IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON NEURAL ENGINEERING INFORMATICS

Received August 22, 2019, accepted September 2, 2019, date of publication September 5, 2019,
date of current version September 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939623

Motor Imagery EEG Signals Classification Based on Mode Amplitude and Frequency Components Using Empirical Wavelet Transform

MUHAMMAD TARIQ SADIQ[1], XIAOJUN YU[1], ZHAOHUI YUAN[1], ZEMING FAN[1], ATEEQ UR REHMAN[2], GUOQI LI[3], AND GAOXI XIAO[4]

[1]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
[2]College of Internet of Things Engineering, Hohai University, Changzhou Campus, Changzhou 213022, China
[3]Center for Brain Inspired Computing Research, Department of Precision Instrument, Tsinghua University, Beijing 100084, China
[4]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798

Corresponding authors: Xiaojun Yu (xjyu@nwpu.edu.cn) and Zhaohui Yuan (yuanzhh@nwpu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61705184 and Grant 51875477, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2018JQ6014, in part by the Fundamental Research Funds for the Central Universities under Grant G2018KY0308, in part by the China Postdoctoral Science Foundation under Grant 2018M641013, and in part by the Seed Foundation of Innovation and Creation for Graduate Students, Northwestern Polytechnical University, under Grant ZZ2019028.

ABSTRACT As one of the key techniques determining the overall system performances, efficient and reliable algorithms for improving the classification accuracy of motor imagery (MI) based electroencephalography (EEG) signals are highly desired for the development of brain-computer interface (BCI) systems. In this study, we propose, for the first time to the best of our knowledge, a novel data adaptive empirical wavelet transform (EWT) based signal decomposition method for improving the classification accuracy of MI based EEG signals. Specifically, to reduce the system complexity and execution time, the proposed method selects 18 electrodes out of 118 to analyze the non-stationary and nonlinear EEG signal behaviors. Meanwhile, the method adopts the Welch power spectral density (PSD) analysis method for single mode selection out of the total 10 for each channel, and the Hilbert transform (HT) method for both instantaneous amplitude (IA) and instantaneous frequency (IF) signal components extraction for each selected mode. With seven commonly used machine-learning classifiers adopted, extensive experiments were conducted with the benchmark dataset IVa from BCI competition III to evaluate the performance of the proposed method. Results show that with the IA and IF component features being tested using the least-square support vector machine (LS-SVM) classifier, the EWT method achieves an average classification accuracy of 95.2% and 94.6% respectively, which is higher as compared with the existing methods. While for every participant, a classification accuracy of at least 80% could be achieved by employing a single feature only. Results also show that a combination of EWT and higher order statistics features, which contain both kurtosis and skewness of the extracted instantaneous components, help achieve a higher success rate. The better performances of EWT over those of the existing methods demonstrate the effectiveness and great potential of EWT for BCI system applications.

INDEX TERMS Brain-computer interface, motor imagery, electroencephalography, empirical wavelet transform, higher order statistics.

I. INTRODUCTION

Brain Computer Interface (BCI) system has numerous applications [1] and it can help disabled people interact with the actual world via thoughts alone [2], [3]. It collects the electrical signals produced from the brain activities and then translates them into outputs for control purposes via nonmuscular channels [4]. Motor imagery (MI) is the utmost common mental rehearsal used in BCI applications [3], [5], wherein participants are directed to imagine themselves performing a specific motor action, like moving a foot or hand, without any activation of muscles [6]. BCIs relying on MI of the users

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

have attracted extensive interests over the past years [7], [8] and various clinical methods, like electroencephalography (EEG) [9], magnetoencephalography (MEG) [10], functional magnetic resonance imaging (fMRI) [11], positron emission tomography (PET) [12], and single-photo emission computed tomography (SPECT) [13], have been used to monitor the brain activities during MI mental rehearsal. Among all these methods, MI-based EEG signal analysis is the most widely used one owing to its cheap setup, easy usage, superb temporal information and harmless nature [7].

Currently, most of the MI-based BCIs work as a pattern recognition systems, which identify the EEG signals with feature extraction and detection schemes first, and then translates these signals into different MI task commands. As an essential step of BCI development, the capturing and classification of EEG signals with high accuracy is thus of critical importance, and different signal processing techniques, which provides the internal mechanisms, procedures, and dynamics of the EEG signals, have been proposed over the past years. These signal processing techniques could be mainly categorized into the Fourier transform (FT) based, autoregressive (AR) model based, common spatial pattern (CSP) based, sparse representation (SR) as well as signal decomposition (SD) based ones. Precisely, the FT based methods are typically utilized for EEG signal power spectral analysis, yet the main drawback of these methods is the lack of time domain information [14], [15]. While the AR model-based methods are usually used for the AR spectrum or AR model coefficient calculations, and are computationally efficient even for online systems, however, they suffer from the artifacts, which reduces the BCI system reliability [16]–[20]. For more details about those FT and AR model-based methods, readers are referred to [21], [22].

As multi-channel BCI systems are attracting increasing interests, various spatial pattern based methods, which adopt suitable spatial filtering techniques for oscillatory EEG components analysis, have also been reported in [23]. A sparse spatial filter optimization (SSFO) method was first introduced to reduce the number of BCI channels based upon the feature variances of the projected signals, and an overall success rate of 73.5% was achieved. However, no automated method was presented for the selection of regularization parameters for this method [24]. While based on the CSP scheme for feature extraction, the regularized CSP (R-CSP) [25], spatially regularization CSP (SRCSP) [26], CSP based z-score linear discriminant classifier (ZLDA) [27], regularized-CSP with aggregation method (R-CSP-A) [28] sparse group representation model (SGRM) [29], temporally constrained sparse group spatial pattern (TSGSP) method [30], as well as the CSP-rank channel selection for multifrequency band EEG (CSP-RMF) [31] have also been proposed. These methods improved the overall EEG signal classification accuracy from 74.2% to 88.5%, yet they are restricted to a small sampling and may suffer from the over-fitting issue.

To furthermore improve the overall classification accuracy, clustering based least square support vector machine (LS-SVM) classifier has been proposed and improved the accuracy up to 88.32% [32]. The main limitation of this method, however, is that the parameters for the LS-SVM classifier were chosen manually. To facilitate the extraction and classification process, an iterative spatio-spectral pattern learning (ISSPL) method was proposed to perform feature extraction and classification within a single step in [33], and an overall classification accuracy up to 94.2% was achieved. However, little attention was paid to the selection of the optimal regularization parameters. Song and Epps extracted covariance based second ordered dynamic features from the motor cortex data of EEG signals according to a Fisher ratio criterion [34], while Wang and James employed a constraint-based independent component analysis to classify power features from rhythmic motor cortex area EEG data for MI tasks in [35]. Qin *et al.* performed a pilot study for the binary classification of MI tasks by employing motor cortex region and achieved a classification accuracy of 80% [36]. It has also shown that higher order statistics (HOS) features containing signal skewness (Sk) and kurtosis (Kt) are gaining popularity for the analysis of biomedical signals [21], [37].

Furthermore, some channel selection methods like correlation-based channel selection [38] results in low classification accuracy due to noise signal components. More recently, some other EEG signal feature extraction and classification techniques adopting the convolutional neural networks (CNN) or the recurrent neural networks (RNN) based deep learning schemes [39], [40], and with the sparse Bayesian extreme learning schemes [41], [42] have also been devised. For recent advancements on deep learning based BCI systems, readers are referred to [43] for more details. Due to the lack of extensive data required for model training, however, the classification accuracies provided by most of those methods were not high enough. The required system resources and high computational loads are two other limitations that may hinder the practical applications of these methods.

Signal decomposition (SD) based methods are relatively new for the classification of MI tasks, and the main idea of these methods is to decompose EEG signals into different sub-bands, and then extract the desired coefficients separately. Three SD based methods, namely empirical mode decomposition (EMD), discrete wavelet transform (DWT) and wavelet packet decomposition (WPD) were proposed in [21], which achieved an overall classification accuracy of 62.8%, 81.1%, and 92.8%, respectively. However, the main drawbacks of these methods are their relatively low classification success rates and the difficulties in selecting the appropriate number of decomposition levels, especially for WPD [21]. Empirical wavelet transform (EWT) is a relatively new SD based method for signal analysis [44] and is gaining increasing popularity in various domains like power engineering, biomedical engineering, civil engineering, electromechanical

engineering, renewable energy, Geoscience, as well as image processing [45]–[47]. However, it has not been utilized in neurocomputing domain for MI based EEG signal classifications so far.

In this paper, for the first time to the best of our knowledge, we evaluate the feasibility of EWT for MI EEG signal classification, and propose a novel EWT based data adaptive decomposition method for analyzing the non-stationary and nonlinear behaviors of MI EEG signals. An EWT based data adaptive decomposition technique is proposed for analyzing the non-stationary and nonlinear behaviors of MI EEG signals. To reduce the processing time and system complexity, 18 motor cortex channels are chosen among the overall 118 channels for analysis. The signal from each channel is decomposed into 10 adaptive frequency modes, and the Welch Power Spectral Density (PSD) is utilized to select the most sensitive one, while the Hilbert Transform (HT) method is applied on each selected mode to acquire the instantaneous amplitude (IA) and instantaneous frequency (IF) components for feature extraction. The extracted IA and IF features are finally provided to the selected classifiers for signal classification accuracy check. Experiments are conducted with the benchmark dataset IVa from BCI competition III to verify the effectiveness of the proposed EWT algorithm. For the sake of simplicity, the abbreviations used in this paper are listed in Table 1.

TABLE 1. Abbreviations used in the paper.

Abbr.	Means	Abbr.	Means
RH	Right Hand	RF	Right Foot
MI	Motor Imagery	LS-SVM	Least Square Support-Vector Machine
BCI	Brain- Computer Interface	PSD	Power Spectral Density
EWT	Empirical Wavelet Transform	HT	Hilbert Transform
EEG	Electroencephalography	HOS	Higher Order Statistics
IA	Instantaneous Amplitude	IF	Instantaneous Frequency
ROC	Receiver Operating Characteristics	EMD	Empirical Mode Decomposition
AUC	Area under the ROC curve		

The rest of the paper is organized as follows. Section II presents EEG dataset and performance metrics utilized in this study. Section III explains the main working steps of the proposed EWT algorithm. Section IV presents the experimental results obtained for evaluating the performances of the proposed EWT method. Section V presents the discussions, and section VI concludes this work.

II. MATERIALS AND PERFORMANCE METRICS

In this study, simulation experiments were conducted with the publicly available benchmark dataset IVa from BCI competition III [48] to evaluate the effectiveness of the proposed EWT method. The dataset was provided by the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen (Head: N.Birbaumer) and Max-Planck-Institute for Biological Cybernetics,Tübingen (Head: B. Schökopf), and the Department of Epileptology,Universität Bonn, and has been broadly utilized in [21], [24], [26]–[28], [32]–[34].

A. DATASET DESCRIPTION

The dataset IVa contains two MI tasks, which are right hand (RH, class1) and right foot (RF, class2) [48]. This dataset was acquired from five healthy participants that were fully relaxed seated in a chair, and they were labeled to be ''aa'',''al'',''av'',''aw'' and ''ay'' respectively. For each individual participant, the recording was made using BrainAmp amplifiers and a 128-channel Ag/AgCl electrode cap from ECI, and 118 EEG channels were measured at positions according to the extended international 10/20-system [49]. The dataset contains information about the MI EEG data at the four initial sessions with no feedback, while for each participant, there were totally 280 trials, with half trials being dedicated to class 1 tasks and the other half to class 2 tasks. Each participant performed either of the two MI tasks for 3.5 seconds(sec), with the training and testing trials being different for each participant. Among the total 280 trials, 168, 224, 84, 56 and 28 are selected to compose the training trials for participants ''aa'',''al'',''av'',''aw'' and ''ay'' respectively, while the rest of those trials compose the testing trials.

The visual cues were shown for 3.5 sec to choose any of the MI tasks. From 1.75 to 2.25 sec of random length periods, the target cues were suspended so that the participant could relax. The visual cues stimulation was comprised of two types of targets. The type 1 targets were indicated by letters appearing behind a fixation cross, whereas the type 2 targets were indicated by a randomly moving object. For participants ''aa'',''av'' and ''ay'' 3 sessions of type 2 and 1 session of type 1 were recorded, while for participants ''al'' and ''aw'' 2 sessions of both types were recorded [48].

B. DATASET PREPARATION

Li and Wen *et al.* proposed CC-LR [50] and a Modified CC-LR algorithms [51], and in all those studies, they used the labeled data (data with categories) for their experiments. In this study, we also used the same labeled data in our experiments, as the proposed algorithm requires a class label at each data sequence. The dataset IVa contains position markers to indicate the data points of 280 trials with 118 EEG electrodes for each participant separately. The original data lengths for both classes of participants ''aa'',''al'',''av'',''aw'' and ''ay'' are 298458×118, 283574×118, 283042×118, 282838×118 and 283562×118 data points, respectively. We considered the EEG segments from both classes which were responsible for specific MI tasks only [21]. For this purpose, the position markers representing the beginning of each trial were used. The original dataset was recorded with a 1000 Hz sampling frequency, whereas we used the available down-sampled version of the dataset at Fs = 100 Hz rate as those [21], [24], [26]–[28], [32]–[34]. We considered 3.5 sec from each EEG

segment for both classes, which resulted in 350 samples (3.5 × Fs). Therefore, we obtained a data matrix of 350 × 118 for each EEG segment of an individual class used for each participant.

C. PERFORMANCE METRICS

For a given dataset, a subset is selected as the test set while the rest of the subsets are gathered to form a training set. The classical K-fold cross-validation [51] method was adopted to divide the dataset into K mutually exclusive subsets, wherein K was determined by the size of the given dataset, and in this study, the parameter K was chosen to be K = 4 as an appropriate value to reduce the number of experiments, computational time as well as the variance of the estimator [50], [51].

To evaluate the performances of the proposed EWT based method, different statistical parameters, like classification accuracy (A_{cc}), true positive rate (TPR) or sensitivity (S_{en}), true negative rate (TNR) or specificity (S_{pe}), as well as area under the receiver operating characteristics (AUC) are used to measure the algorithm performances from 4-folds. These parameters are defined in Eqs. (1), (2) and (3). The overall algorithm classification performance A_{cc} was measured by taking the average of overall folds,

$$A_{cc} = \frac{TP + TN}{P + N} \tag{1}$$

$$S_{en} = \frac{TP}{P} \tag{2}$$

$$S_{pe} = \frac{TN}{N} \tag{3}$$

For the dataset IVa, TP means true positive denoting the number of correctly estimated RH MI tasks, while TN is true negative representing the number of correctly estimated RF MI tasks. P and N denote the total amount of RH and RF classes, respectively. All experiments were carried out on a personal computer with an Intel R Core (TM) i7-7500U CPU @ 2.70 GHz processor, 64-bit OS and 8 GB RAM using MATLAB R2018a. The EWT toolbox v.3.4 and LS-SVM toolbox v.1.8 were utilized for investigation.

III. METHODS

The proposed method consists of six different modules as depicted in Fig. 1, and each of which is explained briefly as follows.

A. MODULE 1:CHANNEL SELECTION

In this study, we propose a novel method based on EWT to improve the MI EEG signal classification accuracy while reducing the algorithm computational load and the hardware complexity. To reduce the computational loads, we adopted the same way as that in [34], [35] to select only 18 channels from the motor cortex region of the brain to perform EEG signal analysis. These 18 channels, which are responsible for monitoring the execution of MI tasks, were chosen manually around the motor cortex region, and they were labeled

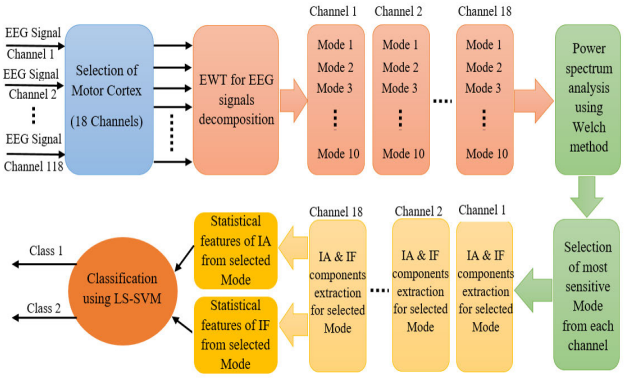

FIGURE 1. Block diagram of the proposed methodology.

to be $C5, C3, C1, C2, C4, C6, CP5, CP3, CP1, CP2, CP4, CP6, P5, P3, P1, P2, P4$ and $P6$, according to the standard 10/20-system [49] in this study.

B. MODULE 2: EWT BASED EEG SIGNALS DECOMPOSITION

In [35], Wang and James proposed to utilize the EWT method to overcome the limitations encountered by the EMD method in signal decomposition and analysis process. By utilizing a filter bank comprised of adaptive wavelet filters, the EWT method is able to decompose any non-stationary signal into different modes, with each being adjusted around an explicit frequency to fulfill the properties of intrinsic mode functions (IMF) [44]. The main steps of the EWT algorithm are presented as follows [44]:

Step1: Fast Fourier Transform (FFT) is performed to obtain the signal spectrum within the range from 0 to π.

Step2: EWT boundary detection method is executed to obtain the segmentation $w_k \mid k = 0 \ldots M$, for the Fourier spectrum.

Step3: Empirical wavelets are applied to all spectrum segmentations as band-pass filters, and a parameter β based on the concept of Meyer's wavelets and Littlewood-Paley theory [52] was defined for a tight frame as,

$$\beta < min_k \left(\frac{w_{k+1} - w_k}{w_{k+1} + w_k} \right) \tag{4}$$

The EWT toolbox v.3.4 was utilized in our experiments. For EEG signal decompositions, the scale space parameter was chosen to perform spectrum boundary detection, while the number of modes were empirically selected to be 10 for each individual channel. Additional details about the EWT method could be found in [44].

C. MODULE 3: POWER SPECTRUM ANALYSIS WITH WELCH PSD METHOD

It is important to mention that all the 10 modes for a particular channel do not contain relevant information related to the MI task. For this purpose, we executed the PSD of each mode and the mode that showed the maximum value of PSD, i.e. the signal power is concentrated in that oscillatory

component [53], is considered as the most sensitive mode for both MI tasks. The Welch PSD method was used for power spectrum analysis of all channels modes, and a brief summary of this method could be described as follows [54]:

Step1: dividing a sequence of length A into B sections with an equal length of C.

Step2: applying a window to each section and calculate the corresponding periodogram.

Step3: estimating the spectral density by averaging over the modified periodograms obtained from B sections. Additional details of the Welch PSD method could be found in [54].

D. MODULE 4: HILBERT TRANSFORM FOR INSTANTANEOUS COMPONENTS EXPLORATION

In this study, HT was adopted in the EWT method to extract the IA and IF components from filtered sub-band signals, and for any time-series signal $\dot{a}_{n+}(t)$ being analyzed, its HT mathematical formulation is given as,

$$\dot{a}_{n+}(t) = \dot{a}_n(t) + jH(\dot{a}_n(t)) \tag{5}$$

The instantaneous amplitude component of $\dot{a}_{n+}(t)$ can be expressed as,

$$\dot{A}_{\dot{a}_n}(t) = \sqrt{(\dot{a}_n(t))^2 + (H(\dot{a}_n(t)))^2} \tag{6}$$

The instantaneous phase component of $\dot{a}_{n+}(t)$ can be computed as follows,

$$\dot{\Theta}_{\dot{a}_n}(t) = arctan\left(\frac{H(\dot{a}_n(t))}{\dot{a}_n(t)}\right) \tag{7}$$

The instantaneous frequency component of $\dot{a}_{n+}(t)$ is obtained by taking the derivative of instantaneous phase component $\dot{\Theta}_{\dot{a}_n}(t)$ as,

$$\dot{F}_{\dot{a}_n}(t) = \frac{d}{dt}\left(\dot{\Theta}_{\dot{a}_n}(t)\right) \tag{8}$$

To show the proposed EWT method graphically, the channel $C5$ of participant "aa" in dataset IVa is randomly chosen as an example. Welch PSD method provides the most sensitive mode while HT is applied to the sensitive mode of $C5$ channel signal to extract the IA and IF components. The graphical demonstration of signals for the proposed methdology is shown in Fig. 2. The Fig. 2(a) represents the typical pattern of RH and RF classes in dataset IVa. The Fig. 2(b) shows 10 modes of channel $C5$ for each class and analysis results show that mode 10 has the highest PSD for both MI task classes, as shown in Fig. 2(c). The Fig. 2 (d) and Fig. 2 (e) presents the IA and IF components of mode 10 for RH and RF classes, respectively. As can be seen in these figures, for any MI task class, the two curves are not the same, instead, there is a significant difference between them. The obtained results indicate that these tasks are statistically independent, and thus, the chance of acquiring good classification results are quite high [55].

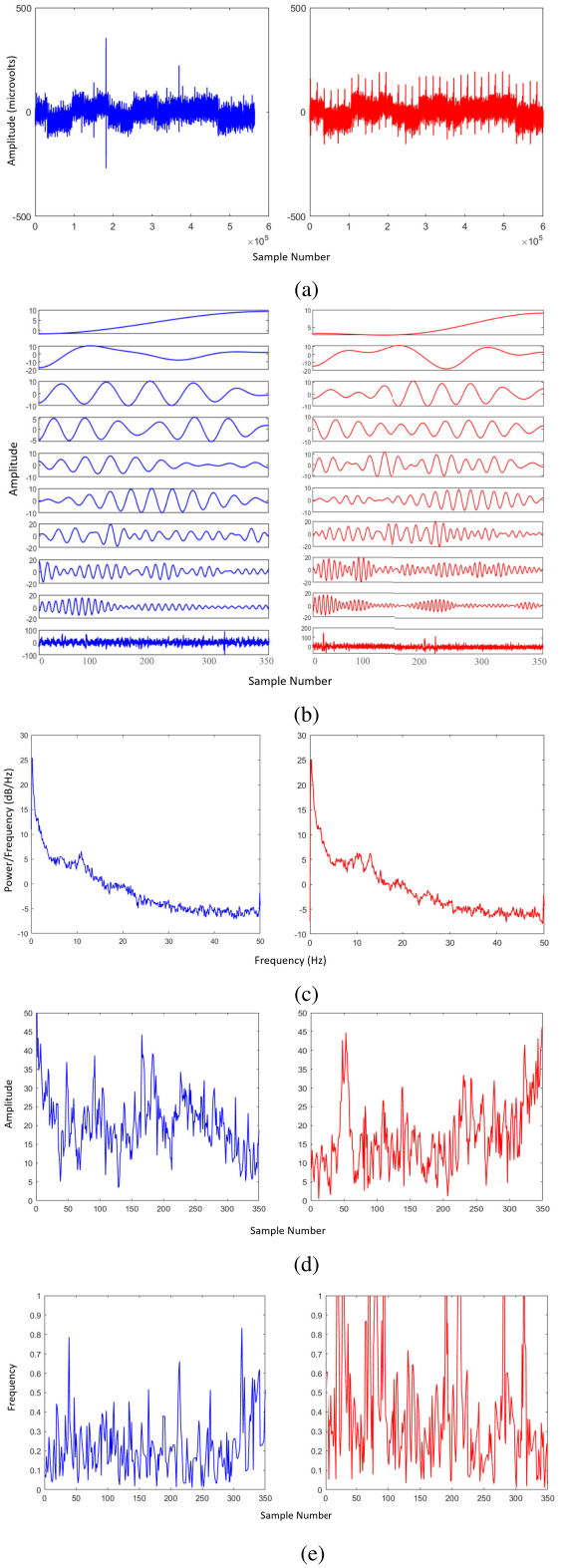

FIGURE 2. Representative pattern of signals for the proposed method, with blue color denoting the RH class while red denoting the RF class. (a) Signals representation for RH and RF class (b) Amplitudes for modes 1 to 10 for RH and RF classes. (c) The mode with maximum PSD for RH and RF classes. (d)-(e): IA and IF components of the mode with the highest PSD for RH and RF classes, respectively.

E. MODULE 5: FEATURE EXTRACTION

Feature extraction plays a vital role in EEG signals classification, since inappropriate features may distress the outcome of the system. Among many others numerous features, the time domain ones are the most appropriate choice for their effectiveness and computational simplicity [56]. In this study, the simple time domain features were extracted from the instantaneous components for characterizing the EEG signals for fast classification.

F. MODULE 6: CLASSIFICATION

The LS-SVM classifier was utilized for extracted features classification for its simplicity [57]. By employing hyperplanes with appropriate margins and solving a combination of linear equations, the LS-SVM classifier is able to distinguish the data of different classes. The decision function as shown in Eq. (9) was utilized for the LS-SVM classier [57],

$$Y(x) = sign\left(\sum_{j=1}^{r} y_j \sigma_i K(x, x_j) + b\right) \tag{9}$$

where y_j is the class label of the j_{th} input feature vector x_j, σ_j denotes the Lagrange multipliers, b is the bias term, r is the dimension of the feature vector and $K(x, x_j)$ is the radial basis function (RBF) kernel [32], which gave the highest accuracy as compared with the other functions tested using the hit and trial method, and it is defined in Eq. (10) as,

$$K(x, x_j) = e^{\frac{-\left(||x-x_j||\right)^2}{2\sigma^2}} \tag{10}$$

with γ and σ^2 being the regularization parameter and bandwidth, respectively. These two parameters could largely determine the classification accuracy of LS-SVM.

G. EXPERIMENTAL SETUP

In this study two experiments were conducted with the proposed EWT algorithm, of which one adopted six commonly used features and the other included four more features, to evaluate its effectiveness by using only 18 channels. The details of these experiments are summarize in following subsections.

1) EXPERIMENT 1 WITH SIX COMMON STATISTICAL FEATURES

In the first experiment, we followed the same way as [55] to extract six typically used statistical features, namely, mean, median, mode, minimum, maximum and standard deviation, from each instantaneous component of the two MI task classes for classification. Explicitly, mean and median are the average and the middle value of the data, while mode is the value that repeats the most within the dataset. In addition, the maximum and minimum values define the data range, while the standard deviation tells how much values spread out from the mean. It is reported that these features could provide a reasonable pattern of the analyzed signals with maximum accuracy [55], [56].

2) EXPERIMENT 2 WITH HOS FEATURES

In the second experiment, four more features, i.e., mean absolute deviation (MAD), interquartile range (IQR), skewness (Sk) and kurtosis (Kt), were included for classification. The MAD is an average of the absolute deviations, measuring the data statistical dispersion, while IQR is used to find the dispersion in a dataset by subtracting the lower and the upper quartile. Both IQR and MAD features have been utilized in [22], [56] for improve MI EEG signal classifications results and gave satisfactory results. The Sk indicates an asymmetric nature of the data distribution around a mean value, and the Kt measures the data flatness with respect to the Gaussian distribution. Both Sk and Kt are also known as HOS features and have been utilized to provide anomalies from linear, stationary and the Gaussian nature of the signals [21], [51]. In this study, we combined these features together to benefit from their effectiveness and efficiency for improving EEG classification accuracy. These particular sets of features in both experiments were extracted from the instantaneous components, and they were used as inputs of the LS-SVM classifier. The mathematical descriptions of these features are listed in Table 2.

IV. RESULTS

In this study, the same training trials as those in [50], [51] are used for experiments as the proposed technique needs class labels at each data point. Results of experiment 1 with six common features are represented by IA1 and IF1, while results of experiment 2 with HOS features are represented by IA2 and IF2. For each participant, all experiments were conducted separately as MI EEG signals are highly participant-dependent for the distinct mental and physical nature of each participant.

A. LS-SVM PARAMETERS SELECTION

The LS-SVM classifier employs the RBF kernel function with its two key parameters γ and σ^2 determining the classification, wherein γ decides the relationship between the training data error reduction and the model simplification issues, and σ^2 provides a mapping between the inputs and the features. As large γ and σ^2 may cause over-fitting problems, a parameter tuning method was adopted to provide an efficient solution. In literature, although various methods could be adopted for parameters tuning, some of those methods may be slow in their convergence, or may converge to non-stationary points, or may even suffer from the speed up issues and result in poor estimations. To overcome those practical issues, the coupled simulating annealing (CSA) method [58] was used for LS-SVM parameters selection, since it is able to provide a better solution together with stable variance control without affecting the convergence speed. The parameters of the RBF kernel were tuned in two steps for each fold of IA1, IF1, IA2, and IF2. Precisely, in the first step, the starting values were determined by the CSA, which uses five multiple starters [58], and the search limit was fixed within [exp(−10), exp(10)]. While in the second step, those starting

TABLE 2. Mathematical description of statistical features.

No.	Feature	Expression	Remarks
1	Mean(μ)	$\mu = \frac{1}{n}\sum_{i=0}^{n} f_i$	n is the number of samples within a dataset
2	Median(M)	$M = (\frac{n+1}{2})^{th} term$	n is the number of samples within a dataset
3	Mode(Mod)	$Mod = L + (\frac{f_s}{f_s+f_p})h$	L is Lower limit of modal class f_s= Frequency of the class succeeding the modal class f_p= Frequency of the class preceding the modal class h is the width of class interval
4	Maximum(Max)	$Max = max(f_i)$	max is the maximum value in the data
5	Minimum(Min)	$Min = min(f_i)$	min is the minimum value in the data
6	Standard Deviation(Std)	$Std = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(f_i - \mu)^2}$	.....
7	Mean Absolute Deviation(MAD)	$MAD = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(\mid f_i - \mu \mid)}$	.....
8	Interquartile Range (IQR)	$IQR = Q_3 - Q_1$	Q_3 is the Third Quartile Q_1 is the First Quartile
9	Skewness (Sk)	$Sk = \sqrt{\frac{1}{n}\sum_{i=0}^{n}\frac{(f_i-\mu)^3}{(Std)^3}}$	.....
10	Kurtosis(Kt)	$Kt = \sqrt{\frac{1}{n}\sum_{i=0}^{n}\frac{(f_i-\mu)^4}{(Std)^4}}$	.....

TABLE 3. Classification outcomes of LS-SVM classifier for experiments 1 and 2 in dataset IVa.

Participant	Experiments	Accuracy for Each fold of 4 folds				Overall Performance
		F1	F2	F3	F4	
"aa"	IA1	100	77.78	88.89	86.11	88.19±9.17
	IF1	100	77.78	100	100	94.44±11.11
	IA2	100	88.89	88.89	100	94.45±6.41
	IF2	100	88.89	100	100	97.22±5.5
"al"	IA1	77.78	88.89	100	100	91.67±10.63
	IF1	77.78	100	66.67	100	86.11±16.66
	IA2	84.78	88.89	97	96	91.67±5.84
	IF2	88.89	88.89	88.89	100	91.67±5.5
"av"	IA1	100	88	100	100	97±6
	IF1	100	100	88.89	100	97.22±5.5
	IA2	97.22	97.22	97.22	97.22	97.22±0
	IF2	100	100	100	100	100±0
"aw"	IA1	100	88.89	77.78	88.89	88.89±9.07
	IF1	100	100	100	100	100±0
	IA2	95	93.45	94	100	95.61±2.99
	IF2	100	100	100	100	100±0
"ay"	IA1	97	96	98	97	97±0.82
	IF1	88.89	66.67	66.67	66.67	72.22±11.11
	IA2	97	96	98	97	97±0.82
	IF2	88.89	86.89	82.89	77.77	84.11±4.91
Mean±Std	IA1					92.55±4.26
	IF1					89.99±11.21
	IA2					**95.19±2.26**
	IF2					**94.60±6.77**

values were provided to the grid search method for better tuning.

B. THE MI CLASSIFICATION RESULTS OF EXPERIMENTS 1 AND 2 WITH LS-SVM

Table 3 presents the classification outcome of each fold for IA1, IF1, IA2, and IF2 in dataset IVa. The overall A_{cc} is listed by the mean ± standard deviation (Std) for all outcomes across the 4 folds. As can be seen, satisfactory classification success rates were achieved for most of the folds. Precisely, experiment IA1 provides an average classification success rate of 88.19%, 91.67%, 97%, 88.89% and 97% across all folds for five participants ''aa'',''al'',''av'',''aw'' and ''ay'' respectively, and the average success rate across all

participants is 92.55% for this case. In contrast, the average classification success rate with IF1 experiment across all folds is 94.44%, 86.11%, 97.22%, 100% and 72.22% for five participants ''aa'',''al'',''av'',''aw'' and ''ay'' respectively, and the average success rate across all participants is 89.99%. These obtained results show that the average classification accuracy with the IA1 experiment is 2.56% higher than that with IF1. Furthermore, it could also be observed that although there are some variations among different folds for IA1 and IF1, the overall success rates are notable without significant discrepancies among those five participants.

In Table 3, it can be observed that with IA2, the EWT algorithm achieves an average A_{cc} of 94.45%, 91.67%, 97.22%, 95.61% and 97% of all different folds for five participants ''aa'',''al'',''av'',''aw'' and ''ay'', respectively, and the average (A_{cc}) with IA2 for all five participants is 95.19%. Whereas IF2 attains an average A_{cc} of 97.22%, 91.67%, 100%, 100%, 84.11% among different folds for five participants ''aa'',''al'',''av'',''aw'' and ''ay'' respectively, and the average A_{cc} with IF2 is 94.60%. Results using IA2 and IF2 show that the accuracy variations both for different folds of a participant and for different participants are smaller or even negligible as compared with those using IA1 or IF1. These results demonstrate that those features utilized in experiment 2, i.e., IA2 and IF2 could help provide higher success rates and accuracy stabilities. It is also worth mentioning that the average classification accuracy using IA2 is 0.59% higher as compared with that using IF2.

C. THE CONTRIBUTION OF INDIVIDUAL FEATURE BY EMPLOYING EWT WITH IA AND IF COMPONENTS

To evaluate the contribution of individual feature in IA and IF for all participants, we define the feature that provides the maximum average classification accuracy with least variations to the most effective one for MI EEG signal classification. When the IA component was utilized, our experimental results presented in Fig. 3(a) show that among all those features, IQR provides the highest average classification outcome of 88.89 ± 8.8, 92.78 ± 9.6, and 93.94 ± 10 for participants ''aa'',''av'' and ''ay'', respectively, while Sk provides the maximum average classification accuracy of 83.35 ± 5.5 and 85.27 ± 8.6 for participants ''al'' and ''aw'', respectively. Whereas if the IF component was utilized, results in Fig. 3(b) show that MAD provides a maximum classification outcome of 88.8 ± 4.6, 87.78 ± 10.8 and 81 ± 9 for participants ''aa'',''av'' and ''ay'' respectively, while IQR provides the highest average classification accuracy of 80.95 ± 7.9 and 95.83 ± 7.2 for participants ''al'' and ''aw'', respectively.

Results in Fig. 3(a) and Fig. 3(b) demonstrate that the features 6 to 10, i.e., MAD, IQR, Sk, and Kt, provide good classification accuracy for EWT using the IA and IF components. Precisely, for all those five participants, at least one feature is able to provide an average classification accuracy over 80% in different cases.

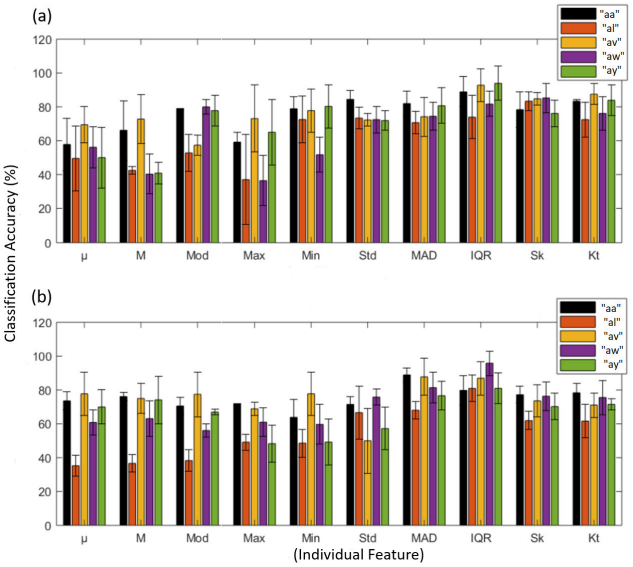

FIGURE 3. Contribution of each feature in MI EEG signal classification. (a) Classification Accuracy of individual feature for the IA approach. (b) Classification Accuracy of individual feature for the IF approach.

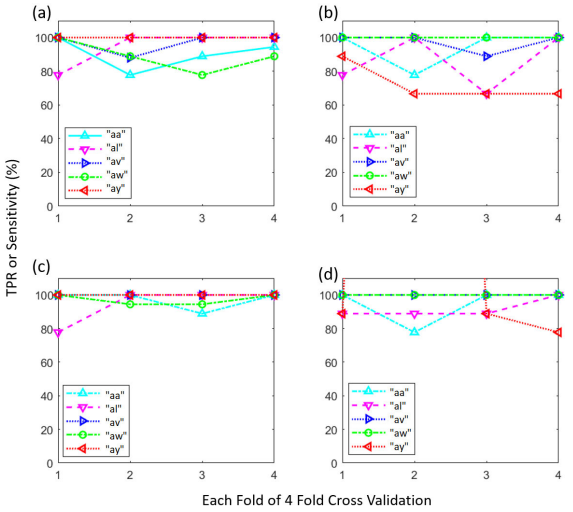

FIGURE 4. Sensitivity for five participants in dataset IVa with EWT using different features. (a) Sensitivity for all participants using IA1. (b) Sensitivity for all participants using IF1. (c) Sensitivity for all participants using IA2. (d) Sensitivity for all participants using IF2.

D. SIGNIFICANCE OF SENSITIVITY AND SPECIFICITY FOR IA AND IF APPROACHES WITH LS-SVM

Fig. 4 demonstrates the sensitivity or TPR for EWT with IA1, IF1, IA2, and IF2 for five participants of dataset IVa. As can be seen, reasonable results with IA1, IF1 as well as admirable results using IA2 and IF2 were achieved. As illustrated in Fig. 4(a), if the IA1 component was used, an average TPR of 94.1% could be accomplished with a std of 4.6% among different participants outcomes, without any significant variations among different folds for participants ''av'' and ''ay''. By comparison, an average TPR of 89.9% with a std of 11.2% is obtained if the IF1 component was used.

The Std value is large among the average TPR values of five participants, but reasonable average results are obtained as shown in Fig. 4(b). In Fig. 4(c) and Fig. 4(d), it could be seen that significant improvements are achieved along with less variation among different average values of TPR for five participants. For example, for IA2 and IF2, a TPR of 97.8% with a Std of 2.3% and a TPR of 95.8% with a std of 5.9% are achieved respectively. The obtained results show that a 3.7% increase on average TPR value with a 2.3% decrease in a Std is achieved with IA2 as compared to IA1. Similarly, for IF2, an average TPR increase of 5.9% with a decrease of 5.3% in Std relative to IF1 is achieved. These findings show that experiment 2 results with fewer variations are more precise, making the proposed EWT method highly stable.

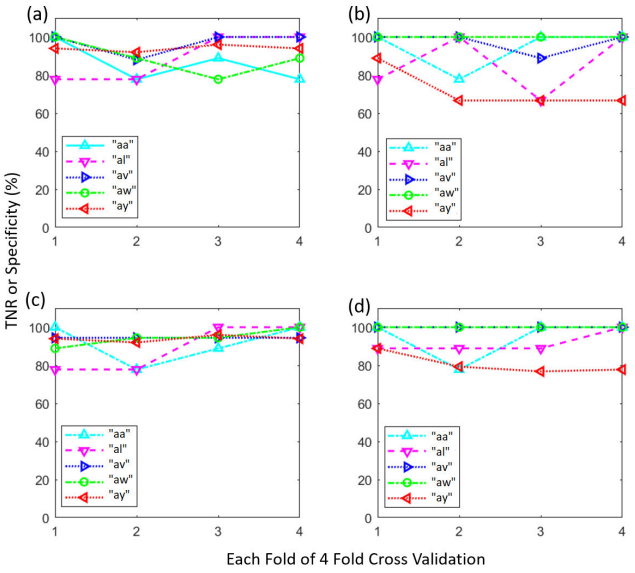

FIGURE 5. Specificity for five participants in dataset IVa with EWT using different features. (a) Specificity for all participants using IA1. (b) Specificity for all participants using IF1. (c) Specificity for all participants using IA2. (d) Specificity for all participants using IF2.

The Specificity or TNR results of IA1, IF1, IA2, and IF2 of each fold for the five participants are shown in Fig. 5. An average TNR value of 90.9% with a Std of 4.4% and 89.9% with a Std 11.2% is obtained for experiments using IA1 and IF1, as shown in Fig. 5(a) and Fig. 5(b) respectively. While an average TNR value of 96.7% with a Std of 2.4% and 93.4 % with a Std of 7.9% is obtained for experiments using IA2 and IF2 as shown in Fig. 5(c) and Fig. 5(d) respectively. While comparing the IA2 strategy to IA1, using IA2 we obtained a 5.8% enhancement for TNR and a 2% decrease in Std value. By comparing IF2 to IF1 for TNR, we have achieved an average TNR value enhancement of 3.5% with a 3.3% decrease in Std value for IF2 strategy. These results conclude that the second experiment using IA2 and IF2 features helps achieve higher classification accuracies with smaller variations than experiment 1.

TABLE 4. Confusion matrix of fold 3 in IA2 approach from participant "av" in dataset IVa.

		Predicted rate		Correct classification(%)
		RH	RF	
Observed values	RH	18	0	100
	RF	1	17	94.44
Average				97.22

E. CONFUSION MATRIX

To demonstrate the prediction and classification results of RH and RF class in any fold, we also computed the confusion matrix, and the Fold 3 of IA2 from participant "av" was randomly chosen to illustrate the outcomes. Results in Table 4 show that no value of the RH class was misclassified to the RF class, while only 1 value of the RF class was misclassified as the RH class. The correct classification rate for RH class is 100% while for RF class is 94.44%, which thus gives a satisfactory average classification accuracy of 97.22% in Fold 3.

F. RECEIVER OPERATING CHARACTERISTICS CURVE FOR EXPERIMENTS 1 AND 2

Both receiver operating characteristics (ROC) curve and Kappa coefficient are typical performance matrics being utilized for evaluating the classification capability of a classifier [59]. Specifically, among these two metrics, the Kappa coefficient is a single-dimensional scalar and usually utilized for multiple class signals, i.e., more than two dimensions. While the ROC curves are basically two-dimensional, which could be able to convey more information as compared to the Kappa coefficient. In this study, since the dataset IVa is a two-class problem, we choose to adopt the ROC curves as a performance metric, as those did in [60], [61], to determine the performance of the proposed method in different cases. To assess the classification capability of an LS-SVM classifier for RH and RF classes, we calculated the area under the ROC curve (AUC). The AUC ranges from 0 to 1, representing a classifier's classification capability, and the higher the AUC is, the better the ability of a classifier to identify two distinct tasks. Fig. 6 presents the ROC curve for experiments 1 and 2 with dataset IVa. As can be seen, the AUC values are more than 0.9 for most participants in IA1, IF1, IA2 and IF2. Specifically for IA2 and IF2 the AUC values for five participants were equal or over 0.95, illustrating the better classification capabilities of LS-SVM classifier for these cases.

Comparing the outcomes in Table 3 and Fig's. 3-6, we could also observe that for all metrics, the average classification accuracy using IA2 and IF2 was better than those using IA1 and IF1. This is because the four additional features of IQR, MAD, and HOS, i.e., Sk and Kt have been included in the signal classification in experiment 2. These four features assist to obtain higher success rates and provide higher

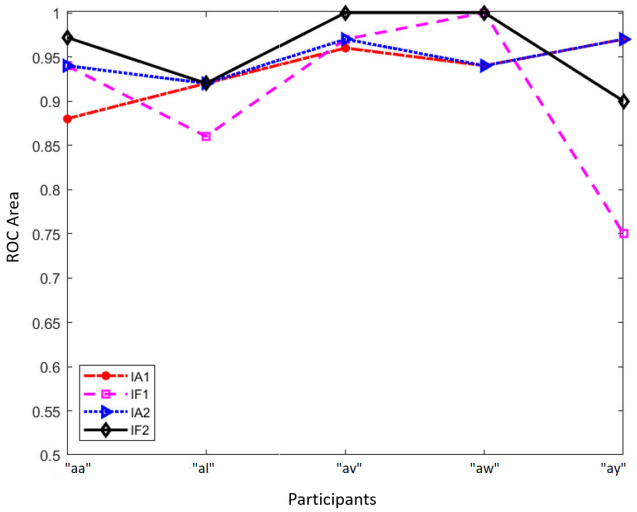

FIGURE 6. The ROC Curve for experiments 1 and 2 with dataset IVa.

stabilities in different cases. It should be noted that both Sk and Kt have already been used for various BCI schemes to achieve good outcomes [62].

G. COMPARISON OF LS-SVM WITH SIX OTHER CLASSIFIERS FOR EXPERIMENTS 1 AND 2

To check if the LS-SVM classifier is a suitable option for the proposed EWT techniques, both experiments 1 and 2 were conducted with six other well-known classifiers, i.e., Logistic regression (LR) [63], Multilayer perceptron neural networks (MLP) [64], Simple logistic (SL) [65], K-star [66], Logistic Model Tree [67] and Random Forest (RF) [68], in different cases. These classifiers can also be found in the software Waikato environment for knowledge analysis (WEKA) [69]. Results for these different classifiers with experiments 1 and 2 are shown in Tables 5 and 6, respectively.

From those two tables, it could be observed that in all IA1, IF1, IA2, and IF2 experiments for dataset IVa, the LS-SVM classifier achieves the best results in terms of average classification accuracy. In particular, the LS-SVM achieves an improvement of 11.4%-15.9%, 8.3%-13.9%, 4.1%-11.9%, and 6.4%-15.7% with IAI, IF1, IA2, and IF2, respectively, as compared to these six other classifiers. Such findings convincingly show that LS-SVM helps achieve the highest classification performance with these extracted features, and therefore, for the proposed EWT based technique, the LS-SVM classifier is an appropriate option for enhancing the classification accuracy of MI signals.

H. TIMING EXECUTION FOR IA1, IF1, IA2 AND IF2

To further evaluate the efficiency of the proposed EWT method, its execution time was also calculated in different cases. This includes the total time required for EEG signal decomposition, appropriate signal mode selection, instantaneous components exploration, feature extraction, K fold training, and testing.

Fig. 7 shows the total execution time required for the five participants in dataset IVa. As can be seen, the total execution time required for most of the cases is less than 12 sec with a common PC and the Matlab software being used. In practice, if more computationally effective software or more powerful PC is utilized, this timing implementation could be further decreased. It is also worth mentioning that those execution

TABLE 5. Comparison between LS-SVM and six other classifiers for experiment 1 with dataset IVa.

	Comparison of LS-SVM with six other classifiers in term of classification accuracy.													
	LS-SVM		LR		MLP		SL		K-Star		LMT		RF	
Participants	IA1	IF1	IA1	IF1	IA1	IF1	IA1	IF1	IA1	IF1	IA1	IF1	IA1	IF1
"aa"	88.1	94.4	72.2	88.8	75	86.1	77.78	80.6	69.4	77.78	75	83.3	77.8	83.3
"al"	91.6	86.4	66.6	55.6	69.4	75	66.7	55.6	77.8	75	66.7	63.8	77.8	61.1
"av"	97	97.2	88.8	94.2	91.7	77.8	91.7	83.3	86.1	66.7	91.7	83.3	94.2	75
"aw"	88.8	100	80.5	100	83.3	100	88.9	94.4	80.6	97.2	86.1	94.4	75	88.9
"ay"	97	72.2	80.5	69.4	75	66.7	77.8	66.7	69.4	80.6	77.8	63.8	80.6	75
Mean	**92.5**	**89.9**	77.8	81.6	78.8	81.1	80.6	76	76.6	79.5	79.5	77.7	81.1	76.7
±Std	± **4.2**	±**11.2**	±8.6	±18.5	± 7.8	± 12.6	± 10	±15	±7.3	±11.2	±9.7	±13.5	± 7.6	± 10.5

TABLE 6. Comparison between LS-SVM and six other classifiers for experiment 2 with dataset IVa.

	Comparison of LS-SVM with six other classifiers in term of classification accuracy.													
	LS-SVM		LR		MLP		SL		K-Star		LMT		RF	
Participants	IA2	IF2	IA2	IF2	IA2	IF2	IA2	IF2	IA2	IF2	IA2	IF2	IA2	IF2
"aa"	94.5	97.2	80.6	94.4	75	97.2	77.78	91.7	69.4	83.3	77.8	91.7	80.6	86.1
"al"	91.7	91.7	88.9	58.3	72.2	86.1	83.3	58.3	83.3	75	88.9	69.4	80.6	66.7
"av"	97.2	100	100	97.2	97.2	86.1	97.2	88.9	88.9	75	97.2	88.9	94.4	75
"aw"	95.6	100	88.9	100	83.3	100	91.7	94.4	86.1	100	88.9	94.4	75	88.9
"ay"	97	84.1	97.2	72.2	88.9	72.2	88.9	69.4	94.4	86.1	88.9	69.4	94.4	77.8
Mean	**95.2**	**94.6**	91.1	84.4	83.3	88.3	87.8	80.5	84.4	83.8	88.3	82.8	85	78.9
±Std	±**2.2**	±**6.7**	±7.7	±18.3	±10	±11	±7.5	±15.8	±9.3	±10.3	±6.9	±12.4	±8.8	±8.9

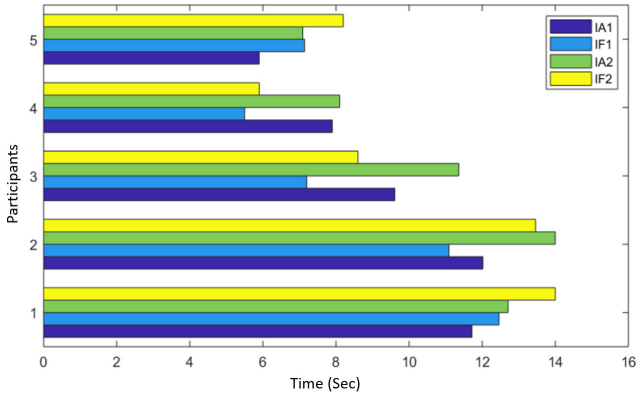

FIGURE 7. The total time of the proposed EWT method with LS-SVM classifier for experiments 1 and 2 using dataset IVa.

time were measured with all trials in dataset IVa taken into account, and thus, for any single trial, the algorithm execution time would be much shorter, illustrating that the proposed algorithm could be utilized for online applications.

Furthermore, since only 18 channels out of 118 channels were utilized in the experiments, both the hardware cost and the system complexity are lower as compared with those of the existing techniques, using all 118 channels. The reduced cost, complexity and execution time convincingly predict the great potential of the proposed EWT algorithm for online clinical applications in the future. It is worth mentioning that, here we presented only the complete exection time of our proposed algorithm. This is because we did not find the execution time of the other algorithms for reasonable comparisons in literature [21], [23], [24], [26]–[28], [32]–[34].

V. DISCUSSIONS

The main objective of this study is to build a simple and computationally efficient algorithm for improving classification accuracy among different MI tasks. For this purpose, the following aims are fulfilled in this study.

1) To avoid using the large amount of data for signal decomposition and feature extraction, we selected only those data samples that contains MI information related to hand/foot movements. In this study, we considered t = 3.5 sec to obtain 350 samples from each EEG segments.
2) To reduce computational load, we selected different channels based on the physiological knowledge to decode appropriate channels related to MI tasks, and thus, only a 18 channels combination around motor cortex region of brain was selected in our study to provide satisfactory classification outcomes among different MI tasks.
3) The EWT algorithm was implemented to analyze the complex, nonlinear and non-stationary behavior of EEG signals, and 10 modes were chosen empirically for each channel to discriminate two MI tasks. However, it does not mean that other modes for each channel are physiologically not significant. To further reduce the system complexity we selected a single mode for each channel based on the power spectral analysis of all modes.
4) Afterwards, the Hilbert transform (HT) was applied on each mode to obtain meaningful statistical features from instantaneous components. Experimental results reveal that a combination of EWT and higher order statistics (HOS) features helps achieve higher success rates and provide higher stabilities in results for the detection of each MI task. The contribution of each feature in classification improvement was also evaluated. Results indicate that there exists at-least one feature that provides more than 80% classification accuracy, which is higher than the desired lowest accuracy of 70% for a BCI system to be controlled [70]. This information can also help to build a participant specific BCI system.
5) All single, six and ten features for IA and IF components obtained from each participant were individually tested on seven classifiers, and the results suggested that LS-SVM classifier with RBF kernenl and parameter tuning method provide the best classification results.
6) Finally, to verify the effectiveness of the proposed EWT method, we also compared the performance of the proposed EWT based method with those of the other eleven best existing algorithms for processing the dataset IVa [21], [23], [24], [26]–[28], [32]–[34]. Table 7 presents such comparisons, wherein the highest classification accuracies for both individual participants and the overall outcomes are represented in bold font. The standard deviation accuracies of various respondents are also listed in the last column.

   As can be seen in Table 7, the ISSPL technique proposed by Wu *et al.* in [33] achieved a classification accuracy of 100% for participant ''al'', yet the overall classification accuracy of our proposed strategy using IA2 is better with the least variation. When compared with the WPD method proposed by Kevric and Subasi [21], our proposed strategy achieves the highest classification accuracy for four participants by using IA2 and also improves the overall classification accuracy. When comparing our EWT based method with the CSP based method proposed in [34], we could find that our proposed EWT methods, i.e., IA2 features with LS-SVM and IA2 features with SVM provide an accuracy improvement of 7.21% and 2.58% as compared with the CSP based dynamically filtered features with SVM, even if the dynamically filtered features may give better classification performances for a dataset IVa as compared with the statistical IA2 features.

   Further investigate Table 7, we could find that higher classification accuracies were achieved for all participants when using the IA2 and IF2 features. Specifically, it could be observed that the IA1 approach together with EWT enables participants ''av'' and ''ay''

TABLE 7. Comparison with other works.

Papers by	Methods employed	Classification accuracy(%)						
		"aa"	"al"	"av"	"aw"	"ay"	Av.	Std
The proposed	EWT+IA1+LS-SVM	88.19	91.67	97	88.89	97	92.55	4.26
	EWT+IF1+LS-SVM	94.44	86.11	**97.22**	**100**	72.22	89.99	11.21
	EWT+IA2+LS-SVM	94.45	91.67	**97.22**	95.61	97	**95.19**	**2.26**
	EWT+IA2+SVM	80.56	88.89	**100**	88.89	94.44	90.56	6.47
	EWT+IF2+LS-SVM	**97.22**	91.67	**100**	**100**	84.11	94.60	6.77
Wu *et al.* [33]	ISSPL	93.57	**100**	79.29	99.64	**98.57**	94.21	8.73
Kevric *et al.* [21]	MSPCA+WPD+HOS+KNN	96	92.3	88.9	95.4	91.4	92.8	2.93
Siuly *et al.* [32]	Clustering+LS-SVM	92.63	84.99	90.77	86.50	86.73	88.32	3.22
Song *et al.* [34]	CSP+Dynamic Features+SVM	87.42	97.35	69.70	96.84	88.57	87.98	11.2
Lu *et al.* [28]	R-CSP+Aggregation	76.8	98.2	74.5	92.2	77	83.74	10.72
Zhang *et al.* [27]	Z-LDA	77.70	**100**	68.4	99.60	59.9	81.1	18.12
Lotte *et al.* [26]	SSRCSP	70.54	96.43	53.57	71.88	75.39	73.56	15.31
	TRCSP	71.43	96.43	63.27	71.88	86.9	77.98	13.38
	WTRCSP	69.64	98.21	54.59	71.88	85.32	75.93	16.55
	WTRCSP	72.32	96.43	60.2	77.68	86.51	78.63	13.78
Yong *et al.* [24]	SSFO	57.5	86.9	54.4	84.4	84.3	73.50	16

to achieve satisfactory classification accuracies, while if IF1 was used, a classification accuracy of 100% could be achieved for participant ''aw'', and an accuracy of 94.44% and 97.22% could also be achieved for participants ''aa'' and ''av'', respectively. Such results indicate that as compared to the existing techniques, EWT with IA2 provides the higher average accuracy of 95.19% with the smallest std of 2.26%, while IF2 ranks the second with an overall classification accuracy of 94.60%. These results convincingly show that the proposed EWT method helps achieve satisfactory results by using the IA and IF components, especially with the HOS features, and could offer an improvement up to 21.69% and 21.10% in classification accuracy as compared with the other existing methods for dataset IVa. Such an accuracy improvement could help end-users to express their MI tasks more clearly. For example, the disabled people could control their wheelchair more accurately, while rehabilitated ones could enhance their rehabilitation therapies with appropriate feedback delivered once they execute the correct movement.

The main benefit of EWT based method proposed in this study is its simplicity owing to the less number of channels being, the appropriate selection of the delicate MI-related mode for each channel, as well as the choice of simple statistical features along with the LS-SVM classifier. Tables 3-7 and Figs. 1-7 show the efficiency and stability of the proposed EWT algorithm, and indicate that it could be implemented for online applications for its efficiency. Such a technique can also be used for big data applications for its low computational expenses.

However, there also exist certain limitations with the proposed EWT method. The first one is that 18 channels were selected manually based on physiological knowledge of the participants according to the standard 10/20 systems, and such a process may effect on the overall accuracy of the participants. In practice, MI EEG signals are highly participant dependent due to the inherent mental and physical nature of each participant, and thus, the channels should be selected by automated participant specific criteria, which in turn would further reduce both the channel number and system complexity, as well as increase the overall accuracy. To address such a limitation, new automated participant dependent channel selection methods are highly desired to make the EWT method more adaptive and efficient for practical applications. The second limitation is that the EEG signals are typically contaminated with participants and equipment noises, which may degrade the classification performance of MI signals, and thus, effective noise removal algorithms could be devised to make current BCI system robust against noise. Last but not least, in our current study, 10 modes for each channel signal were selected in a hit and trial manner for EWT method. Since one has to try different mode combinations manually to choose the one with relevant MI information, this hit and trial scheme is time-consuming, hindering the online applications of the proposed method in practice.

To address these limitations, our next step work is to devise automated channel and mode selection mechanisms such that those signal classification methods could be efficient and adaptive enough for practical applications. Testing those implemented methods with some other participants online, instead of the publically available datasets, is also our next-step work.

VI. CONCLUSION

In summary, we propose a novel method based on EWT for improving the MI task EEG signal classification accuracy in this study. By employing EWT-based signal decomposition technique for feature extraction and LS-SVM classifier

for different MI task discriminations, EWT based algorithms using either IA or IF components are proposed to improve success rate of EEG signal classification. Specifically, to reduce the computational load, only 18 motor cortex area channels among all 118 were selected for experiments. While to further reduce the feature dimension, only one mode is selected from 10 modes for each channel. The selected mode is susceptible to MI data and helps achieve the highest classification accuracy among various MI classes. Experiments with the publically available dataset IVa from BCI competition III were performed to evaluate the performance of algorithm. Results showed that the proposed EWT technique achieves higher classification accuracy as compared with those existing techniques using 118 electrode channels, and an average classification accuracy of up to 95.19% and 94.60% for dataset IVa was achieved using the IA and IF components, respectively. The HOS features could further improve the signal classification accuracy for the EWT algorithm. The overall results convincingly demonstrate that the proposed EWT-based technique is promising for MI task signal classification.

ACKNOWLEDGMENT

(Muhammad Tariq Sadiq and Xiaojun Yu are co-first authors.)

REFERENCES

[1] C. G. Coogan and B. He, "Brain-computer interface control in a virtual reality environment and applications for the Internet of Things," *IEEE Access*, vol. 6, pp. 10840–10849, 2018.

[2] N. Birbaumer, A. R. Murguialday, and L. Cohen, "Brain-computer interface in paralysis," *Current Opinion Neurol.*, vol. 21, no. 6, pp. 634–638, Dec. 2008.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[4] V. Kaiser, I. Daly, F. Pichiorri, D. Mattia, G. R. Müller-Putz, and C. Neuper, "Relationship between electrical brain responses to motor imagery and motor impairment in stroke," *Stroke*, vol. 43, no. 10, pp. 2735–2740, Oct. 2012.

[5] G. Purtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[6] H. Alkadhi, P. Brugger, S. H. Boendermaker, G. Crelier, A. Curt, M.-C. Hepp-Reymond, and S. S. Kollias, "What disconnection tells about motor imagery: Evidence from paraplegic patients," *Cerebral Cortex*, vol. 15, no. 2, pp. 131–140, Feb. 2004.

[7] G. Pfurtscheller, C. Neuper, G. R. Müller, B. Obermaier, G. Krausz, A. Schlogl, R. Scherer, B. Graimann, C. Keinrath, D. Skliris, M. Wortz, G. Supp, and C. Schrank, "Graz-BCI: State of the art and clinical applications," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 1–4, Jun. 2003.

[8] J. Kronegg, G. Chanel, S. Voloshynovskiy, and T. Pun, "EEG-based synchronized brain-computer interfaces: A model for optimizing the number of mental tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 50–58, Mar. 2007.

[9] N. Kosmyna and A. Lécuyer, "A conceptual space for EEG-based brain-computer interfaces," *PLoS ONE*, vol. 14, no. 1, Jan. 2019, Art. no. e0210145.

[10] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler, "An MEG-based brain-computer interface (BCI)," *Neuroimage*, vol. 36, no. 3, pp. 581–593, Jul. 2007.

[11] R. Sitaram, A. Caria, R. Veit, G. Gaber T., G. Rota, A. Kuebler, and N. Birbaumer, "FMRI brain-computer interface: A tool for neuroscientific research and treatment," *Comput. Intell. Neurosci.*, vol. 2007, Apr. 2007, Art. no. 25487.

[12] Y. Zhu, K. Xu, C. Xu, J. Zhang, J. Ji, X. Zheng, H. Zhang, and M. Tian, "PET mapping for brain-computer-interface-based stimulation in a rat model with intracranial electrode implantation in the ventro-posterior medial thalamus," *J. Nucl. Med.*, vol. 25, p. 115, Feb. 2016.

[13] H. Fukuyama, Y. Ouchi, S. Matsuzaki, Y. Nagahama, H. Yamauchi, M. Ogawa, J. Kimura, and H. Shibasaki, "Brain functional activity during gait in normal subjects: A SPECT study," *Neurosci. Lett.*, vol. 228, no. 3, pp. 183–186, Jun. 1997.

[14] G. Rodríguez-Bermúdez and P. J. García-Laencina, "Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces," *J. Med. Syst.*, vol. 36, no. 1, pp. 51–63, Nov. 2012.

[15] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Appl. Math. Comput.*, vol. 187, no. 2, pp. 1017–1026, Apr. 2007.

[16] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–325, Sep. 1998.

[17] A. Schlögl, C. Neuper, and G. Pfurtscheller, "Estimating the mutual information of an eeg-based brain-computer interface," *Biomed. Eng.*, vol. 47, nos. 1–2, pp. 3–8, Jan./Feb. 2002.

[18] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 1, pp. 12–17, Mar. 2005.

[19] B. H. Jansen, J. R. Bourne, and J. W. Ward, "Autoregressive estimation of short segment spectra for computerized EEG analysis," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 9, pp. 630–638, Sep. 1981.

[20] D. J. Krusienski, D. J. McFarland, and J. R. Wolpaw, "An evaluation of autoregressive spectral estimation model order for brain-computer interface applications," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2006, pp. 1323–1326.

[21] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, Jan. 2017.

[22] K. S. Biju, H. A. Hakkim, and M. G. Jibukumar, "Ictal EEG classification based on amplitude and frequency contours of IMFs," *Biocybern. Biomed. Eng.*, vol. 37, no. 1, pp. 172–183, 2017.

[23] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[24] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar./Apr. 2008, pp. 417–420.

[25] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 6599–6602.

[26] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[27] R. Zhang, P. Xu, L. Guo, Y. Zhang, P. Li, and D. Yao, "Z-score linear discriminant analysis for EEG based brain-computer interfaces," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e74433.

[28] H. Lu, H.-W. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[29] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 631–641, Mar. 2018.

[30] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2018.

[31] J. K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, and A. Cichocki, "An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system," *Comput. Intell. Neurosci.*, vol. 2019, May 2019, Art. no. 8068357.

[32] S. Siuly, Y. Li, and P. Wen, "Clustering technique-based least square support vector machine for EEG signal classification," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. 358–372, 2011.

[33] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.

[34] L. Song and J. Epps, "Classifying EEG for brain-computer interface: Learning optimal filters for dynamical system features," *Comput. Intell. Neurosci.*, vol. 2007, Jan. 2007, Art. no. 57180.

[35] S. Wang and C. J. James, "Extracting rhythmic brain activity for brain-computer interfacing through constrained independent component analysis," *Comput. Intell. Neurosci.*, vol. 2007, Jun. 2007, Art. no. 41468.

[36] L. Qin, L. Ding, and B. He, "Motor imagery classification by means of source analysis for brain-computer interface applications," *J. Neural Eng.*, vol. 1, no. 3, pp. 135–141, Sep. 2004.

[37] Y. Kutlu and D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," *Comput. Methods Programs Biomed.*, vol. 105, no. 3, pp. 257–267, 2012.

[38] 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 Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[39] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[40] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 234–239.

[41] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, pp. 1–9, Oct. 2018. doi: 10.1007/s00521-018-3735-3.

[42] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *Int. J. Neural Syst.*, vol. 27, no. 2, Mar. 2017, Art. no. 1650032.

[43] X. Zhang, L. Yao, X. Wang, J. Monaghan, and D. Mcalpine, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," 2019, *arXiv:1905.04149*. [Online]. Available: https://arxiv.org/abs/1905.04149

[44] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[45] P. Maya, S. V. Shree, K. Roopasree, and K. P. Soman, "Discrimination of internal fault current and inrush current in a power transformer using empirical wavelet transform," *Procedia Technol.*, vol. 21, pp. 514–519, Jan. 2015.

[46] W. Liu, S. Cao, and Y. Chen, "Seismic time–frequency analysis via empirical wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 28–32, Jan. 2016.

[47] S. Moushmi, V. Sowmya, and K. P. Soman, "Multispectral and panchromatic image fusion using empirical wavelet transform," *Indian J. Sci. Technol.*, vol. 8, no. 24, Sep. 2015, Art. no. IPL0346.

[48] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, R. M. Jdel, M. Schröder, and N. Birbaumer, "The BCI competition. III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.

[49] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *Neuroimage*, vol. 34, no. 4, pp. 1600–1611, 2007.

[50] Y. Li and P. Wen, "Identification of motor imagery tasks through CC-LR algorithm in brain computer interface," *Int. J. Bioinf. Res. Appl.*, vol. 9, no. 2, pp. 156–172, 2013.

[51] S. Siuly, Y. Li, and P. Wen, "Modified CC-LR algorithm with three diverse feature sets for motor imagery tasks classification in EEG based brain–computer interface," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 767–780, 2014.

[52] I. Daubechies and B. J. Bates, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1993.

[53] S. Dutta, M. Singh, and A. Kumar, "Automated classification of non-motor mental task in electroencephalogram based brain-computer interface using multivariate autoregressive model in the intrinsic mode function domain," *Biomed. Signal Process. Control*, vol. 43, pp. 174–182, May 2018.

[54] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[55] S. Siuly and Y. Li, "Improving the separability of motor imagery EEG signals using a cross correlation-based least square support vector machine for brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, Jul. 2012.

[56] H. R. Al Ghayab, Y. Li, S. Siuly, and S. Abdulla, "A feature extraction technique based on tunable Q-factor wavelet transform for brain signal classification," *J. Neurosci. Methods*, vol. 312, pp. 43–52, Jan. 2019.

[57] S. Siuly and Y. Li, "A novel statistical algorithm for multiclass EEG signal classification," *Eng. Appl. Artif. Intell.*, vol. 34, pp. 154–167, Sep. 2014.

[58] S. Xavier-de Souza, J. A. Suykens, J. Vandewalle, and D. Bollé, "Coupled simulated annealing," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 40, no. 2, pp. 320–335, Apr. 2009.

[59] A. Ben-David, "About the relationship between ROC curves and Cohen's kappa," *Eng. Appl. Artif. Intell.*, vol. 21, no. 6, pp. 874–882, Sep. 2008.

[60] M. A. Joadder, S. Siuly, E. Kabir, H. Wang, and Y. Zhang, "A new design of mental state classification for subject independent BCI systems," *IRBM*, to be published.

[61] N. F. Ince, F. Goksu, A. H. Tewfik, and S. Arica, "Adapting subject specific motor imagery EEG patterns in space–time–frequency for a brain computer interface," *Biomed. Signal Process. Control*, vol. 4, no. 3, pp. 236–246, 2009.

[62] N. K. Qureshi, N. Naseer, F. M. Noori, H. Nazeer, R. A. Khan, and S. Saleem, "Enhancing classification performance of functional near-infrared spectroscopy- brain-computer interface using adaptive estimation of general linear model coefficients," *Frontiers Neurorobot.*, vol. 11, p. 33, Jul. 2017.

[63] A. Subasi and E. Erçelebi, "Classification of EEG signals using neural network and logistic regression," *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, May 2005.

[64] S. N. Oğulata, C. Şahin, and R. Erol, "Neural network-based computer-aided diagnosis in classification of primary generalized epilepsy by EEG signals," *J. Med. Syst.*, vol. 33, no. 2, pp. 107–112, Apr. 2009.

[65] J. H. McDonald, *Handbook of Biological Statistics*. Baltimore, MD, USA: Sparky House, 2009, vol. 2.

[66] J. G. Cleary and L. E. Trigg, "K: An instance-based learner using an entropic distance measure," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 108–114.

[67] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Mach. Learn.*, vol. 59, no. 1, pp. 161–205, May 2005.

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

[69] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[70] N. Birbaumer, "Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.

MUHAMMAD TARIQ SADIQ received the B.Sc. degree (Hons.) in electrical engineering from the COMSATS Institute of Information Technology, Lahore, Pakistan, in 2009, and the M.Sc. degree in electrical engineering from the Blekinge Institute of Technology, Sweden, in 2011. He was an Assistant Professor with the Sharif College of Engineering and Technology (SCET), which is affiliated with the University of Engineering and Technology, Lahore, and a Lecturer with the University of South Asia. He was also a Project Manager with SCET managed final year student's projects, a Patron of IEEE-SCET Student Branch, and a Lifetime Member of the Pakistan Engineering Council, Islamabad, Pakistan. He is currently a Ph.D. Scholar with Northwestern Polytechnical University, Xi'an, China, and an Assistant Professor with the Electrical Engineering Department, The University of Lahore. His current research interests include biomedical signal analysis and classification.

XIAOJUN YU received the Ph.D. degree from Nanyang Technological University, Singapore, in 2015. From 2015 to 2017, he was a Postdoctoral Research Fellow with Nanyang Technological University. He is currently an Associate Professor with Northwestern Polytechnical University, China. His current research interest includes high-resolution optical coherence tomography and its imaging applications.

ZHAOHUI YUAN received the B.S., M.S., and Ph.D. degrees from Northwestern Polytechnical University, China, in 1984, 1987, and 2005, respectively, all in control engineering. He has been with Northwestern Polytechnical University, as a Teaching Assistant, Lecturer, Associate Professor, and a Professor, since 1987. He has coauthored more than 50 articles in technical journals and conferences. His current research interests include high-precision detection instrumentation and system control engineering, hydraulic system control and test, and flow field analysis of the hydraulic systems. He is a Fellow of the Shaanxi Association for Science and Technology.

ZEMING FAN received the Ph.D. degree in mechanical manufacturing and automation from the School of Mechanical and Electrical Engineering, Xi'an University of Technology. From 2002 to 2005, he was a Postdoctoral Research Fellow with Yongji Motor Factory. He is currently a Professor with Northwestern Polytechnical University, China. His current research interests include detection and control technology research of electro-hydraulic servo systems, and detection of the aircraft hydraulic systems and control systems.

ATEEQ UR REHMAN was born in Rawalpindi, Pakistan, in 1987. He received the bachelor's degree in electrical engineering (telecommunication) from the COMSATS Institute of Information Technology, Lahore, Pakistan, in 2009, and the M.S. degree in electrical engineering (telecommunications) from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2011. He is currently pursuing the Ph.D. degree with the College of Internet of Things Engineering, Hohai University, Changzhou Campus, China. Since 2016, he has been a Lecturer with Government College University, Lahore. His current research interests include biomedical signal processing, the Internet of Things, the social Internet of Things, and big data.

GUOQI LI received the B.E. degree from the Xi'an University of Technology, Xi'an, China, in 2004, the M.E. degree from Xi'an Jiaotong University, Xi'an, in 2007, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2011. He was a Scientist with the Data Storage Institute and the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore, from 2011 to 2014. He is currently an Associate Professor with the Department of Precision Instrument, Tsinghua University, Beijing, China. He has published over 70 journal and conference articles. His current research interests include brain-inspired computing, complex systems, neuromorphic computing, machine learning, and system identification.

GAOXI XIAO received the B.S. and M.S. degrees in applied mathematics from Xidian University, Xi'an, China, in 1991 and 1994, respectively, and the Ph.D. degree in computing from The Hong Kong Polytechnic University, in 1998. He was an Assistant Lecturer with Xidian University, from 1994 to 1995. He was a Postdoctoral Research Fellow with Polytechnic University, Brooklyn, NY, USA, in 1999, and a Visiting Scientist with The University of Texas at Dallas, from 1999 to 2001. He joined the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2001, where he is currently an Associate Professor. His research interests include complex systems and complex networks, communication networks, smart grids, and system resilience and risk management. He served/serves as a TPC Member for numerous conferences, including the IEEE ICC and the IEEE GLOBECOM, and as an Associate Editor or a Guest Editor for the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *PLOS ONE*, and *Advances in Complex Systems*.