IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 4, 2019, accepted November 21, 2019, date of publication November 26, 2019,
date of current version December 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956018

Motor Imagery EEG Signals Decoding by Multivariate Empirical Wavelet Transform-Based Framework for Robust Brain–Computer Interfaces

MUHAMMAD TARIQ SADIQ[1,5], XIAOJUN YU[1], ZHAOHUI YUAN[1], FAN ZEMING[1], ATEEQ UR REHMAN[2], INAM ULLAH[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
[5]Department of Electrical Engineering, The University of Lahore, Lahore 54000, Pakistan

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 in Northwestern Polytechnical University under Grant ZZ2019028.

ABSTRACT The robustness and computational load are the key challenges in motor imagery (MI) based on electroencephalography (EEG) signals to decode for the development of practical brain-computer interface (BCI) systems. In this study, we propose a robust and simple automated multivariate empirical wavelet transform (MEWT) algorithm for the decoding of different MI tasks. The main contributions of this study are four-fold. First, the multiscale principal component analysis method is utilized in the preprocessing module to obtain robustness against noise. Second, a novel automated channel selection strategy is proposed and then is further verified with comprehensive comparisons among three different strategies for decoding channel combination selection. Third, a sub-band alignment method by utilizing MEWT is adopted to obtain joint instantaneous amplitude and frequency components for the first time in MI applications. Four, a robust correlation-based feature selection strategy is applied to largely reduce the system complexity and computational load. Extensive experiments for subject-specific and subject independent cases are conducted with the three-benchmark datasets from BCI competition III to evaluate the performances of the proposed method by employing typical machine-learning classifiers. For subject-specific case, experimental results show that an average sensitivity, specificity and classification accuracy of 98% was achieved by employing multilayer perceptron neural networks, logistic model tree and least-square support vector machine (LS-SVM) classifiers, respectively for three datasets, resulting in an improvement of upto 23.50% in classification accuracy as compared with other existing method. While an average sensitivity, specificity and classification accuracy of 93%, 92.1% and 91.4% was achieved for subject independent case by employing LS-SVM classifier for all datasets with an increase of up to 18.14% relative to other existing methods. Results also show that our proposed algorithm provides a classification accuracy of 100% for subjects with small training size in subject-specific case, and for subject independent case by employing a single source subject. Such satisfactory results demonstrate the great potential of the proposed MEWT algorithm for practical MI EEG signals classification.

INDEX TERMS Electroencephalography, multiscale principal component analysis, brain-computer interface, multivariate empirical wavelet transform.

I. INTRODUCTION

Brain-Computer Interface (BCI) is a system being developed to connect the brain and a computer by using individual brain signals [1]. BCI has attracted extensive research interests over the past few years, and also largely improved the living standards for the disabled by providing a communication bridge between the human beings and the external devices, e.g. wheelchair, robotic arm etc, [1], [2]. Motor Imagery (MI)

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

is the most commonly used practical opt for BCI systems for its robust nature [2], and various methods, e.g., positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), electrocorticography (ECoG) and electroencephalography (EEG) [2], [3], have been developed to monitor the MI signals. Among all those techniques, EEG based MI BCI systems are the utmost used owing to their non-invasiveness, and capability of providing excellent temporal information of MI signals at a low cost [2], [4]. A big challenge for any real-time BCI system, however, is to correctly decode different MI EEG signals automatically [5].

An unbiased automated EEG classification system typically consists of three components, i.e., preprocessing, feature extraction and signal classification. Specifically, preprocessing is responsible for noise removal of the input signals, and those noise removal methods, e.g., independent component analysis (ICA), principal component analysis (PCA), canonical correlation analysis [6] are typically used for this purpose. In [7] various versions of ICA-based methods are evaluated for MI EEG signals classification in subject-to-subject transfer BCI systems and results suggested that simplified infomax ICA based approach have great potential for BCI applications. Another handy new method for noise removal is multiscale principal component analysis (MSPCA), which has been widely used for noise removal of non-stationary signals, like Electrocardiography (ECG) [8], EEG [9], Electromyography (EMG) [10], ect.

Feature extraction and classification are key components of an automated system determining the classification results, and a broad range of techniques have been proposed for such purpose [11]. Fourier Transform (FT), autoregressive (AR) and common spatial patterns (CSP) are techniques commonly utilized for MI EEG signal feature extraction, yet they either fail to provide temporal facts for EEG signals [11]–[13], or are sensitive to noise [14], or even suffer from low classification success rates [15], [16]. The statistical clustering [17], iterative spatio-spectral patterns learning (ISSPL) [18], as well as the cross-correlation [5] and optimal allocation (OA) [19] based methods also suffer from the similar issues.

Methods based on signal decomposition (SD) are gaining popularity recently in the classification of MI EEG signals. In study [20], comparison of three SD-based methods for the classification of MI tasks, namely empirical mode decomposition (EMD), discrete wavelet transformation (DWT) and wavelet packet decomposition (WPD), showed that WPD had the best results. Some hybrid methods, like multivariate EMD along with FT [21], the dual-tree complex wavelet transform with machine learning classifiers [22], and intrinsic mode function (IMF) with LS-SVM classifier [23], have also been proposed for MI EEG signal classification. Although such methods provide satisfactory results, they do not provide any solution to selecting a suitable number of signal decomposition levels.

In [24], an instantaneous phase difference method was implemented to obtain phase-based features by extraction of phase synchrony information among different electrodes; however, the success rate could be further improved with more sophisticated extraction techniques. In literature different spectral signal representation techniques have been used to explore discriminative features for MI EEG signal classification and results suggest that power spectral density (PSD) provides reasonable success rates in comparison with energy distribution, atomic decompositions and wavelet based techniques [25]. The recurrent quantum neural network filtering technique is implemented in BCI system with a goal of filtering EEG signals before attributes detection and identification to increase the classification outcome [26]. The multilayer perceptron neural networks (MLP) with stochastic gradient descent algorithm was utilize in [27] to recognize the eye state. Researchers in [28] proposed various algorithms to improve the convergence speed and classification accuracy with neural networks, while many deep learning based approaches have also been suggested in BCI with driver drowsiness detection applications [29].

A large number of channels may result in system complexity due to the redundant information from the irrelevant channels [18]. In previous studies, the sequential channel selection methods have been proposed, and up to 20 channels were chosen [30], [31]. However, such methods resulted in low classification accuracy, and the number of channels is still too many for practical applications. For the review of channel selection methods, readers are referred to [32].

Furthermore, it is conventional to extract features from different channels and time segments to build a large feature vector, and then use it as an input to different machine learning classifiers for classification performance tests. To reduce the dimension of such large feature vector, some feature reduction techniques e.g. PCA or ICA, were utilized in [33], [34], yet the classification results still could not be improved. In [35], [43], different feature combination techniques were also tested to select the best combination for classification performance enhancement, however, several experiments were required for choosing the best features combination. To address such issues, a feature selection method was proposed in [36] to choose the suitable time and spectral domain features, however, the computational load of this method was very high, hindering the practical applications of this method. For a comprehensive review of the feature selection methods, readers are referred to [37] for more details.

In our previous work [43], we proposed empirical wavelet transform based algorithm and achieved reasonable results for MI EEG signals classification yet there are certain limitations in it. *First*, no noise removal algorithm proposed however, EEG signals contaminated with several environmental and cognitive noises. *Second*, we performed univariate analysis however EEG electrodes are closely placed to each other's and EEG signals exhibits nonlinear and non-stationary behavior so to analyze multi-channel information

TABLE 1. Abbreviations utilized in the paper.

Abb.	Definition	Abb.	Definition
RH	Right Hand	RF	Right Foot
MI	Motor Imagery	LS-SVM	Least Square Support-Vector Machine
BCI	Brain- Computer Interface	MLP	Multi-layer perceptron Neural Network
MEWT	Multivariate Empirical Wavelet Transform	LR	Logistic Regression
EEG	Electroencephalography	LMT	Logistic Model Tree
MSPCA	Multiscale Principal Component Analysis	NB	Naïve Bayes
JIA	Joint Instantaneous Amplitude	JIF	Joint Instantaneous Frequency
ME	Mean Energy	MTKE	Mean Teager-Kaiser Energy
ShWE	Shannon Wavelet Entropy	LEn	Log Energy Entropy
AUC	Area under the receiver operating characteristics	KW	Kruskal Wallis

dependency and information available in different frequency bands, multivariate analysis of data is obligatory. *Third*, we selected 18 channels manually based on the physiological knowledge and these numbers of channels are still high for certain BCI applications. *Fourth*, different combinations of features was tested by performing several experiments, no automated feature selection criteria was proposed to select the suitable combination of features and number of features for each subject was too many. *Last*, we utilized only one dataset and previous study was only limited for subject-specific case.

Improving the accuracy of the MI EEG signal classification while reducing the BCI system complexity introduced by the number of channels utilized, we evaluate, for the first time to our knowledge, the effectiveness of multivariate empirical wavelet transform (MEWT) [40] for both channel selection and feature extraction for the multi-channel, non-stationary and nonlinear MI EEG signals. The main contributions of this paper are as follows,

- MSPCA is adopted for EEG signals noise removal.
- The automated channel selection strategy is proposed and analyzed for decoding channel selection.
- The sub-band alignment method is adopted to obtain joint instantaneous amplitude (JIA) and joint instantaneous frequency (JIF) components.
- A correlation-based feature selection algorithm is implemented to get rid of redundant information while improving classification outcomes.

To achieve the above listed goals, experiments are performed on three publicly available benchmark datasets IVa, IVb and IVc from BCI competition III website for both subject-specific and subject independent cases.

Some of the abbreviations (Abb.) utilized in this paper are listed in Table 1.

The rest of the paper is arranged accordingly. Section II presents the MI EEG datasets used in this study. Section III describes the details of our proposed method. Section IV gives the results while section V describes the discussions, and section VI concludes this work.

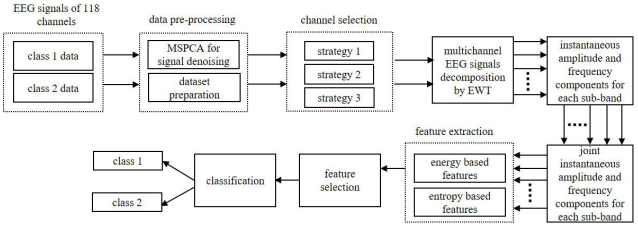


FIGURE 1. Block diagram of the proposed methodology for subject-specific MI EEG signal classification.

II. MATERIALS

The key challenge in BCI application is to provide good classification results even by using small training data. In this study, we used the publicly available dataset IVa from BCI competition III [41], for performance comparisons, and offline experimentations were conducted to obtain a stable assessment of the proposed approach.

The dataset IVa consists of the right hand (RH) and right foot (RF) MI tasks. This dataset was recorded from five healthy subjects (named as ''aa'', ''al'', ''av'', ''aw'' and ''ay'') by placing 118 electrodes according to the 10/20 international system instructions [42]. There were, in total, 280 trials for each subject, with half of them for class 1 tasks and the other half for class 2 tasks. To perform different MI tasks, subjects were shown different visual cues for 3.5 seconds. The original dataset was recorded at a sampling rate of 1000 Hz [41].

The datasets IVb and IVc consists of the two MI tasks, i.e., left hand (LH, class1) and right foot (RF, class2) [41]. They were recorded from a single healthy subject labelled as ivb and ivc in our study with 118 electrodes placed on subject according to the extended international 10/20 system. These datasets have 7 initial sessions without feedback. 210 trials from 118 electrodes were recorded for both MI tasks, and those signals were filtered by a band-pass filter with its lower and upper frequencies being 0.05Hz and 200Hz respectively. The key challenge in dataset IVb is to achieve maximum classification accuracy for continuous EEG data with no cue information, whereas for dataset IVc the main target is to provide successful classification outcome on test data that is recorded after many hours of training data recordings. Further information for all three datasets are available on BCI competition III website (http://www.bbci.de/competition/iii/).

III. METHODS

The proposed automated MEWT algorithm comprises of six distinct modules as shown in Fig. 1, each of which is briefly described as follows.

A. MODULE 1: DATA PRE-PROCESSING

1) DATASETS PREPARATION

We used the same training trials for experiments as we used in our previous study [43] based on the information available in [19], [35], [44]. Particularly, the 100 Hz downsampled version of all three datasets is used in this study. These datasets consists of both labeled and unlabeled data as an

TABLE 2. Dataset IVa information.

Subjects	Data points for both classes	Training trials (labeled data)	Testing trials (unlabeled data)
"aa"	298458 × 118	168	112
"al"	283574 × 118	224	56
"av"	283042 × 118	84	196
"aw"	282838 × 118	56	224
"ay"	283562 × 118	28	252

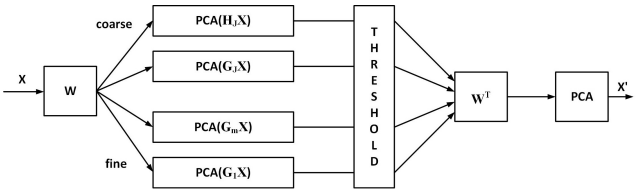


FIGURE 2. Multiscale principal component analysis for signal pre-processing [9], [45].

example data is shown for dataset IVa in Table 2. In this work, we considered labeled data only as class labels for each data segment that is pre-requisite to implement the proposed algorithm. We considered only those EEG segments containing MI tasks information only, and position markers are used to obtain MI data [20]. As discussed, each MI task is 3.5 seconds, and thus, we have $3.5 \times 100 = 350$ samples for each EEG segment of the individual class.

2) MSPCA FOR NOISE REMOVAL

EEG signals suffered with several types of noises and artifacts, and they could be described with a linear model as below,

$$R = P + Q \tag{1}$$

where P represents the EEG data matrix with a clear source and Q denotes the noise information.

The basic aim of any noise removal algorithm is to remove Q from P. For this purpose, PCA has been widely used as it de-correlates correlated variables by providing a linear association between different observations [6]. Moreover, since EEG is a non-stationary and nonlinear signal, the wavelet transform is adopted for signal analysis. By combining the benefits of PCA and wavelets together, the MSPCA, which is illustrated in Fig. 2, can be described by the following steps [9].

- For a data matrix of X, decompose each column at a decomposition level J by using the wavelet transform.
- For $1 \leq m \leq J$, run PCA of the detailed matrices $G_m X$ to choose the meaningful principal components, whose eigenvalues are greater than the mean of all eigenvalues by using the Kaiser rule [45] or reject the detail.
- Run PCA of the approximation matrix $H_J X$ to select useful principal components, whose eigenvalues are greater than the mean of all eigenvalues by using the Kaiser rule [45].
- Perform invert wavelet transform W^T to obtain a new matrix from the reduced detail and approximation matrices.
- Perform the PCA of that new matrix to get $\dot{X}$.

By employing such an MSPCA method explained above, noise could be removed from EEG signals.

B. MODULE 2: CHANNEL SELECTION

A large number of channels for the BCI system results in system complexity since a huge number of features have to be extracted. The foremost objective of this work is to improve the classification accuracy among different MI tasks while minimizing the hardware resources and system complexity. Hence, we performed different experiments by using three different channel selection strategies to select the most suitable channels for the proposed algorithm.

1) STRATEGY 1

In brain science, the motor cortex is well known as a specific region of the brain which is responsible for MI tasks. According to the standard 10-20 systems [42], only 18 electrodes are actually placed around the sensor motor cortex, which gives satisfactory results in EEG signals classification as reported in [46]. Therefore, in the strategy 1 we manually selected 18 channels from the overall 118 channels for all subjects in dataset IVa and label these electrodes according to the standard 10-20 systems [42] to be $C_5, C_3, C_1, C_2, C_4, C_6, CP_5, CP_3, CP_1, CP_2, CP_4, CP_6, P_5, P_3, P_1, P_2, P_4$ and P_6, respectively.

2) STRATEGY 2

The electrodes C_3, C_Z, and C_4 are placed around the sensory motor cortex region according to the standard 10-20 system to record the MI information [42]. In strategy 2, we considered these 3 electrodes for our study since these electrodes contain the most discriminative information related to the hands and foot movements. It is important to mention that the MI movement of RH is usually observed above the left motor cortex around the C_3 electrode and MI foot movements perceived around the C_Z electrode [20].

3) STRATEGY 3

In strategy 3, an automated subject dependent channel selection criteria is used to select the best group of channels for each subject separately. We have $C_k^n = [C_1^n, C_2^n, C_3^n, \cdots, C_{118}^n]$ with n representing class 1 or class 2 tasks, then the Fisher score for mean spectrum power of C_k^n is calculated using the equation below,

$$F(K) = \frac{| p_1^n(K) - p_2^n(K) |^2}{v(p_1^n(K)) + v(p_2^n(K))} \tag{2}$$

where $p_1^n(K)$ and $p_2^n(K)$ represents the mean power spectrum of the Kth channel for class 1 and class 2, respectively, while $v(p_1^n(K))$ and $v(p_2^n(K))$ represent the variances of the Kth channel for class 1 and class 2, respectively. By doing so, $F(K)$ for all channels are calculated, and in our case we

TABLE 3. Channel selection by using strategy 3 for dataset IVa.

Subjects	Best Channel Groups
"aa"	$CCP_5,\ CP_5,\ CP_5$
"al"	$C_3,\ FFC_7,\ CCP_3$
"av"	$FT_9,\ P_8,\ PPO_8$
"aw"	$C_4,\ CCP_6,\ CP_6$
"ay"	$CCP_5,\ C_3,\ CFC_5$

selected 3 channels for each subject based on the top three values. The best channel group for each subject in dataset IVa is given in Table 3 as,

C. MODULE 3: MULTIVARIATE EWT BASED EEG SIGNALS DECOMPOSITION

1) EMPIRICAL WAVELET TRANSFORM

The Empirical Wavelet Transform (EWT) is a fully data-adaptive approach used for the analysis of non-stationary and nonlinear signals, and it decomposes any signal into distinct sub-bands according to the information available in the signal. The brief overview of the EWT is described as follows,

Step 1: Run fast Fourier Transform (FFT) for examined signals to obtain the Fourier spectra of the signals within 0 to π range.

Step 2: Divide the obtained Fourier spectra into N adjacent segments by employing the scale-space boundary detection method described in [40]. For this purpose, $N+1$ boundary frequencies $\{v_k \mid k = 0...M\}$ are needed. The $v_0 = 0$ and $v_M = \pi$ represents the first and last boundary frequencies, respectively, and the remaining $N-1$ boundary frequencies are calculated by using the scale-space representation of Fourier spectra [40].

Step 3: Construct and apply empirical Meyer's wavelets based bandpass filter (BPF) [47] for segmented Fourier spectra. The expressions of empirical scaling and wavelet functions are given by Eq. (3) and Eq. (4) respectively,

$$\hat{S}_k(v) = \begin{cases} 1, & \text{If } |v| \le (1-\beta)v_k \\ \cos\left(\dfrac{\pi\phi(\beta, v_k)}{2}\right), & \text{If } (1-\beta)v_k \\ & \le |v| \le (1+\beta)v_k \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$\hat{W}_k(v) = \begin{cases} 1, & \text{If } (1+\beta)v_k \\ & \le |v| \le (1-\beta)v_{k+1} \\ \cos\left(\dfrac{\pi\phi(\beta, v_{k+1})}{2}\right), & \text{If } (1-\beta)v_{k+1} \le |v| \\ & \le (1+\beta)v_{k+1} \\ \sin\left(\dfrac{\pi\phi(\beta, v_k)}{2}\right), & \text{If } (1-\beta)v_k \le |v| \\ & \le (1+\beta)v_k \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

$$\phi(\beta, v_k) = \alpha\left(\frac{(\mid v \mid -(1-\beta)v_k)}{2\beta v_k}\right) \tag{5}$$

where the parameter β is responsible for avoiding any overlap between functions defined in Eq. (3) and Eq. (4) respectively and it provides a tight frame. For a tight frame Eq. (6) should be met [40],

$$\beta < min_k\left(\frac{v_{k+1} - v_k}{v_{k+1} + v_k}\right) \tag{6}$$

where the arbitrary function $\alpha(z)$ is expressed by Eq. (7) [40],

$$\alpha(z) = \begin{cases} 0, & \text{if } z \le 0 \\ \alpha(z) + \alpha(1-z) = 1, & \text{for all } z[0\ 1] \\ 1, & \text{if } z \ge 1 \end{cases} \tag{7}$$

The detail and approximation coefficients are determined by the dot product of the analyzed signal with the empirical scaling function and the empirical wavelet function, respectively. Readers are referred to [40] for more details.

2) EWT EXTENSION FOR MULTIVARIATE ANALYSIS

EWT extension for multivariate analysis is further divided into three sub-steps,

Step 1: Sub-bands Alignment: To achieve multivariate analysis, all channels should have the equal number of sub-bands, and the same indexed sub-bands for different channels should have the same frequency range. To accomplish these two requirements, we executed mean spectrum magnitude of signals (let say $\dot{x}(t) = [\dot{x}_1(t), \dot{x}_2(t), \cdots \dot{x}_k(t)]$) obtained from channel selection strategies 1, 2 and 3 for each subject separately, and it is given as,

$$\dot{X}(f) = \frac{1}{N_c}\sum_{n=1}^{N_c} \mid \dot{X}_n(f) \mid \tag{8}$$

where $\dot{X}_n(f)$ is the Fourier transform of the $n-th$ channel signal $\dot{x}_n(t)$ and N_c is the total number of signal channels. Afterward boundary detection method as described in [40] is applied onto $\dot{X}(f)$ to obtain adaptive wavelet filters. All channel signals from strategies 1, 2 and 3 are decomposed by using the filter bank to obtain aligned sub-bands.

Step 2: Hilbert Transform for Instantaneous Components Extraction: Hilbert transform is applied onto each narrow sub-band signal to extract the instantaneous components and the analytical representation is given as below [48],

$$\dot{x}_{n+}(t) = \dot{x}_n(t) + jH(\dot{x}_n(t)) \tag{9}$$

The instantaneous amplitude of $\dot{x}_{n+}(t)$ can be computed as,

$$\dot{A}_{\dot{x}_n}(t) = \sqrt{(\dot{x}_n(t))^2 + (H(\dot{x}_n(t)))^2} \tag{10}$$

While the instantaneous phase of $\dot{x}_{n+}(t)$ can be expressed as follows,

$$\dot{\Theta}_{\dot{x}_n}(t) = arctan\left(\frac{H(\dot{x}_n(t))}{\dot{x}_n(t)}\right) \tag{11}$$

The instantaneous frequency of $\dot{x}_{n+}(t)$ is obtained by the rate of change of the instantaneous phase as $\dot{\Theta}_{\dot{x}_n}(t)$,

$$\dot{F}_{\dot{x}_n}(t) = \frac{d}{dt}\left(\dot{\Theta}_{\dot{x}_n}(t)\right) \tag{12}$$

Step 3: Joint Instantaneous Components: Instantaneous components of all sub-bands in channel selection strategies 1, 2 and 3 are combined to obtain joint instantaneous amplitude (JIA) and joint instantaneous frequency (JIF) components, which are mathematically formulated as below [49], [50],

$$\dot{A}^{Joint}_{\dot{x}_n}(t) = \sqrt{\sum_{n=1}^{N}[\dot{A}_{\dot{x}_n}(t)]^2} \tag{13}$$

$$\dot{F}^{Joint}_{\dot{x}_n}(t) = \frac{\sum_{n=1}^{N}[\dot{A}_{\dot{x}_n}(t)]^2\dot{F}_{\dot{x}_n}(t)}{\sum_{n=1}^{N}[\dot{A}_{\dot{x}_n}(t)]^2} \tag{14}$$

We randomly picked channels from strategy 2 for the subject ''aa'' of data set IVa to represent our proposed methodology graphically. To distinguish between different classes, the blue color is used to represent class 1 signals, whereas the red color is used to represent class 2 signals. Furthermore, to show the comparisons clearly, the sub-figures in the same row are kept within the same X-axis and Y-axis ranges.

Fig. 3(a) represents a typical signal pattern from electrodes C_3, C_z and C_4 for RH and RF classes of subject ''aa'' in dataset IVa. From Fig. 3(b) and Fig. 3(c), we can see a significant difference between the JIA and JIF components for two classes, and therefore, there are good chances to achieve satisfactory classification accuracy due to the statistical independence between them [5].

D. MODULE 4: FEATURE EXTRACTION

To classify two MI classes, we extracted both entropy [51] and energy [52] based features from JIA components of each subject in dataset IVa obtained in channel selection strategies 1, 2 and 3, respectively. The description of these features are followed as,

1) MEAN ENERGY

The mean energy (ME) of the EEG signal has been widely used in different SD methods for classification purposes, which provides satisfactory classification results [49], [53]. We adopted this feature in our study, and the mathematical formulation of ME is defined as below,

$$ME_{\dot{x}} = \log(\frac{1}{T}\sum_{t} \mid \dot{x}_n(t)^2 \mid) \tag{15}$$

2) MEAN TEAGER-KAISER ENERGY

The mean Teager-Kaiser energy (MTKE) is a nonlinear attribute sensitive to track even little fluctuations in non-stationary signals. By using MTKE, meaningful information from small fluctuations of amplitude and frequency of EEG signal could be obtained. In this study, we obtained significant values of JIA for all sub-bands of channel selection strategies 1, 2 and 3, and the mathematical expression for MTKE is given as [52], [53],

$$MTKE_{\dot{x}} = \log(\frac{1}{T}\sum_{t} \mid \dot{x}_n(t)^2 - \dot{x}_n(t+1) * \dot{x}_n(t-1) \mid) \tag{16}$$

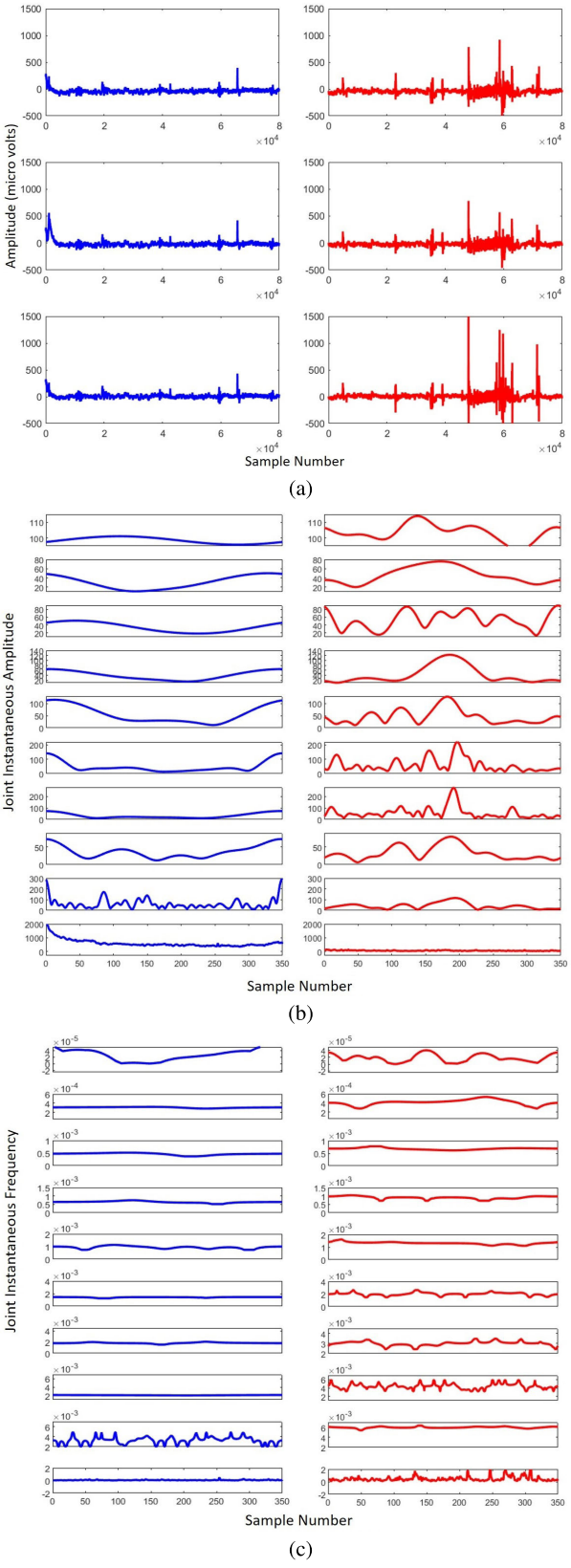


FIGURE 3. Graphical representation of proposed methodology for channel selection strategy 2. (a) Plot for signals C_3, C_Z and C_4 for RH and RF MI classes. (b) JIA for RH and RF MI classes. (c) JIF for RH and RF MI classes.

3) SHANNON WAVELET ENTROPY

Without requiring any parameters during its execution process, Shannon wavelet entropy (ShWE) has been widely utilized for EEG signal classifications for its simplicity and effectiveness both in noise suppression and fine changes for non-stationary signals. In this study, ShWE is also adopted to reduce computational load while achieving satisfactory results, and it can be expressed as [54],

$$ME_{\dot{x}} = \sum_{t} \dot{x}_n(t) \log(\dot{x}_n(t)) \tag{17}$$

4) LOG ENERGY ENTROPY

EEG signal amplitude changes with time instances. The simplest way to obtain complete information of such a signal is to take its square, as it consequently helps achieve better discrimination between different MI classes [55]. The formula for Log energy entropy (LEn) is given as below [55], [56],

$$LEn_{\dot{x}} = \sum_{t} \log(\dot{x}_n(t)^2) \tag{18}$$

E. MODULE 5: FEATURE SELECTION

A typical feature selection algorithm use a search technique for finding suitable feature subsets and perform evaluation measure to rank the different feature subsets. In literature, different feature selection methods are available, including those scheme independent methods (e.g. wrapper methods [38]), which are very slow, and those fast attribute selection methods (e.g. ranking based [39]), which may eliminate the irrelevant attributes yet failed to eliminate redundant attributes. In this study, we used the best first method for searching [57] together with a scheme independent attribute selection method named Correlation-based feature selection (CfsSubsetEval) for appropriate feature subset selection. An attribute subset is considered too good if its attributes are highly correlated with the class attribute yet are less correlated with each other. The mathematical representation is given as follows, [58],

$$\textit{Appropriate attribute subset} = \frac{\sum_{allattributesf} C(f, class)}{\sqrt{\sum_{allattributesf} \sum_{allattributesg} C(f, g)}} \tag{19}$$

where C provides the correlation between two attributes and entropy-based metric called "symmetric uncertainty" is used [58].

F. MODULE 6: CLASSIFICATION

To segregate different MI tasks, six well-known classifiers as below are used.

1) LOGISTIC REGRESSION

The logistic regression (LR) is a popular and powerful classifier been widely used in BCI applications [35], which uses the logit transform to predict class probabilities directly. The logit transform of the LR model can be mathematically formulated as,

$$logit(q) = ln = \acute{\alpha}_0 + \sum_{n=1}^{N} \acute{\alpha}_n f_n \tag{20}$$

where $\acute{\alpha}_0$ is intercept while $\acute{\alpha}_1, \acute{\alpha}_2 ... \acute{\alpha}_n$ are regression coefficients associated with independent variables $f_1, f_2 ... f_n$ and q is considered as a dependent variable belonging to class (y) 0 or 1.

Suppose q belongs to class 0, then mathematically as,

$$q(y = 0 \mid f_1, f_2 ... f_n) = \left(\frac{e^{\acute{\alpha}_0 + \sum_{n=1}^{N} \acute{\alpha}_n f_n}}{1 + e^{\acute{\alpha}_0 + \sum_{n=1}^{N} \acute{\alpha}_n f_n}} \right) \tag{21}$$

Similarly, if q belongs to the class 1, it can be calculated as $1 - q(y = 0 \mid f_1, f_2 ... f_n)$ [59].

2) MULTILAYER PERCEPTRON NEURAL NETWORK

The multilayer perceptron neural network (MLP) is another typical classifier used for EEG signal classification [59] which produces nonlinear decision boundaries, and the weights of the MLP classifier is determined by the backpropagation algorithm.

A simple neural network for linear solution consists of input and output layers, while for nonlinear solution additional layers are used between input and output layers. These additional layers are known as hidden layers because they do not interact with the external environment, and the number of such layers depends on the system nonlinearity and complexity. Each layer contains some processing elements known as neurons, and the number of them for the input and output layers are the same as those of the attributes in the feature vector and the classes, respectively. The number of neurons in the hidden layer is determined empirically as follows [59].

$$\#of\ neurons\ in\ hidden\ layer = \frac{\#of\ features + \#of\ classes}{2} \tag{22}$$

The initialization of neurons in MLP classifier is determined by the backpropagation algorithm.

3) LEAST-SQUARE SUPPORT VECTOR MACHINE

The least-square support-vector machine (LS-SVM) has been used in BCI applications for its simplicity and effectiveness in dealing with both linear and nonlinear problems [5], and the mathematical formulation of the LS-SVM decision function can be expressed as,

$$q(f) = sign\left(\sum_{i=1}^{d} q_i l_i Ker(f, f_i) + b \right) \tag{23}$$

where for i^{th} input feature vector f_i, the class label is q_i, d represents the total dimension of the feature vector, b represents the bias term, while l_i indicates the Lagrange multipliers. The decision accuracy of the LS-SVM classifier is dependent on the kernel choice. In this study, the Radial

basis function (RBF) kernel is empirically selected and can be represented as,

$$Ker(f, f_j) = e^{\frac{-(||f-f_j||)^2}{2\sigma^2}} \tag{24}$$

wherein the RBF kernel has two important parameters, i.e., the regularization parameter γ and the bandwidth σ^2. The values of both parameters should be carefully chosen to avoid over-fitting while achieving maximum accuracy [5].

4) LOGISTIC MODEL TREE

The logistic model tree (LMT) is a simple, compact and precise classifier recently being used for EEG signal classifications [60], [61]. Suppose a feature vector is represented by f, then for S input feature vectors and C classes, a posterior probability for respective class can be formulated as [60], [61],

$$p(c \mid f) = \frac{e^{(F_c(f))}}{\sum_{n=1}^{C} e^{(F_c(f))}} \tag{25}$$

where $F_c(f)$ is the linear regression function to be fitted, e is the natural logarithm and c is the number of classes.

5) Naïve BAYES

The naïve Bayes (NB) is a straightforward classifier recently being used for MI EEG signals classification [21]. In NB classifier, the class with maximum value of posterior probability is considered as the resulting class, while the prediction of an input feature vector class $f = [f_1, f_2, \cdots, f_i]$ is performed by computing the highest probability of respective class (c), and it is formulated as below [19],

$$p(c \mid f) = \frac{p(c) \prod p(f_i \mid c)}{p(f)} \tag{26}$$

IV. RESULTS

In this study, both subject-specific and subject independent experiments are performed as detailed in case 1 and case 2 respectively.

A. CASE 1: RESULTS FOR SUBJECT-SPECIFIC EXPERIMENTS

1) EXPERIMENTAL RESULTS FOR DATASET IVa

a: PERFORMANCE METRICS

The 10-fold cross-validation method is used in the experiments to avoid over-fitting while obtaining the fair output. A feature vector set of each subject is divided into 10 mutually exclusive subsets with an approximately equal size by using the common standard 10 fold cross-validation technique, and the process is repeated 10 times (folds). One of the subsets is used as a test set each time while the other 9 subsets are put together to create a training set. Finally, the average accuracy is calculated throughout all sets [5]. The number of trials for each subject in the experiments are unequal, which are intended to check the performances of the proposed algorithm for different amount of training trials. The performances of MEWT based experiments are measured by average sensitivity (S_{en}), specificity (S_{pe}), classification accuracy (A_{cc}), and area under the receiver operating characteristics (AUC) [62], and they are defined as follows,

$$S_{en} = \frac{TP}{P} \tag{27}$$

$$S_{pe} = \frac{TN}{N} \tag{28}$$

$$A_{cc} = \frac{TP + TN}{P + N} \tag{29}$$

where TP is a true positive represents the amount of accurately calculated MI tasks for the RH class, and TN is true negative refer to the number of accurately calculated MI tasks for the RF class, whereas P and N describe the number of RH and RF classes in total, respectively. AUC ranges from 0 to 1, depicting a classifier classification capacity. The higher the AUC is, the better the capability of the classifier to identify two different tasks [62]. All experiments were performed on a personal computer with 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 the LS-SVM toolbox v.1.8 have been used for experiments.

b: PARAMETERS SELECTION

1. MSPCA: The parameters for MSPCA were chosen empirically. The Kaiser rule was selected in our experiments [45], which retains only those principal components whose eigenvalues are greater than the mean of all Eigenvalues. The wavelet decomposition levels were empirically chosen to be 5 in our experiments.

2. EWT: The parameters for EWT were chosen in hit and trial manners, and 10 signal decomposition levels were chosen in this study. While for Fourier boundary detection, the scale-space method as described in [40] was selected.

3. LR: LR is the simplest classifier since there is no need to tune any parameter as required in support vectors machine classifier. The parameters of LR classifier were automatically selected by the employed built-in maximum likelihood estimation (MLE) method.

4. MLP: The number of neurons in the hidden layer of MLP classifier was selected by choosing the default parameter settings available in Waikato Environment for Knowledge Analysis (WEKA) software, wherein the following formula is utilized to select the number of neurons in the hidden layer of MLP classifier.

$$\#of\ neurons = \frac{\#of\ features + \#of\ classes}{2} \tag{30}$$

To show the number of neurons in a hidden layer in this study, we randomly took channel selection strategy 3 with FS2 as an example. In this case, we considered only a single hidden layer and two neurons for each subject.

5. LS-SVM: The parameters of the RBF kernel employed in LS-SVM were estimated in two steps for each fold in all experiments. In the first step, the initial values of parameters were searched out using coupled simulating

TABLE 4. Tuned parameter values for each fold of LS-SVM classifier by employing FS2 in channel selection strategy 3.

Subjects	"aa"		"al"		"av"		"aw"		"ay"	
Parameters	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2	γ	σ^2
Fold 1	3.67	3.48	29.52	0.34	2.18	1.82	7.08	11.20	4.64	476.21
Fold 2	1.24	7.31	1.40	9.79	12.82	22.25	10.17	11.07	0.86	40.27
Fold 3	7.98	21.85	14.59	80.77	17.71	99.84	5.34	7.24	0.02	2.77
Fold 4	2.97	14.99	37.43	12.27	33.15	30.03	8.35	3.42	8.05	68.8
Fold 5	17.21	31.93	15.09	6.82	5.95	22.89	2.27	7.19	2.63	9.61
Fold 6	4.37	47.13	13.28	2.54	12.29	2.78	13.37	13.73	571.54	10.15
Fold 7	2.39	11.03	592.96	4.48	668.81	4.01	1.71	6.09	798.37	5.84
Fold 8	11.12	123.02	17.93	0.35	19.55	0.37	2.04	9.86	287.82	1.05
Fold 9	6.07	13.86	8.83	0.31	12.34	34.38	3.39	8.94	21.49	1.00
Fold 10	2.95	9.36	20.93	112.62	13.66	92.22	4.32	6.44	309.87	1.71

TABLE 5. P values for strategies 1, 2 and 3.

Subjects	Strategy 1			
	F1	F2	F3	F4
"aa"	8.11×10^{-06}	4.61×10^{-02}	1.00×10^{-02}	6.35×10^{-07}
"al"	8.97×10^{-05}	0.01	1.57×10^{-04}	3.81×10^{-04}
"av"	1.08×10^{-09}	1.79×10^{-07}	5.28×10^{-07}	8.11×10^{-06}
"aw"	1.12×10^{-07}	7.00×10^{-03}	0.01	8.11×10^{-06}
"ay"	2.69×10^{-02}	1.16×10^{-02}	1.09×10^{-02}	8.11×10^{-06}
	Strategy 2			
	F1	F2	F3	F4
"aa"	2.50×10^{-02}	7.36×10^{-04}	1.80×10^{-03}	4.77×10^{-04}
"al"	6.81×10^{-03}	8.30×10^{-03}	0.01	8.97×10^{-05}
"av"	1.08×10^{-09}	1.08×10^{-09}	5.28×10^{-07}	9.11×10^{-02}
"aw"	3.28×10^{-11}	8.81×10^{-04}	2.90×10^{-02}	6.12×10^{-09}
"ay"	5.25×10^{-07}	7.00×10^{-03}	1.76×10^{-02}	4.61×10^{-02}
	Strategy 3			
	F1	F2	F3	F4
"aa"	3.70×10^{-02}	1.84×10^{-04}	3.81×10^{-04}	1.63×10^{-04}
"al"	6.81×10^{-03}	4.60×10^{-02}	1.57×10^{-04}	5.28×10^{-07}
"av"	5.27×10^{-07}	1.17×10^{-05}	1.79×10^{-04}	6.35×10^{-07}
"aw"	3.28×10^{-10}	8.81×10^{-04}	2.90×10^{-02}	6.13×10^{-09}
"ay"	3.58×10^{-07}	4.00×10^{-03}	2.00×10^{-03}	1.74×10^{-04}

annealing (CSA) [63] method, wherein the search limit was fixed in the range of $[exp(-10), exp(10)]$. The values obtained in the first step were provided as input to the grid search method for best results. The parameter values for each fold of channel selection strategy 3 with FS2 are shown in Table 4 for illustration purposes.

6. LMT: The default parameter settings available in WEKA were utilized for the LMT classifier.

7. NB: Like the LR classifier, the parameters for the NB classifier were selected by the MLE method automatically, and thus, there is no need of parameters tuning.

c: STATISTICAL SIGNIFICANCE OF FEATURES FOR CHANNEL SELECTION STRATEGIES 1, 2 AND 3

The significance of features extracted in channel selection strategies 1, 2 and 3 for different MI tasks were evaluated by using Kruskal Wallis (KW) test, which is a non-parametric test without requiring input data to follow the normal distribution [51]. The probability (P) values for all features obtained by the KW test are presented in Table 5. For simplicity, all features are represented by notations as F1=ME, F2=MTKE, F3=ShanEnt, F4=Len.

TABLE 6. Feature selection for channel selection strategies 1, 2 and 3.

Subjects	Strategy 1	Strategy 2	Strategy 3
"aa"	**F4, F1**	**F4, F2**	**F4, F2**
"al"	**F4**	**F4**	**F4**
"av"	**F1, F2, F3**,F4	**F1, F2, F3**,F4	**F1**, F2, F3,**F4**
"aw"	**F1, F4**,F2	**F1, F4**,F2	**F1, F4**
"ay"	**F1**	**F1**	**F1**

d: FEATURE SELECTION FOR CHANNEL SELECTION STRATEGIES 1, 2 AND 3

The feature selection method explained in section III is used for each subject in dataset IVa. In Table 6, the features presented in bold font are the appropriate features, which show the strongest correlation with the class and lowest correlation among each other, being selected by CfsSubsetEval for different cases.

e: THE MI CLASSIFICATION RESULTS FOR THREE CHANNEL SELECTION STRATEGIES BY ALL FEATURES IN DATASET IVa

Table 7 and Fig. 4 represents the classification outcomes obtained by all features (FS1) for different channel selection strategies. It is apparent that for the three strategies, notable classification accuracies could be achieved by using all classifiers. Our experimental results evidently show that MLP and LS-SVM classifiers achieved the highest results, succeeding by LR, LMT and NB classifiers for different MI tasks in three strategies. Moreover, we observed in Table 7 that except NB, we obtained 100% sensitivity, specificity, and classification accuracy figures by using all classifiers for the subject "aw".

For strategy 1 with FS1, we observed few variations among different subject classification accuracies for all classifiers, and in most of the cases specificity numbers (detection capabilities of classifiers for RF class) are higher as compared to the sensitivity values (detection capabilities of classifiers for RH class). Moreover, fewer scores were achieved for subject "ay" by using all classifiers, concluding that strategy 1 is less stable and less unbiased.

For strategy 2 with FS1, the average classification accuracies for all classifiers are slightly better as compared to the strategy 1 with a significant decrement in variations among different subject accuracies obtained by all classifiers. Meanwhile, less difference in numbers among RH and RF class detection capabilities could also be observed, which illustrates that strategy 2 is more stable and more unbiased relative to strategy 1.

In comparison with strategies 1 and 2, strategy 3 showed remarkable results with the maximum sensitivity, specificity and classification accuracies, which result in more stable and less biased towards detection among different classes, proving strategy 3 to be a suitable choice for MI EEG signal classification.

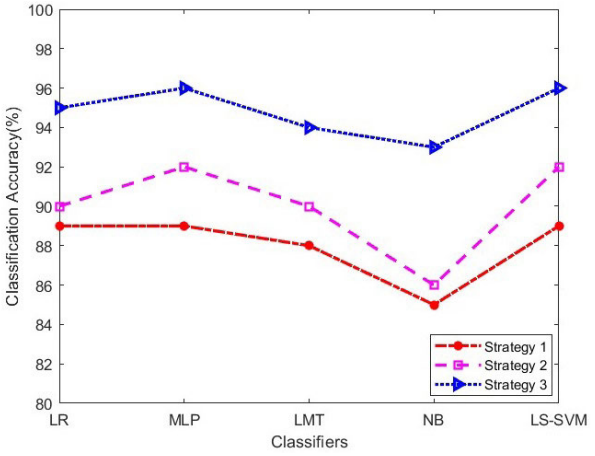


FIGURE 4. Comparison of different classifiers with FS1 for three channel selection strategies.

f: THE MI CLASSIFICATION RESULTS FOR CHANNEL SELECTION STRATEGIES 1, 2 AND 3 BY FEATURE SELECTION STRATEGY

In this section, we discuss the results achieved by employing a feature selection (FS2) strategy discussed in Section III. As indicated in Table 8 and Fig. 5, the classification accuracies for different classifiers increased, this is because in this situation the redundant features were removed and only relevant features for a specific subject in different channel strategies were chosen. These results also suggest that by employing appropriate features for an individual subject, the classification accuracy of LMT could be improved by a significant amount.

For strategies 1 and 2, MLP along with LS-SVM classifier ranked the first in terms of separation abilities, LMT ranked the second, the LR ranked the third and NB the last. For strategy 3, MLP, LS-SVM and LMT classifiers topped as they provide maximum classification accuracy of 97% with a standard deviation of 2.7%. LR provides a maximum classification accuracy of 96% with variations of 2.2% only, whereas NB achieved a classification accuracy of 93% among different MI classes with the least variations. In terms of sensitivity, strategy 3 achieved 97% average amount among different subjects by using MLP, LS-SVM, LMT and LR classifiers, whereas NB classifier provides 92% on average. Likewise, for specificity outcomes, strategy 3 obtained 97% average amount among five subjects by utilizing MLP, LS-SVM and LMT classifiers where the average value of 96% and 95% is obtained by employing the LR and NB classifiers.

Furthermore, we provide a comparison between three strategies for FS1 and FS2 in Fig. 6 to show the importance of feature selection. As illustrated, for every classifier, FS2 provides higher or equal classification accuracy as compare with FS1, which concludes that FS2 is more useful features in different cases and considered to be the relevant information source to provide better discrimination abilities among MI classes.

TABLE 7. Classification outcome (%) for FS1 in strategies 1, 2 and 3.

Subject	Classifier	Strategy 1			Strategy 2			Strategy 3		
		Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec	Acc
"aa"	LR	80	100	90	80	90	90	90	100	95
	MLP	80	100	90	80	100	90	90	100	95
	LMT	80	90	85	80	90	85	100	100	100
	NB	90	90	90	80	80	85	80	100	90
	LS-SVM	80	90	90	80	100	90	90	100	95
"al"	LR	100	100	100	90	90	80	90	90	90
	MLP	100	100	100	90	95	90	90	90	90
	LMT	100	90	95	90	95	90	80	80	80
	NB	90	90	90	80	80	85	80	100	90
	LS-SVM	100	100	100	90	95	90	90	90	90
"av"	LR	70	90	80	85	90	90	100	90	95
	MLP	80	90	80	90	90	90	100	90	95
	LMT	70	100	85	95	90	90	100	90	95
	NB	85	85	85	90	90	90	100	90	95
	LS-SVM	80	100	80	90	90	90	100	90	95
"aw"	LR	100	100	100	100	100	100	100	100	100
	MLP	100	100	100	100	100	100	100	100	100
	LMT	100	100	100	100	100	100	100	100	100
	NB	85	85	85	90	85	85	100	90	95
	LS-SVM	100	100	100	100	100	100	100	100	100
"ay"	LR	75	75	75	90	90	85	95	95	95
	MLP	75	75	75	85	85	90	100	100	100
	LMT	75	75	75	85	85	85	95	95	95
	NB	75	75	75	85	85	85	95	95	95
	LS-SVM	75	75	75	85	85	90	100	100	100
mean ± std	LR	85 ± 14.1	93 ± 10.9	89 ± 11.7	89 ± 7.4	92 ± 4.5	90 ± 7.1	95 ± 5.0	95 ± 5.0	95 ± 3.5
	MLP	$\mathbf{87 \pm 12}$	$\mathbf{93 \pm 10.9}$	$\mathbf{89 \pm 11.4}$	$\mathbf{89 \pm 7.4}$	$\mathbf{94 \pm 6.5}$	$\mathbf{92 \pm 4.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{96 \pm 4.2}$
	LMT	85 ± 14.1	91 ± 10.2	88 ± 9.7	90 ± 7.9	92 ± 5.7	90 ± 6.1	95 ± 8.7	93 ± 8.4	94 ± 8.2
	NB	85 ± 6.1	85 ± 6.1	85 ± 6.1	85 ± 5	84 ± 4.2	86 ± 2.2	91 ± 10.2	95 ± 5	93 ± 2.7
	LS-SVM	$\mathbf{87 \pm 12}$	$\mathbf{93 \pm 10.9}$	$\mathbf{89 \pm 11.4}$	$\mathbf{89 \pm 7.4}$	$\mathbf{94 \pm 6.5}$	$\mathbf{92 \pm 4.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{96 \pm 4.2}$

From Tables 6-8 and Figs. 4-6 we conclude that the strategy 3 with FS2 achieved better results by removing redundant information as we obtained maximum classification results for subjects "aa", "al", "av", "aw" and "ay" by using only two, one, two, one and one feature respectively. The automated channel selection strategy, reduced set of features, better results, fewer variations among different subject's accuracies and unbiased nature for the detection of RH and RF classes are the key highlights of strategy 3, demonstrating its great potential for online clinical applications.

g: TIMING EXECUTION FOR THREE CHANNEL SELECTION STRATEGIES FOR FS1 AND FS2 BY USING ALL CLASSIFIERS

To check the applicability of this study for real-time applications, we calculated the execution time, which includes the time for channel extraction, EWT signal decomposition, multivariate procedure, feature extraction (FS1 and FS2) as well as the time required to build a model for each classifier associated with each subject, of this method. The executed time for those five subjects is shown in Fig. 7. As shown in all three cases, the execution time for FS2 is shorter as compared with FS1, which indicates the significance of relevant features selection for reducing the computational time. Moreover, it is evident that if more channels are used like in case 1 (18 channels), the time required to obtain a classification for both FS1 and FS2 are much longer (around 20 seconds) compared with that of strategies 2 and 3. This signifies the importance of appropriate channel selection for reducing computational load.

It is noteworthy that all implementation times were evaluated with all trials considered in dataset IVa, and hence the implementation time of the algorithm would be much shorter

TABLE 8. Classification outcome (%) for FS2 in strategies 1, 2 and 3.

Subject	Classifier	Strategy 1			Strategy 2			Strategy 3		
		Sen	Spec	Acc	Sen	Spec	Acc	Sen	Spec	Acc
"aa"	LR	90	100	90	90	90	90	95	100	95
	MLP	100	100	100	90	100	95	95	95	95
	LMT	80	100	90	90	100	95	95	95	95
	NB	80	100	85	80	90	85	85	100	90
	LS-SVM	100	100	100	90	100	95	95	95	95
"al"	LR	90	100	95	90	90	85	100	90	95
	MLP	90	100	95	90	90	90	95	95	95
	LMT	90	100	95	90	90	90	95	95	95
	NB	90	100	85	85	90	85	80	100	90
	LS-SVM	90	100	95	90	90	90	95	95	95
"av"	LR	90	90	90	90	90	90	95	95	95
	MLP	90	100	95	95	100	95	95	95	95
	LMT	90	100	95	90	100	95	95	95	95
	NB	90	100	90	90	100	90	100	90	95
	LS-SVM	90	100	95	95	100	95	95	95	95
"aw"	LR	100	100	100	100	100	100	100	100	100
	MLP	100	100	100	100	100	100	100	100	100
	LMT	100	100	100	100	100	100	100	100	100
	NB	90	90	90	90	90	90	100	90	95
	LS-SVM	100	100	100	100	100	100	100	100	100
"ay"	LR	75	75	75	80	90	85	95	95	95
	MLP	75	75	75	90	90	90	100	100	100
	LMT	75	75	75	80	80	80	100	100	100
	NB	75	75	80	80	80	80	95	95	95
	LS-SVM	75	75	75	90	90	90	100	100	100
mean ± std	LR	89 ± 8.9	93 ± 10.9	90 ± 9.4	90 ± 7.1	92 ± 4.5	90 ± 6.1	97 ± 2.7	96 ± 4.2	96 ± 2.2
	MLP	$\mathbf{91 \pm 10.2}$	$\mathbf{95 \pm 11.2}$	$\mathbf{93 \pm 10.4}$	$\mathbf{93 \pm 4.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{94 \pm 4.2}$	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$
	LMT	87 ± 9.7	95 ± 11.2	91 ± 9.6	90 ± 7.1	94 ± 8.9	92 ± 7.6	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$
	NB	85 ± 7.1	93 ± 10.9	86 ± 4.2	85 ± 5	90 ± 7.1	86 ± 4.2	92 ± 9.1	95 ± 5	93 ± 2.7
	LS-SVM	$\mathbf{91 \pm 10.2}$	$\mathbf{95 \pm 11.2}$	$\mathbf{93 \pm 10.4}$	$\mathbf{93 \pm 4.5}$	$\mathbf{96 \pm 5.5}$	$\mathbf{94 \pm 4.2}$	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$	$\mathbf{97 \pm 2.7}$

for any single trial, showing that the proposed algorithm could be used for online clinical application. It should also be noted that in this research, experiments are performed in MATLAB and the implementation period was evaluated without any software optimization. However, in reality, if the software implementation is optimized, this algorithm implementation period could be decreased further. Moreover, since the execution time for all the other methods listed in Table 12 were not provided in their studies, the time comparisons between our proposed algorithm and those in the previous studies are not shown in the study.

h: ROC CURVE FOR CHANNEL SELECTION STRATEGY 3 WITH FS2 BY USING ALL CLASSIFIERS

The AUC of channel selection strategy 3 obtained by FS2 for different classifiers related to five subjects is shown in Fig. 8. As noted, for most of the classifiers, the obtained AUC value is near to 1, which indicates the better discrimination power of classifiers for different MI tasks. Precisely we achieved an AUC value between 0.98 to 1 for all subjects in the case of MLP and LS-SVM classifiers as shown in Fig. 8.

i: ROBUSTNESS OF THE PROPOSED APPROACH AGAINST NOISE

As mentioned earlier EEG signals contain several noises and artifacts, it is necessary to remove those artifacts to obtain a clean signal for further processing. To observe the influence of noise on overall classification accuracy, we performed experiments with and without MSCPA.

Fig. 9 presents the classification accuracy of channel selection strategy 3 by employing FS2 with and without MSPCA.

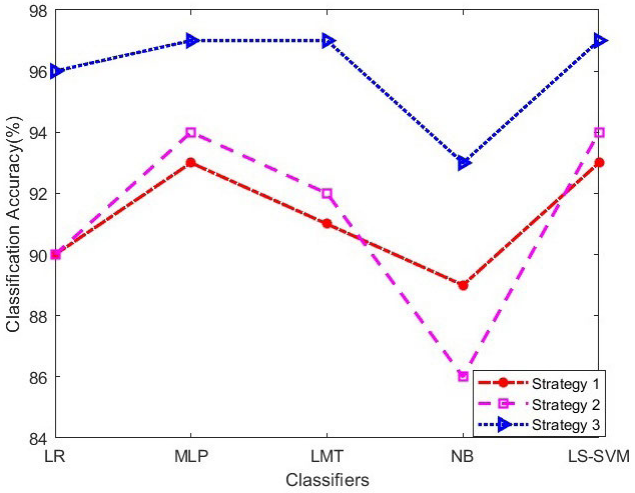


FIGURE 5. Comparison of different classifiers with FS2 for three channel selection strategies.

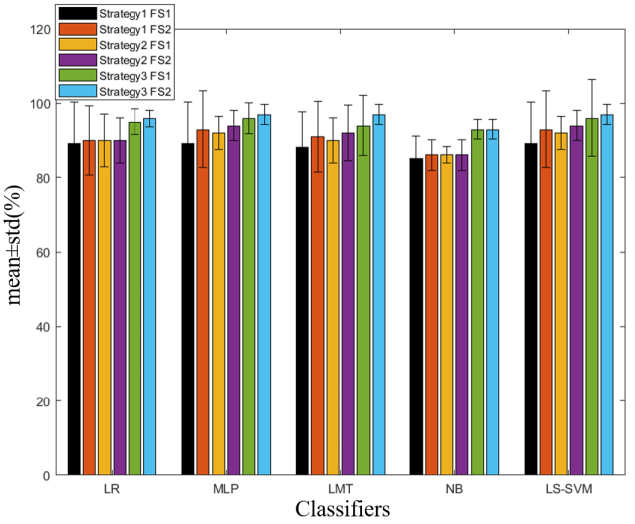


FIGURE 6. Comparison of different classifiers with FS1 and FS2 for three channel selection strategies.

As seen, without MSPCA we obtained 88%, 90%, 88%, 90% and 90% classification accuracies for subjects ''aa'', ''al'', ''av'', ''aw'' and ''ay'' respectively and the average classification accuracy was 89.2%. When MSPCA is applied for the noise removal, the classification accuracy of each subject increased drastically and we obtained 95%, 95%, 95%, 100% and 100% classification accuracies for subjects ''aa'', ''al'', ''av'', ''aw'' and ''ay'' respectively, while the average classification accuracy was 97%. The classification accuracy improvement of 7.8% with MSPCA demonstrated that our proposed method is robust against noise.

2) APPLICABILITY OF PROPOSED MEWT METHOD FOR DIFFERENT DATASETS (IVb AND IVc)

To show the applicability of proposed method for different subjects, we have performed the experiments on datasets IVb

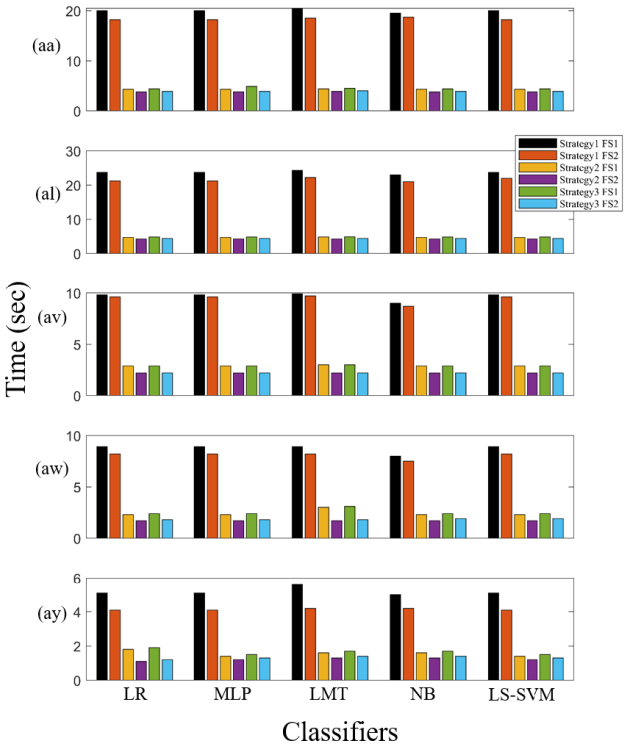


FIGURE 7. Timing execution with FS1 and FS2 among three channel selection strategies with all classifiers, for five subjects respectively.

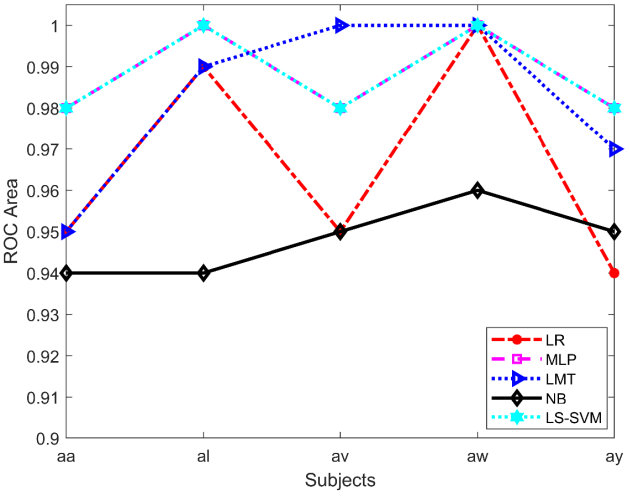


FIGURE 8. ROC Area of different classifiers with FS2 for channel selection strategy 3.

and IVc, which are openly available on BCI competition III website (http://www.bbci.de/competition/iii/). The results obtained for these two datasets are available in Table 9 respectively for both all and selected features.

As seen in Table 9, an average sensitivity, specificity and classification accuracy of 100%, 100% and 100% was achieved by employing FS2 (F4) and all classifiers with the channel selection strategy 3 (C_3, CCP_5,C_4) for dataset IVb, while similar measures of 97% was achieved for dataset IVc

TABLE 9. Classification outcomes of datasets IVb and IVc for subject-specific case.

Subject	Classifier	Strategy 1				Strategy 2				Strategy 3			
		Sen	Spec	Acc	AUC	Sen	Spec	Acc	AUC	Sen	Spec	Acc	AUC
Dataset IVb with all Features (FS1)	LR	90	90	90	0.9	100	100	100	1	100	100	100	1
	MLP	95	95.5	95	0.96	100	100	100	1	100	100	100	1
	LMT	90	91.7	90	0.91	100	100	100	1	100	100	100	1
	NB	95	95.5	95	0.96	95	95.5	95	1	95	95.5	95	1
	LS-SVM	95	95.5	95	0.96	95	95.5	95	0.95	95	95.5	95	0.97
Dataset IVb with Feature selection (FS2)	LR	95	95.5	95	0.96	100	100	100	1	**100**	**100**	**100**	**1**
	MLP	95	95.5	95	0.96	100	100	100	1	**100**	**100**	**100**	**1**
	LMT	95	95.5	95	0.96	100	100	100	1	**100**	**100**	**100**	**1**
	NB	95	95.5	95	0.96	100	100	100	1	**100**	**100**	**100**	**1**
	LS-SVM	95	95.5	95	0.96	100	100	100	1	**100**	**100**	**100**	**1**
Dataset IVc with all Features (FS1)	LR	85	85.4	85	0.9	90	90	90	0.9	90	90	90	0.9
	MLP	85	85.4	85	0.9	90	90	90	0.9	90	90	90	0.9
	LMT	85	85.5	85	0.9	90	90	90	0.9	90	90	90	0.9
	NB	80	80	80	0.8	90	90	90	0.9	90	90	90	0.9
	LS-SVM	85	85.4	85	0.9	90	90	90	0.9	90	90	90	0.9
Dataset IVc with Feature selection (FS2)	LR	90	90	90	0.9	97	97	97	1	**97**	**97**	**97**	**1**
	MLP	90	90	90	0.9	97	97	97	1	**97**	**97**	**97**	**1**
	LMT	90	90	90	0.9	97	97	97	1	**97**	**97**	**97**	**1**
	NB	90	90	90	0.9	97	97	97	1	**97**	**97**	**97**	**1**
	LS-SVM	90	90	90	0.9	97	97	97	1	**97**	**97**	**97**	**1**

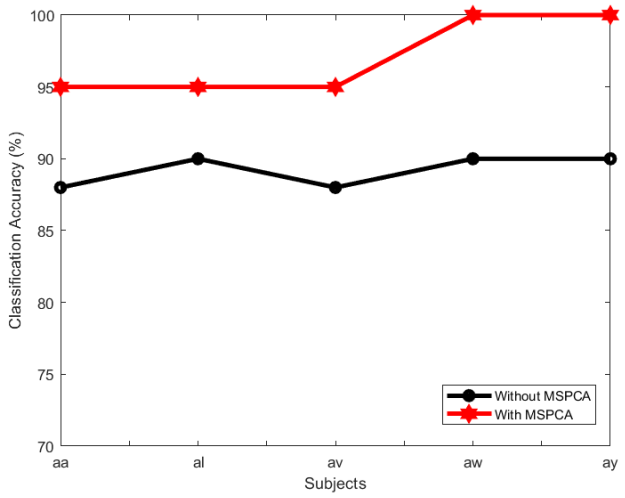


FIGURE 9. Robustness check of MEWT based method against noise with FS2 for channel selection strategy 3.

with channel section strategy 3 (C_3, CCP_3, C_4) for FS2 (F4). In addition, the AUC value for most of the cases is equal or near to 1 which indicates the better classification performance for different classifiers. These results conclude that proposed MEWT based approach is versatile and can be used for subjects with different mental and physical nature.

B. CASE 2: RESULTS FOR SUBJECT INDEPENDENT EXPERIMENTS

1) EXPERIMENTAL RESULTS FOR DATASET IVa

Due to the nonlinear and non-stationary properties of EEG signals, classifiers are trained and tested using the same subject data, and thus, subject-specific MI EEG signal classification algorithms have been proposed in literature [5], [15]–[20], [46], [64]–[66]. In practice, however, it is extremely difficult and time-consuming for stroke patients to do lengthy training sessions before using a specific device, and thus, Joadder *et al.* [67] proposed a subject independent (SI) algorithm for MI EEG signal classification recently. Yet it is worth noting that the classification accuracy of this method is low, a large number of channels and training samples are also required as source subjects.

In this study, we also tried to explore the feasibility of our proposed MEWT based algorithm for SI EEG signal classification. The main building blocks of the SI algorithm are shown in Fig. 10. As seen, there are two main streams with source and target EEG data. In our study, the source subject data was used for training, while the data of target subjects was used as test data. The test data was rotated by using cross-validation, and among all the five subjects, any subject could be chosen to act as a target subject, while the other subjects could participate in the source domain.

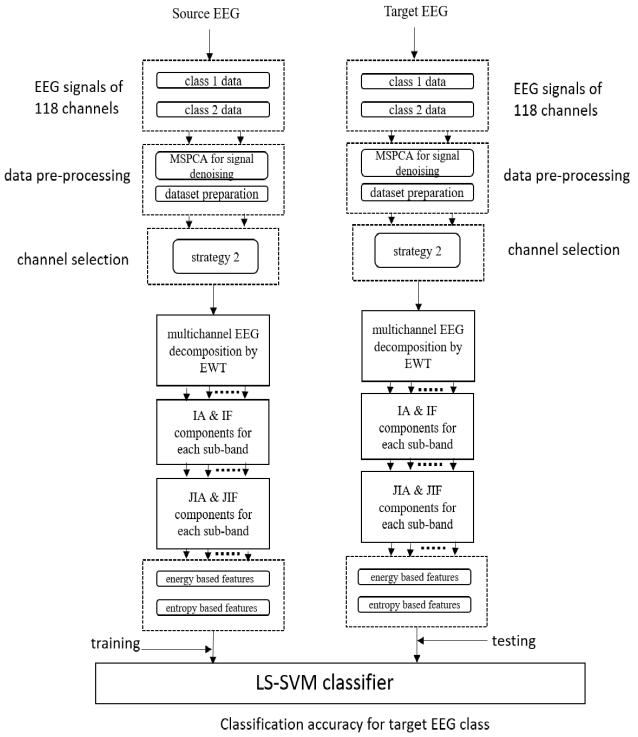


FIGURE 10. Block diagram of the proposed methodology for subject independent MI EEG signal classification.

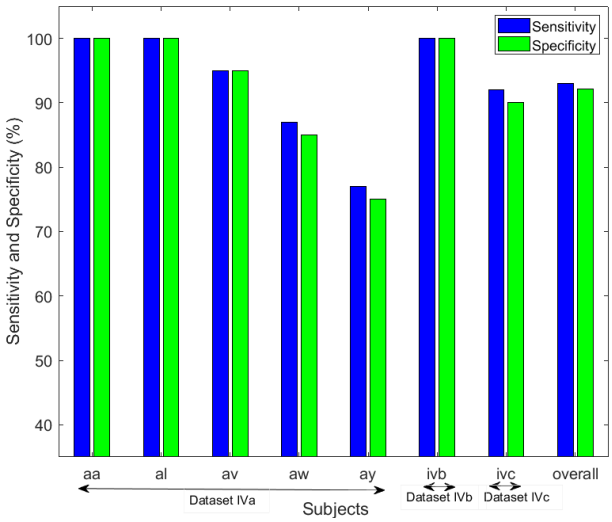


FIGURE 11. The sensitivity and specificity values for three datasets in subject independent case.

To show the effectiveness of our proposed SI MEWT based approach, we choose channel selection strategy 2, i.e., C_3, C_Z and C_4 channels and LS-SVM classifier were utilized for experimentation. From dataset IVa, we formed five sub-datasets, each of which has four source subjects in the source domain and one subject in the target domain. Since a subject who participated in the target domain was not involved in the source domain, each subject in the target domain corresponds to 15 different source domains.

As shown in Table 10, subjects "aa" and "al" in target domain achieved a highest classification accuracy of 100% with source domain subjects of "al" and "aa", respectively. The subject "av" in the target domain achieved a highest classification accuracy of 95% corresponding to the most of subjects in the source domain, especially with the single source domain subjects of "aa" and "al". Likewise, subjects "aw" and "ay" in the target domain obtained a classification accuracy of 85% and 70% with the source domain subjects to be "aa+al+av+ay" and "aa+al+av+aw" respectively. The overall classification accuracy for target domain is 90%, which concludes that the MEWT based approach could also be utilized for SI MI EEG signal classification.

2) EXPERIMENTAL RESULTS FOR DATASETS IVb AND IVc:

As seen in last two columns of Table 10, the classification results of datasets IVb and IVc for subject independent case are presented. We have considered dataset IVb and IVc subject as a target case and the different combination of subjects in dataset IVa was considered as a source subjects. As understood in Table 10, a classification accuracy of 100% and 90% was achieved for datasets IVb and IVc by considering several source combinations as shown in first column.

3) THE STATISTICAL ANALYSIS FOR SUBJECT INDEPENDENT CASE FOR THREE DATASETS

Like subject-specific case, we calculated the p values, sensitivity, specificity and AUC for different subject independent cases. Due to space constraints in following discussion, only best cases are consider. These best cases are already represented by bold text in Table 10. The p values for best cases of three datasets are shown in Table 11 and it could be observed that in most of the cases the p values are very small, indicating the significance of features utilized for those cases.

The sensitivity and specificity values for each subject in three datasets as well as average of such measures for all datasets are shown in Fig. 11. We have attained sensitivity values of 100%, 100%, 95%, 87%, 77%, 100% and 92% for subjects "aa", "al", "av", "aw" "ay", "ivb" and "ivc" by employing source subjects of "al", "aa", "aa", "aa+al+aw+ay", "aa+al+av+aw", "aa" and "aa+av+aw", and the overall sensitivity value is 93%. Similarly, the specificity values are 100%, 100%, 95%, 85%, 75%, 100% and 90% for subjects "aa", "al", "av", "aw" "ay", "ivb" and "ivc" by employing source subjects of "al", "aa", "aa", "aa+al+aw+ay", "aa+al+av+aw", "aa" and "aa+av+aw" with an average specificity value of 92.1% These results show that the variations among sensitivity and specificity for each subject are very small, and demonstrate that our proposed method is highly stable in detection of class 1 and class 2 for three datasets.

To show the classification capability of LS-SVM classifier for subject independent cases, the AUC are calculated and shown in Fig. 12. The AUC values of 1, 1, 0.95, 0.9,

TABLE 10. MI EEG signal classification accuracy (%) for different experiments in subject independent case.

	Target						
Source	"aa"	"al"	"av"	"aw"	"ay"	"Dataset IVb"	"Dataset IVc"
"aa"	—	**100**	**95**	65	60	**100**	**90**
"al"	**100**	—	95	60	60	100	90
"av"	65	70	—	60	60	65	60
"aw"	65	60	60	—	65	80	75
"ay "	75	60	75	65	—	50	60
"aa+al"	—	—	60	60	60	100	90
"aa+av"	—	60	—	50	55	100	90
"aa+aw"	—	60	95	—	60	95	80
"aa+ay"	—	60	95	60	—	65	65
"al+av"	55	—	—	65	55	90	80
"al+aw"	60	—	95	—	60	80	75
"al+ay"	60	—	95	60	—	80	75
"av+aw"	60	60	—	—	65	90	80
"av+ay"	60	60	—	70	—	90	80
"aw+ay"	60	60	95	—	—	90	80
"aa+al+av"	—	—	—	55	60	70	65
"aa+al+aw"	—	—	95	—	60	80	70
"aa+al+ay"	—	—	95	55	—	90	90
"aa+av+aw"	—	60	—	—	60	90	90
"aa+av+ay"	—	60	—	80	—	90	90
"aa+aw+ay"	—	60	95	—	—	85	80
"al+av+aw"	60	—	—	—	60	90	90
"al+av+ay"	60	—	—	65	—	90	90
"al+aw+ay"	65	—	95	—	—	95	90
"av+aw+ay"	60	60	—	—	—	65	55
"al+av+aw+ay"	70	—	—	—	—	100	90
"aa+av+aw+ay"	—	65	—	—	—	90	80
"aa+al+aw+ay"	—	—	95	—	—	80	90
"aa+al+av+ay"	—	—	—	**85**	—	95	90
"aa+al+av+aw"	—	—	—	—	**70**	100	90
"aa+al+av+aw+ay"	—	—	—	—	—	100	90

0.85,1 and 0.97 are obtained for subjects ''aa'', ''al'', ''av'', ''aw'', ''ay'', ''ivb'' and ''ivc'' where an overall AUC value for three datasets is 0.95, indicating that LS-SVM is a suitable classifier for subject independent case.

V. DISCUSSIONS

The purpose of this research is to design a robust and computationally effective algorithm to improve the accuracy of identification among various MI tasks. In this research, the following objectives are achieved for this intent.

1) The MSPCA is applied to remove noise from the EEG signals, which makes the proposed system robust against noise. As described earlier, MSPCA combines wavelets and PCA, instead of using all principal components, and the principal components were chosen by Kaiser rule to achieve the maximum classification accuracy. It is worth mentioning that, for careful selection of the appropriate method in the preprocessing module, we have also tested other traditional methods BPF, temporal filtering, and spatial filtering, and found that MSPCA provides the best results in our case.
2) The next finding of this study is to decode appropriate channels related to MI tasks, which helped us to achieve the maximum classification accuracy. For this purpose, three different channel selection strategies are proposed. Especially, for the first two-channel selection strategies, 18 and 3 respective channels are manually

TABLE 11. P values of features for three datasets features in subject independent case.

Source	Target	P Values
"aa"	"al"	3.5×10^{-04}
"al"	"aa"	3.5×10^{-04}
"aa"	"av"	0.0420
"aa+al+av+ay"	"aw"	0.080
"aa+al+av+aw"	"ay"	0.24
"aa"	"Dataset IVb"	2×10^{-04}
"aa"	"Dataset IVc"	1×10^{-03}

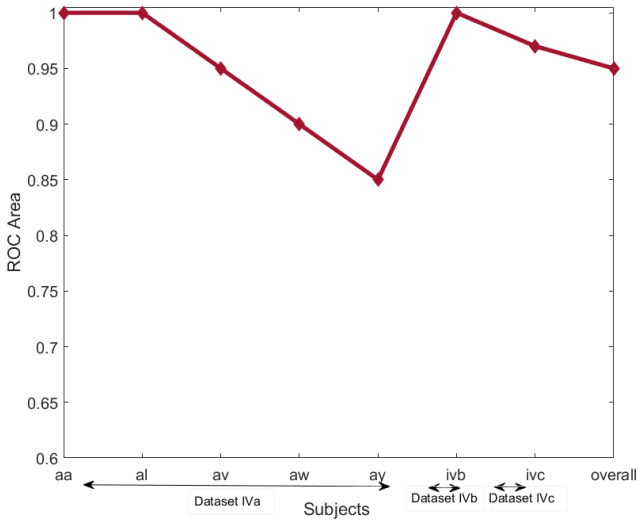


FIGURE 12. The area under the ROC curve for three datasets in subject independent case.

selected by using physiological knowledge, whereas for the third approach, an automated channel selection method is proposed to obtain an effective combination of channels, and 3 channels are selected for each subject in dataset IVa. All proposed channel selection strategies provide adequate results in the classification of various MI tasks.

3) We implement the multivariate extension of the EWT algorithm to evaluate the complicated, nonlinear and non-stationary nature of EEG signals. The MEWT results in different sub-bands for each EEG channel with the same indexed sub-bands are aligned in frequency scale among three channel selection strategies. It is important to mention that 10 subbands are chosen empirically for each channel for further analysis. This does not imply, however, that other sub-bands are not physiologically important for each channel. The instantaneous components of empirically selected aligned sub-bands are combined with different channels to obtain joint instantaneous components, where from each JIA component meaningful features are extracted for better classification accuracy for different MI tasks.
4) A robust feature selection scheme named CfsSubsetEval is employed to select a suitable feature set for each subject for three channel selection strategies. All these selected features are tested on five classifiers, and the best classifiers are highlighted.
5) The performances of our experiments are compared with those of the other fourteen methods in terms of classification accuracy with dataset IVa. As shown in Table 12, it is clear that our proposed method with automated channel selection strategy achieved an average classification accuracy of 97% by employing MLP, LS-SVM and LMT classifiers. Most importantly, a classification success rate of 100% was achieved for the subjects "aw" and "ay". As compared with the results obtained by using the cross-correlation and clustering based techniques and the LS-SVM classifier [5], [17], the accuracy is improved while the number of channels is largely reduced. Our results demonstrate that the classification success rate of the LS-SVM classifier can be improved by removing noise from EEG signals, and utilizing subject-specific channel combination with a robust feature extraction strategy.

Further observing in Table 12, we conclude that our proposed method ranked the number one in terms of overall classification accuracy for different MI tasks, while OA+NB method [19] ranked the second, CS+SVM method [64] ranked the third, CC+LS-SVM method [5] ranked the fourth whereas the SSFO based method [66] ranked the last. In addition, those studies reported in [19], [64], and [5], a number of 118, 33 and 118 channels were utilized respectively, which may largely limit the applicability of those methods due to high computational loads, noise and outliers. While in our study, only 3 channels have been utilized to achieve the highest classification accuracy as compared with other studies. Results summarized that our proposed method achieved a classification accuracy improvement of 0.64%-23.50% as compared with the other fourteen methods applied on dataset IVa for MI EEG signal classification.

In addition, we compared the proposed MEWT method with other methods implemented on datasets IVb and IVc for subject-specific case. Siuly *et. al* utilized dataset IVb in studies CC-LSSVM [5], CC-LR [44], OA [19], Modified CC-LR [35] and achieved 97.88%, 93.6%, 97.39% and 91.97% classification outcomes. Those studies were developed on 118 number of channels with 6,6,11 and 11 features respectively whereas in our proposed study we use only 3 channels with single feature to obtain 100% classification accuracy with dataset IVb. The proposed method yields classification improvement of up to 8.03% in comparison with other studies. For dataset IVc, few methods [72] and [73] are available in literature for comparison, and overall classification accuracy of 73.71% and 96.07% is obtained

TABLE 12. Comparison of subject-specific case in term of classification accuracy (%) with other works.

Papers by	Methods employed	Classification accuracy (%)					
		"aa"	"al"	"av"	"aw"	"ay"	Average
The proposed	MEWT +JIA+MLP	95	95	95	**100**	**100**	**97**
	MEWT +JIA+LS-SVM	95	95	95	**100**	**100**	**97**
	MEWT +JIA+LMT	95	95	95	**100**	**100**	**97**
Siuly *et al.* [19]	OA+NB	**97.92**	97.88	98.26	94.47	93.26	96.36
Ince *et al.* [65]	CS+SVM	95.60	99.70	90.50	98.40	95.70	96
Siuly *et al.* [5]	CC+LS-SVM	97.88	99.17	**98.75**	93.43	89.36	95.72
Wu *et al.* [18]	ISSPL	93.57	**100**	79.29	99.64	98.57	94.21
Kevric *et al.* [20]	MSPCA +WPD+HOS + k-NN	96	92.3	88.9	95.4	91.4	92.8
Siuly *et al.* [17]	Clustering with LS-SVM	92.63	84.99	90.77	86.50	86.73	88.32
Song *et al.* [47]	CSP with SVM	87.42	97.35	69.70	96.84	88.57	87.98
Lu *et al.* [16]	R-CSP through aggregation	76.8	98.2	74.5	92.2	77	83.74
Zhang *et al.* [66]	Z-LDA	77.70	**100**	68.4	99.60	59.9	81.1
Lotte *et al.* [15]	SSRCSP	70.54	96.43	53.57	71.88	75.39	73.56
	TRCSP	71.43	96.43	63.27	71.88	86.9	77.98
	WTRCSP	69.64	98.21	54.59	71.88	85.32	75.93
	WTRCSP	72.32	96.43	60.2	77.68	86.51	78.63
Yong *et al.* [67]	SSFO	57.5	86.9	54.4	84.4	84.3	73.50

whereas our proposed method with 3 channels and single feature results in 97% of classification accuracy.

6) There are quite a few studies on the subject independent experiments using dataset IVa available in the literature. In this study, we showed a comparison between our proposed MEWT based approach with those in studies [68]–[70], which adopted different variations of the CSP method to explore the inter-subject information for enhancing the MI EEG signal classification. Results in Fig. 13 show that, among those methods, the proposed SI MEWT based algorithm achieved an average classification accuracy of 90% and ranked the number one in terms of overall performances. Specifically, although authors in [67]–[71] had proposed algorithms to transfer inter-subject information, those methods involved learning from multiple source subjects to decode MI information of a single subject, and thus, data acquisition from multiple subjects is time-consuming. It is also restricted by the availability of the subjects. On the contrary, in our study, the MI information of a single subject is decoded by using a single source and multiple source subjects respectively. The overall classification accuracy of 100%, 100%, and 95% were achieved for subjects "aa", "al" and "av" by employing a single source subject only. Compared with studies [68]–[71], our proposed algorithm improves the average classification accuracy by up to 18.14%. It is also worth noting that a number of 68, 118, 22 and 118 channels were utilized in studies [68]–[71], whereas in our study, only 3 channels were utilized to achieve the best classification accuracy. Such results convincingly demonstrate that our proposed algorithm could be used for subject independent MI EEG signal classification. It is important to note that datasets IVb and IVc were not used for subject independent case in literature according to the best of our knowledge, so we are unable to provide comparison for subject independent case for those datasets.

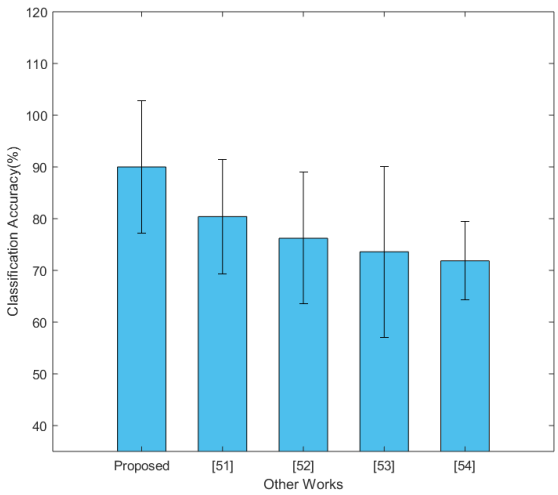


FIGURE 13. Comparison of subject independent case in term of classification accuracy (%) with other works.

Consequently, the advantages of the proposed subject specific method is to use three subject-specific channels with less number of features for each subject, which thus result in less computational time, making it suitable for online applications. Another interesting fact is that our algorithm

provides stable classification outcomes for each subject by using three different classifiers, whereas, in most of the studies mentioned in Table 12, only one classifier is utilized to achieve the maximum classification outcomes. It is also worth mentioning that since the time complexity of those methods listed in Table 12 is not available, we are unable to compare the time complexities among different algorithms. For subject independent case, our proposed algorithm provide highest classification accuracy by employing 3 channels and single-source subject only. In future an automated mode selection criteria will make this study more interesting.

VI. CONCLUSION

In summary, we proposed a simple and robust MEWT based algorithm for different MI EEG signal classification. With the MSPCA method adopted for noise removal, while the subject-specific automated channel selection method and the sub-bands alignment in frequency scale for feature extraction, the proposed method achieves higher classification accuracy with a reduced number of channels and features. Specifically, for subject-specific case, remarkable success rates for subjects with small training trials were obtained, while for subject independent cases, much higher classification accuracies were also achieved by employing only single-source subject. The simple classifiers and low computational load, and thus less processing time make our system a perfect candidate for online clinical applications, and our future work is to implement this method for online applications.

ACKNOWLEDGMENT

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

REFERENCES

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

[2] 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.

[3] 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.

[4] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni, "Non-invasive brain–computer interface system: Towards its application as assistive technology," *Brain Res. Bull.*, vol. 75, no. 6, pp. 796–803, 2008.

[5] 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.

[6] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: A review," *Sensors*, vol. 19, no. 5, p. 987, 2019.

[7] W. Xiaopei, B. Zhou, Z. Lv, and C. Zhang, "To explore the potentials of independent component analysis in brain-computer interface of motor imagery," *IEEE J. Biomed. Health Inform.*, to be published.

[8] E. Alickovic and A. Subasi, "Effect of multiscale PCA de-noising in ECG beat classification for diagnosis of cardiovascular diseases," *Circuits, Syst., Signal Process.*, vol. 34, no. 2, pp. 513–533, Feb. 2015.

[9] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.

[10] E. Gokgoz and A. Subasi, "Effect of multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders," *J. Med. Syst.*, vol. 38, no. 4, p. 31, 2014.

[11] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.

[12] 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, 2007.

[13] 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, 2012.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] S. Siuly, H. Wang, and Y. Zhang, "Detection of motor imagery EEG signals employing Naïve Bayes based learning process," *Measurement*, vol. 86, pp. 148–158, May 2016.

[20] 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.

[21] S. K. Bashar and M. I. H. Bhuiyan, "Classification of motor imagery movements using multivariate empirical mode decomposition and short time Fourier transform based hybrid method," *Eng. Sci. Technol., Int. J.*, vol. 19, no. 3, pp. 1457–1464, 2016.

[22] S. K. Bashar, A. B. Das, and M. I. H. Bhuiyan, "Motor imagery movements detection of EEG signals using statistical features in the dual tree complex wavelet transform domain," in *Proc. Int. Conf. Elect. Eng. Inf. Commun. Technol. (ICEEICT)*, May 2015, pp. 1–6.

[23] S. Taran, V. Bajaj, D. Sharma, S. Siuly, and A. Sengur, "Features based on analytic IMF for classifying motor imagery EEG signals in BCI applications," *Measurement*, vol. 116, pp. 68–76, 2018.

[24] S. Kumar, T. Reddy, and L. Behera, "EEG based motor imagery classification using instantaneous phase difference sequence," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 499–504.

[25] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 4, pp. 317–326, Aug. 2008.

[26] V. Gandhi, G. Prasad, D. Coyle, L. Behera, and T. M. McGinnity, "Quantum neural network-based EEG filtering for a brain–computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 278–288, Feb. 2014.

[27] T. K. Reddy and L. Behera, "Online eye state recognition from EEG data using deep architectures," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 712–717.

[28] T. K. Reddy, V. Arora, and L. Behera, "HJB-equation-based optimal learning scheme for neural networks with applications in brain–computer interface," *IEEE Trans. Emerg. Topics Comput. Intell.*, to be published.

[29] T. K. Reddy, V. Arora, S. Kumar, L. Behera, Y.-K. Wang, and C.-T. Lin, "Electroencephalogram based reaction time prediction with differential phase synchrony representations using co-operative multi-task deep neural networks," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 5, pp. 369–379, Oct. 2019.

[30] O. W. Samuel, X. Li, Y. Geng, P. Feng, S. Chen, and G. Li, "Motor imagery classification of upper limb movements based on spectral domain features of EEG patterns," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2976–2979.

[31] X. Li, O. W. Samuel, X. Zhang, H. Wang, P. Fang, and G. Li, "A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, 2017, Art. no. 2.
[32] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, 2015.
[33] X. Yu, P. Chum, and K.-B. Sim, "Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system," *Optik*, vol. 125, no. 3, pp. 1498–1502, Feb. 2014.
[34] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI competition 2003-data set IIb: Enhancing P300 wave detection using ICA-based subspace projections for BCI applications," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1067–1072, Jun. 2004.
[35] S. Siuly, Y. Li, and P. 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.
[36] O. W. Samuel, Y. Geng, X. Li, and G. Li, "Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors," *J. Med. Syst.*, vol. 41, no. 12, p. 194, Oct. 2017.
[37] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Comput. Biol. Med.*, vol. 112, Sep. 2019, Art. no. 103375.
[38] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, nos. 1–2, pp. 273–324, 1997.
[39] S. B. Hariz and Z. Elouedi, "Ranking-based feature selection method for dynamic belief clustering," in *Proc. Int. Conf. Adapt. Intell. Syst.* Springer, 2011, pp. 308–319. [Online]. Available: https://link.springer.com/chapter/10.1007%2F978-3-642-23857-4_31, doi: 10.1007/978-3-642-23857-4_31.
[40] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.
[41] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, 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.
[42] 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.
[43] M. T. Sadiq, X. Yu, Z. Yuan, Z. Fan, A. U. Rehman, G. Li, and G. Xiao, "Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform," *IEEE Access*, vol. 7, pp. 127678–127692, 2019.
[44] 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.
[45] J. Kevric and A. Subasi, "The effect of multiscale PCA de-noising in epileptic seizure detection," *J. Med. Syst.*, vol. 38, no. 10, p. 131, Oct. 2014.
[46] L. Song and J. Epps, "Classifying EEG for brain-computer interface: Learning optimal filters for dynamical system features," *Comput. Intell. Neurosci.*, vol. 2007, Apr. 2007, Art. no. 3. [Online]. Available: https://www.hindawi.com/journals/cin/2007/057180/abs/, doi: 10.1155/2007/57180.
[47] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61. Philadelphia, PA, USA: SIAM, 1992.
[48] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[49] S. Saleem, S. S. Naqvi, T. Manzoor, A. Saeed, N. U. Rehman, and J. Mirza, "A strategy for classification of 'vaginal vs. cesarean section' delivery: Bivariate empirical mode decomposition of cardiotocographic recordings," *Frontiers Physiol.*, vol. 10, p. 246, Mar. 2019.
[50] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.
[51] R. Sharma, R. Pachori, and U. Acharya, "An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures," *Entropy*, vol. 17, no. 12, pp. 5218–5240, Jul. 2015.
[52] N. Sriraam, K. Tamanna, L. Narayan, M. Khanum, S. Raghu, A. S. Hegde, and A. B. Kumar, "Multichannel EEG based inter-ictal seizures detection using teager energy with backpropagation neural network classifier," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 4, pp. 1047–1055, 2018.
[53] 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.
[54] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, and J. E. Koh, "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review," *Knowl. Based Syst.*, vol. 88, pp. 85–96, Nov. 2015.
[55] S. Raghu, N. Sriraam, and G. P. Kumar, "Classification of epileptic seizures using wavelet packet log energy and norm entropies with recurrent Elman neural network classifier," *Cogn. Neurodyn.*, vol. 11, no. 1, pp. 51–66, 2017.
[56] S. Aydin, H. M. Saraoğlu, and S. Kara, "Log energy entropy-based EEG classification with multilayer neural networks in seizure," *Ann. Biomed. Eng.*, vol. 37, no. 12, p. 2626, Dec. 2009.
[57] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.
[58] M. A. Hall, "Correlation-based feature selection for machine learning," Dept. Comput. Sci., Univ. Waikato, Hamilton, New Zealand, Tech. Rep., 1999. [Online]. Available: https://www.cs.waikato.ac.nz/~mhall/thesis.pdf
[59] 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.
[60] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Mach. Learn.*, vol. 59, no. 1, pp. 161–205, May 2005.
[61] W. Chen, H. Shahabi, A. Shirzadi, T. Li, C. Guo, H. Hong, W. Li, D. Pan, J. Hui, M. Ma, M. Xi, and B. B. Ahmad, "A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment," *Geocarto Int.*, vol. 33, no. 12, pp. 1398–1420, 2018.
[62] T. Fawcett, "An introduction to ROC analysis tom," Tech. Rep., 2005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S016786550500303X, doi: 10.1016/j.patrec.2005.10.010.
[63] S. Xavier-de-Souza, J. A. K. Suykens, J. Vandewalle, and D. Bollé, "Coupled simulated annealing," *IEEE Trans. Syst., Man, Cybern., B. Cybern.*, vol. 40, no. 2, pp. 320–335, Apr. 2010.
[64] 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.
[65] 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, 2013, Art. no. e74433.
[66] 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.
[67] M. A. M. Joadder, S. Siuly, E. Kabir, H. Wang, and Y. Zhang, "A new design of mental state classification for subject independent BCI systems," *IRBM*, vol. 40, no. 5, pp. 297–305, 2019.
[68] W. Samek, F. C. Meinecke, and K.-R. Müller, "Transferring subspaces between subjects in brain–computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2289–2298, Aug. 2013.
[69] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.
[70] D. Devlaminck, B. Wyns, M. Grosse-Wentrup, G. Otte, and P. Santens, "Multisubject learning for common spatial patterns in motor-imagery BCI," *Comput. Intell. Neurosci.*, vol. 2011, Jan. 2011, Art. no. 8. [Online]. Available: https://www.hindawi.com/journals/cin/2011/217987/, doi: 10.1155/2011/217987.
[71] A. Atyabi, M. Luerssen, S. P. Fitzgibbon, T. Lewis, and D. M. W. Powers, "Reducing training requirements through evolutionary based dimension reduction and subject transfer," *Neurocomputing*, vol. 224, pp. 19–36, Feb. 2017.
[72] C.-Y. Kee, S. G. Ponnambalam, and C.-K. Loo, "Binary and multi-class motor imagery using Renyi entropy for feature extraction," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2051–2062, 2016.
[73] Y. Shin, S. Lee, M. Ahn, H. Cho, S. C. Jun, and H.-N. Lee, "Simple adaptive sparse representation based classification schemes for EEG based brain–computer interface applications," *Comput. Biol. Med.*, vol. 66, pp. 29–38, Nov. 2015.

MUHAMMAD TARIQ SADIQ received the B.Sc. degree (Hons.) in electrical engineering from the COMSATS Institute of Information Technology (CIIT), Lahore, Pakistan, in 2009, and the M.Sc. degree from the Blekinge Institute of Technology (BTH), 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 (UET), Lahore, and a Lecturer with the University of South Asia. He is currently a Ph.D. Scholar with Northwestern Polytechnical University, Xi'an, China, and an Assistant Professor with the Electrical Engineering (EE) Department, The University of Lahore (UOL). His research interest includes biomedical signal analysis and classification. He was a Project Manager with SCET to manage final year student's projects, Patron of the IEEE-SCET Student Branch, and a Lifetime Member of Pakistan Engineering Council (PEC), Islamabad, Pakistan.

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

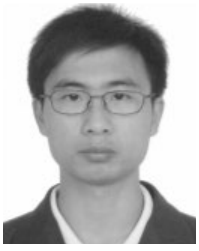
ZHAOHUI YUAN received the B.S., M.S., and Ph.D. degrees in control engineering from Northwestern Polytechnical University, China, in 1984, 1987, and 2005, respectively. He has been with Northwestern Polytechnical University, as a Teaching Assistant, a Lecturer, an Associate Professor, and a Professor, since 1987. He has coauthored more than 50 articles in technical journals and conferences. His 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 also a Fellow of the Shaanxi Association for Science and Technology.

FAN ZEMING received the Ph.D. degree in mechanical manufacturing and automation from the School of Mechanical and Electrical Engineering, Xi'an University of Technology. From November 2002 to October 2005, he worked as a Postdoctoral Research Fellow with Yongji Motor Factory. He is currently a Professor with Northwestern Polytechnical University, China. His main 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 (telecommunication) engineering from the COMSATS Institute of Information Technology (CIIT), Lahore, Pakistan, in 2009, and the M.S. degree in electrical engineering (telecommunications) from the Blekinge Institute of Technology (BTH), Karlskrona, Sweden, in 2011. He is currently pursuing the Ph.D. degree with the College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, China. Since 2016, he has been a Lecturer with Government College University (GCU), Lahore. His research interests include biomedical signal processing, the Internet of Things (IoTs), the social Internet of Things (SIoTs), and big data.

INAM ULLAH received the B.Sc. degree in electrical engineering (telecommunication) from the Department of Electrical Engineering, University of Science and Technology Bannu (USTB), Pakistan, in 2016, and the master's degree in information and communication engineering from the Department of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, China, in 2018, where he is currently pursuing the Ph.D. degree.

His research interests include wireless sensor networks (WSNs), underwater sensor networks (USNs), and edge/fog computing. He received the Best Student Award from the University of Science and Technology Bannu (USTB), in 2015, the Top-10 Students Award of the College of Internet of Things (IoT) Engineering, Hohai University, in 2019, and the Top-100 Students Award of Hohai University, in 2019.



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 papers. 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. He served/serves as an Associate Editor or a Guest Editor for the IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, *PLOS ONE*, and *Advances in Complex Systems*.