, "Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals," *J. Integrative Neurosci.*, vol. 17, no. 1, pp. 17–30, Feb. 2018.
[52] H. Chen, W. Chen, Y. Song, L. Sun, and X. Li, "EEG characteristics of children with attention-deficit/hyperactivity disorder," *Neuroscience*, vol. 406, pp. 444–456, May 2019.
[53] B. Abibullaev and J. An, "Decision support algorithm for diagnosis of ADHD using electroencephalograms," *J. Med. Syst.*, vol. 36, no. 4, pp. 2675–2688, Aug. 2012.
[54] A. Tenev, S. Markovska-Simoska, L. Kocarev, J. Pop-Jordanov, A. Muller, and G. Candrian, "Machine learning approach for classification of ADHD adults," *Int. J. Psychophysiol.*, vol. 93, no. 1, pp. 162–166, Jul. 2014.
[55] H. Xing, Z. Xiao, D. Zhan, S. Luo, P. Dai, and K. Li, "SelfMatch: Robust semisupervised time-series classification with self-distillation," *Int. J. Intell. Syst.*, vol. 37, no. 11, pp. 8583–8610, Nov. 2022.
[56] X. Zhang, R. Ma, M. Li, X. Li, Z. Yang, R. Yan, and X. Chen, "Feature enhancement based on regular sparse model for planetary gearbox fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.
[57] M. Maniruzzaman, M. J. Rahman, B. Ahammed, M. M. Abedin, H. S. Suri, M. Biswas, A. El-Baz, P. Bangeas, G. Tsoulfas, and J. S. Suri, "Statistical characterization and classification of colon microarray gene expression data using multiple machine learning paradigms," *Comput. Methods Programs Biomed.*, vol. 176, pp. 173–193, Jul. 2019.
[58] 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, Sep. 2019.
[59] A. Ekhlasi, A. M. Nasrabadi, and M. Mohammadi, "Classification of the children with ADHD and healthy children based on the directed phase transfer entropy of EEG signals," *Frontiers Biomed. Technol.*, vol. 8, pp. 115–122, Jun. 2021.
[60] H. Helgadottir, O. O. Gudmundsson, G. Baldursson, P. Magnusson, N. Blin, B. Brynjolfsdottir, A. Emilsdottir, G. B. Gudmundsdottir, M. Lorange, P. K. Newman, G. H. Johannesson, and K. Johnsen, "Electroencephalography as a clinical tool for diagnosing and monitoring attention deficit hyperactivity disorder: A cross-sectional study," *BMJ Open*, vol. 5, no. 1, Jan. 2015, Art. no. e005500.
[61] A. Muller, S. Vetsch, I. Pershin, G. Candrian, G.-M. Baschera, J. D. Kropotov, J. Kasper, H. A. Rehim, and D. Eich, "EEG/ERP-based biomarker/neuroalgorithms in adults with ADHD: Development, reliability, and application in clinical practice," *World J. Biol. Psychiatry*, vol. 21, no. 3, pp. 172–182, Mar. 2020.
[62] S. Kim, J. H. Baek, Y. J. Kwon, H. Y. Lee, J. H. Yoo, S.-H. Shim, and J. S. Kim, "Machine-learning-based diagnosis of drug-naive adult patients with attention-deficit hyperactivity disorder using mismatch negativity," *Transl. Psychiatry*, vol. 11, no. 1, p. 484, Sep. 2021.
[63] M. Ahmadlou and H. Adeli, "Wavelet-synchronization methodology: A new approach for EEG-based diagnosis of ADHD," *Clin. EEG Neurosci.*, vol. 41, no. 1, pp. 1–10, 2010.
[64] A. M. Nasrabadi, M. Samavati, and M. R. Mohammadi, "EEG data for ADHD/control children," *IEEE Dataport*, 2020, doi: 10.21227/rzfh-zn36.
[65] V. Harpale and V. Bairagi, "An adaptive method for feature selection and extraction for classification of epileptic EEG signal in significant states," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 33, no. 6, pp. 668–676, Jul. 2021.
[66] Y. Chen, R. Chang, and J. Guo, "Emotion recognition of EEG signals based on the ensemble learning method: AdaBoost," *Math. Problems Eng.*, vol. 2021, pp. 1–12, Jan. 2021.
[67] A. Khaleghi, A. Sheikhani, M. R. Mohammadi, A. M. Nasrabadi, S. R. Vand, H. Zarafshan, and M. Moeini, "EEG classification of adolescents with type I and type II of bipolar disorder," *Australas. Phys. Eng. Sci. Med.*, vol. 38, no. 4, pp. 551–559, Dec. 2015.
[68] J. Neumann, C. Schnörr, and G. Steidl, "Combined SVM-based feature selection and classification," *Mach. Learn.*, vol. 61, nos. 1–3, pp. 129–150, 2005.
[69] M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, "Support vector machine and random forest modeling for intrusion detection system (IDS)," *J. Intell. Learn. Syst. Appl.*, vol. 6, no. 1, pp. 45–52, 2014.
[70] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Netw.*, vol. 17, no. 1, pp. 113–126, Jan. 2004.

[71] M. Nariai, H. Abe, Y. Hoshioka, Y. Makino, and H. Iwase, "Biomarker profiling of postmortem blood for diabetes mellitus and discussion of possible applications of metabolomics for forensic casework," *Int. J. Legal. Med.*, vol. 136, no. 4, pp. 1075–1090, 2022.

[72] F. Jalali-Najafabadi, M. Stadler, N. Dand, D. Jadon, M. Soomro, P. Ho, H. Marzo-Ortega, P. Helliwell, E. Korendowych, M. A. Simpson, and J. Packham, "Application of information theoretic feature selection and machine learning methods for the development of genetic risk prediction models," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, Art. no. 23335.

[73] H. Naseri, E. O. D. Waygood, B. Wang, Z. Patterson, and R. A. Daziano, "A novel feature selection technique to better predict climate change stage of change," *Sustainability*, vol. 14, no. 1, pp. 40–62, 2021.

[74] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[75] S. M. Kim, Y. Kim, K. Jeong, H. Jeong, and J. Kim, "Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography," *Ultrasonography*, vol. 37, no. 1, pp. 36–42, Jan. 2018.

[76] S.-H. Yu, J.-H. Cai, D.-L. Chen, S.-H. Liao, Y.-Z. Lin, Y.-T. Chung, J. J. Tsai, and C. C. Wang, "LASSO and bioinformatics analysis in the identification of key genes for prognostic genes of gynecologic cancer," *J. Pers. Med.*, vol. 11, no. 11, pp. 1177–1190, 2021.

[77] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, vol. 32. Cambridge, MA, USA: MIT Press, 2006, p. 68.

[78] M. Maniruzzaman, N. Kumar, S. Islam, H. S. Suri, A. S. El-Baz, J. S. Suri, and M. M. Abedin, "Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm," *Comput. Methods Programs Biomed.*, vol. 152, pp. 23–34, Dec. 2017.

[79] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[80] M. Hiba, A. F. Ibrahim, S. Elkatatny, and A. Ali, "Prediction of cohesion and friction angle from well-logging data using decision tree and random forest," *Arabian J. Geosci.*, vol. 15, no. 1, pp. 1–11, Jan. 2022.

[81] L. Torre-Tojal, A. Bastarrika, A. Boyano, J. M. Lopez-Guede, and M. Graña, "Above-ground biomass estimation from LiDAR data using random forest algorithms," *J. Comput. Sci.*, vol. 58, Feb. 2022, Art. no. 101517.

[82] E. Fix and J. L. Hodges Jr., "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *Int. Stat. Rev.*, vol. 57, no. 3, pp. 238–247, 1989.

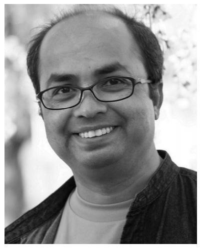
[83] P. Hall, B. U. Park, and R. J. Samworth, "Choice of neighbor order in nearest-neighbor classification," *Ann. Statist.*, vol. 36, no. 5, pp. 2135–2152, 2008.

[84] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, nos. 5–6, pp. 183–197, Jul. 1991.

[85] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[86] F. Khan, S. Kanwal, S. Alamri, and B. Mumtaz, "Hyper-parameter optimization of classifiers, using an artificial immune network and its application to software bug prediction," *IEEE Access*, vol. 8, pp. 20954–20964, 2020.

MD. MANIRUZZAMAN received the B.Sc., M.Sc., and M.Phil. degrees in statistics from the Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh, in 2013, 2014, and 2021, respectively. He is currently pursuing the Ph.D. degree with the Pattern Processing Laboratory, School of Computer Science and Engineering, The University of Aizu, Japan, under the direct supervision of Prof. Dr. Jungpil Shin and Prof. Dr. Md. Al Mehedi Hasan. He became a Lecturer and an Assistant Professor with the Statistics Discipline, Khulna University, Khulna, Bangladesh, in September 2018 and November 2020, respectively. He has coauthored more than 30 publications published in widely cited journals and conferences. His research interests include bioinformatics, artificial intelligence, pattern recognition, medical image, signal processing, machine learning, data mining, and big data analysis.



MD. AL MEHEDI HASAN received the B.Sc., M.Sc., and Ph.D. degrees in computer science and engineering from the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh, in 2005, 2007, and 2017, respectively. He became a Lecturer, an Assistant Professor, an Associate Professor, and a Professor with the Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology (RUET), Rajshahi, Bangladesh, in 2007, 2010, 2018, and 2019, respectively. Recently, he has completed his postdoctoral research with the School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Japan. He has coauthored more than 100 publications published in widely cited journals and conferences. His research interests include bioinformatics, artificial intelligence, pattern recognition, medical image, signal processing, machine learning, computer vision, data mining, big data analysis, probabilistic and statistical inference, operating systems, computer networks, and security.

NOBUYOSHI ASAI received the B.Eng. degree in information science and the M.Eng. and Ph.D. degrees in engineering from the University of Tsukuba, Japan, in 1991, 1993, and 1996, respectively. He was with Wave Front Company Ltd., Japan, in 1996. He has been an Associate Professor and a Senior Associate Professor with the School of Computer Science and Engineering, The University of Aizu, Japan, since 2000 and 2010, respectively. He was also a Visiting Researcher with the National Institute of Environmental Study, Japan, in 2002 and 2003. His research interests include applied mathematics, numerical analysis, quantum algorithms, high-performance computing, synesthesia, and signal processing on EEG or motion. He is a member of JSIAM and IPSJ. He also served as the general chair for the IEEE international conference. He served as the Editor for the *Journal of IPSJ* and currently serves as the Editor-in-Chief.

JUNGPIL SHIN (Senior Member, IEEE) received the B.Sc. degree in computer science and statistics and the M.Sc. degree in computer science from Pusan National University, South Korea, in 1990 and 1994, respectively, and the Ph.D. degree in computer science and communication engineering from Kyushu University, Japan, in 1999, under a scholarship from the Japanese Government (MEXT). He was an Associate Professor, a Senior Associate Professor, and a Full Professor with the School of Computer Science and Engineering, The University of Aizu, Japan, in 1999, 2004, and 2019, respectively. He has coauthored more than 300 published papers for widely cited journals and conferences. His research interests include pattern recognition, image processing, computer vision, machine learning, human–computer interaction, non-touch interfaces, human gesture recognition, automatic control, Parkinson's disease diagnosis, ADHD diagnosis, user authentication, machine intelligence, as well as handwriting analysis, recognition, and synthesis. He is a member of ACM, IEICE, IPSJ, KISS, and KIPS. He serves as the Editor for IEEE journals, *Sensors* (MDPI), *Electronics* (MDPI), and Tech Science. He serves as a reviewer for several major IEEE and SCI journals.