

RESEARCH

Open Access



Multiclass EEG signal classification utilizing Rényi min-entropy-based feature selection from wavelet packet transformation

Md. Asadur Rahman^{1*} , Farzana Khanam², Mohiuddin Ahmad³  and Mohammad Shorif Uddin⁴ 

Abstract

This paper proposes a novel feature selection method utilizing Rényi min-entropy-based algorithm for achieving a highly efficient brain–computer interface (BCI). Usually, wavelet packet transformation (WPT) is extensively used for feature extraction from electro-encephalogram (EEG) signals. For the case of multiple-class problem, classification accuracy solely depends on the effective feature selection from the WPT features. In conventional approaches, Shannon entropy and mutual information methods are often used to select the features. In this work, we have shown that our proposed Rényi min-entropy-based approach outperforms the conventional methods for multiple EEG signal classification. The dataset of BCI competition-IV (contains 4-class motor imagery EEG signal) is used for this experiment. The data are preprocessed and separated as the classes and used for the feature extraction using WPT. Then, for feature selection Shannon entropy, mutual information, and Rényi min-entropy methods are applied. With the selected features, four-class motor imagery EEG signals are classified using several machine learning algorithms. The results suggest that the proposed method is better than the conventional approaches for multiple-class BCI.

Keywords: Electro-encephalogram (EEG), Brain–computer interface (BCI), Feature extraction, Wavelet packet transformation (WPT), Shannon entropy, Mutual information, Rényi min-entropy, Machine learning algorithms

1 Introduction

Brain–computer interface (BCI) is a modern notion that enables the way of generating communication between computer and brain functionalities. It is not only expected, but also satisfactorily achieved technology that could be a nice solution to assist the physically challenged people to control devices utilizing their brain functionalities. The brain functionality can be assessed in two different ways: electrical activities (suggested modalities are electro-corticogram (ECoG), electro-encephalogram (EEG), and magneto-encephalogram (MEG)) and hemodynamics (functional near-infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI),

etc.) [1, 2]. Based on the electrical activities, EEG is the most familiar, non-invasive, cheapest, and fastest modality for functional brain signal recording [3]. Among many brain stimuli, motor imagery (MI) movement is the highest choice for the researchers [2]. MI has a special benefit because it needs no additional setup like visual stimuli [3].

For providing the MI movement-based stimuli a candidate imagines the pattern of the real executive movements. During such MI movement, the EEG signals are recorded from the scalp of the participant that can be used to control the switches through a computer-based signal processing and this is broadly called BCI. There are different types of MI EEG-based BCI such as lifting hands and feet [4], simple–compound upper limb MI [5], uninterrupted hand movements [6] and finger movements [7], etc. Different stimuli show different classification performances. Especially, the upper limb shows

*Correspondence: bmeasadur@gmail.com

¹ Department of Biomedical Engineering, Military Institute of Science & Technology (MIST), Mirpur Cantonment, Dhaka 1216, Bangladesh
Full list of author information is available at the end of the article

more activeness than the lower limb MI movements. In addition, most of the EEG-MI research works [8–13] are related to two-class or three-class problem. Multiclass, i.e., more than 3-class classification for the EEG-MI signal is very challenging because of their non-discriminative features. To meet the challenge of 4-class EEG-MI, a very handful research works [14–16] were proposed where two hands MI, feet MI, and tongue movements were considered but their classification accuracy is not promising. This implies that they could not extract the required discriminative features in EEG signal through multiple stimuli. A very intelligent feature extraction and selection method could be an ultimate solution to meet the challenge of multiclass EEG signals.

There are a number of ways to extract features from EEG signals, such as autoregressive (AR) methods [17, 18], phase-space reconstruction approach [10], CSP-based methods [14, 16, 19], empirical mode decomposition [20–22], multivariate empirical mode decomposition-based methodologies [23, 24], channel correlation and feature optimization-based model [25], PCA-*t*-statistics-based feature extraction and selection method [26], wavelet transforms (WT) method [17, 27–30], etc. For a wide range of pattern recognition, wavelet packet transformation (WPT) provides excellent time–frequency features and therefore this approach is widely accepted feature extraction method for the EEG-MI movement classification. The WPT-based feature extraction has two important limitations: (i) structuring the features and (ii) selection of the bases. The features are structured by WPT coefficients those are considered to yield the significant pattern of the different classes EEG signal. Besides the feature structuring, proper base selection is the other step by which the structured features can show the highest discriminative characteristics among the classes. To overcome the limitations of WPT, Shannon entropy-based joint best basis method was proposed in [31], which is also questionable. This is because the proposed joint best basis method is effective in compression instead of classification [32]. To overcome the existing limitations in [31], a symmetric relative entropy-based local discriminant basis algorithm was proposed in [33]. It is reported in [34] that both methods rely on the signal's energy level which exhibits hampering effect to achieve high classification for multiple classes. Eventually, it can be summarized that measuring the distance alone may not be wise consideration to judge the ability of the features for discriminating different classes [35]. Recent feature selection-based research works [23–25] reported their proposal on the BCI competition-IV dataset [36]. Correlation-based channel selection method with regularized common spatial pattern proposed in [25] did not cover the time–frequency characteristics of

the EEG signal and on the other hand, the proposal given in [23, 24] utilized the time–frequency domain features but the feature selection method based on the Riemannian geometry could be further improved to achieve more accuracy in the classification of the MI events from the BCI competition-IV dataset. In this consequence, a more powerful feature selection algorithm is required to identify the meaningful content among the different features.

As it is explained in [37, 38] that the WPT computes more effectively than the WT and EMD in case of large size signal, we can choose WPT for feature extraction. In conventional approaches, Shannon entropy and mutual information method are often used to select the features found from the WPT coefficients. This work proposes to utilize the concept of Rényi min-entropy-based approach with a slight modification for the feature selection from the WPT coefficients. This method chooses features from a large feature set based on a special form of entropy compared to the Shannon entropy and mutual information method to attain the higher classification accuracy. Therefore, the main contributions of this work are:

- To extract WPT features from the EEG signals of four-class MI dataset.
- Selecting the features utilizing the proposed modified Rényi min-entropy-based approach.
- Comparing the classification performance of the proposed method with the conventional feature selection methods.
- To evaluate the classification performances of the proposed method using different classifiers.
- To compare the classification accuracy of the proposed method with the recent published proposals.

The dataset of BCI competition-IV (contains 4-class MI EEG signal) is used for this experiment. The data are preprocessed and separated in classes and used for the feature extraction using WPT. Then, for feature selection Shannon entropy, mutual information, and Rényi min-entropy methods are applied. With the selected features of 4-class motor imagery EEG signals are classified using several machine learning algorithms such as support vector machine (SVM), random forest, k-nearest neighbor (k-NN), multi-layer perceptron artificial neural network (MLP-ANN), logistic regression (LR), etc. Obtained results confirmed that the accuracy of the proposed method is higher than that of the Shannon entropy and mutual information. In addition, the proposed method also outperforms the recent state-of-the-art methods related to the applied dataset.

The rest of the paper is organized as follows: the materials and methods of this work are described with necessary steps in Sect. 2. In Sect. 3, the results are presented

with required discussions; and research outcomes are concluded in Sect. 4.

2 Materials and methods

2.1 Data collections

A multiclass valid and widely accepted datasets are taken in this study which is well known as BCI competition-IV. This dataset includes 9 healthy participants with no history of diseases and medications as well as they were verbally informed about the acquisition procedure and the possible outcomes of this research. Each participant performs different tasks based on a visual cue given on a screen. This cue-assisted data acquisition paradigm consisted of four motor imagery tasks: the imagination of movement of the left hand, right hand, both feet, and tongue. For each participant, two sessions (one for training and another for testing) on different days were recorded. Each session has 6 runs while one run contains 12 trials in each class. Therefore, one run produces $(12 \times 4 =)$ 48 trials and each session contains total $(48 \times 6 =)$ 288 trials. As one session is considered as training and another session is considered as testing, so we have 72 $(= 288/4)$ trials of each class for training and the same number of trials for testing. Each data set is strongly influenced by EOG. Data were recorded using 25 channels (22 EEG channels and 3 EOG channels) with a sampling rate of 250 Hz. The data acquisition schedule is illustrated in Fig. 1. The original EEG data are downsampled to 100 Hz. The detailed explanation of the dataset is available in <http://www.bbc.de/competition/iv/>.

2.2 Data preprocessing

The sampling rate of the EEG signals of the used dataset is 250 Hz and the signals were previously filtered using a bandpass filter of frequency band within 0.5 and 100 Hz. The sensitivity of the amplifier was considered 100 mV. In addition, a 50-Hz notch filter was utilized to remove the line noise. Among 25 channels, 10 significant channels (channel no. 2, 3, 4, 5, 6, 8, 9, 10, 11, and 12) were considered as they represent the central and frontal region of the brain (the detailed positions of the channels are given in Fig. 2). Also, the effect of eye blink and EOG were removed utilizing the enhanced automatic wavelet independent component

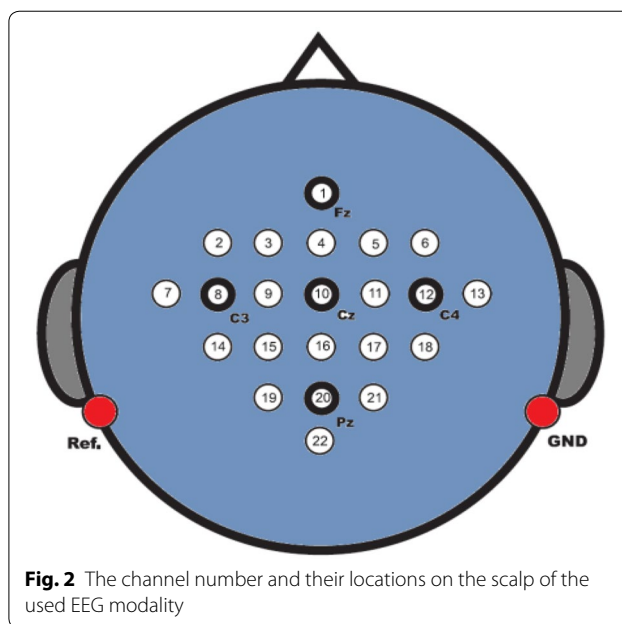


Fig. 2 The channel number and their locations on the scalp of the used EEG modality

analysis (EAWICA) toolbox [39]. Finally, the EEG signals were separated according to the schedule of the tasks.

2.3 Feature extraction using wavelet packet transformation

WPT differs from conventional wavelet transformation as it decomposes both approximate coefficients and detailed coefficients. We can compare WPT to a subspace tree. The original signal space represents the root node of the tree and it can be denoted as $\Pi_{0,0}$. The general form of this equation is $\Pi_{j,k}$, where notation j and k indicate the scale and the sub-band space. This original signal equation becomes $\Pi_{j,k} \rightarrow \Pi_{j+1,2k}$ when it is decomposed into approximation space. In case of detailed space, the equation is $\Pi_{j,k} \rightarrow \Pi_{j+1,2k+1}$. The space decomposition idea is derived from the concept of dividing the orthogonal basis function of the original signal. Here, $\{\varphi_j(t - 2^j k)\}_{k \in \mathbb{Z}}$ denoting orthogonal basis function is transformed into two new orthogonal bases: (i) $\{\varphi_{j+1}(t - 2^{j+1} k)\}_{k \in \mathbb{Z}}$ of approximate space $\Pi_{j+1,2k}$ and (ii) $\{\psi_{j+1}(t - 2^{j+1} k)\}_{k \in \mathbb{Z}}$ of detailed space $\Pi_{j+1,2k+1}$. Here $\Pi_{j,k}(t)$ and $\Psi_{j,k}$ represent the scaling and wavelet functions, respectively. These functions are equated as [32]:

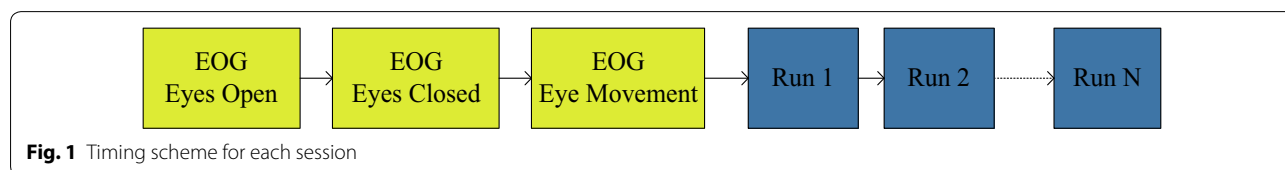


Fig. 1 Timing scheme for each session

$$\varphi_{j,k}(t) = \frac{1}{\sqrt{|2^j|}} \varphi\left(\frac{t - 2^j k}{2^j}\right), \tag{1}$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|2^j|}} \psi\left(\frac{t - 2^j k}{2^j}\right). \tag{2}$$

The scaling or compression degree of the original signal is calculated by the scaling parameter 2^j . Moreover, $2^j k$ of these equations is named as a location parameter or translation parameter, which indicates the time location of the wavelet. The mentioned process can be repeated J times, where J must be less than $\log_2 N$. N of $\log_2 N$ denotes the total number of samples in the original signal $\Pi_{j,k}$. The final outcome of WPT is $J \times N$ coefficients. Thus, the tree has $N / (2^j)$ number of coefficient blocks at any level of calculation $j [j = 1, 2, \dots, J]$. The iterative process adds more tree nodes to the WPT tree, where the nodes represent the subspace of different frequency localization characteristics. The corresponding decomposition procedure can be presented as Fig. 3 [4, 32, 40].

2.4 Evolution of Rényi entropies

According to the usual definition, the Rényi entropies [41] are a type of functions that represent the uncertainty related to a random variable. The Rényi entropy is regarded as a non-negative real number (suppose, the order of the number is $\delta; \delta \neq 1$), with $\delta \neq 1$, and is defined mathematically as:

$$H_\delta(X) \stackrel{\text{def}}{=} \frac{1}{1 - \delta} \log \left(\sum_i p(x_i)^\delta \right). \tag{3}$$

It is clear to understand that in case of uniform value of $p(\cdot)$ the Rényi entropies are equal to $\log |X|$. Otherwise, the process will be as decreasing form in δ . Particularly, we can define the Shannon and min-entropy as:

$$\delta \rightarrow 1 H_1(X) = - \sum_x p(x) \log p(x) \quad (\text{Shannon entropy}),$$

$$\delta \rightarrow \infty H_\infty(X) = -p(x) \log p(x) \quad \text{min-entropy}.$$

According to the Shannon approach, the conditional entropy of X with respect to the given Y provides the mean residual entropy of X if the Y value is given and mathematically we can represent it as:

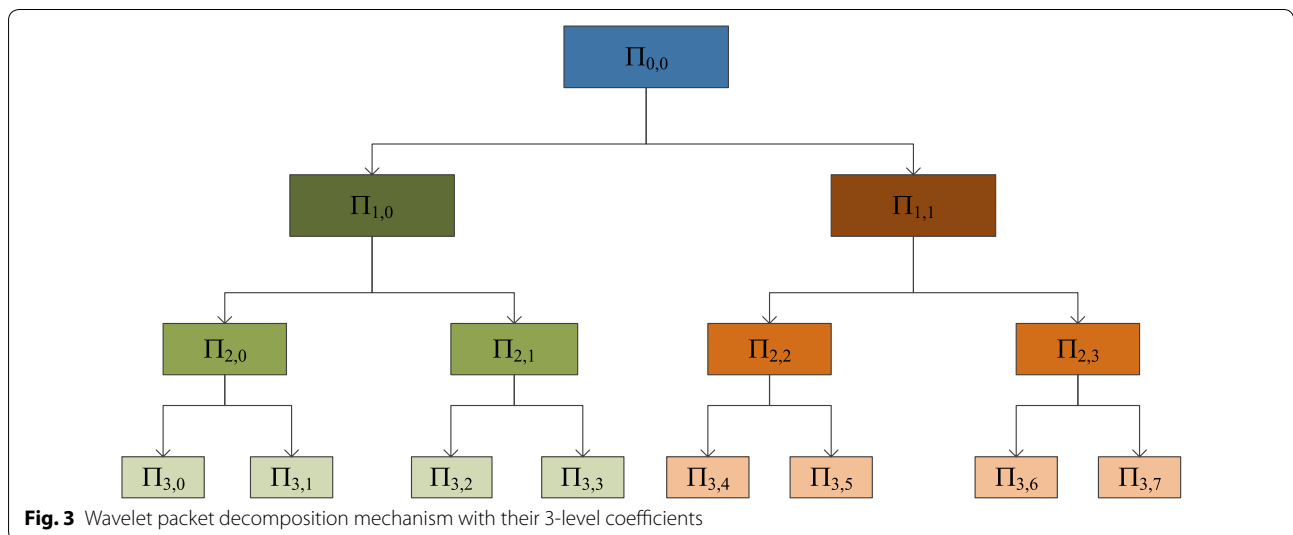
$$H_1(X|Y) \stackrel{\text{def}}{=} - \sum_{xy} p(x,y) \log p(x|y) = H_1(X, Y) - H_1(Y). \tag{4}$$

In (4), $H_1(X, Y)$ = entropy of $(X \cap Y)$. On the other hand, mutual information of Shannon approach regarding the previously proposed X and Y represents the correlation of information between X and Y , and we can define it as:

$$I_1(X; Y) \stackrel{\text{def}}{=} H_1(X) - H_1(X|Y) = H_1(X) + H_1(Y) - H_1(X, Y). \tag{5}$$

We can easily show that $I_1(X; Y) \geq 0$ along with $I_1(X; Y) = 0$ if we found that both X and Y are independent of each other, and hence $I_1(X; Y) = I_1(Y; X)$.

The fundamental proposal of Rényi was not supposed to define the conditional entropy as well as mutual information for the basic δ . A nice proposal with conditional



min-entropy associating the Rényi’s algorithm is given in [42] and the approach can be presented mathematically as:

$$H_{\infty}(X|Y) \stackrel{\text{def}}{=} -\log \sum_y \max_x p((y|x)p(x)). \tag{6}$$

It can be shown that this proposal is related to the Bayes risk. According to the method of Bayes risk, it is the error of guessing X with the given value of Y which can be defined mathematically as:

$$\beta(X|Y) \stackrel{\text{def}}{=} 1 - \sum_y p(y) \max_x p(x|y). \tag{7}$$

Also, we define the mutual information as,

$$I_{\infty}(X; Y) \stackrel{\text{def}}{=} H_{\infty}(X) - H_{\infty}(X|Y). \tag{8}$$

We can easily show that $I_{\infty}(X; Y) \geq 0$ along with $I_{\infty}(X; Y) = 0$ if we found that both X and Y are independent of each other (the reverse may not be true, that means I_{∞} is not symmetric). Therefore, the conditional mutual information in the case of this approach can be defined as, $I_{\infty}(X; Y|Z) \stackrel{\text{def}}{=} H_{\infty}(X|Z) - H_{\infty}(X|Y, Z)$, which is also analogous to conditional mutual information of Shannon approach.

2.5 Proposed feature selection algorithm

Suppose that the feature and class set are F and C , respectively. Since the proposed algorithm is designed considering (i) forward feature selection and (ii) dependency maximization, it builds a continuously incremental sequence $\{Q^t\}_{t>0}$ of subsets of F . According to the necessity, at every phase, the subset Q^{t+1} is calculated by adding the next feature. It should be noted that the consideration of the “**order of importance**” is based upon the conditional min-entropy. An interactive test based on the stopping criteria is supposed to be performed for sequence construction on the achieved accuracy through the current subset for the multiclass problem. While we achieve the required accuracy level, the algorithm stops itself and provides the resulting subset Q^t . In this case, the accuracy level $1 - \varepsilon$ will possibly be found if the Bayes risk function gives as $\beta(C|F) < \varepsilon$.

We define the series $\{Q^t\}_{t>0}$ and $\{f^t\}_{t>0}$ inductively as given below:

$$Q^0 \stackrel{\text{def}}{=} \varphi, \tag{9}$$

$$f^{t+1} \stackrel{\text{def}}{=} \arg \min_{f \in F \setminus Q^t} H_{\infty}(C|f, Q^t), \tag{10}$$

$$Q^{t+1} \stackrel{\text{def}}{=} S^t \cup \{f^{t+1}\}. \tag{11}$$

The proposed algorithms in [43] and [44] are also similar although they used Shannon entropy. According to the mentioned algorithms in [43, 44], f^{t+1} is defined based on the maximization of the mutual information rather than the conditional entropy minimization. According to the proposal of [45] this is irrelevant because $I_{\infty}(C; f|Q^t) = H_{\infty}(C|Q^t) - H_{\infty}(C|f, Q^t)$. Therefore, maximizing $I_{\infty}(C; f|Q^t)$ with respect to f is similar to minimizing $H_{\infty}(C|f, Q^t)$ with respect to f . This condition holds similar action in Shannon entropy. It can be proved that this proposal is locally optimal and the proof is given in [Appendix](#).

3 Results and discussion

As the approach of the proposal, 4-class motor imagery EEG signals are collected from the BCI competition-IV. Although these signals were preprocessed, the signal was filtered with a 50-Hz notch filter and after that, the signals were again filtered to remove EOG effect by EWICA toolbox as described in preprocessing subsection of this article. The stepwise filtering effect from the original EEG signals is presented in Fig. 4. After that, the EEG signals are considered for dual-tree WPT. A randomly selected EEG signal is illustrated up to a 3-level WPT in Fig. 5. This figure illustrates the different frequency content-based signals of the used EEG signal. Since the EEG signals are separated according to the schedule of the previously mentioned 4-class MI tasks, the separated EEG signals were considered for feature extraction by WPT. With the help of WPT all EEG signals are decomposed up to 5 levels and extracted four different features (Energy, Variance, Standard Deviation, and Waveform Length) as proposed in [46]. Therefore, for level 5 decomposition, we can find $2^5 = 32$ -type of features for each class. Eventually, every EEG signal will provide $32 \times 4 = 128$ features.

According to the claim of this work, all these features are not necessary for the classifier. Therefore, we need to choose the correct features from them. In conventional procedures like Shannon theory and mutual information theory, we can reduce the feature dimensions. Having some limitations of the conventional procedure (explained in the previous subsections), the proposed work utilizes the method of Rényi’s min-entropy-based algorithm for feature selection. In the feature space, it is found that the proposed method extracts significant distinguished patterns from each other among the features.

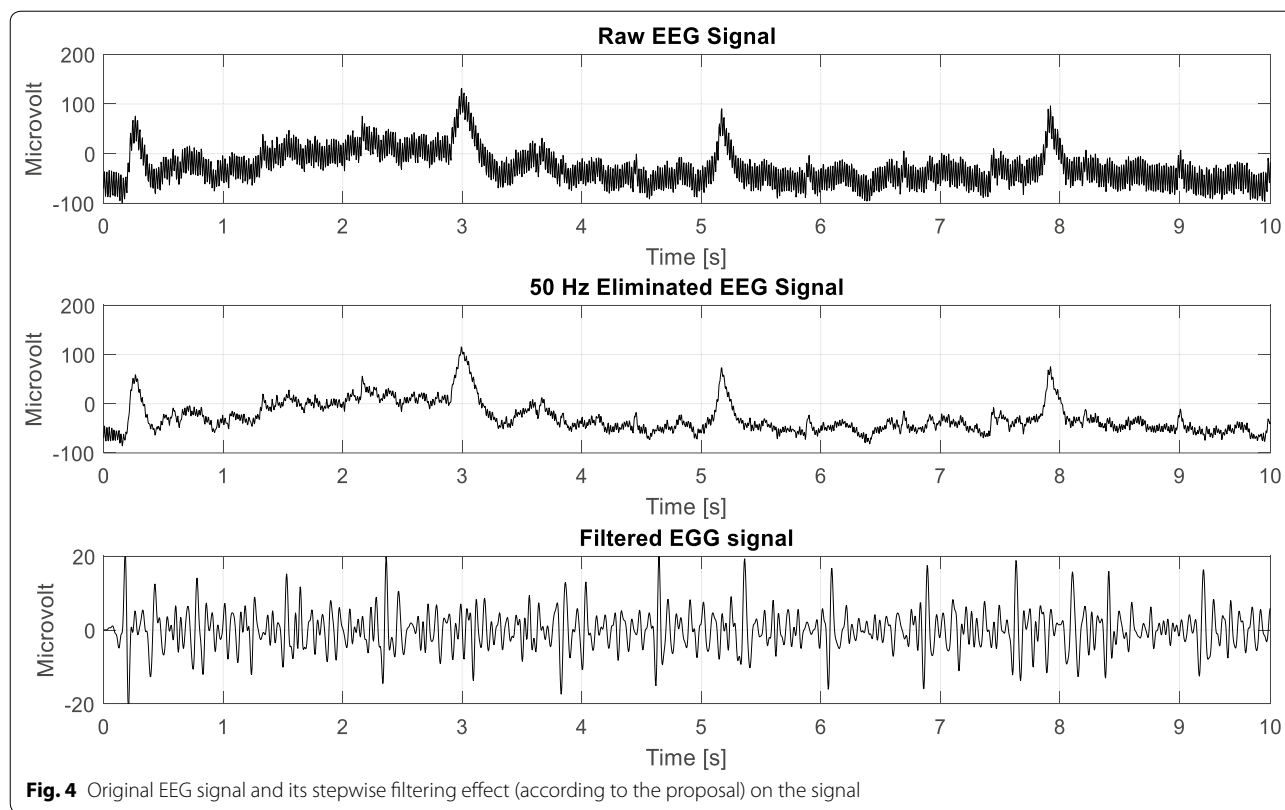
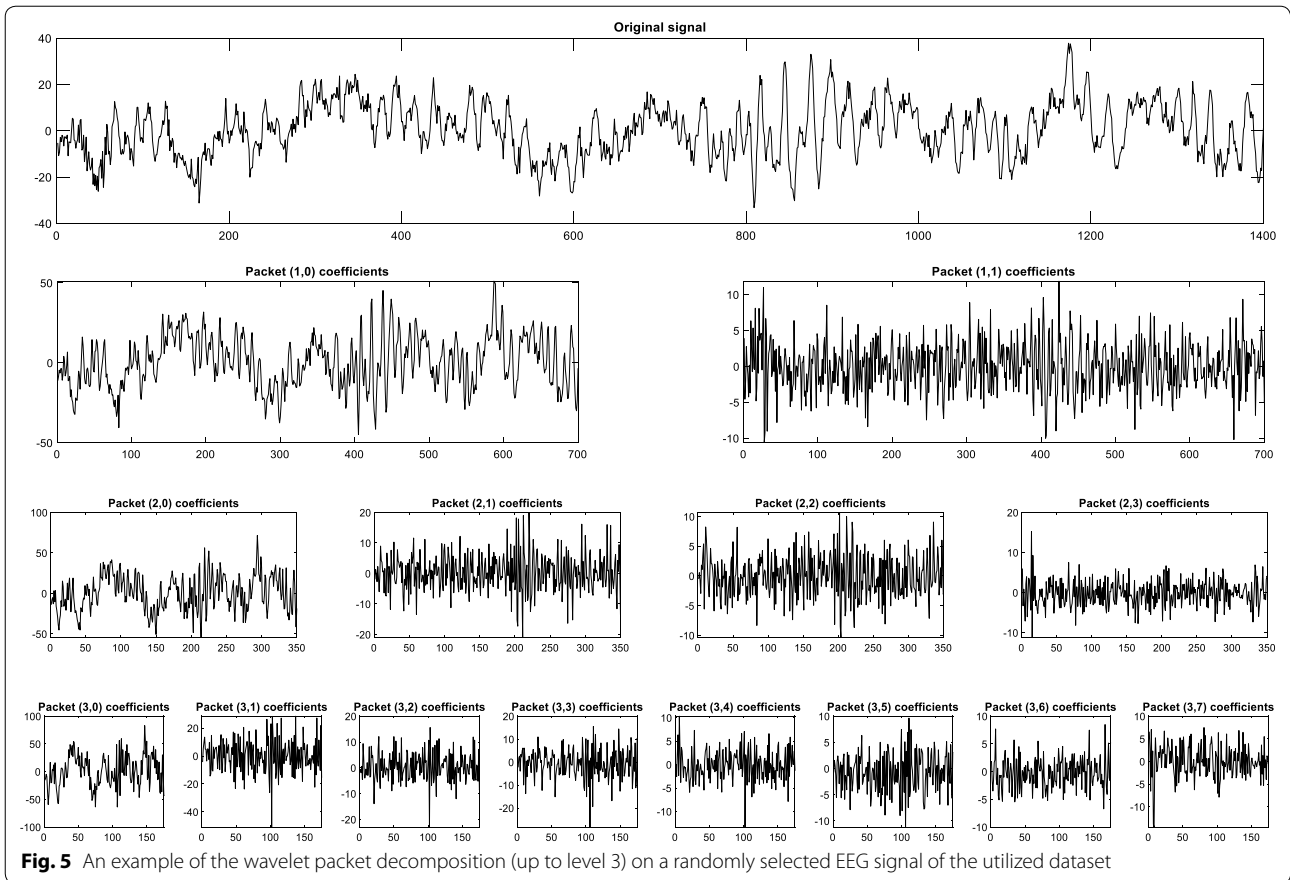


Figure 6 shows the features of the four classes in the feature space. From the figure, we get that the differentiability among the features in Rényi’s min-entropy is larger than the conventional methods (Shannon entropy and mutual information).

This representation of the features of different classes is the justification of the higher efficiency of the proposed method compared to the other conventional methods. Therefore, the training efficiency of the classifier would be better in the case of the proposed method and it can be predicted that the classification accuracy could also be higher than the other feature selection methods.

Since the training and testing data were previously separated in the used dataset, the features from training data were used to train the SVM classifier and the selected features of the testing EEG signals were utilized to test the classification accuracies. The average classification accuracies of SVM corresponding to the above-mentioned feature selection algorithms are given in Table 1. The results claim that the proposed Rényi’s min-entropy-based feature selection algorithm shows significantly higher classification accuracy than that of the conventional feature selection algorithms.

For thorough experimentation, we have taken the SVM classifier due to its wide acceptance. Furthermore, there are some commonly used classifiers such as random forest, k-NN, MLP-ANN, and LR and these classifiers could be used to justify their applicability and performances using our proposed feature selection algorithm. So, we did the experimentations. Every classifier is trained with the training dataset and tested 5 times with the testing dataset and taken the average classification accuracy. This result is given in Table 2. From this result, we get that random forest and MLP-ANN provides slightly higher classification accuracy than the SVM but k-NN and LR provides inferior results than the SVM. The results of the proposed work are also compared with the previous recent work that dealt with the four-class problem of BCI IV dataset and the result is given in Table 3. From the results, we have found that the proposed method shows higher accuracy than the previous work. There are some works [23, 24] those also dealt with BCI IV dataset, but their presentation of the classification accuracy is in the binary approach and six-tuple presentation (Left vs. Right, Left vs. Foot (LvF), Left vs. Tongue (LvT), Right vs. Foot



(RvF), Right vs. Tongue (RvT), Foot vs. Tongue (FvT)) of the four-class classification problem. Therefore, is not possible to compare this result with the proposed work directly.

4 Conclusions

The dual-tree wavelet decomposition of the EEG signals is a nice way to extract features for the EEG-based motor imagery-related task classification. However, more levels of decomposition create a number of features for multiple classes that become a burden for a classifier and hence the resulting classification accuracy reduces. Therefore, an intelligent feature selection algorithm is necessary to reduce the feature number,

and consequently, it would be necessary to increase the discriminating power of the features. This research work has proposed and utilized the Rényi's min-entropy algorithm along with a slight modification to select the WPT features for getting the higher classification accuracies. A four-class MI EEG signal of BCI competition-IV dataset is used to justify the proposed work, and from the results we found that the proposed method outperforms 18% and 6% increment in classification accuracy (in average) than the Shannon entropy and mutual information methods, respectively, in case of SVM classifier. On the other hand, applying the random forest and MLP-ANN the classification accuracy could be increased up to 8% with respect to mutual

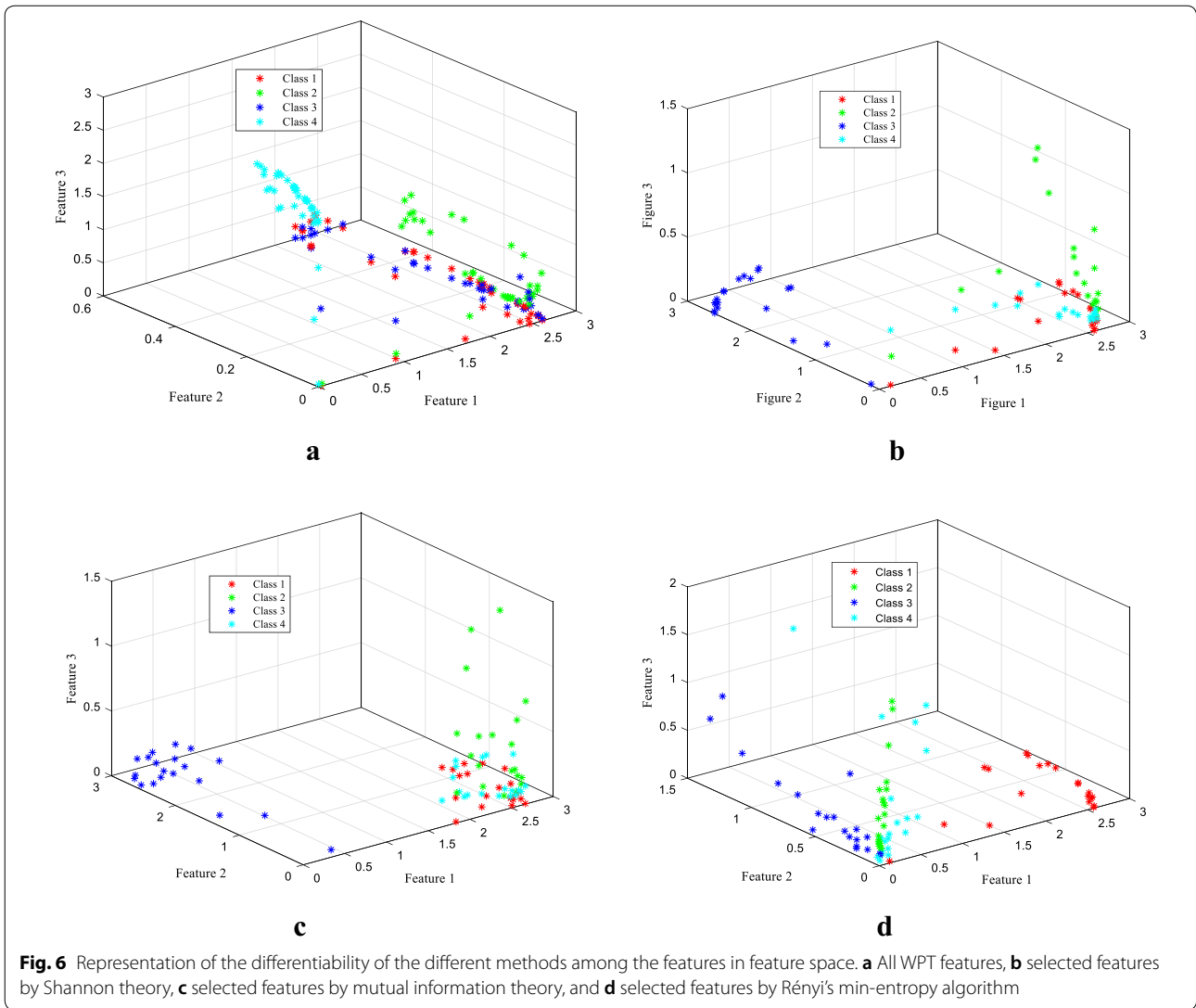


Table 1 The classification accuracy of the proposed algorithm compared to the others

Subject ID	Classification accuracy (%) of only WPT features	Classification accuracy (%) of selected WPT features by Shannon theory	Classification accuracy (%) of selected WPT features by mutual information theory	Classification accuracy (%) of selected WPT features by Rényi's min-entropy algorithm
1	57	60	68	82
2	52	65	76	86
3	70	50	71	78
4	70	65	72	73
5	43	62	74	79
6	64	67	71	81
7	48	53	76	75
8	62	62	72	79
9	69	64	72	75
Average ± std	59.44 ± 10.02	60.88 ± 5.75	72.44 ± 2.55	78.66 ± 4.03

Table 2 The classification accuracy of the different classifiers after adopting the proposed algorithm

Subject ID	SVM	Random forest	k-NN	MLP-ANN	LR
1	82	82	74	80	74
2	86	86	72	82	72
3	78	84	68	84	78
4	73	78	70	74	64
5	79	79	69	78	68
6	81	86	70	85	72
7	75	75	65	85	74
8	79	80	72	79	69
9	75	75	68	74	60
Average ± std	78.66 ± 4.03	80.55 ± 4.24	69.77 ± 2.68	80.11 ± 4.28	70.11 ± 5.53

Table 3 The average classification accuracy of the proposed work with the recent state-of-the-art

Proposed algorithm + random forest	Proposed algorithm + MLP-ANN	Jin et al. [25]
80.55	80.11	78%

information methods. Since the performance is too convincing, this intelligent feature selection algorithm will hopefully open a new pathway to implement multiple-class BCI in practice.

Acknowledgements

The authors would like to thank Rasel Ahammed, Lecturer, Bangabandhu Sheikh Mujibur Rahman University of Science and Technology for his guidelines to develop the Matlab code for this work.

Authors' contributions

MAR and FK developed the idea, prepared Matlab codes used in the algorithm, and wrote the paper. MA and MSU supervised the work. All authors reviewed the language and grammatical structures of the article and gave consent for publication. All authors read and approved the final manuscript.

Funding

This work was done without any financial support.

Availability of data and materials

This work utilizes the open data source BCI competition-IV which is available in [36].

Competing interests

As far as the knowledge of the authors, there is no competing interest to the other regarding this research work.

Author details

¹ Department of Biomedical Engineering, Military Institute of Science & Technology (MIST), Mirpur Cantonment, Dhaka 1216, Bangladesh. ² Department of Biomedical Engineering, Jashore University of Science and Technology (JUST), Jashore 7408, Bangladesh. ³ Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh. ⁴ Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh.

Appendix: Proof of the proposed method as an optimal feature selection algorithm

The previously stated set Q^{t+1} reduces to the Bayes risk related to the classification among the probabilities those are of the form $Q^t \cup \{f\}$, known mathematically as:

$$\forall f \in F \beta(C|Q^{t+1}) \leq \beta(C|Q^t \cup \{f\}). \tag{12}$$

Proof: Suppose that, $\bar{\theta}$, θ , θ' represent common value tuples and values of Q^t , f , and f^{t+1} , respectively,

$$\sum_{\bar{\theta}, \theta} \max_c (p(\bar{\theta}, \theta|c)p(c)) \leq \sum_{\bar{\theta}, \theta'} \max_c (p(\bar{\theta}, \theta'|c)p(c)). \tag{13}$$

Using the Bayes theorem (12), we get

$$\sum_{\bar{\theta}, \theta} p(\bar{\theta}, \theta) \max_c p(c|\bar{\theta}) \leq \sum_{\bar{\theta}, \theta'} p(\bar{\theta}, \theta') \max_c p(\bar{\theta}, \theta'|c). \tag{14}$$

Then, from the definition of Bayes risk, we get the following relation (15):

$$\beta(C|Q^t \cup \{f^{t+1}\}) \leq \beta(C|Q^t \cup \{f\}). \tag{15}$$

Received: 2 January 2020 Accepted: 10 June 2020
Published online: 16 June 2020

References

1. Rahman MA, Uddin MS, Ahmad M (2019) Modeling and classification of voluntary and imagery movements for brain-computer interface from

- fNIR and EEG signals through convolutional neural network. *Health Inf Sci Syst* 7:22. <https://doi.org/10.1007/s13755-019-0081-5>
2. Rahman MA, Rashid MA, Ahmad M (2019) Selecting the optimal conditions of Savitzky-Golay filter for fNIRS signal. *Biocybernetics Biomed Eng* 39(3):624–637. <https://doi.org/10.1016/j.bbe.2019.06.004>
 3. Nicolas-Alonso LF, Gomez-Gil J (2012) Brain computer interfaces, a review. *Sensors* 12:1211–1279
 4. Rahman MA, Hossain MK, Khanam F, Alam MK, Ahmad M (2019) Four-class motor imagery EEG signal classification using PCA, wavelet, and two-stage neural network. *Int J Adv Comput Sci Appl* 10:5. <https://doi.org/10.14569/ijacsa.2019.0100562>
 5. Weibo Y, Qiu S, Qi H, Zhang L, Wan B, Ming D (2013) EEG feature comparison and classification of simple and compound limb motor imagery. *J NeuroEng Rehabil* 10:106. <https://doi.org/10.1186/1743-0003-10-106>
 6. Seok WJ, Muller KR, Lee SW (2015) Classifying directions in continuous arm movement from EEG signals. In: 3rd International winter conference on brain computer interface (BCI), pp. 1–2
 7. Liao K, Xiao R, Gonzalez J, Ding L (2014) Decoding individual finger movements from one hand using human EEG signals. *PLoS ONE* 9(1)
 8. Ma Y, Ding X, She Q, Luo Z, Potter T, Zhang Y (2016) Classification of motor imagery EEG signals with support vector machines and particle swarm optimization. *Comput Math Methods Med* 2016:1–8
 9. Tang Z, Li C, Wu J, Liu P, Cheng S (2018) Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI. *Front Inf Technol Electron Eng*
 10. Djemal R, Bazyed AG, Belwafi K, Gannouni S, Kaaniche W (2016) Three-class EEG-based motor imagery classification using phase-space reconstruction technique. *Brain Sci* 6(36):1–19
 11. Petersen J, Iversen HK, Puthusserypady S (2018) Motor imagery based brain-computer interface paradigm for upper limb stroke rehabilitation. In: International conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, pp 1960–1963
 12. Li M, Zhu W, Liu H, Yang J (2017) Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-isomap. *Appl Sci* 7(390):1–18
 13. Cao L, Xia B, Maysam O, Li J, Xie H, Birbaumer N (2017) A synchronous motor imagery based neural physiological paradigm for brain-computer interface speller. *Front Hum Neurosci* 11(274):1–9
 14. Mahmood A, Zainab R, Ahmad RB, Saeed M, Kamboh AM (2017) Classification of multi-class motor imagery EEG using four band common spatial pattern. In: annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, pp 1034–1037
 15. Ge S, Wang R, Yu D (2014) Classification of four-class motor imagery employing single-channel electroencephalography. *PLoS ONE* 9(6):1–7
 16. Wang L, Wu X (2008) Classification of four-class motor imagery EEG data using spatial filtering. In: International conference on bioinformatics and biomedical engineering, Shanghai, pp 2153–2156
 17. Zhang Y, Liu B, Ji X, Huang D (2017) Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neural Process Lett* 45(2):365–378
 18. Zhang Y, Ji X, Zhang Y (2015) Classification of EEG signals based on AR model and approximate entropy. In: International joint conference on neural networks (IJCNN), Killarney, pp 1–6
 19. Saha PK, Rahman MA, Mollah MN (2019) Frequency domain approach in CSP based feature extraction for EEG signal classification. In: International conference on electrical, computer and communication engineering (ECCE), 7–9 February 2019, Cox's Bazar, Bangladesh
 20. Munoz-Gutierrez PA, Giraldo E, Bueno-Lopez M, Molonas M (2018) Localization of active brain sources from EEG signals using empirical mode decomposition: a comparative study. *IJES* 12(55):1–14
 21. Sokhal J, Garg B, Aggarwal S, Jain R (2017) Classification of EEG signals using empirical mode decomposition and lifting wavelet transforms. In: international conference on computing, communication, and automation (ICCCA), Greater Noida, pp 1197–1202
 22. Nazneen T, Rahman MA, Mollah MN (2019) Towards the effective intrinsic mode functions for motor imagery EEG signal classification. In: International conference on electrical, computer and communication engineering (ECCE), 7–9 February 2019, Cox's Bazar, Bangladesh
 23. Gaur P, Pachori R, Wang H, Prasad G (2018) A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Expert Syst Appl* 95:201–211
 24. Gaur P, Pachori RB, Wang H, Prasad G (2016) A multivariate empirical mode decomposition based filtering for subject independent BCI. In: 2016 27th Irish signals and systems conference (ISSC), Londonderry, pp 1–7
 25. Jing J, Miao Y, Daly I, Zuo C, Huc D, Cichocki A (2019) Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw* 118:262–270
 26. Rahman MA, Hossain MF, Hossain M, Ahmmed R (2019) Employing PCA and *t*-statistical approach for feature extraction and classification of emotion from multichannel EEG signal. *Egypt Inform J*. <https://doi.org/10.1016/j.eij.2019.10.002>
 27. Amin HU, Malik AS, Ahmad RF, Badruddin N, Kamel N, Hussain M, Chooi W-T (2015) Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australas Phys Eng Sci Med* 38(1):139–149
 28. Sherwani F, Shanta S, Ibrahim BSKK, Huq MS (2016) Wavelet-based feature extraction for classification of motor imagery signals. In: IEEE EMBS conference on biomedical engineering and sciences (IECBES), Kuala Lumpur, pp 360–364
 29. Sharmila A, Mahalakshmi P (2017) Wavelet-based feature extraction for classification of epileptic seizure EEG signal. *J Med Eng Technol* 41(8):670–680
 30. Rahman MA, Rashid MMO, Khanam F, Alam MK, Ahmad M (2019) EEG based brain alertness monitoring by statistical and artificial neural network approach. *Int J Adv Comput Sci Appl* 10:1
 31. Coifman RR, Wickerhauser MV (1992) Entropy-based algorithms for best basis selection. *IEEE Trans Inf Theory* 38(2):713–718
 32. Khushaba RN, Kodagoda S, Lal S, Dissanayake G (2011) Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Trans Biomed Eng* 58(1):121–131
 33. Saito N (1994) Local feature extraction and its applications using a library of bases. Ph.D. Thesis, Department of Mathematics, Yale Univ., New Haven, CT
 34. Deqiang L, Pedrycz W, Pizzi NJ (2005) Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Trans Biomed Eng* 52(6):1132–1139
 35. Schalko R (1992) Pattern recognition: statistical, structural and neural approaches. Wiley, New York
 36. Graz University BCI Competition data sets IV. <http://www.bbci.de/competition/iv/>
 37. Muo UE, Madamedon M, Ball AD, Gu F (2017) Wavelet packet analysis and empirical mode decomposition for the fault diagnosis of reciprocating compressors. In: 23rd international conference on automation and computing (ICAC), Huddersfield, pp 1–6
 38. Wang D, Yue C, Wei S, Lv J (2017) Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting. *Algorithms* 10(3):1–25. <https://doi.org/10.3390/a10030108>
 39. Mammone N, Morabito FC (2014) Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy* 16(12):6553–6572. <https://doi.org/10.3390/e1612.6553>
 40. Wali MK, Murugappan M, Ahmmad B (2013) Wavelet packet transform based driver distraction level classification using EEG. *Math Probl Eng* 2013, Article ID 297587

41. Rényi A (1961) On measures of entropy and information. In: Proceedings of the 4th Berkeley symposium on mathematics, statistics, and probability, pp 547–561, <https://projecteuclid.org/euclid.bsm/1200512181>
42. Smith G (2009) On the foundations of quantitative information flow. In: de Alfaro L (ed) Foundations of software science and computational structures. FoSSaCS 2009. Lecture Notes in Computer Science, vol 5504. Springer, Berlin. https://doi.org/10.1007/978-3-642-00596-1_21
43. Brown G, Pocock AC, Zhao MJ, Luján M (2012) Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *J Mach Learn Res* 13:27–66
44. Vergara JR, Estévez PA (2014) A review of feature selection methods based on mutual information. *Neural Comput Appl* 24(1):175–186
45. Palamidessi C, Romanelli M (2018) Feature selection with Rényi min-entropy. In: Artificial neural networks in pattern recognition—8th IAPR TC3 workshop (ANNPR 2018), Sep 2018, Siena, Italy. Springer, Lecture Notes in Computer Science, 11081, pp 226–239
46. Khushaba R (2019) Feature extraction using multisignal wavelet transform decomposition. MATLAB Central File Exchange. <https://www.mathworks.com/matlabcentral/fileexchange/37950-feature-extraction-using-multisignal-wavelet-transform-decomposition>. Accessed 29 Dec 2019

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
