

REVIEW

Open Access



A review of channel selection algorithms for EEG signal processing

Turky Alotaiby^{1*}, Fathi E Abd El-Samie^{2,3}, Saleh A Alshebeili⁴ and Ishtiaq Ahmad⁵

Abstract

Digital processing of electroencephalography (EEG) signals has now been popularly used in a wide variety of applications such as seizure detection/prediction, motor imagery classification, mental task classification, emotion classification, sleep state classification, and drug effects diagnosis. With the large number of EEG channels acquired, it has become apparent that efficient channel selection algorithms are needed with varying importance from one application to another. The main purpose of the channel selection process is threefold: (i) to reduce the computational complexity of any processing task performed on EEG signals by selecting the relevant channels and hence extracting the features of major importance, (ii) to reduce the amount of overfitting that may arise due to the utilization of unnecessary channels, for the purpose of improving the performance, and (iii) to reduce the setup time in some applications. Signal processing tools such as time-domain analysis, power spectral estimation, and wavelet transform have been used for feature extraction and hence for channel selection in most of channel selection algorithms. In addition, different evaluation approaches such as filtering, wrapper, embedded, hybrid, and human-based techniques have been widely used for the evaluation of the selected subset of channels. In this paper, we survey the recent developments in the field of EEG channel selection methods along with their applications and classify these methods according to the evaluation approach.

Keywords: EEG signals; Channel selection; Seizure detection; Sleep state classification; Motor imagery classification; Emotion classification; Mental task classification

1 Review

1.1 Introduction

Digital processing of EEG signals plays an important role in a variety of applications, e.g., seizure detection/prediction, sleep state classification, and motor imagery classification. Digital processing of EEG signals consists of different components: signal acquisition unit, feature extraction unit, and a decision algorithm as shown in Fig. 1. The input to the system in Fig. 1 is an EEG signal acquired from the scalp, brain surface, or brain interior. The signal acquisition unit is represented by electrodes whether they are invasive or non-invasive. The feature extraction unit is a signal processing unit aiming to extract discriminative features from channel(s). The decision unit, in brain computer interface (BCI) for example, is a hybrid unit with the purpose of classification, decision-making, and passing the decisions to external devices outputting the intention of the subject [1].

As mentioned above, the interface between the brain and the computer (or a device) could be through invasive or non-invasive technologies. Although invasive technologies have recently shown some promises in different applications for their large accuracy and low noise [2], non-invasive technologies are still used extensively for safety purposes with some additional signal processing tasks to compensate for the noise and resolution limitations. Scalp EEG acquisition devices are generally preferred due to their low-cost, ease of use, portability, and high temporal resolution. The scalp EEG signals can be recorded by different modes such as unipolar and bipolar modes. In the former mode, the voltage differences between all electrodes and a reference one are recorded, where a channel is formed by an electrode-reference pair. On the other hand, in the bipolar mode, the voltage differences between two specified electrodes are recorded, where each pair forms a channel. An electrode placement scheme on scalp, known as International 10–20 system, was recommended by the International Federation of

* Correspondence: totaiby@kacst.edu.sa

¹KACST, King Abdulaziz City, Kingdom of Saudi Arabia

Full list of author information is available at the end of the article

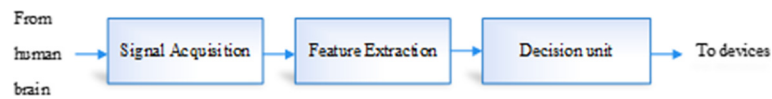


Fig. 1 Processing of EEG signals

Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) [3]. Figure 2 shows the 10–20 EEG electrode positions for the placement of electrodes from the left and the top of the head. These electrodes (channels) show the activities of different brain areas. Figure 3 shows the brain areas.

Most of the useful information about the functional state of a human brain lies in five major brain waves distinguished by their different frequency bands. These frequency bands are delta band (0–4 Hz), theta band (3.5–7.5 Hz), alpha band (7.5–13 Hz), beta band (13–26 Hz), and gamma band (26–70 Hz) [4]. Delta waves are related to the deep sleep state. Theta waves are related to the deepest state of mediation (body asleep/mind awake). Alpha waves are related to the case of dreaming and relaxation. Beta waves are the dominant with the waking state with large attention. Gamma waves are highly related to the decision-making mode of the brain. When dealing with mental illnesses states, unexpected disturbances of the brain waves occur leading to the need of considerable signal processing burdens for diagnosis of abnormal states [4].

The acquired EEG signals are generally of multi-channel nature. To classify these signals, for example, we have two choices: to work on a subset of channels selected based on certain criteria or to work on all channels [5]. Figure 4 gives an illustration for the general process of EEG signal

classification based on channel selection. In this signal processing setting, reducing the number of channels is needed because the setup process with a large number of channels is time-consuming and causes subject inconvenience. In addition, it adds to the computational complexity of the system, which is required to be low in certain applications.

Another example where channel reduction is of a potential value is in seizure detection and prediction. In particular, there is a great interest from the industry and scientific community in the development of portable medical support systems that incorporate algorithms capable of detecting early onset of epileptic seizures or even predicting them hours before they occur, as this will help to alert ambulatory patients or caregivers before seizure occurs to avoid injury [6, 7]. The development of such portable systems should be based on computationally efficient prediction algorithms that make use of as minimum number of channels as possible to reduce system power consumption, a necessary step to maintain longer time of operation.

Various techniques have been investigated for channel selection in the processing of EEG signals. This paper presents a survey for the recent developments in this field. Several flowcharts and tabular forms are presented to enable the reader to explore the different channel selection algorithms, to determine their classification

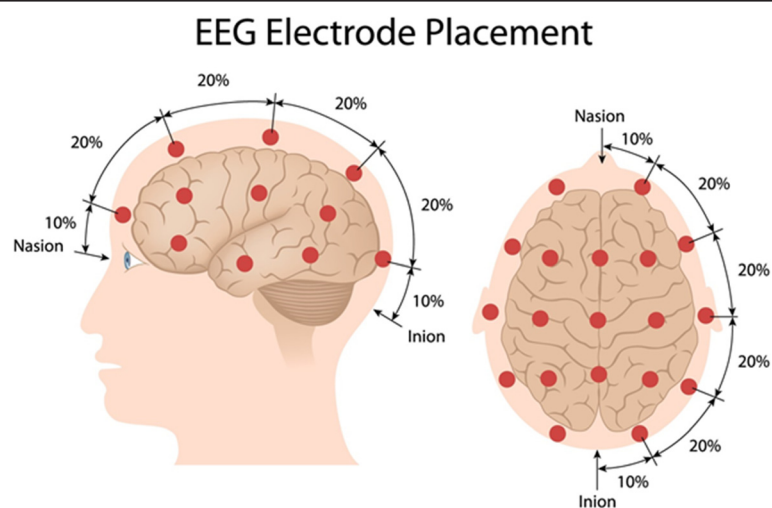
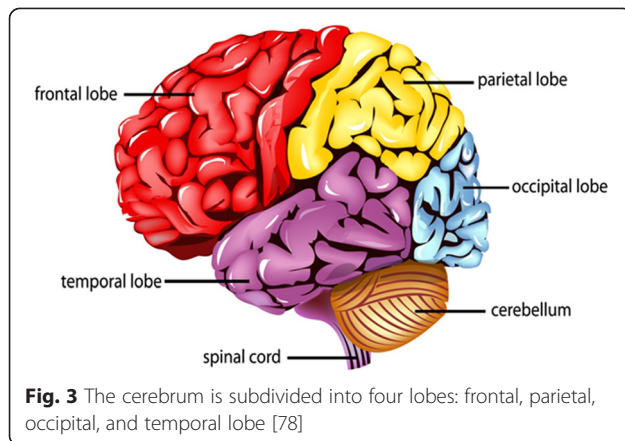


Fig. 2 The international 10–20 system. The *left* image shows the left side of the head, and the *right* image presents the view from above the head [78]



according to the evaluation algorithms, and to bring attention to directions of research in EEG channel selection for different applications. The performance of different methods, if available, is given in terms of classification/detection accuracy and probability of false alarm to produce clear and informative comparisons among the channel selection approaches. In addition, this survey may assist designers to choose the appropriate algorithms that suite intended applications. Furthermore, this work is expected to help newcomers to the field to determine the limitations associated with the available channel selection methods and to pave the road for the development of new channel selection designs.

The rest of this paper is organized as follows. Section 2 covers the selection techniques in general. Section 3 discusses channel selection for seizure detection/prediction. Section 4 is devoted for channel selection for motor imagery classification. Sections 5 and 6 cover the topics of emotion classification and mental task classification with channel selection strategies. Section 7 discusses channel selection for the task of sleep state analysis. Section 8 discusses the channel selection process for drug effects diagnosis. Finally, concluding remarks are given in Section 9.

2 Channel selection techniques

During the last decades, EEG-based processing has become a highly attractive research field. The large number of channel recordings due to the availability of low-cost interfaces led to the evolution of channel selection algorithms. The objectives of channel selection are manifold: improving model performance, providing faster processing

and dimensionality reduction, and identifying brain area that generates class-event activity.

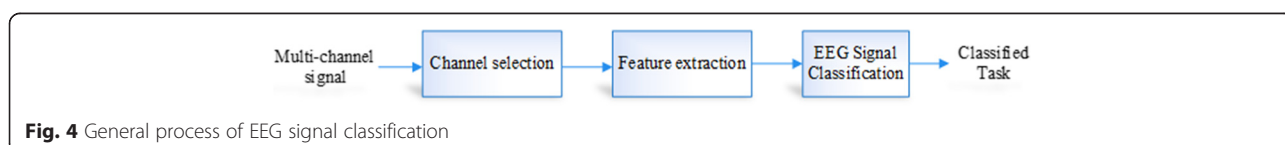
Based on the literature, feature selection algorithms were used for EEG channel selection [8–10]. In this section, we show how to adapt such techniques for channel selection. The main steps of channel selection are illustrated in Fig. 5 for a set of EEG channels. The subset generation step is a heuristic search process to present a candidate for evaluation based on a search strategy such as complete search, sequential search, or random search. In some applications, a trained specialist selects a subset of channels based on his experience. There are five main categories of candidate evaluation strategies, namely, filtering, wrapping, embedded, hybrid, and human-based techniques. These techniques are used for subset evaluation. The process of channel subset generation and evaluation is terminated when a stopping criterion is satisfied (search is completed or a threshold is reached). In the last step, the selected channel subset is validated via prior knowledge about the data. The evaluation techniques are discussed in the next subsections.

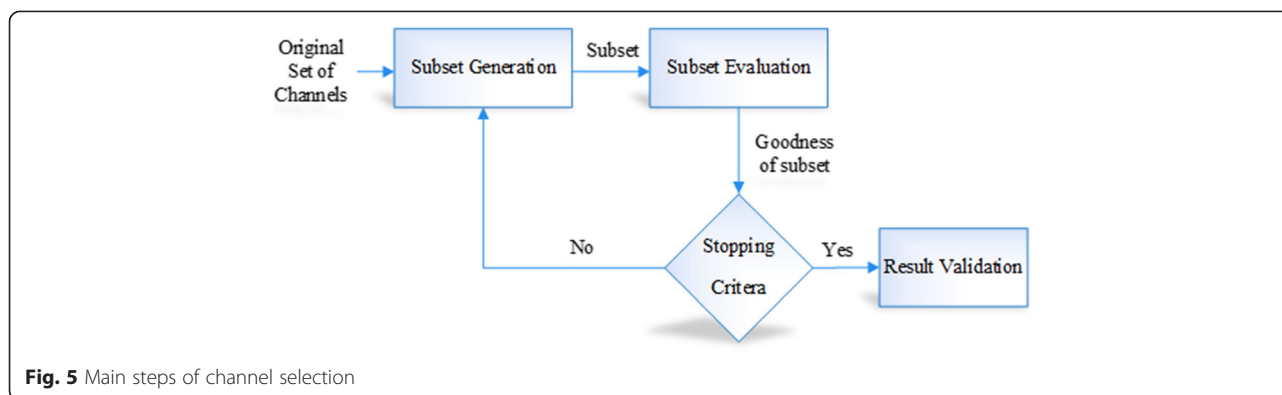
2.1 Filtering techniques

Filtering techniques use an independent evaluation criterion such as distance measure, information measure, dependency measure, and consistency measure to evaluate the candidate channel subsets, which are generated using a search algorithm. Filtering techniques have some advantages among which are the high speed, independence from the classifier, and scalability [10], but they suffer from the low accuracy, since they do not consider the combinations of different channels. Figure 6 shows a general flowchart for the filtering techniques. In this flowchart, S_0 represents the initial subset and S_{best} represents the selected best subset of channels. Also, $D(C_0, \dots, C_{n-1})$ represents a pool of n channels for selection, and M refers to an independent evaluation criterion. The γ represents the value of the evaluation criterion for each subset of channels. The “evaluate” function refers to an evaluation process.

2.2 Wrapper techniques

In case of wrapper techniques, a classification algorithm is used to evaluate the candidate channel subsets, which are generated by a search algorithm as shown in Fig. 7, in which A denotes a classifier, and γ_{best} represents the best value of the evaluation criterion. The evaluation of every candidate is obtained by training and testing a





specific classification algorithm [10]. Consequently, they are more computationally expensive than filtering techniques and they are prone to overfitting.

2.3 Embedded techniques

In the embedded techniques, the channels are selected based on criteria generated during the learning process of a specific classifier because the selection is included into the classifier construction [9]. Embedded techniques achieve an interaction between the channel selection and the classification. They are computationally less expensive and less prone to overfitting. They are based on recursive channel elimination to keep only channels with appreciated magnitude.

2.4 Hybrid techniques

A hybrid technique is a combination of a filtering technique and a wrapper technique attempting to take advantage of both in avoiding the pre-specification of a stopping criterion (see Fig. 8). Generally, hybrid techniques utilize both an independent measure and a mining algorithm for evaluation of the available channel subsets [10]. Two threshold values are evaluated: γ_{best} corresponding to the case with a classifier and θ_{best} corresponding to the case without a classifier. The independent measure is used to select the best subset for a given size Cr (cardinality), and then the mining algorithm is used to select the final best subset across cardinalities.

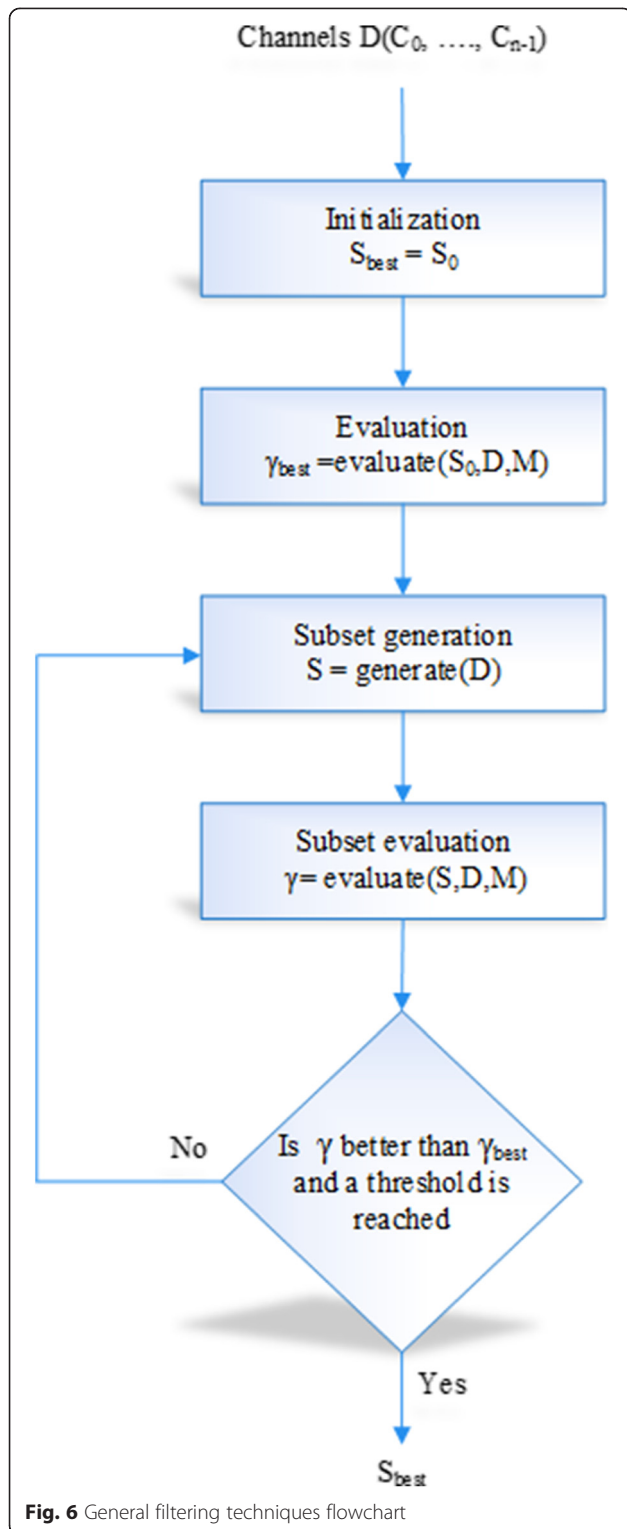
2.5 Human-based techniques

In some applications, a well-trained observer evaluates the outcome of a specific application, like seizure detection, on the selected channels with any of the abovementioned subset generation techniques based on his experience. Thus, the findings of the human-based techniques can be used in a feedback manner to refine the channel selection process.

3 Channel selection for seizure detection/prediction

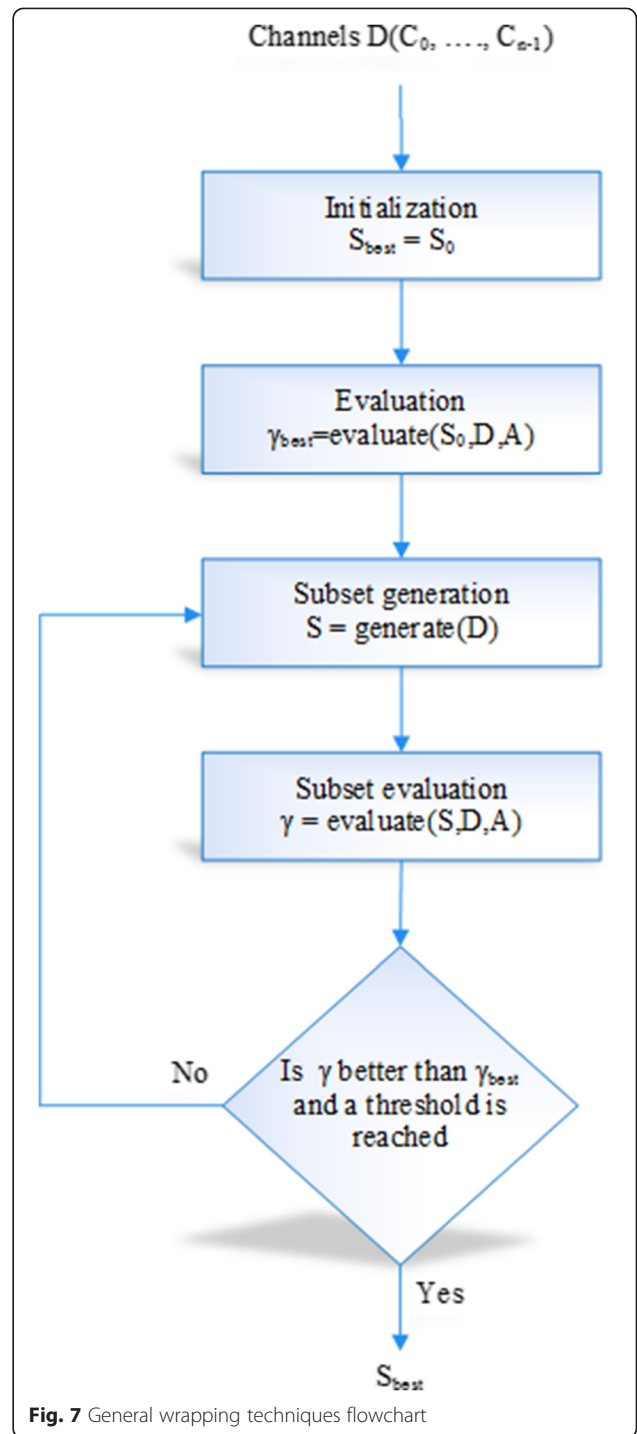
Epilepsy is well known as the second most prevalent brain disorder (after stroke) characterized with unexpected occurrence of seizures. The International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) presented a definition for the epileptic seizure as “a transit occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain” [11]. Epilepsy affects around 1 % of the world population, and based on the seizure statistics of the Epilepsy Foundation of America (EFA), about 200,000 cases of epilepsy are diagnosed per year. The primary tool for diagnosis and management of epilepsy is through EEG signals.

In general, EEG recordings have different channels for signals acquired from different spots of the human brain. In certain applications, there is a need to select some of these channels for EEG seizure detection/prediction because the computational load required for a seizure detection/prediction algorithm increases as a function of the number of channels [12]. Reducing the number of channels is of an utmost importance, for example, in the development of portable medical support systems for epilepsy patients, as reducing algorithmic computational complexity will lead to faster real-time response and lower power consumption to maintain longer time for operation. In addition, the lower the number of channels is, the more convenience the patient would have and the lower the setup time required to fix gel-based EEG electrodes. Another factor that needs to be considered carefully in seizure detection/prediction is the overfitting effect due to the utilization of a large number of redundant channels. Therefore, channel selection could be used to reduce the feature pattern size and lower the computational cost of feature extraction and classification. In what follows, we cover some of the channel selection techniques utilized for seizure detection/prediction and classify them according to the channel subset evaluation techniques given in Section 2.



3.1 Filtering techniques

For EEG seizure detection and prediction, signal statistics such as variance and entropy can be used for channel selection. This subsection presents four trends for channel selection for EEG seizure detection and prediction, with a



common thread that they are all based on signal statistics. Duun-Henriksen et al. [12] investigated different channel selection schemes based on different statistical criteria as follows:

3.1.1 Selection based on variance

The variance of ictal data of all available channels is estimated with the equation:

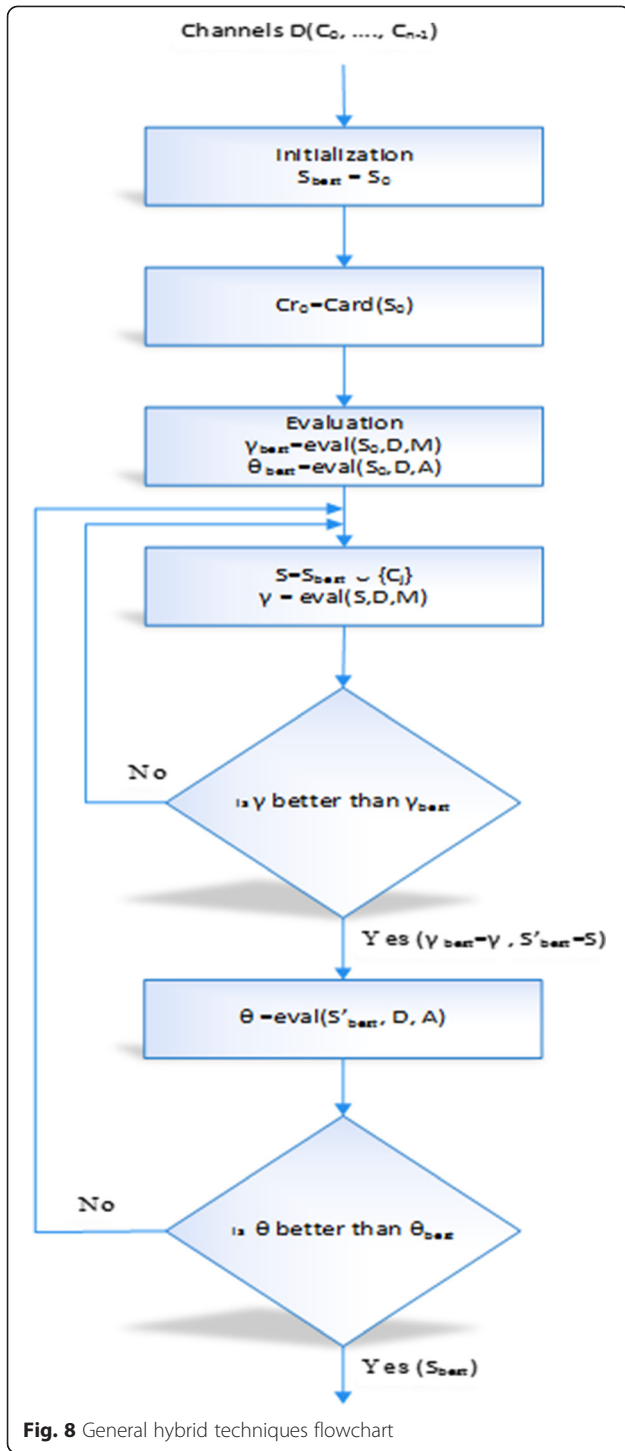


Fig. 8 General hybrid techniques flowchart

$$V_{ict}(c) = \frac{1}{k} \sum_{i=1}^k (x_c(i) - \mu_c)^2 \tag{1}$$

where x_c , μ_c , and k are the seizure data, mean, and number of samples of the training seizure data of channel c , respectively. The *three channels* with maximum variance are selected for signal classification [12].

3.1.2 Selection based on difference in variance

The difference in variance is calculated as follows:

$$V_{diff}(c) = V_{ict}(c) - V_{non-ict}(c) \tag{2}$$

where $V_{non-ict}$ is the non-ictal (non-seizure) training data, which represents the background variance that would be deduced (variance of the normal state). The N input channels are selected based on the minimum difference of variance.

3.1.3 Selection based on entropy

The entropy of an EEG channel is a measure of uncertainty, where the EEG signal is considered as a random variable. The entropy of channel c is defined as:

$$H(c) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \tag{3}$$

where $p(x_i)$ is the probability mass function of the channel having n samples. The N channels with the highest entropy are chosen as input to the automatic seizure detector.

In the method developed by Duun-Henriksen et al. [12], after channel selection, a feature extraction process utilizing wavelet transform is performed on the selected channels. A support vector machine (SVM) was used as a classifier for seizure detection, with a non-linear radial basis kernel having a regularization term of 0.5 and a cost factor varying between 0.05 and 0.5. A recorded iEEG data with 59 seizures and 1419 h from 10 patients were used for training and testing. It was found that the best channel selection method is based upon maximum variance during seizure, which led to a seizure detection sensitivity of 96 % and false detection rate of 0.14/h using three channels. The work of Duun-Henriksen et al. [12] falls in the class of filtering techniques, as it used an independent measure for channel selection, which is the variance. The authors did not use a search strategy.

Faul [13] used another statistical measure, which is the probability of a seizure in a channel produced by the real-time EEG analysis for event detection (REACT) system of Temko et al. [14] as a tool for channel selection to change the behavior of the system aiming to reduce the computational effort. The SVM output in this system is treated as a probability, which is further passed through a sigmoid function. He used a waiting time for each channel to be incorporated in the channel selection process, and this time is computed according to the probability of seizure in that channel. He reported that the computational efforts can be reduced by up to 65 % with no effect on the seizure detection performance of the REACT system (96 % for neonatal and 94 % for adult databases).

Faul and Marnane [15] developed statistics-based dynamic EEG channel selection methods to reduce the power consumption in seizure detection, which can be classified as filtering techniques. In their methods, a number predefined or primary screening channels (PSCs) are chosen in the following approaches depending on the probability provided by the REACT system, and then the channels are added or removed in the next epoch. The first approach is location spread based on some seed PSCs with their neighborhood channels. Channels are added or excluded based on probability of seizure. If the probability of seizure in any of the PSCs exceeds a threshold, the nearest channel to the PSC is added to the analysis and remains until its probability drops below a threshold. Their results show that with two channels {2,7}, the performance is 95.74 % with 66.76 % computational load saving, while with four channels {2,4,6,7}, the performance is 96.55 % with 43.11 % computational cost reduction.

The second approach is idling (single/twin) based on the two brain hemispheres. For idling, channels in the PSCs are analyzed sequentially in alternate epochs, and when the channel is activated, it is then analyzed continuously until deactivation. Both activation and deactivation are based on the probability of seizure. In twin idling, the PSCs in each hemisphere are idled, separately. The performance of the idling approach, without dynamic channel selection, is 91 and 91.48 % and the computational saving is 87.5 and 75 % for single and twin idle, respectively. There is an increase in computational saving with single idling of approximately 10 and 30 % with two PSC and four PSC configurations, respectively. Location spread with twin idling is another approach for the authors, but it did not achieve an appreciable gain in terms of performance over the single idling approach with approximately 10 % drop in computational cost. The results of these approaches have been compared with the all-channel (eight channels) REACT system [14] results leading to a feasibility of the suggested approaches by Faul and Marnane [15].

Atoufi et al. [16] investigated the prediction ability of neuro-fuzzy models in different states of EEG signals: normal, pre-ictal, and ictal. Although the main objective of this study was to improve the model prediction accuracy using information fusion, the main challenge was selecting the channels that should be used to construct the predictor. A selection algorithm was used to select the channels with the largest amount of information about the target (channel whose signal is to be predicted) but with the least information about each other. This technique can be classified as a filtering technique with a greedy sequential search strategy because it depends on independent channel evaluation criteria. An information theoretic criterion (mutual information (MI)) was used to select a group of channels for multi-

channel prediction. The MI of two random variables X and Y is defined as [16]:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X; Y) \end{aligned} \quad (4)$$

where $H(X)$ and $H(Y)$ are the entropies of the variables X and Y and $H(X|Y)$ is the conditional entropy. $H(X; Y)$ is the joint entropy of the two variables X and Y . The authors evaluated their method on two patients' data, of Freiburg database, which is a publicly available intracranial EEG database containing six channels (three focal and three extrafocal) ECoG recordings of 21 patients with a 256 Hz sampling rate [15]. This method achieved 60.6 and 60 % success rates in three-channel cases of ECoG and EEG datasets, respectively. The authors reported that the prediction accuracy using the selection algorithm with multi-channels has been improved with a noticeable improvement in pre-state detection over the single channel.

3.2 Wrapper techniques

In this subsection, we try to bring together the channel selection techniques for EEG seizure detection and prediction that can be classified as wrapper techniques and show why they fall into this category. Shih et al. [17] presented a machine learning-based approach to construct detectors that use fewer channels for seizure onset detection. For selecting channels, the authors used an instance of the wrapper approach, which is a feature selection algorithm with backward elimination. This approach reduced the average number of channels required to detect the seizure onset from 18 to 4.6, while the mean fraction of seizure detection decreases from 99 to 97 %. In addition, the average number of false events per hour decreased from 0.35 to 0.19. An increase of average detection latency from 7.8 to 11.2 s with average of 69 % of energy saving was achieved. When this approach is combined with a patient-specific screening detector, an additional energy saving of 16 % was achieved. Those results were compared to the 18-channel [18] results revealing the feasibility of the channel reduction with this approach.

Glassman and Guttag [19] presented another method that uses recursive feature elimination to design patient-specific SVM detectors that use small numbers of electrodes. The recursive feature elimination uses the SVM to rank the contributions of each selected channel. They used a leave-one-out cross validation principle to estimate the performance of the detectors as illustrated in Table 1. The main idea of the process is to find the smallest number of channels n , such that the average cross validation performance of detectors built using n channels is at least as good as the average cross validation for the 21-channel detector.

Table 1 Using cross validation to select channels for a patient [19]

```

//“Full montage average performance evaluation”
init(All)
for s = each seizure in set of seizures S
    C = 21 channels
    d = Detectorsetup (C, S - {s})
    Modify (All, d, s)
end
//“Subset channel selection performance evaluation”
Numchannels = 21
for n = 20 to 1
    Init(Subset)
    for s = each seizure in set of seizures S
        S' = S - {s}
        C = RFE(n, S') “Find n best channels”
        d = Detectorsetup (C, S')
        Modify (Subset, d, s)
    end
    if Subset >= All
        Numrequired = n
    end
end
return Numrequired

```

The `Detectorsetup (C, S)` is a function to build an SVM detector using the channels in C . It uses S seizures as a training set and `REF(n , S)` as a recursive feature elimination function used to find the best n channels when training on S . The function `Modify(Subset, d , s)` is a function used to calculate the performance of the detector d when used on the file S and update the average performance. This method is based on a wrapper algorithm with a sequential search strategy. The authors evaluated their method using 21-channel scalp EEG recordings of 10 patients. The method achieved 7.1 average channels, 0.011 average false negative, 0.48 average false positive, and 9.54 s average latency time.

Statistical metrics have also been used for wrapper channel selection techniques. Mirowski et al. [20] presented a method based on computing bivariate features of EEG synchronization. These features are cross correlation, non-linear independence, dynamic entrainment, and wavelet synchrony. They computed the features on 21 patients from Freiburg dataset [21]. They used time aggregation for features before classification. Chang et al. [22] proposed a channel selection method to reduce the feature pattern size produced from Mirowski et al. work [20] for seizure prediction. Their work can be classified as a wrapper technique with a pre-specified subset of channels. Their method requires computation of features from pairs of channels of the available EEG signal. They investigated the performances of using all

channel pair combinations with the number of channels from two to six in the case of electrocorticography (ECoG) dataset and 75 combinations of fixed channel pairs in case of EEG dataset. This method aims to reduce the computed wavelet coherence (localized correlation coefficient in time-frequency space) values for a given channel pair over non-overlapping 5 s and frequency bands after decomposing the channels into sub-bands. The features are aggregated into patterns. Then, the SVM is used to classify the patterns into pre-ictal and inter-ictal states as shown in Fig. 9. The authors investigated the performance of all channel pair combinations in the ECoG database and 75 combinations of fixed channel pairs in the EEG database. They evaluated their method using three datasets: Freiburg database [21], CHB-MIT database (6 patients: 1, 3, 6, 7, 10, 22) which is a scalp EEG database with a 256 Hz sampling rate and more than 22 channels for most of them [23], and National Taiwan University Hospital database (one patient) which is also a scalp EEG database with 200 Hz sampling rate and 18 channels [24]. The method achieved 60.6 and 60 % success rate in three-channel cases of ECoG and EEG datasets, respectively. Also, the method achieved more than 93.73 % of computational saving compared to the full 22-channel case.

Greene et al. [25] developed another non-patient-specific statistical method for automated neonatal channel selection and seizure detection based on a regularized discriminant classifier. This method can be classified as a wrapper method with a pre-specified subset channel selection scheme. Seven features were extracted from non-overlapping 8 s, which are spectral entropy, Shannon entropy, spectral edge frequency, non-linear energy, line length, wavelet energy, and root mean square (RMS) amplitude. They compared the effectiveness of their method on a single channel with the training performed on multi-channel EEG. The authors evaluated their method on 17 recordings from 17 neonates with a 251.9 total number of hours and 411 seizures with a 256 Hz sampling rate. Each recording contained 7–11 EEG channels and 1 ECG channel. They examined the performance of nine single channels, which are C4-C3, C3-T3, C4-T4, F3-C3, F4-C4, Cz-C4, Cz-C3, C4-02, and C3-01. Channel C3-C4 gave the best seizure detection performance, when compared to other single channels. It achieved a 90.77 % correct detection and 9.43 % false detection rate, respectively, while the multi-channels achieved an 81.03 % detection rate and a 3.82 % false detection rate.

Another statistical approach was presented by Temko et al. [26]. They presented an online neonate seizure detection framework based on EEG channel weighting and moving average filtering as illustrated in Fig. 10. The authors computed the channel weights on the fly using patient specific history and clinically derived

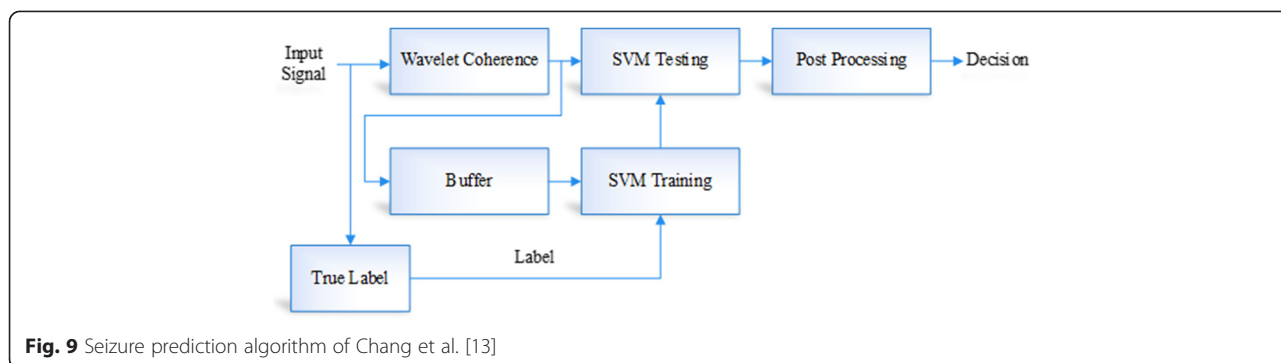


Fig. 9 Seizure prediction algorithm of Chang et al. [13]

prior channel importance. The moving average filtering is used to smooth the SVM output which is interpreted as probabilities. The dataset used consists of 17 newborns EEG recordings obtained from Cork University Maternity Hospital database, Ireland [27]. A total of 36 channels were used for recording the data. Eight channels were selected for further processing using channel weighting. From each channel, 55 features were extracted which are then fed to an SVM classifier. The output of the classifier was then smoothed with a moving average filter and converted to probability-like values using a Bayesian probabilistic framework. These values were then compared with a threshold in the interval [0,1], and based on this comparison, decisions about the presence of seizures per channel were taken. Area under the precision-recall (PR) curve [28] was used as a metric in this work. The authors have shown that with their proposed channel weighting technique, the PR area has increased up to 25 % with the average increase from 81.0 to 84.42 %. Furthermore, it was reported that the performance of the channel weighting algorithm was proportional to the subject observation time.

3.3 Human-based techniques

In this subsection, we explore two human-based techniques for channel selection in EEG signal detection and prediction. Zimbric et al. [29] compared a collection of 3 channels with a collection of 21 channels for the detection of neonatal seizures and quantification of seizure burden. Tracking were analyzed in the three-channel montage for

seizure number, duration, and quantification of seizure burden before reanalysis with the full 21-channel neonatal minimal placement montage. Seizures were identified using standard definitions of EEG seizure. Analysis of the results was performed by two independent readers, and hence, this method can be classified as a human-based technique with a pre-specified subset of channels. Evaluation metrics such as sensitivity, specificity, and reliability were calculated. They evaluated their method using 35 EEG recordings from 28 infants with a total of 1389 min. The sensitivity and specificity of three-channel montage for detecting seizures >10 s were 91 and 100 % for reader 1 and 82 and 96 % for reader 2, respectively.

Tekgual et al. [30] presented a comparison study of reduced electrode montage (9 electrodes) with full 10/20 electrode montage (19 electrodes) considering detection and characterization of neonatal seizures and background EEG characteristics. Three independent readers reviewed EEG recordings for number, duration, and topography of seizures and background features. Hence, we can consider this approach as a human-based approach with a pre-specified subset channel selection. The reviewers started with reduced montage and then the full montage. A total of 151 EEG recordings from 139 infants, obtained from Bio-logic System of the Clinical Neurophysiology Laboratory, Children’s Hospital, Boston [30], were reviewed by the reader on both montages. The sensitivity and specificity of the reduced montage for seizure detection were 96.8 and 100 %, respectively.

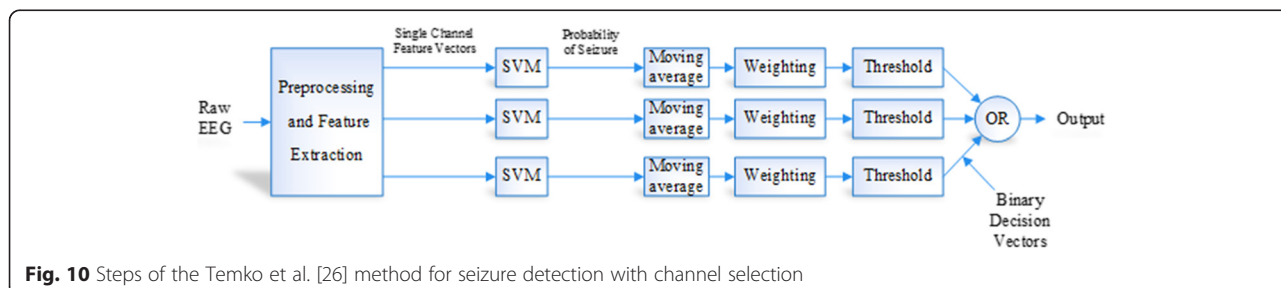


Fig. 10 Steps of the Temko et al. [26] method for seizure detection with channel selection

4 Channel selection for motor imagery classification

Motor imagery is a mental task in which the subject imagines that he is doing an action. Motor imagery classification is very important for certain patients. This task can be performed with EEG signals and may require channel selection to choose the most related channels to the cortical activity patterns and to reduce the computation time as well.

4.1 Filtering techniques

The common thread between the channel selection techniques adopted for motor imagery classification and categorized as filtering techniques is that they are based on EEG signal statistics. He et al. [31] presented a statistical channel selection method for classifying motor imagery. This method used Bhattacharyya bound of common spatial patterns [32] as an optimal index and fast forward search to find the optimal combination of channels. It is a filtering method with sequential search strategy for subset channel selection. Then, it uses Bayes algorithm [33] as a classifier. The authors utilized four EEG recordings for subjects a, b, d, and e of dataset 1 of BCI competition IV to evaluate the performance of their method, each of which contains 200 trials. A total of 59 channels were used for the recordings. They reported that the classification accuracies obtained by their method, which is $\sim 95\%$ with ~ 33 average number of channels, is higher than those obtained by all channels, but still their solution is a sub-optimal solution.

Another statistical method was presented by Tam et al. [34], who proposed a channel selection method for motor imagery classification based on the sorting of common spatial pattern (CSP) filter coefficients. Their method, called CSP-rank, is based on a filtering approach with a sequential search strategy for subset channel selection. It uses two CSP filters for two classes corresponding to motor imagery and immobilization. It firstly sorts the absolute value of the filter coefficients in each filter and then selects the electrode with the next largest coefficient in turn from the two spatial filters. They utilized 64-channel EEG recordings from five chronic stroke patients through 20 sessions and each session consisted of 80 trials. They compared CSP-rank with support vector machine recursive feature elimination (SVM-RFE) [35, 36] and random selection. CSP-rank was able to maintain an average classification accuracy rate above 90% for 8–38 electrodes. It obtained the highest average classification accuracy rate of 91.7% with 22 electrodes. The SVM-RFE maintained average classification accuracy rate above 90% for 12–28 electrodes and achieved the highest average classification accuracy rate of 90.7% with 14 electrodes. Random selection maintained an average classification accuracy rate above 85% for 10–50 electrodes and obtained the highest

average classification accuracy rate of 89.6% with 32 electrodes.

Yong et al. [37] presented another statistical channel selection method for classifying two motor imageries based on introducing l_1 norm regularization term in the CSP algorithm which supports sparsity in the weights of the spatial filter. This method adopts a filtering approach with a pre-specified subset channel selection strategy based on experience. The EEG data used for evaluating this method was recorded from five subjects (aa, av, al, aw, ay) using 118 channels and a 1 kHz sampling rate, provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and University of Medicine Berlin (Neurophysics Group) with two classes: right hand and right foot motor imageries [38]. Each class consisted of 140 trials. The method was able to reduce the number of channels on average to 13 electrodes (of 118 electrodes), while the average classification accuracy rate dropped from 77.3 to 73.5% only. The value of the regularization parameter is subject-specific and was selected manually. Therefore, it needs to be chosen automatically to produce reasonable results.

Meng et al. [39] presented an automated channel selection method based on CSP in BCI systems. The CSP algorithm is commonly used to derive spatial filters for the multi-channel EEG signals. However, it is known that the performance of the CSP degrades due to the overfitting problem, when the number of channels is large. Therefore, to reduce the number of channels, the authors used a heuristic algorithm, namely, l_1 norm, to select the most useful channels, and then extract the features from the selected channels using the CSP. They initially applied the CSP to the datasets and then scored the channels based on their l_1 norm. The channels with the highest scores were selected for further processing, while the others were excluded. Using the CSP, features are extracted only from the selected channels and are forwarded to the classifier. This algorithm adopted a filtering approach. It was evaluated on datasets provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and University of Medicine Berlin (Neurophysics Group). A total of 118 electrodes were placed on the scalp to record the data with a sampling rate of 1 KHz for five subjects [38]. The algorithm was compared with a commonly used γ^2 algorithm for channel selection [40]. It was shown that, with Meng et al. algorithm, the classification accuracies increased from 80.8 to 82.4, 97.5 to 98.6, 72.2 to 76.8, 93.6 to 94, and 92.1 to 96.6% for all the five subjects with 20 channels.

Wang et al. [41] proposed a channel reduction method in motor imagery, in which the prominent channels in this method were selected using the maximum of spatial pattern vectors obtained with the CSP algorithm. Event-

related desynchronization (ERD) and readiness potential (RP) of the selected EEG channels were used as features. Using these features, EEG signals were classified using a Fisher discriminant (FD) classifier [42–44]. This method was evaluated on a datasets provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and University Medicine Berlin (Neurophysics Group) [38]. It was shown that the classification accuracies with four channels were 93.45 and 91.88 % for two subjects. Increasing the channels to eight increased the classification accuracies to 96.68 and 93.25 %, respectively, at the expense of decreasing the convenience of the system user.

The effect of increasing the number of channels on the classification accuracy of EEG-based motor imagery has been presented by Shan et al. [45]. They adopted a filtering approach with a sequential search strategy for subset channel selection. They used two different datasets, an imagery-based cursor movement control dataset and a motor imagery tasks dataset for comparison. In the first dataset, 64 channels were used for recording the data with a sampling rate of 200 Hz. Similarly, in the second dataset, 59 channels were used for recording data with a sampling rate of 100 Hz. A modified time-frequency-spatial synthesized method was used for right and left motor imagery classification. It was observed that increasing the number of channels increases the classification accuracy in the first dataset, while it is not the case in the second dataset in which the optimum accuracy is achieved at a subset of channels. Increasing the number of channels from two to all in the first dataset increased the training and testing classification accuracies from 68.7 to 90.4 and 63.7 to 87.7 %, while in the second dataset, it was observed that the classification accuracy increased till 16 channels and then significantly decreased from 81.3 to 68.9 % for all channels. It was concluded that the performance of online BCI systems increases by increasing the number of channels in contrast to the offline motor imagery tasks paradigm.

EEG patterns in a BCI system vary from the first session to subsequent sessions on other days due to several subjects' preconditions. Therefore, there is a need for a robust and stable channel selection algorithm across different sessions. Arvaneh et al. [46] presented a robust channel selection approach across sessions in BCI system involving stroke patients adopting a filtering approach with pre-specified subset channel selection based on experience. They proposed a robust sparse common spatial pattern (RSCSP) algorithm for optimal EEG channel selection across different sessions, where the estimates of the covariance matrices of EEG measurements are replaced with the robust minimum covariance determinant (MCD) estimates. The stability and robustness of this algorithm were evaluated by comparison with five existing channel selection algorithms across 12 different sessions of motor

imagery-based datasets from 11 stroke patients. A total of 27 channels were used for recording the data with a sampling rate of 250 Hz. Eight channels were selected using the RSCSP algorithm from the first session and were evaluated on the 11 subsequent sessions. The results showed that the RSCSP algorithm outperformed other algorithms like SCSP, CSP, MI, Fisher criterion (FC), and SVM by an average accuracy of 0.88, 2.85, 2.69, 4.85, and 4.58 %, respectively.

He et al. [47] presented a Rayleigh coefficient (RC) maximization-based genetic algorithm (GA) for channel selection in motor-imagery BCI system adopting a filtering approach with a random search strategy for subset channel selection. This algorithm uses the CSP to diagonalize the covariance matrices and maximize the difference of variances of two classes. On the other hand, RC maximization is performed not only for maximizing the difference of covariance of two classes but also for minimizing the sum of these two covariance matrices. Hence, the RC features can be more discriminating than CSP. However, like CSP, the performance of the RC maximization is deteriorated with the redundant electrode channels. Therefore, the authors proposed an improved GA for channel selection based on RC maximization. Using this algorithm, Fisher ratios for every single channel were calculated and ranked in descending order. The first selection of the subset of channels was made through the maximum Fisher ratios of the channels. An improved GA based on RC maximization was then applied on the selected channels to get the optimum subset of channels. This algorithm was evaluated on two datasets. In the first dataset, 118 channels were used to record the data with a sampling rate of 100 Hz, and in the second dataset, 32 channels were used for recording the data with a sampling rate of 250 Hz. It was observed that the RC-GA achieved high classification accuracy with lower computational cost. The average accuracies are 88.2 and 89.38 % for the first and second datasets, respectively. The performance of this algorithm was also compared with other channel selection algorithms like SVM-GA, Sequential Forward Search (SFS), and Sequential Backward Search (SBS) algorithms. It was shown that RC-GA provided more compact selected channels, while acquiring higher classification accuracy than the other mentioned algorithms.

Arvaneh et al. [48] proposed an SCSP algorithm for subject-dependent channel selection in BCI systems adopting a filtering approach with a pre-specified subset channel selection scheme. They formulated the SCSP algorithm as an optimization problem to select the minimum number of channels within a constraint of classification accuracy. The CSP is usually used to derive spatial filters for the multi-channel EEG signals. However, the weights of the CSP are very dense. The CSP algorithm is

sparsified by adding a l_p norm, $0 < p < 1$, regularization term into the optimization problem. The performance of this algorithm was evaluated using two datasets, Dataset IIa [49] from BCI competition IV with 22 channels (four motor imagery tasks; left hand, right hand, feet, or tongue) and Dataset IVa [50] from BCI competition III with 118 channels (two motor imagery tasks; right hand and foot) recorded from 14 subjects. It was shown that the SCSP algorithm achieved the best classification accuracy by reducing the number of channels and an improvement of 10 % in classification accuracy compared to the three channels case (C3, C4, and Cz). The average accuracy rates for SCSP1 (maximizing the accuracy by removing noisy and irrelevant channels) were 81.63 and 82.28 % for Dataset IIa and Dataset IVa with average number of channels 13.22 and 22.6, respectively. The average accuracy rates for SCSP2 (minimizing the selected channel(s) while maintaining the accuracy comparable to all channels' accuracy) were 79.07% with 8.55 average number of channels and 79.28% with 7.6 average number of channels for the first and second datasets, respectively. It is also shown that this algorithm outperforms other existing channel selection algorithms based on Fisher criterion, mutual information, SVM, CSP, and RCSP, in classification accuracy.

4.2 Wrapper techniques

Some of the adopted channel selection techniques for motor imagery classification are categorized as wrapper techniques. Yang et al. [51] presented a subject-specific channel selection method based on criteria derived from Fisher's discriminant analysis to measure the discrimination power of time-domain parameter (TDP) features extracted from different channels and different time segments for classification of two motor imagery tasks, right hand and right foot. This method adopts a wrapper approach with pre-specified subset channel selection depending on experience. The authors utilized the dataset IVa from BCI competition III [52], which consists of EEG recordings from five subjects using 118 electrodes. The subjects performed 280 trials of cue-drive motor imagery (right hand, 140 trails; right foot, 140 trails) and each trial lasted for 3.5 s. This method reduced the number of channels from 118 to no more than 11 channels without a significant decrease in the accuracy rate (78 % mean accuracy rate).

Wei and Wang [53] presented a method for channel selection during the classification of motor imagery of left hand, right hand, and foot based on a binary multi-objective particle swarm optimization algorithm. This method adopted a wrapper approach with a random search strategy for subset channel selection. It extended the particle swarm optimization algorithm shown in Fig. 11 to handle two objectives: minimizing the number

of selected channels and maximizing the sum of three mutual information metrics. The classification accuracy rate was calculated with three different classifiers: KNN, SVM, and back-propagation (BP) network. This method utilized EEG recordings from five health subjects from 22 channels with 256 Hz sampling rate. The experiment consisted of six runs separated by 5 min break, and each run included 60 trails (120 trails for each class). The results showed that the highest accuracy rate was around 91 % with nine channels in subject two, while the accuracy rate with all channels was around 92 %. Similarly, the data set of subject 3 showed the lowest accuracy of around 75 % with 14 channels and around 76 % with all channels.

Zhou and Yedida [54] presented a method for the reduction of the number of channels for the task of classifying mental states for shoulder and elbow movement intentions for healthy and stroke patients. Their method is based on combining the support vector channel selection with a weighted time-frequency synthesis classification algorithm [54]. It is classified as a wrapper method with a sequential search strategy for subset channel selection. The authors evaluated their method using EEG recordings from two able-bodied (healthy) and one stroke subjects. A total of 131 channels were used, and the sampling rate was 256 Hz. This method was able to achieve higher than 90 % classification accuracy rate for

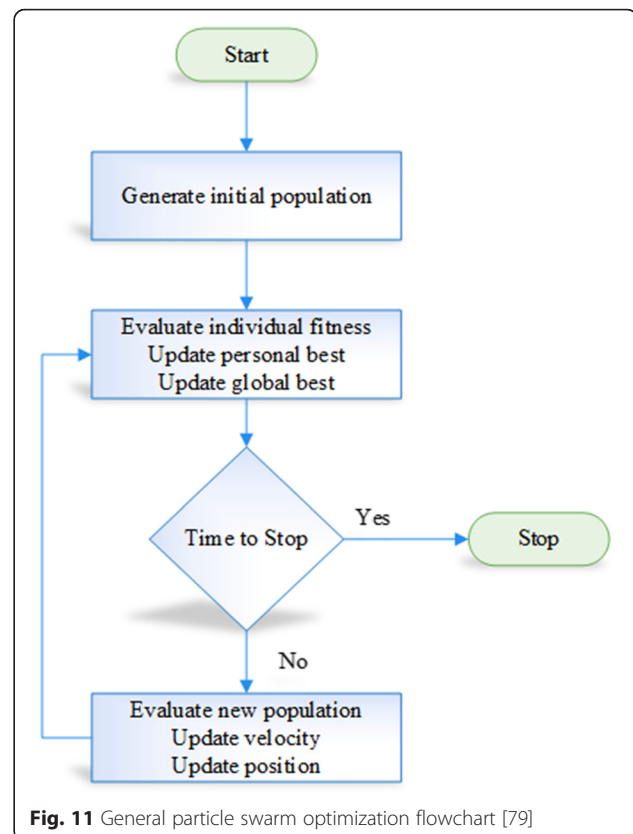


Fig. 11 General particle swarm optimization flowchart [79]

the healthy subjects when the total number of channels N was $20 < N < 131$ and $50 < N < 131$, for the first and second subjects, respectively. On the other hand, the classification accuracy rate was below 85 % for the subject with stroke.

Kamrunnahar et al. [55] presented a systematic optimization algorithm for the optimization of the number and locations of electrodes in BCI systems adopting a wrapper approach with a complete search strategy for subset channel selection. A human scalp EEG data were recorded in response to cue-based motor imagery tasks. A total of 19 channels were used for recording the data with a sampling rate of 256 Hz that was passed through a band-pass Butterworth filter with cutoff frequencies of 0.5 and 60 Hz. To increase the spatial resolution of the recorded data and decrease its dependence on the reference location, the authors used two techniques: Laplacian derivation [56–59] and common average reference (CAR) [55]. They used a model-based autoregressive technique to extract the features. For selecting the optimum number of channels, all possible combinations of channels were calculated. Task discrimination errors were calculated using linear discriminant analysis (LDA) [44] for each combination. The channel combination with the lowest discrimination error was selected as the optimal selection for a specific subject. The average classification errors for subject one were ~ 21.75 with four channels and ~ 28.28 with three channels for tasks one and two, respectively. The performance of this algorithm was evaluated by comparison with another feature selection algorithm, namely, forward stepwise feature selection [60, 61].

Yang et al. [62] presented an artificial neural network and genetic algorithm approach for channel selection and classification of EEG signals in BCI systems adopting a wrapper approach with a random search strategy for subset channel selection. Conventional ANN-based approaches have problems of the lack of explicit input optimization, and their learning results are not easily understood. Therefore, the authors proposed a generic neural mathematic method (GNMM) for EEG channel selection and classification problems, aiming to focus on the issues above. The GNMM consists of three steps [63, 64]: channel selection based on the GA, pattern classification using multi-layer perceptron (MLP), and rule-extraction based on mathematical programming. The channel appearance percentage was used in the GA to optimize the input channel selection. After channel selection, the MLP was used for pattern classification, and finally, regression rules were extracted so that training results can be easily implemented. This technique was evaluated on two datasets. The first dataset contains ECoG signals recorded using an 8×8 electrode grid in touch with the brain at a sampling rate of 1000 Hz, and

the subject had to imagine movements of either the little figure or the tongue. In the second dataset, 32 channels were used for recording EEG signals with a sampling rate of 256 Hz, where the participants had to execute left-hand or right-hand button press. Using the GNMM proposed by the authors, 10 channels were selected in the first dataset which achieved a classification accuracy of about 80 %. For the second dataset, six channels were selected with which they achieved a classification accuracy of 86 %.

4.3 Embedded techniques

Lal et al. [40] adopted feature selection algorithms, recursive feature elimination (RFE), and zero-norm optimization based on the training of SVMs for channel selection and demonstrated the usefulness of these operations on motor imagery classification. Their work adopted an embedded approach with a sequential search strategy for subset channel selection. The authors evaluated their method utilizing 39 EEG channel recordings from five subjects (A, B, C, D, and E). A band-pass filter with cutoff frequencies 0.1 and 40 Hz was applied, and the sampling rate was 256 Hz. With every subject, they recorded 400 trials, and each trial lasted for 9 s tasking every subject to imagine left versus right hand movements during each trial. They found that the RFE and zero-norm optimization are capable of reducing the number of channels without increasing the error. The average error rate for 17 channels (located over or close to the motor cortex) over the five subjects using RFE was reported to be 23 %, while the average error rate using 12 channels was 24 %.

Schroder et al. [65] presented a robust EEG channel selection algorithm across subjects in BCI systems adopting an embedded approach with a sequential search strategy for subset channel selection. They tried to investigate whether channels selected for one subject are useful to the others as well. Data were recorded from eight male subjects using 17 EEG channels with a sampling rate of 256 Hz. For each subject, Welch's method [66] was used to extract the features, which were then fed to the linear SVMs for classification. The authors used a recursive channel elimination (RCE) method for the cross subjects channel selection. Using the RCE, the importance of the channel is determined by its influence on the margin of a trained SVM. After applying the RCE on the datasets from different subjects for cross channel selection, it was observed that it cannot only be used successfully for channel selection in individual subjects but also proved helpful in channel selection across different subjects with low error rates. The average error rate is 26.9 % with more than 32 channels.

4.4 Hybrid techniques

Li et al. [67] proposed a method for selecting suitable channels for classification of two motor imagery tasks: right hand and right foot based on a common spatial pattern algorithm as shown in Fig. 12. This method is based on a wrapper approach with a complete search strategy for subset channel selection. The l_1 norms of common spatial pattern features are used to compute the contribution scores $D_{\text{foot_score}}^i$ and $D_{\text{hand_score}}^i$ for the i th channel. The channels with larger values of $D_{\text{foot_score}}^i$ and $D_{\text{hand_score}}^i$ are used to obtain channel rankings A and B , where A is the channel ranking under right-hand motor imagery task and B is the channel ranking under foot imagery task. The first $m \leq 59$ channels are selected from A and removed from B to obtain channel ranking C . After that, the first n ($n \leq 118 - m$) channels of C are selected. The channel combination is obtained by

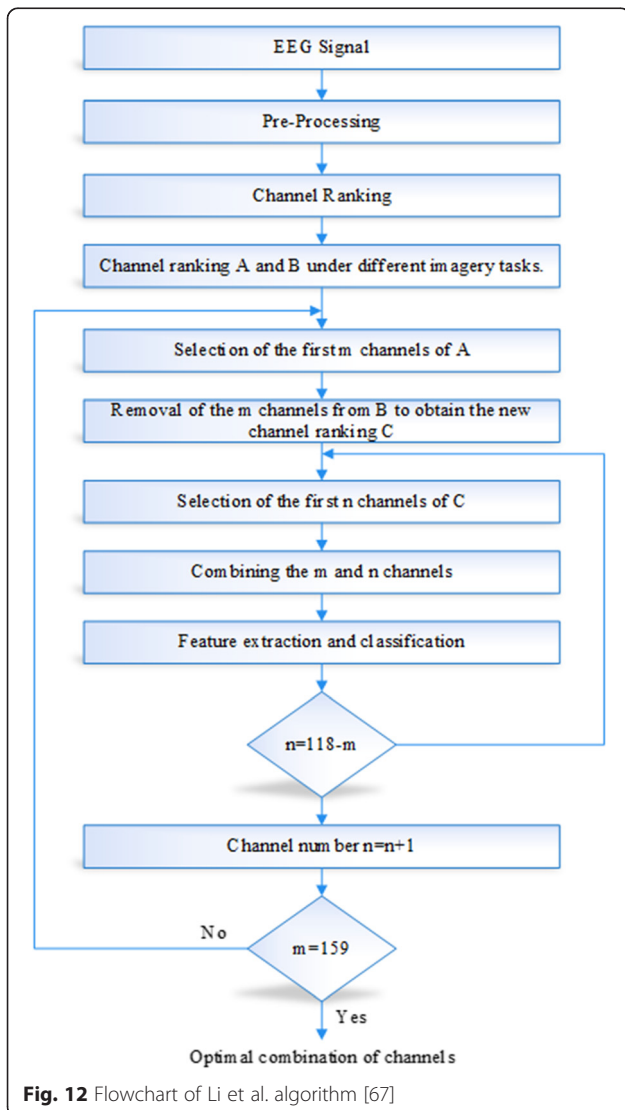


Fig. 12 Flowchart of Li et al. algorithm [67]

combining both m and n . Finally, the optimal combination of channels is selected by comparing the classification accuracy rates using an SVM with all combinations. The authors evaluated their method on the datasets of two subjects: “aa” and “a1” from the dataset IVa from BCI competition III using 118 electrodes [52].

For the dataset “aa,” the highest classification accuracy rate was 92.34 % using seven channels, while the classification accuracy rate with all channels was 90.54 %, respectively. Similarly, for the data set “a1,” the highest classification rate was 94.63 % using eight channels, while it was 90.82 % with all channels.

5 Channel selection for emotion classification

Human emotions are thought to be discrete in nature with distinguishable EEG signals. The process of emotion classification based on EEG signals may require some sort of channel selection to save computation time. In addition, there is a certain area in the brain that is concerned with emotions, which makes channels from other areas unrelated to emotion classification. Channel selection approaches adopted for emotion classification can be categorized to filtering and wrapper techniques.

Rizon et al. [68] proposed an asymmetric ratio (AR) (asymmetric variance ratio and amplitude asymmetric ratio) based channel selection method for human emotion recognition from EEG signals as illustrated in Figs. 13 and 14 and Table 2. The ratio of variances between hemisphere channels was used as an indicator for assessing the region of the brain and the channels associated with emotion detection. The spectral power ratios between hemisphere channels are used to precisely estimate the electrical activity. The asymmetric variance ratio (AVR) is defined as [68]:

$$\text{AVR} = \frac{V(i) - V(j)}{V(i) + V(j)} \quad (5)$$

where $V(i)$ is the variance of left hemisphere channel, $V(j)$ is the variance of the right hemisphere channel, $i = 0, 1, 2, \dots, N$, $j = 0, 1, 2, \dots, N$, and N is the number of homogeneously distributed electrodes on left and right hemispheres.

The amplitude asymmetric ratio (AAR) is given by [68]:

$$\text{AAR} = \frac{P(i) - P(j)}{P(i) + P(j)} \quad (6)$$

where $P(i)$ is the spectral power of left hemisphere channel, $P(j)$ is the spectral power of right hemisphere channel, $i = 0, 1, 2, \dots, N$, $j = 0, 1, 2, \dots, N$, and N is the number of electrodes on left and right hemispheres.

The method of Rizon’s et al. is a filtering approach with a pre-specified subset of channels selected by a

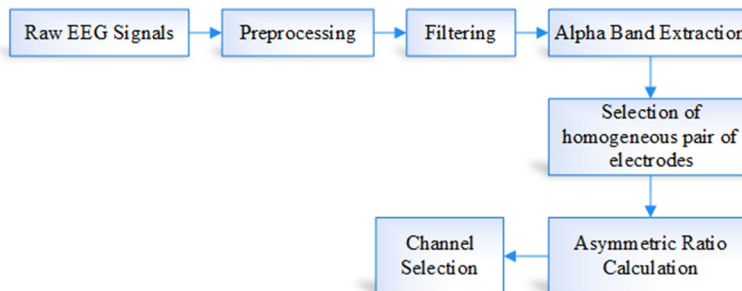


Fig. 13 Channel selection with asymmetric ratio calculation

human expert. It was evaluated using 63 channel EEG recordings (28 pairs, seven center electrodes) from five healthy subjects with a 256 Hz sampling rate and a band-pass filter between 0.05 and 70 Hz with five different classes of emotions (disgust, happy, surprise, sad, and fear). In this method, features are extracted from the wavelet domain using Daubechies 4 (db4) wavelet transform. The results show that this method reduced the 28 pairs of channels to 2. For validating the method, the authors employed a fuzzy C-Means clustering algorithm to classify the emotions [68]. Its results support their findings.

Jatupaiboon et al. [69] proposed a method to classify two emotions based on EEG signals, which are positive and negative emotions elicited by pictures. They extracted the power spectrum from five bands and used SVM as a classifier in a wrapper channel selection evaluation approach. In their experiment, they used a manual approach for reducing the number of channels. They utilized EEG recordings from 11 participants, whom have been shown 100 pictures (50 positives and 50 negatives) from Geneva Affective Picture Database (GAPED) [70]. The authors used EMOTIVE headset with 14 channels [71] for recording with 4-s epochs and 50 % overlapping. They achieved an 85.41 % accuracy rate with seven pairs (14 channels: full) and 84.18 % accuracy rate with five pairs, respectively. The authors found also that frontal pairs of channels and high-frequency bands give higher accuracy than other pairs of channels and lower frequency bands.

6 Channel selection for mental task classification

Mental task classification is a new and challenging trend in EEG signal processing. The main objective of

this classification process is enabling a patient to communicate with the outside world without physical movement. This classification process may require channel selection as a pre-processing step to reduce the computation time.

6.1 Filtering techniques

Lan et al. [72] presented an ambulatory cognitive state classification system to assess the user's mental load based on EEG measurements. Their work focused particularly on dimensionality reduction (channel selection and feature projection) utilizing mutual information techniques as shown in Fig. 15. This work is based on a filtering approach with a sequential search strategy for subset channel selection. In order to select an effective subset of the available channels after pre-processing by artifact removal and band-pass filtering, the authors used a forward incremental method. Three classifiers, Gaussian mixtures model (GMM), K-nearest neighbor (KNN), and Parzen, were used to classify the feature vector based on majority voting in a fusion filtering process. The authors used 32 channels of EEG recordings from three subjects performing two mental tasks (n-back, Larson) at two difficulty levels (low, high) with 256 Hz sampling rate to evaluate the method. The n-back task is a continuous performance task used to measure a part of working memory. The Larson task requires the subjects to maintain a mental count according to the presented configuration of images on the monitor. For performance evaluation, the data was divided into five sets and each set was saved for testing and the other four were used for training. The average classification accuracy was around 80 % for all subjects

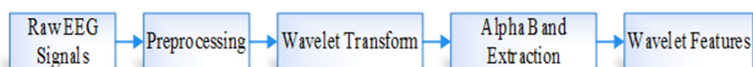


Fig. 14 Feature extraction in Rizon's method

Table 2 Rizon's channel selection algorithm for emotion classification [68]

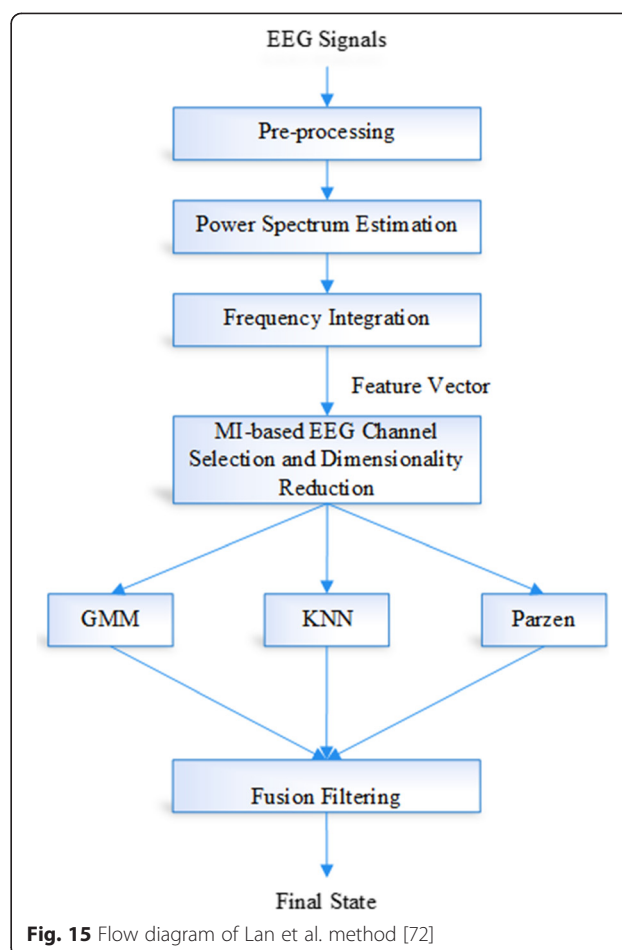
1. The raw EEG signals from five subjects over five discrete emotions are collected using 63 electrodes which are placed through standard International 10/20 system on the scalp.
2. The signals are pre-processed by mean removal and variance normalization.
3. The signals are filtered using 5th order band-pass filter at a cutoff frequency of 0.05 Hz – 45 Hz.
4. The signals are divided into five different EEG frequency bands using 5th order Butterworth filter.
5. Alpha band is used for channel selection.
6. From the 63-channel EEG signals, only 28 homogeneous channel-pairs are separated out for calculating the asymmetric ratio for channel selection.
7. For each subject, the values of AVR and AAR are calculated for all the 28 homogeneous pairs of electrodes using the equations (6) and (7).
8. If all values of AR are positive or negative for five emotions this leads to a rank of 5, for four emotions this leads to a rank of 4, etc.
9. The channel pairs of higher ranks of AAR and APR are sorted as the dominant channels for emotion recognition.

when the number of channels was 7, 10, or 7 channels selected based on literature suggestions.

6.2 Wrapper techniques

Chai et al. [73] presented a method for EEG mental task classification using a genetic algorithm-based neural network classifier. The method of Chai et al. is a wrapper-based method with a pre-specified subset of channels. They used six non-imagery tasks which are arithmetic (math) by imagining and solving simple multiplication, letter composing by mentally composing simple words, Rubik's cube rolling by imaging a Rubik's cube rolled forward, visual counting by mentally counting numbers from one to nine, ringtone by imagining a familiar mobile ringtone, and spatial navigation by moving around and scanning the surroundings in a familiar location in mind. Two methods of feature extraction were used and compared: power spectral density (PSD) and Hilbert Huang transform (HHT). For recording, they used a monopolar EEG system from Compumedic company with 256 Hz sampling rate. Five participants were involved in this experiment with 10 sessions recording for each mental task. The accuracy rate for classifying three mental tasks using the original eight channels is between 76 and 85 % using PSD feature extractor. In case of two channels with PSD feature extractor, the accuracy rate was between 65 and 79 %, and with HHT feature extractor, the accuracy rate was between 70 and 84 %.

Tavakolian et al. [74] presented a channel reduction method for classifying three mental tasks (baseline,

**Fig. 15** Flow diagram of Lan et al. method [72]

multiplication, and geometric figure rotation) based on genetic algorithms for subset generation as shown in Fig. 16. This method is based on the wrapper approach with random search strategy for subset channel selection. They used a feed-forward neural network as a classifier, and its outputs were averaged and considered as the performance function of the genetic algorithm. The genetic algorithm was used to find the best six channel combinations of 19 channels. The method was evaluated using 19 channel EEG recordings from five subjects with a 250 Hz sampling rate. In each session, every task was repeated two times and each time lasted for 10 s. The results showed that the classification accuracy rates were 100, 99.6, 96.66, and 88 % for subjects 1, 3, 5, 2, and 4, respectively.

7 Channel selection for sleep state classification

Sleep state classification is very important for infants as well as adults. This classification process can be performed with EEG signals. This field has been the subject of interest of several neuroscience researchers. It requires also some sort of channel selection to obtain

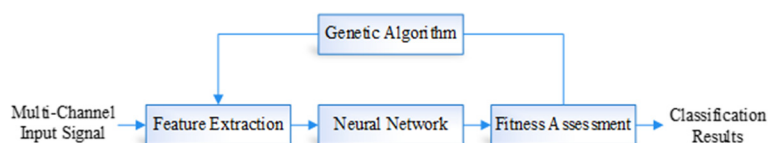


Fig. 16 Tavakolian et al. method [74]

robust classification results because at least one EEG channel combined with one electromyography (EMG) and one electrooculography (EOG) are required for manual scoring. Piryatinska et al. [75] presented a channel selection method for neonate EEG sleep state classification using a multivariate analysis approach adopting filtering with a complete search strategy for subset channel selection. It has two main stages: scoring of sleep stages based on each combination of EEG channels and selection of the optimal channel combination. The latter consists of three steps: producing two rankings, one for the full term and one for the preterm neonates, of the channel combinations; selecting the channels that appear most often in the top channel combinations; and validating these selections with a cross validation methodology. They used EEG sleep signals from 36 neonates (20 full term and 16 preterm) recorded from 14 channels at a sampling rate 64 Hz. This method achieved 87.20 % a mean agreement percentage (MAP) with five channels (compared to physician's scores) and 87.41 % MPA with four channels for full term and preterm, respectively.

8 Channel selection for drug effect classification

Ong et al. [76] presented a channel selection algorithm for visual evoked potentials (VEP) based on principle component analysis (PCA). The VEP is a small electrical potential originating from the brain in response to a visual stimulus. This algorithm was used to classify alcoholic and non-alcoholic subjects. PCA transforms the dataset into a new set of variables called principle components to be ranked from the highest to the lowest bases. The first few principle components usually contain most of the variation present in the original dataset. The performance of this algorithm was evaluated on a VEP dataset recorded from 20 subjects: 10 are alcoholic and 10 are non-alcoholic. A total of 61 channels (variables) were used with a signal being sampled at 256 Hz. The authors selected 16 optimal channels, since they contributed 98.563 % of the total variance. Gamma band powers of the selected channels were used as features for classification. An MLP neural network was used as a classifier to classify the alcoholic and non-alcoholic subjects. The gamma band power

was used as an input feature to the neural network. It was concluded that the classification performance using all the channels was 95.83 % and using 16 channels was 94.06 %, which are very close.

9 Conclusion and future research directions

This paper explored some EEG channel selection techniques for different applications taking into consideration the different criteria developed in the literature for channel selection evaluation and search strategy. The paper introduced the basic notations and procedures of the channel selection process. It presents a description of channel selection approaches for a variety of applications. The comprehensive study in this paper has revealed that it is possible, without much loss in the performance of the classification/detection tasks, to make use of a small set of EEG channels ranging from 10 to 30 % of the available channels. This will in turn reduce the processing complexity with less setup time and maintain the subject's convenience by having less electrodes. In some applications, such as sleep state classification, there are dominant channels responsible for the activity of concern and need to be determined. In some other applications, such as seizure detection and prediction, the use of all channels may lead to an overfitting effect during the classification process. Table 3 summarizes the channel selection techniques surveyed in Sections 3–8 as contrasted to each application. These techniques have been tested using different databases. Therefore, an extensive study is needed to determine the channel selection technique that gives the highest performance score, when all techniques belonging to a specific application are applied to a unified database. Another important issue is to investigate channel selection techniques for emerging applications based on visual and auditory-evoked potential [77]. Finally, it is observed that channel selection algorithms are in general based on features extracted from the EEG signals. Finding features well representing all EEG signal states is still a challenging task that needs further research. It is observed from this study that channel selection has been investigated intensively in motor imagery classification with a variety of techniques. So, extending these techniques to other applications will be useful. For wrapper, hybrid, and embedded channel selection techniques, the performance

Table 3 Summary of channel selection methods

Technique	Subset channel	Evaluation method	Performance metrics	Application
	Selection strategy			
Duun-Henriksen et al. [12]	Statistical criteria	Filtering	Sensitivity, 96 %; false detection rate, 0.14/h (with three channels)	Seizure detection
Faul [13]	Statistical criteria		Average accuracy, 96 % on neonatal database; average accuracy, 94 % on adult database; computational effort saving, 65 %	
Faul and Marnane [15]	Statistical criteria		Average accuracy rate, 95.74 % (location spread), 91 % (single idle), and 91.48 % (twin idle), (with 2 channels)	
Atoufi et al. [16]	Sequential search		Average accuracy rate, 60 % (for EEG dataset with three channels)	
Shih et al. [17]	Sequential search	Wrapper	Average accuracy rate, 97 %; average detection latency, 11.2 s (with 4.6 average number of channels)	
Glassman and Guttag [19]	Sequential search		Average false negative, 0.011; average false positive, 0.48; and average latency time, 9.54 s (with 7.1 average number of channels)	
Chang et al. [22]	Pre-specified		Average accuracy rate, 70 % (for EEG dataset with three channels); average energy saving, 93.73 %	
Greene et al. [25]	Pre-specified		Average accuracy rate, 90.77 % (with single channel, C3-C4)	
Temko et al. [26]	Channel weighting		Average precision-recall, 84.42 (with 8 channels)	
Zimbric et al. [29]	Pre-specified		Average sensitivity, 86.5 %; average specificity, 98 % (with 3 channels)	
Tekgul et al. [30]	Pre-specified	Human-based	Sensitivity, 96.8 %; specificity, 100 % (with 9 channels)	
He et al. [31]	Sequential search	Filtering	Average accuracy rate, ~95 % (with ~33 average number of channels)	Motor imagery Classification
Tam et al. [34]	Sequential search		Highest average accuracy rate, 91.7 % (with 22 channels)	
Yong et al. [37]	Pre-specified		Average accuracy rate, 73.5 % (with 13 average number of channels)	
Meng et al. [39]	Heuristic algorithm		Average accuracy rate, 89.68 % (with 20 channels)	
Wang et al. [41]	Maximum of spatial pattern vectors		Average accuracy rate, 92.66 % (with 4 channels) and 94.96 % (with 8 channels)	
Shan et al. [45]	Sequential search		Accuracy rate, 63.7 (for first dataset with 2 channels) and 81.3 % (for second dataset with 16 channels)	
Arvaneh et al. [46]	Pre-specified		Average accuracy rate, 70.47 % (with eight channels)	
He et al. [47]	Genetic algorithm		Average accuracy rate, 88.2 % (for first dataset) and 89.38% (for second dataset)	
Arvaneh et al. [48]	Pre-specified		Average accuracy rates (SCSP1), 81.63 % (for Dataset IIa with 13.22 average channels) and 82.28 % (Dataset IVa with 22.6 average number of channels)	
			Average accuracy rates (SCSP2), 79.07 % (for Dataset IIa with 8.55 average channels) and 79.28 % (for Dataset IVa with 7.6 average channels)	
Yang et al. [51]	Pre-specified	Wrapper	Average accuracy rate, 78 % (with 11 channels)	
Wei and Wang [53]	Random search		Accuracy rate, 83 % (for S1 with 8 channels), 91 % (for S2 with 9 channels), 75 % (for S3 with 14 channels), 86 % (for S4 with 8 channels), and 87 % (for S5 with 7 channels)	
Zhou and Yedida [54]	Sequential search		Average accuracy rate for healthy subjects >90 % (with 90 channels), accuracy rate for stroke subject <85 % (with 110 channels)	
Kamrunnahar et al. [55]	Complete search		Average classification errors, ~21.75 and 28.28 % (for subject 1 with 4 and 3 channels for tasks 1 and 2)	
Yang et al. [62]	Genetic algorithm		Average accuracy rate, 80 % (for 10 channels with the first dataset) and 86 % (for 6 channels with the second dataset)	
Lal et al. [40]	Sequential search	Embedded	Average error rate, 23 % (for 17 channels) and 24 % (12 channels)	
Schroder et al. [65]	Sequential search		Average error rate, 26.9 % (with ≥ 32 channels)	
Li et al. [67]	Complete search	Hybrid		

Table 3 Summary of channel selection methods (*Continued*)

			Average accuracy rate, 92.34 % (7 channels with dataset "aa") and 94.63 % (8 channels with dataset "a1")	
Rizon et al. [68]	Pre-specified	Filtering	Minimum values of FPI = 0.150051, PME = 0.154724, and SD = 0.328312 (with 4 channels)	Emotion classification
Jatupaiboon et al. [69]	Pre-specified	Wrapper	Accuracy rate, 84.18 % (with 5 pairs) and 85.41 % (with 7 pairs)	
Lan et al. [72]	Sequential search	Filtering	Average accuracy rate, ~80 % (with 7, 10 channels)	Mental task classification
Chai et al. [73]	Pre-specified	Wrapper	Accuracy rate, 65–79 % (2 channels with PSD) and 70–84 % (2 channels with HHT)	
Tavakolian et al. [74]	Genetic algorithms		Average accuracy, 96.85 (best 6 channels' combinations)	
Piryatinska et al. [75]	Complete search	Filtering	Mean agreement percentage, 87.41 % (with 4 channels) and 87.2 % (with 5 channels)	Sleep state classification
Ong et al. [76]	Principal component analysis	Filtering	Average accuracy, 94.06 % (16 channels), 86.01 % (8 channels), and 75.13 % (4 channels)	Drug effect classification

sensitivity should be studied with different types of classifiers. With channel selection, we may still work on a multi-channel basis, so the development of a framework containing channel selection and decision fusion is an open area for further investigation.

Competing interests

The authors declare that they have no competing interests.

Author details

¹KACST, King Abdulaziz City, Kingdom of Saudi Arabia. ²KACST-TIC in Radio Frequency and Photonics for the e-Society (RFTONICS), King Saud University, Riyadh, Saudi Arabia. ³Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt. ⁴Electrical Engineering Department, KACST-TIC in Radio Frequency and Photonics for the e-Society (RFTONICS), King Saud University, Riyadh, Saudi Arabia. ⁵Institute for Telecommunications Research, University of South Australia, Adelaide, Australia.

Received: 4 January 2015 Accepted: 15 July 2015

Published online: 01 August 2015

References

- JR Wolpaw, N Birbaumer, DJ McFarland, G Pfurtscheller, TM Vaughan, Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**(6), 767–791 (2002)
- N Birbaumer, Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* **43**, 517–532 (2006)
- J. S. Ebersole, T. A. Pedley, *Current Practice of Clinical Electroencephalography*, 3 edn. (Lippincott Williams & Wilkins, 2003), p. 72–99
- B. J. Baars, N. M. Gage, *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*, 2nd edn. (Elsevier, 2010)
- D Garrett, DA Peterson, CW Anderson, MH Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11**(2), 141–144 (2003)
- RS Fisher, Therapeutic devices for epilepsy. *Ann. Neurol. Am. Neurol. Assoc.* **71**(2), 157–168 (2012)
- D. Chandler, J. Bisasky, J. L.V.M. Stanislaus, T. Mohsenin, Real-time multi-channel seizure detection and analysis hardware. *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, (2011)
- M Dash, H Liu, Feature selection for classification. *Intell. Data Anal.* **1**(3), 131–156 (1997)
- I Guyon, A Elisseeff, An introduction to variable and feature selection. *J. Machine Learning Res.* **3**, 1157–1182 (2003)
- H Liu, L Yu, Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. Knowledge Data Eng.* **17**(4), 491–502 (2005)
- RS Fisher, W v E Boas, W Blume, C Elger, P Genton, P Lee, J Engel, Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* **46**(4), 470–472 (2005)
- JD Henriksen, TW Kjaer, RE Madsen, LS Remvig, CE Thomsen, HB Sorensen, Channel selection for automatic seizure detection. *Clin. Neurophysiol.* **123**(1), 84–92 (2012)
- S. D. Faul, Dynamic channel selection to reduce computational burden in seizure detection. *32nd Annual International Conference of the IEEE EMBS (San Diego, California USA, 2010)*
- A Temko, E Thomas, W Marnane, G Lightbody, G Boylan, EEG-based neonatal seizure detection with support vector machines. *Clin. Neurophysiol.* **122**(3), 464–473 (2011)
- SD Faul, W Marnane, Dynamic, location-based channel selection for power consumption reduction in EEG analysis. *Comput. Methods Programs Biomed.* **108**, 1206–1215 (2012)
- B. Atoufi, C. Lucas, A. Zakerolhosseini, A survey of multi-channel prediction of EEG signal in different EEG state: normal, pre-seizure, and seizure. *Proceedings of the Seventh International Conference on Computer Science and Information Technologies*, (Yerevan, Armenia, 28 Sept. - 2 Oct., 2009)
- E. Shih, A. Shoeb, J. Guttag, Sensor selection for energy-efficient ambulatory medical monitoring. *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, 2009
- K. M. Ong, K. H. Thung, C. Y. Wee, R. Paramesran, Selection of a subset of EEG channels using PCA to classify alcoholics and non-alcoholics. *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (Shanghai, China, 2005)*
- E. L. Glassman, J. V. Guttag, Reducing the number of channels for an ambulatory patient-specific EEG-based epileptic seizure detector by applying recursive feature elimination. In *proc. of the 28th annual Int'l IEEE conf.*, (New York City, Aug 30 – Sept 3, 2006)
- P. Mirowski, D. Madhavan, Y. LeCun, R. Kuzniecky, Classification of patterns of EEG synchronization for seizure prediction. *Clin. Neurophysiol.* **120**(11), 1927–1940 (2009)
- Freiburg invasive EEG database, Epilepsy Center of the University Hospital of Freiburg, available at <http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database>
- N. F. Chang, T. C. Chen, C. Y. Chiang, L. G. Chen, Channel selection for epilepsy seizure predication method based on machine learning. *34th Annual International Conference of the IEEE EMBS*, (San Diego, California USA, 2012)
- A. L. Goldberger, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, **101** e215–e220 (2000)
- National Taiwan University Medical Library, 2013. http://ntuml.mc.ntu.edu.tw/index_en.asp. Accessed 15 Sept 2014
- B. R. Greene, G. B. Boylan, W. P. Marnan, G. Lightbody, S. Cannolly, Automated single channel seizure section in the neonate. *Proceedings of the 30th annual Int'l IEEE EMBS conf.*, (Vancouver, British Columbia, Canada, 2008)

26. A. Temko, G. Lightbody, G. Boylan, W. Marnane, Online EEG channel weighting for detection of seizures in the neonate. Proceedings of the 33rd Annual International Conference of the IEEE EMBS (Boston, Massachusetts USA, 2011)
27. Cork University Maternity Hospital database, Ireland. <http://www.nbrg.ie/1c9fonyf1j> Accessed 15 Sept 2014
28. J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning, (Pittsburgh, PA, 2006)
29. MR Zimbric, CM Sharpe, KC Albright, MP Nespeca, Three-channel electroencephalogram montage in neonatal seizure detection and qualification. *Pediatr. Neurol.* **44**(1), 31–34 (2011)
30. H Tekgul, BFD Bourgeois, K Gauvreau, AM Bergin, Electroencephalography in neonatal seizures: comparison of a reduced and a full 10/20 montage. *Pediatr. Neurol.* **32**(3), 155–161 (2005)
31. L. He, Z. Yu, Z. Gu, Y. Li, Bhattacharyya bound based channel selection for classification of motor imageries in EEG signals. Proceedings of Chinese Control and Decision Conf., (Guilin, China, 17–19 June, 2009)
32. A. Bhattacharyya, A., On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **35**, 99–109 (1943)
33. L. Devroye, L. Györfi, G. Lugosi, *A probabilistic theory of pattern recognition* (Springer, 1996)
34. W. K. Tam, Z. Ke, K. Y. Tong, Performance of common spatial pattern under a smaller set of EEG electrodes in brain-computer interface on chronic stroke patients: a multi-session dataset study. Proceedings of the 33rd annual Int'l IEEE EMBS conf., (Boston, Massachusetts USA, Aug. 30 – Sept. 3, 2011)
35. E Cortes, V Vapnik, Support-vector networks. *Machine Learning* **20**, 273–297 (1995)
36. C. W. Hsu, C. C. Chang, C. J. Lin, *A practical guide to support vector classification*. Department of Computer science (2010)
37. X. Yong, R. K. Ward, G. E. Birch, Sparse spatial filter optimization for EEG channel reduction in brain-computer interface. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (Las Vegas, Nevada USA, 2008)
38. Fraunhofer FIRST (Intelligent Data Analysis Group) and University of Medicine Berlin (Neurophysics Group). http://www.bbc.de/competition/iii/desc_IVa.html
39. J. Meng, G. Liu, G. Huang and X. Zhu, Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system. Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics, (Guilin, China. Dec. 19–23, 2009)
40. T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, B. Schölkopf, Support vector channel selection in BCI. *IEEE Trans. Biomed. Eng.* **51**(6), (2004)
41. Y. Wang, S. Gao, X. Gao, Common spatial pattern method for channel selection in motor imagery based brain-computer interface. Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, (Shanghai, China, Sept. 1–4, 2005)
42. R. O. Duda, P. E. Hart, D. H. Stork, *Pattern Classification*, 2nd edn. (Wiley Interscience, 2000)
43. J. M. Hilbe, *Logistic Regression Models*. (Chapman & Hall/CRC Press, 2009)
44. S. Mikat, G. fitscht, J. Weston, B. Schölkopf, K. R. Mullert, Fisher discriminant analysis with kernels. *IEEE Conference on Neural Networks for Signal Processing*, 41–48 (1999)
45. H. Shan, H. Yuan, S. Zhu, B. He, EEG-based motor imagery classification accuracy improves with gradually increased channel number. Proceeding of the 34th Annual International Conference of the IEEE EMBS (San Diego, California USA, 28 Aug. - 1 Sept. 2012)
46. M. Arvaneh, C. Guan, K. K. Ang, C. Quek, Robust EEG channel selection across sessions in brain-computer interface involving stroke patients. Proceeding of the WCCI 2012 IEEE World Congress on Computational Intelligence (WCCI), (Brisbane, Australia June10-15, 2012)
47. L. He, Y. Hu, Y. Li, D. Li, Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *J. Neurocomputing* **121**, 423–433 (2013)
48. M Arvaneh, C Guan, KK Ang, C Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Trans. Biomed. Eng.* **58**(6), 1865–1873 (2011)
49. M Naeem, C Brunner, R Leeb, B Graimann, G Pfurtscheller, Separability of four-class motor imagery data using independent components analysis. *J. Neural Eng.* **3**, 208–216 (2006)
50. G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans. Biomed. Eng.* **51**(6), 993–1002 (2004)
51. Y. Yang, O. Kyrgyzov, J. Wiat, I. Bloch, Subject-specific channel selector for classification of motor imagery electroencephalographic data. Proceedings of the 38th Int'l Conf. on Acoustics, Speech, and Signal Processing, (Vancouver Canada, 2013)
52. Dataset IVa from BCI competition III. <http://www.bbc.de/competition/iv/>
53. Q. Wei, Y. Wang, Binary multi-objective particle swarm optimization for channel selection in motor imagery based brain-computer interfaces. Proceedings of the 4th Biomedical Engineering and Informatics (BMEI) Conf., (Shanghai, China, 15–17 Oct, 2011)
54. J. Zhou, S. Yedida, Channel selection in EEG-based prediction of shoulder/elbow movement intentions involving stroke patients: a computational approach. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2007)
55. M. Kamrunnahar, N. S. Dias, S. J. Schiff, Optimization of electrode channels in brain computer interfaces. Proc. Of the 31st Annual International Conference of the IEEE EMBS (Minneapolis, Minnesota, USA, Sept. 2–6, 2009)
56. P. L. Nunez, R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*, 2nd edn. (Oxford University Press, New York, 2006)
57. P. L. Nunez, *Electric fields of the brain*, (Oxford University Press, 1981)
58. R. Srinivasan, Methods to improve the spatial resolution of EEG. *Int. J. Bioelectromagnetism* **1**(1), 102–111 (1999)
59. R. Srinivasan, W. R. Winter, P. L. Nunez, Source analysis of EEG oscillations using high-resolution EEG and MEG, "Event-Related Dynamics of Brain Oscillations. Progress in Brain Research, C. Neuper and W. Klimesch, eds., 29–42 (2006)
60. N. S. Dias, M. Kamrunnahar, P. M. Mendes, S. J. Schiff, J. H. Correia, Comparison of EEG pattern classification methods for brain-computer interfaces. Proceedings of the 29th Annual International Conference of the IEEE EMBS (Cité Internationale, Lyon, France, 2540–2543 (2007))
61. N. S. Dias, M. Kamrunnahar, P. M. Mendes, S. J. Schiff, J. H. Correia, Customized linear discriminant analysis for brain-computer interfaces. Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering (Kohala Coast, Hawaii, USA, pp. 430–433, 2007)
62. J. Yang, H. Singh, EL Hines, F Schlaghecken, DD Iliescu, MS Leeson, NG Stocks, Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artif. Intell. Med.* **55**, 117–126 (2012)
63. C. R. Reeves, J. E. Rowe, *Genetic Algorithms: Principles and Perspectives: A Guide to GA Theory*, (Boston Kluwer Academic Publishers; 2003)
64. SY Yuen, CK Chow, A genetic algorithm that adaptively mutates and never revisits. *IEEE Trans. Evol. Comput.* **13**(2), 454–472 (2009)
65. M. Schroder, T. N. Lal, T. Hinterberger, M. Bogdan, N. J. Hill, N. Birbaumer, W. Rosenstiel, Bernhard Schölkopf, Robust EEG channel selection across subjects for brain-computer interfaces. *EURASIP J. Appl. Signal Process.* **19** 3103–3112 (2005)
66. PD Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**(2), 70–73 (1967)
67. M. Li, J. Ma, S. Jia, Optimal combination of channels selection based on common spatial pattern algorithm. Proceedings of the IEEE Int'l conf. on Mechatronics and Automation, (Beijing, China, Aug. 7–10, 2011)
68. M Rizon, M. Murugappan, R. Nagarajan, S. Yaacob, Asymmetric ratio and FCM based salient channel selection for human emotion detection using EEG. *WSEAS Trans. Signal Process.* **4**(10), (2008)
69. N. Jatupaiboon, S. Pan-ngum, P. Israsena, Emotion classification using minimal EEG channels and frequency bands. Proceedings of 10th Int'l Joint conf. on Computer Science and Software Engineering (JCSSE 2013) (Khon Kaen, Thailand, 2013)
70. S. Elise, D. Glauser, KR Scherer, The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance. *Behav. Res. Methods* **43**(2), 468–477 (2011)
71. Emotiv EEG Neuroheadset. Available online: <http://emotiv.com/upload/manual/EEGSPECIFICATIONS.pdf> [Access 12 Nov. 2013], <http://www.emotiv.com>

72. T. Lan, D. Erdogmus, A. Adami, S. Mathan, M. Pavel, Channel selection and feature projection for cognitive load estimation using ambulatory EEG. *Computational Intelligence and Neuroscience* (2007)
73. R. Chai, S. H. Ling, G. P. Hunter, H. T. Nguyen, Toward fewer EEG channels and better feature extractor of non-motor imagery mental tasks classification for a wheelchair thought controller. Proceedings of the 34th annual Int'l IEEE EMBS conf., (San Diego, California USA, 2012)
74. K. Tavakolian, A. M. Nasrabadi, S. Rezaei, Selecting better EEG channels for classification of mental tasks. Proceedings of Int'l symposium on Circuits and Systems, (Vancouver, Canada, 2004)
75. A. Piryatinska, WA Woyczynski, MS Scher, KA Loparo, Optimal channel selection for analysis of EEG-sleep patterns of neonates. *J. Comput. Methods Programs Biomed.* **106**(1), 14–26 (2012)
76. K. M. Ong, K. H. Thung, C. Y. Wee, R. Paramesranle, Selection of a subset of EEG channels using PCA to classify alcoholics and non-alcoholics. Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference (Shanghai, China, Sept. 1–4, 2005)
77. S. Gao, Y. Wang, X. Gao, B. Hong, Visual and auditory brain–computer interfaces. *IEEE Trans. Biomed. Eng.* **61**(5), 1436–1447 (2014)
78. B. Graimann, B. Allison, G. Pfurtscheller, *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*, (Springer 2010)
79. V. Kachitvichyanukul, Comparison of three evolutionary algorithms: GA, PSO, and DE. *Ind. Eng. Manag. Syst.* **11**(3), 215–223 (2012)

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

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

Submit your next manuscript at ▶ springeropen.com
