Expert Systems With Applications 213 (2023) 118901

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Unsupervised feature extraction with autoencoders for EEG based multiclass motor imagery BCI

Souvik Phadikar [a,*], Nidul Sinha [b], Rajdeep Ghosh [c]

[a] Neurology Department, University of Wisconsin-Madison, Madison, WI, USA
[b] Department of Electrical Engineering, National Institute of Technology Silchar, Silchar, Assam, India
[c] School of Computing Science and Engineering, VIT Bhopal University, Kotri Kalan, Madhya Pradesh 466114, India

ARTICLE INFO

Keywords:
Autoencoder
BCI
Electroencephalogram
EEG Signal Transformation
Feature Extraction
Motor Imagery
Wavelet Transform

ABSTRACT

Decoding of motor imagery (MI) from Electroencephalogram (EEG) is an important component of BCI system that helps motor-disabled people interact with the outside world via external devices. One of the main issues associated with the multiclass classification of MI based EEG is the informative confusion due to non-stationary characteristics of EEG data. In this work, an innovative idea of transforming EEG signal into a new domain, weight vector of autoencoder, unsupervised neural network, is proposed for the first time to solve that confusion. These weight vectors are optimized according to that particular EEG signal. The features: autoregressive coefficients (ARs), Shannon entropy (SE) and wavelet leader were extracted from the weight vector. A rectangular windowing-based feature extraction technique is implemented to capture the local features of the EEG data. Finally, extracted features were used in the support vector machine (SVM) as a classifier network. The proposed method is implemented on two openly available EEG dataset (BCI competition-III and Competition-IV) to validate the effectiveness and superiority of the proposed methodology over the newly reported methods. For four-class EEG based MI classification, the proposed technique has achieved an average test accuracy of 95.33% and 97% for dataset-IIIa from BCI-III and dataset-IIa from BCI-IV respectively. The experimental results reveal that, the proposed technique is a promising solution to improve the decoding performance of BCIs.

1. Introduction

BCI creates a communication pathway between neuronal system and external devices to decode the intent of operator into computer instruction by distinguishing a task associated with neuronal activity (Das et al., 2016). Several techniques are used to capture the neuronal activity inside the brain. EEG is one of them and widely used due to its non-invasive nature and high temporal resolution. It captures the electrochemical fluctuations inside the brain by means of electrodes on the scalp. The MI classification through EEG is one of widely used BCI applications. The cerebral activities of MI can be triggered when a person imagines any movement of his body parts. If these cerebral activities are properly translated, then the findings can be used to interact with external devices such as BCI based wheel chair for physically challenged patients and service robot for several motor neuron diseases (e.g. Poliomyelitis, Parkinson disease, etc.) (Ang & Guan, 2017; Sun et al., 2019). The EEG pattern identification, therefore, plays a key role in the applications of BCIs for MI.

In the MI based BCI systems, initially features are extracted from the EEG signal. These features are then integrated into a feature array, and these arrays are further used to train the machine learning models. The performance of these machine learning models depends on two major factors (i) discriminative feature extraction and (ii) training algorithms. The best performance will be achieved if these features are discriminating for each class of the data. However, in EEG based BCI system, three main challenges such as: artifacts, non-stationarity and mislabelling of training feature sets can degrade the performance of the systems. Eye blink and facial muscular activities, considered as biological artifacts, are unavoidable contaminations that are measured together with neural activities and thereby distort valuable information (Jafarifarmand et al., 2017). The variations seen in different neurophysiological conditions of subjects in terms of degree of alertness while the signal recording sessions are occurring can cause the immense non-stationarity in the EEG signals (Krauledat, 2008). The improper imagination of mental tasks and inappropriate class lsabelling result in deviation in the training data. Various intelligent methods have been

* Corresponding author.
E-mail addresses: phadikar@wisc.edu (S. Phadikar), nidulsinha@ee.nits.ac.in (N. Sinha), rajdeep.ghosh@vitbhopal.ac.in (R. Ghosh).

https://doi.org/10.1016/j.eswa.2022.118901
Received 9 August 2020; Received in revised form 26 August 2022; Accepted 22 September 2022
Available online 26 September 2022

Table 1
Different feature extraction technique used in EEG signals for MI classification.

References	Feature Extraction Technique	Classifier	Dataset Used	Mean Accuracy* (%)
Ang et al., (2012)	CSP Features	Naïve Bayesian Parzen Window (NBPW)	BCI-IV (IIa)	71.73
Kam et al., (2013)	Logarithmic values of normalized variances using non-homogeneous spatial filter	LDA	BCI-IV (IIa)	60
Aghaei et al. (2016)	Spatio-Spectral features using Separable Common Spatio-spectral Patterns (SCSSP)	Linear minimum mean distance classifier	BCI-IV (IIa)	65.4
Rathee et al., (2017)	Spectral features using filter bank common spatial pattern (FBCSP)	SVM	BCI-IV (IIa)	58.14
Jafarifarmand et al., (2017)	Artifact Rejected CSP (AR-CSP)	Adaptive resonance theory (ART) based neuro-fuzzy classifier	BCI-IV (IIa)	85.49
Dose et al., (2018)	Principle component analysis (PCA) with convolutional neural network (CNN)	CNN	The Physionet dataset (Goldberger et al., 2000)	68.51
Amin et al., (2019)	Spectral and temporal features using multi-layer CNN (MCNN)	CNN	BCI-IV (IIa)	75.7
Ai et al., (2019)	CSP and local characteristics scale decomposition (LCD)	Spectral regression discriminant analysis (SRDA)	BCI-IV (IIa)	79.7
Zhang et al., (2019)	One-versus-rest FBCSP	Long short-term memory (LSTM)	BCI-IV (IIa)	83
Sreeja & Samanta, (2019)	Wavelet energy	Sparsity based classifier	BCI-III (IIIa)	91.84
Jafarifarmand & Badamchizadeh, (2020)	Modified CSP with joint approximate diagonalization (JAD)	ART based neuro-fuzzy classifier	BCI-IV (IIa)	62.66
Ma et al., (2020)	Power spectral density (PSD)	Visual geometric based CNN	BCI-IV (IIa)	96.26
Wang et al., (2020)	CSP + Auto regressive	Naïve Bayesian Classifier (NBC)	BCI-IV (IIa)	86.01
Fadel et al., (2020)	Azimuthal equidistant projection and Clough-Tocher algorithm	CNN	The Physionet dataset	70.64
Wu et al., (2021)	Tangent Space Mapping (TSM)	SVM	BCI-IV (IIa)	77.33
Hou et al., 2022	Bispectrum, Entropy and CSP (BECSP)	SVM	BCI-IV (IIa)	71.61
Chen et al., 2022	Filter bank channel group attention (FB-CGANet)	Neural Network	BCI-IV (IIa)	79.4

* All the results in the table are presented from the original paper.

developed in the literature to deal with the EEG artifacts (Ghosh et al., 2019; Phadikar et al., 2020a; Phadikar et al., 2020b). The mislabelling features can be figure out with the help of expert EEG specialist and carefulness during recording and labelling the features. However, to deal with the non-stationarity of EEG signals is a challenging task in EEG pattern recognition. Several strategies such as transforming the EEG signals into frequency domain, time–frequency domain etc. are employed in feature extraction stage to deal with challenges in EEG based motor imagery recognition.

As this paper focuses on multiclass MI classification, several research articles on multiclass MI classification have been surveyed and presented in Table 1. However, among all the methods, the common spatial pattern (CSP) is most frequently used as a technique of feature extraction for MI classification. Initially it was developed for two class data classification, and then modified for multiclass classification problem. Basically, in CSP, the data of two classes are spatially filtered to maximize the difference in variance between two classes. Before implementing the CSP to decode the MI, EEG signals are bandpass filtered between a wide frequency band (4 to 40 Hz). The performance of CSP is highly depending on the selection of frequency band. It is therefore necessary to select a frequency band or a wide frequency band for the specific characteristics, but this process is very inconvenient. If the frequency band is incorrectly used, then the performance of BCI system degrades. To overcome this problem, several extensions of CSP have been proposed in the literature such as filter bank common spatial pattern (FBCSP). In FBCSP, the EEG (4 to 40 Hz) is divided into multiple smaller frequency bands, and then CSP is applied for feature extraction (Rathee et al., 2017). Recently, various methods are employed to transform the EEG signals into images to extract the prominent features for MI tasks (Fadel et al., 2020). The recordings of all the electrodes are converted into 2-D images, thus the problem of EEG signal classification becomes an image classification task.

From the previous research, it is observed that the CSP and transform were widely used as feature extractors in the four-class MI classification. The commonly used CSP algorithm, however, recognizes only spatial-based features while paying no attention to the spectral properties of EEG signals. The performance depends on the spectral filter for which the frequency band is usually predetermined and fixed manually. In AAR-based approaches, time information is ignored. In wavelet transform based techniques, considering coefficient vectors at different level as a feature of EEG signal increase the computational complexity of the system. However, challenges remain in achieving higher classification accuracy to improve the performance of real-time MI-based BCI.

It is quite difficult to decode four-class MI tasks from the EEG data because of its non-stationary nature. Hence, the derived features may not be entirely discriminative. To address the challenge of finding discriminatory features for four-class MI classification because of highly random and non-stationary EEG signal, an innovative idea of transforming the EEG signal in a new domain i.e., weight vector of unsupervised neural network has been proposed in this paper for four-class MI classification for the first time. An efficient feature extraction method is developed rather than improving the classification algorithm to enhance the performance of BCI.

In view of the above, the main features of the innovative method proposed in this paper are:

a) It is a fully automated, unsupervised, and data-driven feature extraction method using individual autoencoders for each EEG channels, which does not require any help of experts or prior knowledge.
b) The proposed novel feature extraction method can adaptively capture the intention of motor movements from the EEG data.
c) It extracts discriminative feature sets in new domain in which the classifier can achieve higher classification accuracy for four-class MI data.

A new hypothesis is proposed in this paper as: when an EEG signal is fed to train the neural network, is it possible to represent the EEG signal in terms of their weight vector from the input layer to hidden layer of a neural network? Because, the weight vectors of the neural network are optimized according to the input data fed to the network. Hence, the input signal can be represented as optimized weight vectors of the

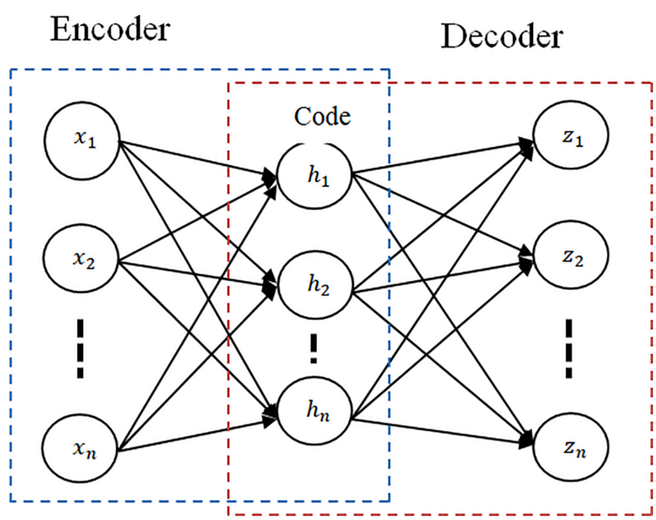


Fig. 1. The architecture of autoencoder. X_i, h_i and Z_i denotes the input node, hidden nodes and output nodes respectively.

network. If it is possible then that the feature values can be extracted from a particular weight vector characterizing a particular EEG signal in new domain.

To test the hypothesis, an autoencoder neural network is used to find the unique weight vector for each MI EEG signal. The autoencoder has been used in this paper as it produces the same output as the input. Hence, it is a data driven network. As a result, the weight vectors are updated according to EEG data. Once the weight vectors are computed, then spectral features are extracted through a slid-windowing method using rectangular window function. The windowing-based feature extraction technique is implemented to capture the local features of EEG data. The window size is selected through proper technique. Then all the windowed feature values are concatenated to form a feature vector. Finally, these feature vectors along with feature labels are fed to the SVM network as a classifier. Once the training of classifier network is completed, test EEG signals are used to test the network and fed to autoencoder to compute its unique weight vector. Now the feature values are extracted and predicted through the previously trained SVM network. This method is fully automatic and unsupervised. The proposed system is compared with the conventional MI based BCI systems (Jafarifarmand et al., 2017; Ai et al., 2019; Zhang et al., 2019; Sreeja & Samanta, 2019; Ma et al., 2020; Wang et al., 2020; Wu et al., 2021; Chen et al., 2022) and also achieved the highest classification accuracy. The proposed system is validated with publicly available two datasets. Same feature sets are also extracted from the original EEG data (without transforming the EEG) and the performance is compared with the proposed method. The proposed system does not use any channel selection method for achieving higher accuracy than the conventional methods. To the best of our knowledge, this is the maiden attempt of feature extraction methods for the EEG based MI classification. The benefits of the stated model are established through the exploratory outcomes as explained in this paper.

2. Materials and tools

2.1. Autoencoder

An autoencoder is a neural network, based on unsupervised machine learning algorithm that captures the signature inside the EEG data (Liou et al., 2014). It mainly has two parts: encoder that maps the input into a code, and a decoder that maps the code to reconstruct the original input. Weights between layers are updated according to the training algorithm to minimize the reconstruction error. As the reconstruction error is small enough, the code could be assumed to incorporate most of the information of the input vector. The architecture of autoencoder is shown in Fig. 1.

By taking $\times$ as input vector, autoencoder maps the input vector to a hidden layer representation y with deterministic mapping as:

$$y = f_{encoder}\left(W_e^T x + b_e\right) \tag{1}$$

where, W_e and b_e is the weight and bias vector respectively of the encoder and $f_{encoder}$ is the activation function of the neurons in the encoder. The output of the autoencoder z, has the same phase as x, is then extracted by mapping the hidden layer representation or code y using the transformation.

$$z = f_{decoder}\left(W_d^T y + b_d\right) \tag{2}$$

where W_d and b_d are characterized as the weight vector and bias vector respectively in the decoder. $f_{decoder}$ is the activation function of neurons in the decoder. In autoencoder network, training can be done by minimizing the reconstruction error which is measured as squared error. The autoencoder finds the appropriate parameters $\theta = \{W_e, W_d, b_e, b_d\}$ through minimizing the cost function.

$$E(\theta) = L(x,z) + \lambda\|W\|^2$$

$$= \sum_i^n \|x_i - z_i\|^2 + \lambda(\|W_e\|^2 + \|W_d\|^2) \tag{3}$$

where, $\lambda\|W\|^2$ is a regularized parameter to avoid the over-fitting by minimizing the L2 norm of parameters, and $L(x,z)$ is the reconstruction error. The weight vector $W = W_e(:)$ can represent the MI signal in new space.

2.2. Feature extraction

Although several feature extraction methods have been developed for MI EEG classification, selecting a suitable method for selection of features for higher classification accuracy is still a thought-provoking task for effective EEG classification. Finding a good feature set for a binary classification task may not be so difficult but for a complex multi task classification, getting discriminative feature set is really a challenging task. In the proposed work, four feature values are extracted from the weight vector for the corresponding EEG data.

2.2.1. Autoregressive (AR) coefficients

The AR model is a representation of a type of random process in signal processing; as such, it is used to characterize such time-varying processes (Neumaier & Schneider, 2001). In parametric method, it estimates the PSD of an EEG. Therefore, there are no chances of spectral leakage. Hence, it yields better frequency resolution. The PSD is estimated by measuring the parameters of the linear systems under consideration. The signal can be represented as a linear combinations of p previous values of the same signal. The signal x[n] at the time instant n can be modelled as:

$$x[n] = -\sum_{i=1}^{p} a[i]x[n-i] + e[n] \tag{4}$$

where, e[n] is a zero-mean white noise, $a[i]$ is the i^{th} coefficient of the model with order p. A total p number of AR coefficients are used as feature values in this work. However, the selection of order p is very sensitive because, the estimates generally improve with increase in order but at higher computational cost.

- Burg's Method to Estimate the AR Model

Several strategies have been proposed for the estimation of AR models. The Burg's method is widely used in EEG based classification of mental state. It directly measures the reflection coefficients without using the autocorrelation function (Al-Fahoum & Al-Fraihat, 2014). This approach estimates data records of PSD that exactly look like the original. For a detail description, interested readers may refer to (Stoica & Moses, 2005).

- Determination of AR Model Order

While designing an AR model, it is important to define the order of the model which best fits the data. It depends on the data sampling rate because, the AR model estimates the present value of data using the number of past data samples. Sum-squared error (SSE) is a widely used tool for determining the order of an AR model. The lower SSE indicates the order of the model that best fits the data (Anderson et al., 1998). As suggested by Anderson et al. (1998), for EEG based mental state classification, AR coefficients of order 6 best fit the data.

2.2.2. Wavelet packet entropy

The wavelet transform is widely used in feature extraction due to its capability of capturing the local features of an EEG signal. It is difficult to use such coefficients directly as features because of its wider length. Hence, for better classification, some higher-level features may be extracted from these coefficients. Entropy is a tool to capture the uncertainty of a given system and commonly used in information theory and signal processing (Li & Zhou, 2016). In this work, Shannon entropy (SE) is directly calculated from the wavelet packet decomposition (WPD) coefficients of the weight vector. The WPD is an extension of discrete wavelet transform (DWT). The key difference between WPD and DWT is that, it decomposes not only the approximation coefficients but also the detail coefficients simultaneously. As a result, the WPD has same frequency bandwidth in each resolution while DWT does not. For an EEG signal *x(t)*, the coefficients can be derived as:

$$\begin{cases} d_{0,0}(t) = x(t), \\ d_{i,2j-1}(t) = \sqrt{2}\sum_{k} h(k) d_{i-1,j}(2t-k), \\ d_{i,2j}(t) = \sqrt{2}\sum_{k} g(k) d_{i-1,j}(2t-k) \end{cases} \tag{5}$$

where $h(k)$ and $g(k)$ denote high pass and low pass filters respectively, and $d_{i,j}$ is the WPD coefficients at the i^{th} level and j^{th} node. The energy at i^{th} level and j^{th} node can be derived by wavelet-energy, defined as:

$$E_{i,j} = \sum_{k=1}^{N} \|d_{i,j,k}\|^2 \tag{6}$$

where, N denotes total number of coefficients in the corresponding node. The SE of j^{th} node at i^{th} level is calculated based on the probability distribution of energy as:

$$SE_{i,j} = -\sum_{k=1}^{N} P_{i,j,k} * \log(P_{i,j,k}) \tag{7}$$

where $P_{i,j,k}$ is the probability of the k^{th} coefficient at its corresponding node and is defined as:

$$P_{i,j,k} = \frac{\|d_{i,j,k}\|^2}{E_{i,j}} \tag{8}$$

Finally, the SE feature vector is computed by cascading all the SEs from every node of level M.

$$SE = (SE_{i,1}, SE_{i,2}, \cdots, SE_{i,2^M})_{i=M} \tag{9}$$

- Selection of Base Wavelet

The selection of appropriate base wavelet (mother wavelet) may affect the calculation of SE feature vector in wavelet domain. Hence, a cross-correlation based approach is proposed to check the performance of all the available wavelet bases for EEG based MI signal classification. The cross correlation between MI signal and the wavelet functions are calculated and the function is selected which gives the maximum value. The correlation X_{corr} between the EEG signal of interest X and the mother wavelet function Y is.

$$X_{corr} = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2(Y-\overline{Y})^2}} \tag{10}$$

2.2.3. Wavelet fractal estimates

Two fractal parameters from DWT coefficients are estimated and used as features. The width of singularity spectrum and the second cumulant of the scaling exponents are obtained as features. The width of singularity spectrum derives the multi-fractal nature of the EEG signal. The scaling exponents are scale based exponents describing power-law behaviour in the signal at different resolution. The second cumulant broadly represents the departure of the scaling exponents from linearity (Leonarduzzi et al., 2010). Both the features are calculated from wavelet leaders. The wavelet leaders estimate the multifractal spectrum based on wavelet transform.

Let ψ be a wavelet function having various null moments and fast decay and dilated by scale 2^j and translated to time position $2^j k$. It can be assumed that, each wavelet coefficient C_{jk} corresponding to the wavelet transform of the series *{x(i)}* is localized on the dyadic interval (Leonarduzzi et al., 2010), $I_{jk} = \left[\frac{k}{2^j}, \frac{k+1}{2^j}\right]$. Then the dilated intervals can be computed as:

$$3I_{jk} = \left[\frac{k-1}{2^j}, \frac{k+2}{2^j}\right] \tag{11}$$

The wavelet leaders d_{jk} are computed as:

$$d_{jk} = \sup\{|C_{lh}| : I_{lh} \subset 3I_{jk}\} \tag{12}$$

The most important key factor about the wavelet leader is the search for the greatest wavelet coefficients in a narrow time neighbourhood for a given time and scale (Leonarduzzi et al., 2010). The singularity spectrum (SS) determines how many singularities are there. However, the SS can be easily computed from the structure function (SF). The SF is computed from the wavelet leader as:

$$S(q, 2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} |d_{jk}|^q \tag{13}$$

If the signal *{x(i)}* shows some form of self-similarity, the SF decays as power laws of the scales. The exponents of these power laws are called scaling exponents (SE) and are computed as:

$$SE(q) = \liminf_{j\to 0}\left(\frac{\log_2(S(q,2^j))}{j}\right) \tag{14}$$

Finally, the SS can be obtained from the SE via Legendre transform (LT) as:

$$D(h) = {}_{q}^{\inf}(1 + qh - SE(h)) \tag{15}$$

The width of the SS is measured as the difference between maximum and minimum value in the D(h) and the second cumulant of the SE are used as feature values.

All the feature values are extracted from the autoencoder weight vector using non-overlapping slid windowing technique using rectangular window function. Here, p number of AR coefficients, 2^M number of Shannon entropy and two wavelet fractal estimates are extracted from each window. Finally, $(p + 2^M + 2) * S$ number of feature values are concatenated to form the feature vector, where S is the total number of rectangular windows to cover the whole weight vector.

2.3. Multiclass SVM classifier

The SVM is the most widely used classifier that utilizes the strategy of supervised machine learning. SVM can competently categorize non-linear data by means of kernel trick. The SVM uses training data to create an optimal hyper plane, with the aim to classify test data (Hsu & Lin, 2002). Some unique features of different datasets are obtained and provided to the classifier. The optimal hyper plane, known as support vectors, is made to get a decision boundary from the adjacent samples of different datasets. If the datasets are linearly indistinguishable in the original finite dimensional, then the data can be re-mapped into sufficiently higher dimensional space to reduce the non-linearity. To address the higher dimensionality, kernel trick is used for better classification and less effort. However, SVMs were originally designed for two-class classification problem. Some extensions of SVM have been proposed in the literature for multiclass classification problem. These extensions are 1-against-1, 1-against-all, DAGSVM etc. (Hsu & Lin, 2002).

2.4. Datasets

In this investigation, two datasets were used: one for validation and another for comparison with other recently reported methods of EEG based MI Classification.

2.4.1. Dataset IIIa from BCI competition III

Dataset IIIa from BCI competition III (Blankertz et al., 2006) has been used to validate the proposed system. The data consist of recording from the three subjects (k3b, k6b, I1b). The subjects performed four MI tasks according to a cue. The subjects were imagining the movements of left hand (class-1), right hand (class-2), tongue (class-3), and foot (class-4) while relaxing in a chair with armrests. The experiment consisted of several runs with 40 trials each. After the trial began, the 2 s were silent, an acoustic stimulus started at $t = 2$ s to indicate the beginning of the trial, and a cross "+" was displayed. Then, from $t = 3$ s, an arrow to the left, right, up or down was displayed for 1 s and at the same time, and the subjects were asked to imagine a movement of left hand, right hand, tongue or foot according to the arrow until the cross disappeared at $t = 7$ s. Each of the four cues is displayed ten times within each run in a randomized order.

2.4.2. Dataset IIa from BCI competition IV

To demonstrate the superiority of the proposed model over other recently developed methods, the dataset IIa from BCI competition IV (Tangermann et al., 2012) has been adapted. This dataset consists of EEG data from 9 subjects (A01 – A09). The subjects were imagining the movements of left hand (class 1), right hand (class 2), tongue (class 3), and foot (class 4) while relaxing in a chair with armrests. The recorded data consisted of 22-channel EEG signals. The data were sampled at 250 Hz.

2.5. Performance metrics

The following parameters are used in this paper to evaluate the performance of the proposed methodology for four class MI EEG classification. The metrics are used after separating the dataset into training and test dataset through holdout cross validation (CV) technique. The CV partition usually divides the datasets into training and test (holdout) set. This partition depends on a scalar r, called holdout parameter. When $0 < r < 1$, this approach randomly selects r*n observations for the test set, where n is the total number of observations. When r is integer, this method randomly selects r observations for the test set.

$$Accuracy(Acc) = \frac{TP + TN}{P + N} \quad (16)$$

$$Sensitivity(Sen) = \frac{TP}{P} \quad (17)$$

$$Specificity(Spe) = \frac{TN}{N} \quad (18)$$

$$precision = \frac{TP}{TP + FP} \quad (19)$$

where, TP (true positive) denotes the number of correctly predicted positive MI class and TN (true negative) represents the number of correctly predicted negative MI class. P denotes the total number of positive class present in the observation and N denotes the total number of negative class present in the observation.

3. Proposed methodology

This paper proposes a novel method of unsupervised feature extraction for EEG based MI classification. The basic architecture of the proposed system is presented in Fig. 2. At first, the MI related EEG signal is fed to an autoencoder and the weight vector which minimizes the reconstruction error in the autoencoder is extracted. Then the $S * p$ number of AR coefficients, $S * 2^M$ number of wavelet packet entropy and $2 * S$ number of wavelet fractal estimates are extracted from the weight vector of corresponding MI data using a non-overlapping rectangular window. Finally, the feature-label pairs are fed to a multiclass SVM classifier to train the model. The proposed method is implemented in MATLAB R2019a and evaluated for BCI-III and BCI-IV dataset. The proposed work achieves a higher classification accuracy as compared to the conventional methods (Jafarifarmand et al., 2017; Ai et al., 2019; Zhang et al., 2019; Sreeja & Samanta, 2019; Ma et al., 2020; Wang et al., 2020; Wu et al., 2021; Chen et al., 2022) reported for EEG based MI classification.

Stages of the proposed model.

- Four class MI data were segmented from the dataset and a new dataset were created.
- The EEG data from each class from the new dataset are fed to the autoencoder. After minimizing the training error of the network, the weight vector from input-to-hidden-layer is extracted. This weight vector represents the corresponding EEG data from each class.
- The weight vectors (representing the original EEG signals) were partitioned into a training set and test set using the CV partition technique with holdout parameter $p = 0.3$. 70 % of the data were selected randomly for the training purpose and 30 % of the data were selected randomly for the testing purpose.
- A feature extraction window of size 1 s (250 samples) is slid forward in each weight vector (training set) to extract the above-mentioned feature values, i.e. AR coefficients, wavelet packet entropy, wavelet fractal estimates. Hence, total $(p + 2^M + 2) * S$ number of feature values are concatenated to form the feature vector, where S is the total number of windows to cover the whole weight vector.

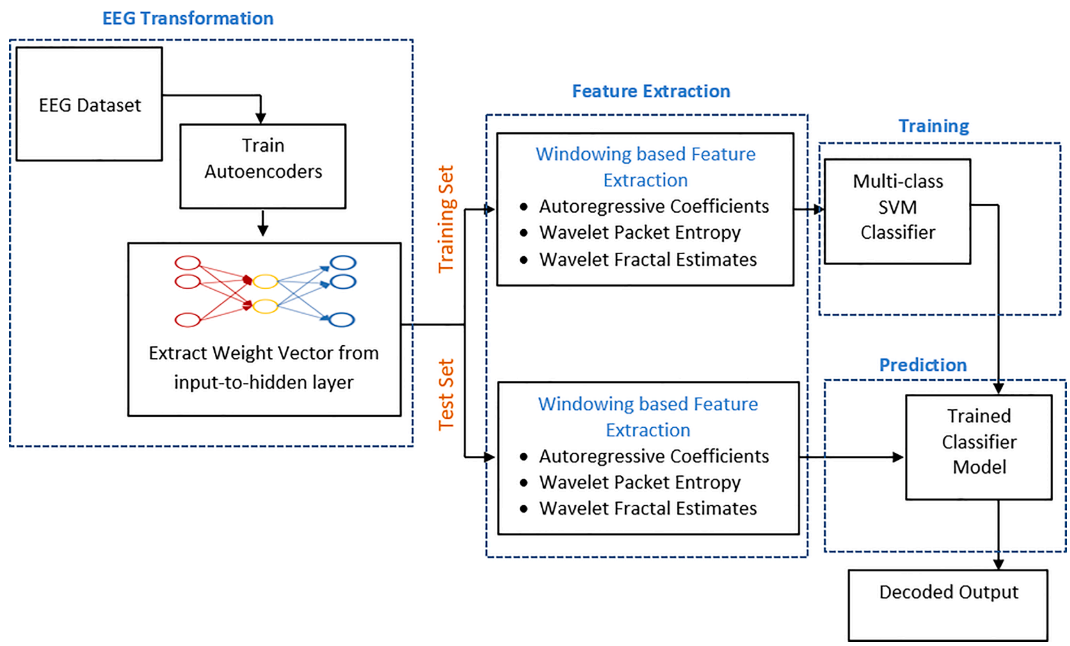


Fig. 2. The basic architecture of proposed system.

Likewise, test feature sets were also extracted using the same method.

- Finally, the training feature-label pairs were fed to a multiclass SVM to train the classifier network.
- To test the trained classifier network, test feature sets are used and the performance is evaluated.

3.1. Data preprocessing

The data were recorded in 6 runs (sessions) separated by short breaks for each subject. One run consists of 40 trials (10 for each of the four possible classes), yielding a total of 240 trials per session. In the proposed method, single trial for each of the four possible classes from 6 runs (sessions) was used to create the main dataset for the proposed algorithm. Hence, $4 \times 6 = 24$ trials (out of 240 trials) were extracted from each subject. The data were recorded from the three subjects, as a result $3 \times 24 = 72$ trials were extracted and the main dataset were created.

The EEG data is very sensitive to various artifacts such as eyeblink, muscle etc. These artifacts should be removed otherwise result in misclassification. In this paper, the eyeblink artifacts are removed through the method described in (Phadikar et al., 2020a). At first the identified corrupted EEG was decomposed into wavelet coefficients up to level 6 using Daubechies wavelet (vanishing moment as 8). Both the decomposition level and wavelet function were selected through the techniques described in Phadikar et al., (2020a). Then the approximate coefficients (ACs) were thresholded in backward manner using the optimum threshold values followed by inverse wavelet transform (IDWT) to reconstruct the original EEG signal. The AC at level 6 was thresholded and used in IDWT together with the un-thresholded detail coefficient (DC) at the same level (i.e., level 6) to get back the AC at level 5. Then, AC at level 5 was thresholded using different threshold value and used in IDWT with DC at the same level (i.e., level 5) to get back the AC at level 4. Likewise, the backward thresholding of the ACs followed by IDWT is continued till the artifact free EEG signal is reconstructed at level 1. The optimum threshold values for different level were calculated through grey wolf optimizer (GWO). Because for simultaneous optimization of number of parameters (here thresholds at different levels) *meta*-heuristic optimizers like GWO is the best option. The muscle artifacts are removed through the method described in Phadikar et al. (2022). After successfully removing the artifacts from the created dataset, the MI related EEG data were fed to the autoencoder for the following steps.

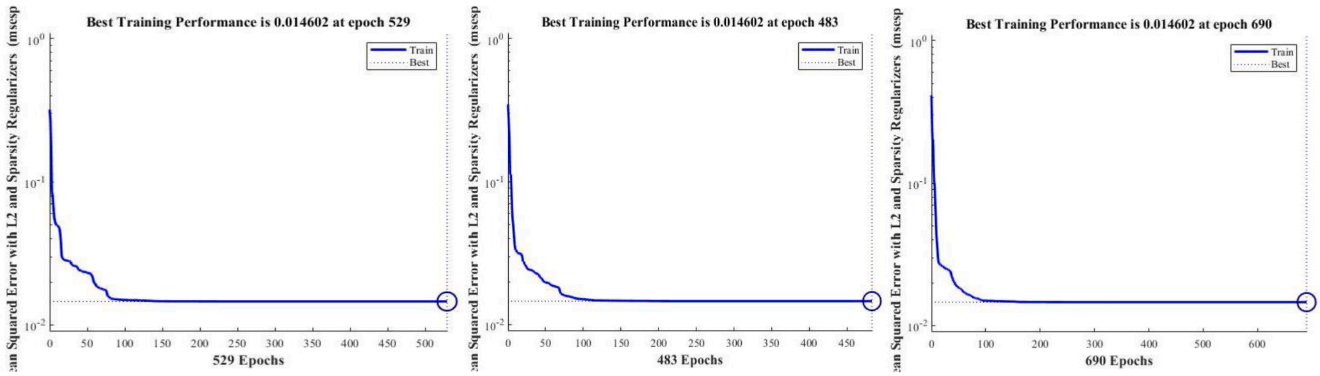
3.2. Extraction of weight vectors from the autoencoder

Before partitioning the data into training and test dataset, the main dataset (containing 72 trials) were fed to the autoencoder. Individual autoencoder was trained for particular EEG signals (for example: if there are 10 EEGs, ten individual autoencoders were trained). When all the autoencoders were trained, their optimized weight vectors corresponding to the particular EEG signals were extracted. Hence, the proposed technique transformed the EEG signals and represented them in a new domain/space (i.e., weight vector). Hence, the autoencoder acted as a transformer (not as a classifier). The activation function of encoder network is selected as log-sigmoid function. For the given input, X_i, the encoder output will be y_i,

$$y_i = logsig(W_i X_i + b_e) \tag{20}$$

$$logsig(z) = \frac{1}{1+e^{-z}} \tag{21}$$

To minimize the error defined in equation (3), scaled conjugate gradient (SCG) algorithm (Møller, 1993) is used as a training algorithm in autoencoder. The weight vectors are selected, when the autoencoder reconstruction error is minimized. The number of hidden nodes in the hidden layer were selected using trial and error method. However, this EEG segment is linearly separable or not, is not known. Hence the number of hidden nodes is selected as 30, 50 and 100 and the training errors of the autoencoder are compared in the Fig. 3. From the figure, it is evident that, the error is the minimum with fast convergence when the number of hidden nodes is 50. After extracting all the weight vectors



(a) Training error plot for N=30 (b) Training error plot for N = 50 (c) Training error plot for N = 100

Fig. 3. Performance of the autoencoder with different configuration.

(W_i) from 72 trials, the weight vectors are separated into training and test set using the holdout CV algorithm. In the CV partition algorithm, the holdout parameter, p is selected as 0.3 and as a result, 70 % of the total weight vectors were randomly chosen to form train set and 30 % weight vectors were kept as test set. Finally, the data were paired with data labels (four motor imagery tasks) and is ready to be applied for the next steps.

3.3. Feature extraction from the weight vectors

A windowing feature extraction method is proposed in this work. The rectangular window slides across the weight vector and the features are computed. Each window will calculate the p number of AR coefficients, 2^{M} number of wavelet packet entropy and two numbers of wavelet fractal estimates as a feature values from the EEG segments. Hence, each EEG signal will be represented as ($S\times r$) number of feature values. Where r is the total number of feature values extracted from a window.

$$r = \left(\mathrm{p} + 2^{M} + 2\right) \tag{22}$$

- In computing AR coefficients, the selection of order p is a key factor. As the researchers in (Leonarduzzi et al., 2010) suggested that, the order 6 is significant in AR model for mental task classification. Hence, the order of AR model p = 6 is selected in the proposed methodology. Also, the Burg's method (described in section II) is used to estimate the AR coefficients due its low computational cost.
- Then to calculate the wavelet packet entropy, SE is calculated from the wavelet coefficients of windowed EEG signals. One major issue in the computation of wavelet-based SE is the selection of proper wavelet function. Several wavelet functions are available in the wavelet families, and every wavelet function has different characteristics. As a result, entropy will be different for different wavelet functions. However, in this work, a proper technique for selection of appropriate wavelet function is adopted. The correlation X_{corr} between the EEG signal of interest X and the mother wavelet function Y is calculated using equation (10) to match the best suitable wavelet function for EEG based MI analysis. The wavelet function is selected which gives the maximum value. In this work, total 30 wavelet functions (Haar, Daubechies with vanishing moment ranging from 2 to 12, Coiflets with vanishing moment ranging from 1 to 5, Symlets with vanishing moment ranging from 2 to 8, Fejer-Korovkin wavelets with vanishing moment ranging from 4 to 22) are compared with MI related EEG data. For each wavelet function, the average X_{corr} is shown in Table 2. From the table, it is evident that, the Haar provides the maximum correlation among all the wavelet functions. Hence, Haar is selected as a mother wavelet function to decompose the windowed EEG data using WPD up to level 4. Hence M = 4, and from the equation (9), $2^{M} = 16$ wavelet packet entropy will be computed from a window. Finally, the second cumulant of the SE and the width of the SS is measured from the equation (14) and (15) respectively from a window.
- In total (6 AR coefficients + 16 wavelet packet entropy + 1 SE + 1 SS) = 24 features are extracted from a window. Now, the window is slid forward to cover all the samples of the weight vector. Finally, all the feature values computed from every window are concatenated to form the feature vector, which consists of (150 * 24) = 3600 feature values.

Table 2
Wavelets vs Average cross correlation.

Wavelets	Xcorr	Wavelets	Xcorr	Wavelets	Xcorr
Haar	0.12	db11	0.02	coif2	0.02
db2	0.05	db12	0.02	coif3	0.02
db3	0.04	sym2	0.05	coif4	0.02
db4	0.04	sym3	0.04	coif5	0.02
db5	0.03	sym4	0.03	fk4	0.06
db6	0.04	sym5	0.03	fk6	0.04
db7	0.02	sym6	0.02	fk8	0.03
db8	0.02	sym7	0.02	fk14	0.03
db9	0.01	sym8	0.02	fk18	0.01
db10	0.02	coif1	0.04	fk22	0.02

db: Daubechies; *sym*: Symlets; coif: Coiflets; fk: Fejer-Korovkin wavelets.

3.4. Classification

On the classification stage, 1-against-all SVM classifier is used in this work for four class MI data classification. Then the feature-label pairs were fed to the classifier to train the network. Additionally, the Gaussian kernel function is used in the classifier for better classification.

4. Experimental results and discussion

The created dataset was fed to the autoencoders. Individual autoencoder was trained for particular EEG signals. When all the autoencoders were trained, their optimized weight vectors corresponding to the particular EEG signals were extracted. After successfully partitioning the weight vectors into training set and test set, the features were extracted through a rectangular slid window from both the weight vectors. The selection of size of the window is very sensitive because lower size results in a large number of feature values which may be redundant and higher size result in a smaller number of feature values which may result in poor classification. The window slides forward across the weight vector. So, depending on the size of weight vectors, the window size is selected as 125, 250, 375, 500, 750, 1000, 1250 and 1500

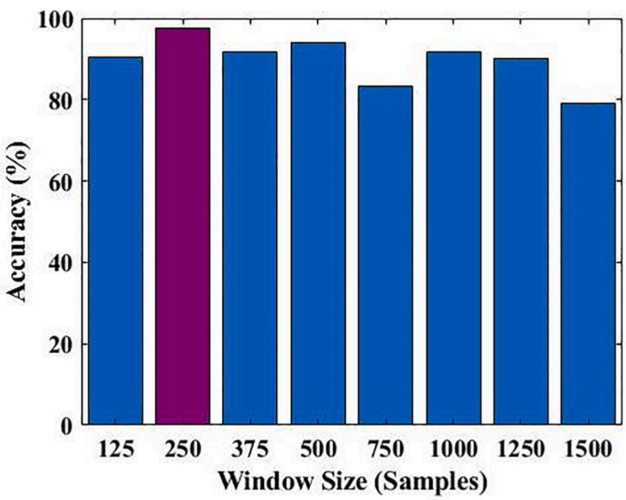


Fig. 4. Comparison of different window sizes for feature extraction.

Table 3
Comparison among several case studies (on dataset IIIa BCI-III).

Methods	Classification Accuracy (%)
Case-1	85
Case-2	91.87
Case-3	95.39

samples. The performance of the window sizes is compared in Fig. 4. It is evident from the figure that, the window size of 250 samples has achieved the highest classification accuracy. Finally, the feature-label pairs were fed to the classifier to train the network. After successfully training of the classifier, the trained model has been used to predict the test dataset.

Three cases have been studied to check the superiority of the proposed model. In the first case, the same feature sets were extracted directly from the information signal and fed to the classifier (case-1). In the second case, window-based features were extracted from the information signal (case-2). In the third case, features were extracted using proposed methodology i.e. from autoencoder weight vectors (case-3). For all three cases, the performance is compared in Table 3. It is evident from the table that, the classification accuracy using proposed model (case-3) is higher than the other two cases. The multiclass SVM classifier has been adopted through comparison with other classifiers. Linear Discriminant Analysis (LDA) network also has been used to evaluate the proposed method. Their performance is shown in Table 4. From the table, it is evident that, the SVM shows the highest performance in terms of precision, sensitivity, specificity and model accuracy.

The proposed system is also evaluated on BCI-IV-2a dataset. Table 5 represents the performance of the proposed system on the dataset IIa from BCI competition IV. Among all the 9 subjects, subjects: A02, A03, A04 and A08 the prediction is 100 % accurate for four-class MI EEG data. The performance of proposed system is also compared with the other recently reported methods (Jafarifarmand et al., 2017; Ai et al., 2019; Zhang et al., 2019; Sreeja & Samanta, 2019; Ma et al., 2020; Wang et al., 2020; Wu et al., 2021; Chen et al., 2022) and shown in Table 6. It is evident from the table that, the proposed approach gives better results than other methods in terms of classification accuracy.

Table 5
Subject wise performance comparison on dataset IIa from BCI-IV.

Subjects	Classification Accuracy (%)	Average Accuracy (%)
A01	93	**97**
A02	100	
A03	100	
A04	100	
A05	97	
A06	93	
A07	93	
A08	100	
A09	97	

Table 4
Performance comparison of the proposed system on dataset IIIa from BCI-III.

					Sub_1					
Parameters (%)			LDA					SVM		
	Class-1	Class-2	Class-3	Class-4	Mean	Class-1	Class-2	Class-3	Class-4	Mean
Sensitivity	100	83	62	69	78.5	94	93	94	100	95.25
Specificity	89	90	95	97	92.75	99	98	100	97	98.5
Precision	76	73	84	86	79.75	97	93	100	90	95
Model accuracy			78 %					95%		
					Sub_2					
Parameters (%)			LDA					SVM		
	Class-1	Class-2	Class-3	Class-4	Mean	Class-1	Class-2	Class-3	Class-4	Mean
Sensitivity	84	66	97	69	79	100	97	100	96	98.25
Specificity	94	96	85	98	93.25	99	99	100	100	99.5
Precision	84	83	72	90	82.25	97	97	100	100	98.5
Model accuracy			80 %					98 %		
					Sub_3					
Parameters (%)			LDA					SVM		
	Class-1	Class-2	Class-3	Class-4	Mean	Class-1	Class-2	Class-3	Class-4	Mean
Sensitivity	68	93	79	77	79.25	97	90	94	92	93.25
Specificity	97	82	98	96	93.25	98	97	100	97	98
Precision	88	63	93	83	81.75	94	90	100	89	93.25
Model accuracy			79 %					93 %		

Table 6
Performance comparison of several methods on dataset IIa from BCI-IV.

Performance	TSM (Wu et al., 2021)	FB-CGANet (Chen et al., 2022)	CSP-LCD (Ai et al., 2019)	One-vs-rest FBCSP (Zhang et al., 2019)	AR-CSP (Jafarifarmand et al., 2017)	CSP-Auto Regressive (Wang et al., 2020)	Wavelet Energy (Sreeja & Samanta, 2019)	PSD (Ma et al., 2020)	Proposed method
Average accuracy (%)	77.33	79.4	79.7	83	85.49	86.01	91.84	96.26	97

5. Conclusions and future scope

In this paper, a new method of feature extraction is proposed for four class MI EEG classification. The main challenge in classification of EEG signal is its non-stationary characteristics which results in non-separability or MI classes. To solve this problem, the EEG signal was transformed into another domain or space i.e. the weight vector of autoencoder. At the minimum reconstruction error of the network, the weight vectors were used to represent the EEG signals in new domain. In the proposed method, the values of three features, AR coefficients, wavelet packet entropy, wavelet fractal estimates of the EEG signal in new domain were used to train an autoencoder neural network. A windowing-based feature extraction technique was implemented to capture the local features of EEG signals from the transformed EEG (i.e. weight vectors). Results reveal that, an EEG signal can be represented in weight vector of autoencoder neural network. The performance of the proposed method was compared with the existing conventional methods. The proposed method successfully predicts the mental task with higher accuracy. It can be concluded that, the proposed method can be regarded as a powerful tool to improve the performance of MI EEG-based BCIs.

CRediT authorship contribution statement

Souvik Phadikar: Conceptualization, Investigation, Methodology. **Nidul Sinha:** Supervision, Validation. **Rajdeep Ghosh:** Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is publicly available online

References

Aghaei, A. S., Mahanta, M. S., & Plataniotis, K. N. (2016). Separable common spatio-spectral patterns for motor imagery BCI systems. *IEEE Transaction on Biomedical Engineering., 63*(1), 15–29.

Ai, Q., Chen, A., Chen, K., Liu, Q., Zhou, T., Xin, S., & Ji, Z. (2019). Feature extraction of four-class motor imagery EEG signals based on functional brain network. *Journal of Neural Engineering, 16*(2), Article 026032.

Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices, 2014*.

Amin, S. U., Alsulaiman, M., Muhammad, G., Mekhtiche, M. A., & Hossain, M. S. (2019). Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. *Future Generation Computer Systems, 101*, 542–554.

Anderson, C. W., Stolz, E. A., & Shamsunder, S. (1998). Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering, 45*(3), 277–286.

Ang, K. K., & Guan, C. (2017). EEG-based strategies to detect motor imagery for control and rehabilitation. *IEEE Transaction On Neural System and Rehabilitation Engineering., 25*(4), 392–401.

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2012). Mutual information based selection of optimal spatial-temporal patterns for single-trial EEG based BCIs. *Pattern Recognition, 45*(6), 2137–2144.

Blankertz, B., Muller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlogl, A., … Birbaumer, N. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions On Neural Systems and Rehabilitation Engineering, 14*(2), 153–159.

Chen, J., Yi, W., Wang, D., Du, J., Fu, L., & Li, T. (2022). FB-CGANet: Filter bank channel group attention network for multi-class motor imagery classification. *Journal of Neural Engineering, 19*(1), Article 016011.

Das, A. K., Suresh, S., & Sundararajan, N. (2016). A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification. *Expert Systems with Applications, 64*, 375–384.

Dose, H., Møller, J. S., Iversen, H. K., & Puthusserypady, S. (2018). An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Systems with Applications, 114*, 532–542.

Fadel, W., Kollod, C., Wahdow, M., Ibrahim, Y., & Ulbert, I. (2020). Multi-class classification of motor imagery EEG signals using image-based deep recurrent convolutional neural network. In *2020 8th International Winter Conference on Brain-Computer Interface (BCI)* (pp. 1–4). IEEE.

Ghosh, R., Sinha, N., & Biswas, S. K. (2019). Automated eye blink artefact removal from EEG using support vector machine and autoencoder. *IET Signal Processing, 13*(2), 141–148.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., … Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), e215–e220.

Hou, Y., Chen, T., Lun, X., & Wang, F. (2022). A novel method for classification of multi-class motor imagery tasks based on feature fusion. *Neuroscience Research, 176*, 40–48.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 13*(2), 415–425.

Jafarifarmand, A., & Badamchizadeh, M. A. (2020). Real-time multiclass motor imagery brain-computer interface by modified common spatial patterns and adaptive neuro-fuzzy classifier. *Biomedical Signal Processing and Control, 57*, Article 101749.

Jafarifarmand, A., Badamchizadeh, M. A., Khanmohammadi, S., Nazari, M. A., & Tazehkand, B. M. (2017). A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI. *IEEE Transactions on Fuzzy Systems, 26*(3), 1485–1497.

Kam, T. E., Suk, H. I., & Lee, S. W. (2013). Non-homogeneous spatial filter optimization of electroencephalogram (EEG) based motor imagery classification. *Neurocomputing., 108*, 58–68.

Krauledat, J. M. (2008). *Analysis of nonstationarities in EEG signals for improving brain-computer interface performance*. Berlin, Germany: Technische Universitat Berlin. Ph. D. dissertation.

Leonarduzzi, R. F., Schlotthauer, G., & Torres, M. E. (2010). August). Wavelet leader based multifractal analysis of heart rate variability during myocardial ischaemia. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 110–113).

Li, T., & Zhou, M. (2016). ECG classification using wavelet packet entropy and random forests. *Entropy, 18*(8), 285.

Liou, C. Y., Cheng, W. C., Liou, J. W., & Liou, D. R. (2014). Autoencoder for words. *Neurocomputing, 139*, 84–96.

Ma, X., Wang, D., Liu, D., & Yang, J. (2020). DWT and CNN based multi-class motor imagery electroencephalographic signal recognition. *Journal of Neural Engineering, 17*(1), Article 016073.

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*(4), 525–533.

Neumaier, A., & Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS), 27*(1), 27–57.

Phadikar, S., Sinha, N., & Ghosh, R. (2020a). Automatic eyeblink artifact removal from EEG signal using wavelet transform with heuristically optimized threshold. *IEEE Journal of Biomedical and Health Informatics, 25*(2), 475–484.

Phadikar, S., Sinha, N., & Ghosh, R. (2020b). Automatic EEG eyeblink artefact identification and removal technique using independent component analysis in combination with support vector machines and denoising autoencoder. *IET Signal Processing, 14*(6), 396–405.

Phadikar, S., Sinha, N., Ghosh, R., & Ghaderpour, E. (2022). Automatic muscle artifacts identification and removal from single-channel EEG using wavelet transform with meta-heuristically optimized non-local means filter. *Sensors, 22*(8), 2948.

Rathee, D., Raza, H., Prasad, G., & Cecotti, H. (2017). Current source density estimation enhances the performance of motor-imagery-related brain-computer interface. *IEEE Transaction on Neural Systems and Rehabilitation Engineering., 25*(12), 2461–2471.

Sreeja, S. R., & Samanta, D. (2019). Classification of multiclass motor imagery EEG signal using sparsity approach. *Neurocomputing, 368*, 133–145.

Stoica, P., & Moses, R. L. (2005). *Spectral analysis for signals*. Upper Saddle River, NJ, USA: Prentice Hall.

Sun, L., Feng, Z., Lu, N., Wang, B., & Zhang, W. (2019). An advanced bispectrum features for EEG-based motor imagery classification. *Expert Systems with Applications, 131*, 9–19.

Tangermann, M., Müller, K. R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., … Nolte, G. (2012). Review of the BCI competition IV. *Frontiers in neuroscience, 6*, 55.

Wang, J., Feng, Z., Ren, X., Lu, N., Luo, J., & Sun, L. (2020). Feature subset and time segment selection for the classification of EEG data based motor imagery. *Biomedical Signal Processing and Control, 61*, Article 102026.

Wu, F., Gong, A., Li, H., Zhao, L., Zhang, W., & Fu, Y. (2021). A new subject-specific discriminative and multi-scale filter bank tangent space mapping method for recognition of multiclass motor imagery. *Frontiers in Human Neuroscience, 15*, Article 595723.

Zhang, R., Zong, Q., Dou, L., & Zhao, X. (2019). A novel hybrid deep learning scheme for four-class motor imagery classification. *Journal of Neural Engineering, 16*(6), Article 066004.