Expert Systems With Applications 83 (2017) 273–282

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Dynamically identifying relevant EEG channels by utilizing channels classification behaviour

Ahmed Al-Ani[a,*], Irena Koprinska[b], Ganesh Naik[a]

[a] Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[b] School of Information Technologies, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 2 October 2016
Revised 5 March 2017
Accepted 20 April 2017
Available online 23 April 2017

Keywords:
EEG channel classification behaviour
Dynamic channel identification
Channel selection
Alertness state classification

ABSTRACT

It is well established that multiple EEG channels are required for various brain functionality studies, including classification tasks. Yet, due to the curse of dimensionality problem, the analysis of multiple channels may not lead to the desired performance. Accordingly, a number of static channel selection algorithms have been proposed to identify the most relevant subset of channels. However, static methods select a fixed subset of channels that is unchanged when processing new data, and hence cannot adapt to changes in data. In this paper, we propose a novel algorithm that utilizes the dynamic classification behaviour of channels in selecting the channel that is most relevant for each time segment of the signal. The main idea is to identify for each time segment of every channel of the signal (testing sample) the closest training samples. These training samples are used to estimate the local accuracy of each channel. The best performing channel for that time segment will then be identified as the relevant one. Results obtained using EEG data of a four-class alertness state classification problem, with two different feature sets, reveal that the proposed approach is capable of achieving competitive performance compared to a traditional static channel selection based method. More importantly, the evaluation of the selected channels reveals that our approach is able to select relevant channels for each of the four alertness states. The proposed algorithm is expected to make a valuable contribution to the field of multichannel biomedical signal classification.

1. Introduction

Recognizing patterns in Electroencephalography (EEG) recordings is important for a number of applications, including diagnosis of sleep disorders (Koprinska, Pfurtscheller, & Flotzinger, 1996; Platt & Gernot, 2011), predicting epileptic seizures (Mirowski, Madhavan, LeCun, & Kuzniecky, 2009), building brain-computer interfaces (Siamac et al., 2012), detecting drivers' drowsiness (Gianluca, Astolfi, Vecchiato, Mattia, & Babiloni, 2014) and recognising alertness states (Al-Ani, Mesbah, Van Dun, & Dillon, 2013; Ning-Han, Chiang, & Hsu, 2011). To achieve good classification results and identify relevant clinical patterns, most of the EEG applications require signals from multiple electrodes that are distributed across the scalp. However, the usage of a large number of EEG channels might have negative impact on the performance of automated EEG analysis systems due to the increased complexity of such systems, compared to those with limited number of channels. This study deals with the problem of dynamically identifying relevant channels, i.e., selecting the most informative subset of channels for each time segment of the signal in order to achieve the best possible performance for a given EEG classification task. The identification of relevant channels may help clinicians to ascertain which regions of the brain are responsible for certain brain functionalities, especially for new EEG applications.

The existing EEG channel selection methods can be categorized into static and dynamic, with most of the researches focused on static channel selection, such as Ansari, Karim, and Thierry (2007), Arvaneh, Guan, Ang, and Quek (2011), Duun-Henriksen et al. (2012), He, Hu, Li, and Li (2013), Lal et al. (2004), Lan, Erdogmus, Adami, Mathan, and Pavel (2007), Piryatinska, Woyczynski, Scher, and Loparo (2012), Schrder et al. (2005) and Wang, Shangkai, and Xiaorong (2006). A more recent work that proposed a multi-objective approach for channel selection is presented in Torres-Garcia, Reyes-Garcia, Villasenor-Pineda, and Garcia-Aguilar (2016).

Static methods select a fixed subset of channels that will not be changed when processing new (unseen) data. One of the

* Corresponding author:
E-mail addresses: ahmed.al-ani@uts.edu.au (A. Al-Ani), irena.koprinska@sydney.edu.au (I. Koprinska), ganesh.naik@uts.edu.au (G. Naik).

http://dx.doi.org/10.1016/j.eswa.2017.04.042

advantages of static approaches is that the saving in computational effort is known a priori. Although static channel selection may enable the detection of redundant channels and the identification of channels that are highly influenced by noise, it is difficult to ascertain whether a subset of EEG channels selected for one subject is also useful for other subjects, or to draw generalized conclusions about the most relevant channels for a given classification task across subjects (Thulasidas, Guan, & Wu, 2006). Also, since static methods cannot adapt to the specific EEG data, valuable information may be lost or undetected (Gonzalez, Nambu, Hokari, & Wada, 2014).

Dynamic channel identification, on the other hand, aims to dynamically select which channel to analyse for each time segment. This provides the ability to adapt to changes in the EEG data by selecting the most relevant, informative and noise free channel for each time segment of the data. However, the implementation of dynamic channel selection is far more challenging than its static counterpart. There have been limited attempts to investigate dynamic channel selection in the literature. Recently, Faul and Marnane (2012) proposed a classification based dynamic channel selection algorithm for power consumption reduction in EEG analysis. The main purpose of the method was to reduce the computational cost, and hence reduce power consumption when implemented on a Blackfin microprocessor. Kota, Gupta, Molfese, , and Vaidyanathan (2009) presented a dynamic channel selection strategy to classify event related potentials into pre-defined brain related activities. They used statistical approach to select channels that have different Gaussian densities across the event related potential categories. These two methods, however, have not presented thorough classification comparisons with static channel subsets of varying sizes. In addition, the two methods did not analyse the existence of noise among the selected channels for each time segment, and have not validated the relevance of selected channels to brain regions that are known to have an impact on the corresponding classification tasks.

In this paper we consider the task of dynamically finding the best channels for the classification of EEG data based on four alertness states. We propose a new dynamic channel identification algorithm that utilizes the concept of multiple classifier behaviour to select which channel to analyse for each time segment. The advantage of the proposed approach is that it is able to adapt to changes in the EEG data by selecting the location of the channel to be analysed, automatically avoiding noisy portions of the data.

2. The proposed dynamic channel identification algorithm

The proposed algorithm is inspired by the concept of dynamic classifier selection, and in particular the multiple classifier behaviour approach. Below is a brief description of the dynamic classifier selection concept. The proposed algorithm is explained in the subsequent Section.

2.1. Dynamic classifier selection

Unlike classifier combination that attempts to combine the results of multiple classifiers, dynamic classifier selection aims to select a classifier for each unknown test sample x, and assign to x the label produced by that classifier (Huang & Suen, 1995).

Most of the dynamic classifier selection methods that have achieved competitive results attempt to estimate the relevance of each classifier in a "local region" of the training data for each unseen test sample (Giacinto & Roli, 2001; Woods, 1997). The K nearest neighbours are usually used to identify the local region for each test sample. However, existing methods differ in the evaluation of classifiers' relevance. Woods (1997) presented a method that he named dynamic classifier selection using local accuracy estimates. His method uses the percentage of the local training samples assigned to a given class that are correctly labelled to estimate the local accuracy of each classifier. Giacinto and Roli (2001) utilized the concept of multiple classifier behaviour (MCB) to compute the classifiers' local accuracy. MCB of sample x is defined by a vector whose elements are the class labels assigned to x by the L classifiers. The similarity between the MCB of an unseen test sample and each of its neighbouring training samples is used to identify the neighbours that influence the computation of classifiers' local accuracy.

2.2. The proposed algorithm

The proposed dynamic channel identification algorithm (termed here as DCI) uses the following parameters:

- ch: one of the EEG channels.
- $C_j(z)$: the class label produced by classifier j when fed with the features of sample z. $j = 1 : L$, where L is the number of classifiers.
- K: number of nearest neighbour samples to be used in calculating the local accuracy of each channel.
- x: a testing sample, and y_i^x: a training sample that is one of the $\acute{K}$ nearest neighbours to x. $i = 1 : \acute{K}$, where $\acute{K} = 3 \times K$.

The implementation procedure of the proposed algorithm is illustrated in Fig. 1 and explained below:

- Choose L classifiers. Classifiers of different types can be used, such as linear discriminant analysis (LDA), support vector machine (SVM), Bayesian classifier, etc.
- For each EEG channel, ch, train the classifiers and record the decisions for the training and testing samples (time windows, as will be explained in Section 3). The decisions for sample z are: $C_1^{ch}(z), C_2^{ch}(z), \ldots, C_L^{ch}(z)$.
- For each test sample, x
  - for each channel, ch
    - * Identify the $\acute{K}$ nearest neighbours in the training samples, y_i^x $(i = 1 : \acute{K})$ and record the distance $Dist^{ch}(x, y_i^x)$.
    - * Calculate the similarity between x and y_i^x

$$S^{ch}(x, y_i^x) = \frac{1}{L} \sum_{j=1}^{L} T_j^{ch}(x, y_i^x) \tag{1}$$

$$T_j^{ch}(x, y_i^x) = \begin{cases} 1 & \text{if } C_j^{ch}(x) = C_j^{ch}(y_i^x), \\ 0 & \text{if } C_j^{ch}(x) \neq C_j^{ch}(y_i^x). \end{cases} \tag{2}$$

    - * The updated distance $DistU^{ch}(x, y_i^x)$ is calculated according to the following equation:

$$DistU^{ch}(x, y_i^x) = Dist^{ch}(x, y_i^x) \times \left(2 - S^{ch}(x, y_i^x)\right) \tag{3}$$

    - * Identify the K nearest neighbours that have the lowest updated distances. Discard the remaining $\acute{K} - K$ neighbours.
    - * Calculate the accuracy of each training sample, y_i^x

$$A_{ch}(y_i^x) = \frac{1}{L} \sum_{j=1}^{L} D_j(y_i^x) \tag{4}$$

$$D_j(y_i^x) = \begin{cases} 1 & \text{if } C_j(y_i^x) = O(y_i^x), \\ 0 & \text{if } C_j(y_i^x) \neq O(y_i^x). \end{cases} \tag{5}$$

      where $O(y_i^x)$ is the true class label of input y_i^x.
    - * The channel's local accuracy of the current window w is

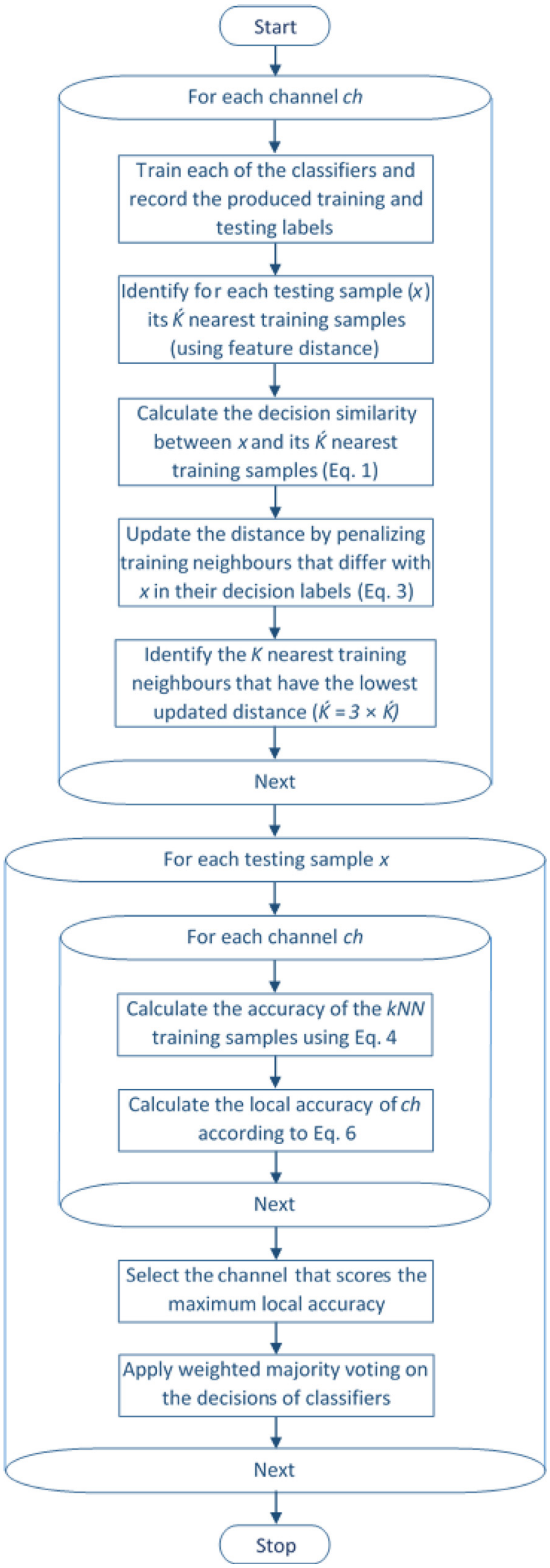

Fig. 1. Flow chart of the proposed algorithm.

calculated using the accuracy of the current and past M windows according to the following equation

$$LA^{w}(ch) = \frac{1}{KM} \sum_{m=0}^{M} \sum_{i=1}^{K} A_{ch}^{w-m}(y_i^x) \tag{6}$$

– Select the channel $\widehat{ch}$ that scores the maximum local accuracy.
– Apply a weighted majority voting to the decisions of the L classifiers of channel $\widehat{ch}$, $(C_j^{\widehat{ch}}(x), j = 1 : L)$. The weights are measured based on the training accuracy of each classifier. The class label with the majority votes will be assigned as the label of the test sample x.

There are two reasons behind using multiple classifiers. Firstly, they help to enhance the identification of the K nearest neighbours by considering the decision similarity, and secondly they contribute towards achieving a higher classification accuracy through the fusion of the classification results. In regards to the first reason, neighbouring samples are expected to produce similar decision labels, and hence the feature space based distance will be penalized if the labels are different. When the corresponding decision labels of x and y_i^x are exactly the same for all L classifiers, then $S^{ch}(x, y_i^x) = 1$, which makes the updated distance defined in Eq. (3) equal to the original distance. On the other hand, when $S^{ch}(x, y_i^x) = 0$, i.e., x and y_i^x differ in the labels of all classifiers, the updated distance is double that of the original one. If $0 < S^{ch}(x, y_i^x) < 1$, then the updated distance will fall between those two limits. In other words, the updated distance is introduced to reduce the sensitivity to the distance measure, and hence identify the true neighbours for each test window.

As it would be valuable to consider the temporal aspect of the EEG signal, the similarity and accuracy of the past M windows were incorporated in calculating the current window's local accuracy. It is important to mention that this approach can be applied to individual channels or collection of channels. For the latter case, a matrix can be formed for this purpose, where each row represents one of the possible combinations, while the columns contain the channels of each combination. For each possible combination, features of the individual channels will be concatenated to form a single feature vector. These issues will be evaluated in Section 4.1.

In regards to the computational complexity of the algorithm, it is mainly influenced by the training of the L classifiers. If this process is to be performed off-line and the trained classification models are to be saved along with their classification outcomes, then it will be possible to use this algorithm in on-line applications, where implementation of the remaining steps (including the execution of Eqs. (1)–(6)) is not computationally expensive.

3. Experimental setup

The proposed algorithm is tested using an alertness classification dataset. Ten adult subjects, 4 females and 6 males, with an age range of 24–53 years (mean age = 33.4 years and standard deviation = 8.4 years), participated in the experiment. The data was anonymized by concealing the identities of the subjects when constructing the dataset. The EEG data was recorded using a 64 channel Neuroscan system (Compumedics, Australia), with the reference electrode chosen close to Cz (vertex). The subjects were asked to press one of three buttons every 30 s to indicate their perceived level of alertness, i.e., *engaged*, *calm* (but not drowsy), and *drowsy*. Each recording session lasted one hour. The one-hour recording was divided into 6 divisions of 10 min each. If the subject did not provide an input for more than 3 min, he/she was considered to have fallen *asleep*.

Each of the recorded signals was divided into windows of 5 s with overlap of 3 s. The windows that fall on the edge between two classes (alertness states) were removed. In order to have a better evaluation of the performance of the proposed algorithm, we considered using two feature extraction methods to represent each window of the signal. The first one uses a filter-bank to produce five features that correspond to the average energy of the five EEG frequency bands; δ (up to 4 Hz), θ (4 - 8 Hz), α (8 - 13 Hz), β (13 - 30 Hz), and γ (30 - 100 Hz). The second feature extraction method produces nine features that correspond to the energy of dyadic wavelet transform scales. Each 10 consecutive time windows were grouped to form an epoch, given that they all belong to a same alertness state. The epoch ratios (and numbers) of the four classes are: *sleep*: 5.2% (96), *drowsy*: 35.9% (678), *calm*: 20.7% (387) and *engaged*: 38.2% (711). The reason behind this choice (1 epoch = 10 windows) is that a too big epoch will reduce the overall number of epochs and increase the chance of having epochs across the boundaries between alertness states. On the other hand, a small epoch (small number of windows per epoch) will increase the chance of overlap between training and testing samples, where as mentioned earlier, adjacent windows overlap by 3 s.

A four-fold cross-validation approach has been adopted, where for each fold, 75% of the epochs of each subject were used for training and the remaining 25% for testing. This procedure of selecting epochs was chosen to prevent unrealistically biasing the classifier toward high testing accuracy. As stated in Section 2, L classifiers are required for the implementation of the algorithm. We chose to use five classifiers, which are support vector machine (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Bayes classifier (Bay), and extreme learning machine (ELM). We used the LIBSVM matlab toolbox[1] for the SVM classifier, the matlab implementation of LDA, QDA and Bayes classifiers, and the NTU Matlab toolbox implementation of ELM.[2] Those classifiers are chosen because they represent different learning paradigms, and they are widely used in the literature. We will consider processing each channel with three of its neighbours as this is expected to provide the classifiers with more informative inputs, which would in turn lead to better classification accuracy. In such case, the classifiers are provided with 20 EEG rhythm features (or 36 wavelet features). Fig. 2 shows channels FP1 and C4 and their corresponding three neighbours. Hence, each channel is processed along with its right horizontal, lower vertical, and lower right diagonal channels. As for the remaining edge channels, they are processed with either their horizontal or vertical neighbouring channels. The processing of channels along with their neighbours is also expected to reduce fluctuation in performance between neighbouring channels.

4. Results and discussion

The classification performance of the proposed dynamic channel identification algorithm (DCI) is evaluated using three scenarios. In the first scenario, only two classes were considered, which are *engaged* and *drowsy*. In the second scenario, three classes were considered; *engaged, drowsy* and *sleep*. In the third scenario, all four alertness states, *engaged, calm, drowsy* and *sleep*, were considered.

This Section will first study the effect of: (i) the number of temporal windows that are required, and (ii) the inclusion of features of neighbouring channels versus representing each channel by its own features only. It will then evaluate three aspects of the proposed algorithm: (i) comparison with static channel selection in terms of classification performance, (ii) ability to avoid noisy channel segments and (iii) relevance of the selected channels to each of

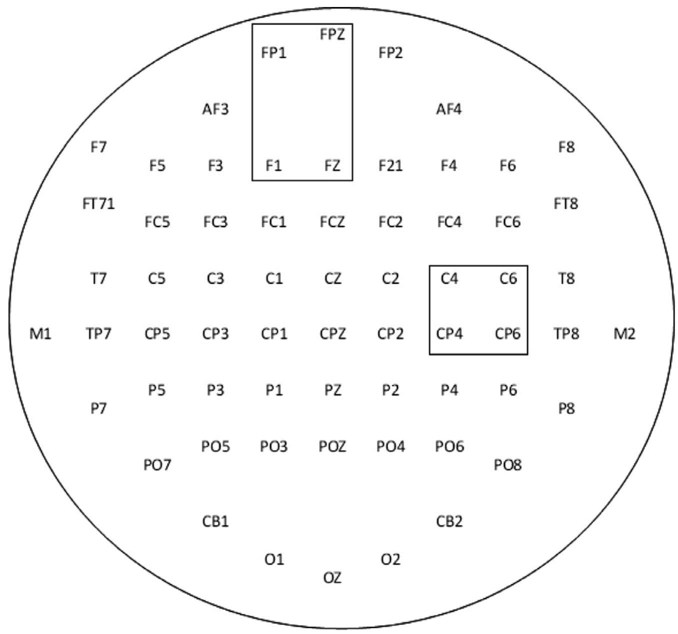

Fig. 2. EEG montage that shows distribution of the 64 channels. Neighbourhoods of channels FP1 and C4 are also shown.

the alertness states with respect to the properties of these states as described in the literature.

4.1. Issues affecting performance of the proposed algorithm

The temporal effect is evaluated by varying the number of time windows between 1 and 20 (1 corresponds to the current window only, while 20 corresponds to the current window and 19 previous windows). We processed each channel along with its three neighbours when conducting this experiment. The results of the three classification scenarios (2, 3, and 4 classes) for both types of features are shown in Fig. 3. All three graphs show that using 7 or more time windows would produce good results for all three classification scenarios and both types of features. This choice is logical, as for most time varying signals temporal information could contribute positively towards enhancing the classification results. Hence, we fixed the number of windows to 10 for the remaining experiments.

We looked at the issue of adding the features of neighbouring channels, as explained in Section 3, when training the five classifiers. Results of utilizing and not utilizing neighbourhood channel features shown in Fig. 4 and Table 1 indicate that the neighbouring channels provide useful information that help in enhancing the performance of classifiers. Accordingly, features of the neighbouring channels are utilized in the experiments presented in the next Section. The above results also indicate that the wavelet features consistently achieved better performance than their EEG rhythm counterparts. We also evaluated another channel neighbourhood that also considers subsets of four channels, but with the reference channel being located in the lower right corner of the rectangle, i.e., the neighbours of channel C4 are {FC4, FC2, C2}. Results show that the performance of the two neighbourhood combinations is comparable.

The K nearest neighbours are identified by measuring the "Euclidean" distance between a test sample (window) and the training samples. In order to evaluate the sensitivity to the type of distance measure, we considered three other distance measures; namely Mahalanobis, Minkowski and Chebychev. Fig. 5 shows the performance of those distance measures for the EEG rhythm fea-

[1] https://www.csie.ntu.edu.tw/cjlin/libsvm/.
[2] http://www3.ntu.edu.sg/home/egbhuang/elm-codes.html.

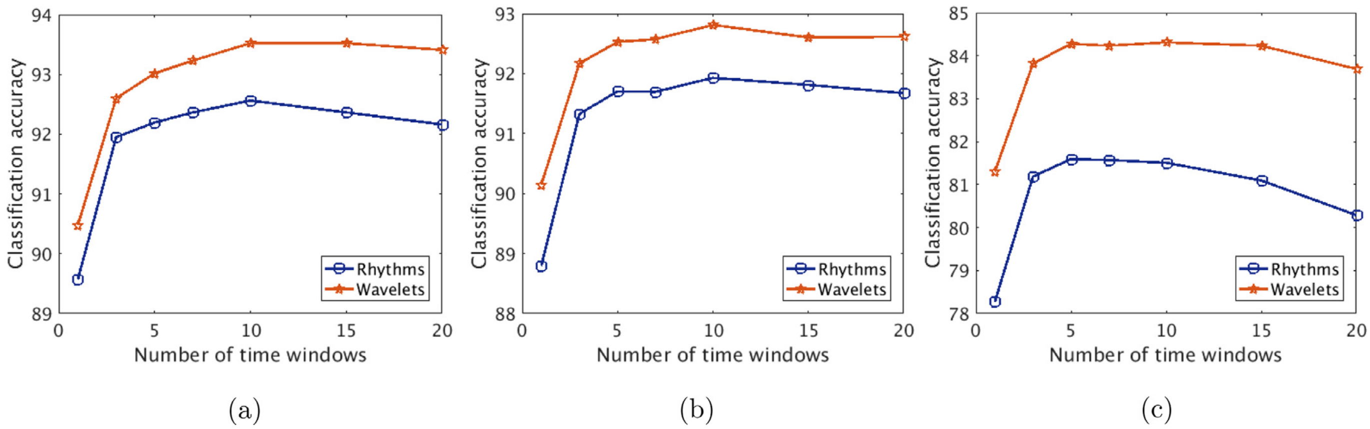

Fig. 3. Effect of number of time windows used to calculate channels' local accuracy. (a) 2 classes, (b) 3 classes, and (c) 4 classes.

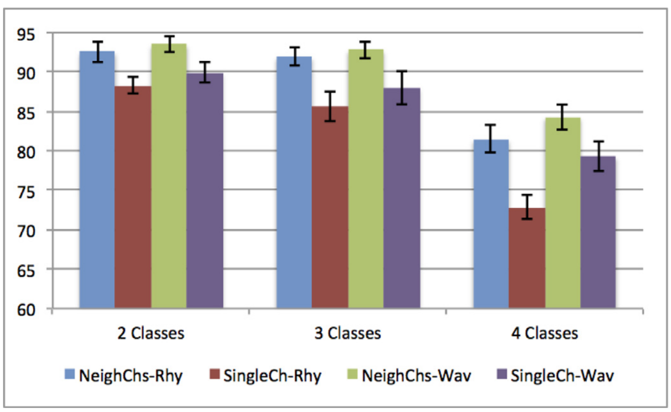

Fig. 4. Evaluation of inclusion of features of neighbouring channels for the three classification scenarios.

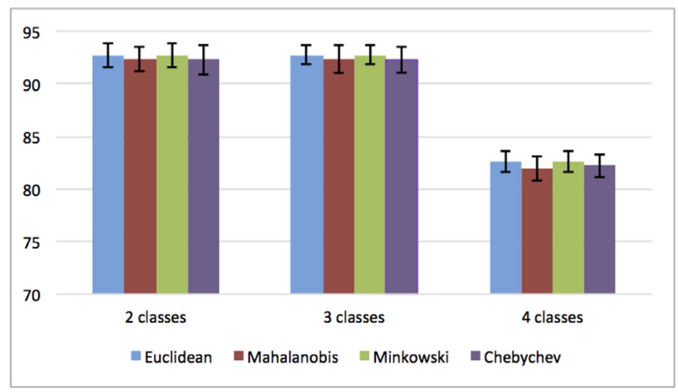

Fig. 5. Evaluation of distance measures for the three classification scenarios.

Table 1
Performance with and without the inclusion of neighbourhood channels' features.

	EEG Rhythms		Wavelet features	
	Neigh-Channels	Single-Channel	Neigh-Channels	Single-Channel
2 Classes	92.56 ± 1.33	88.30 ± 1.06	93.52 ± 0.99	89.96 ± 1.29
3 Classes	91.93 ± 1.10	85.57 ± 1.89	92.81 ± 1.08	88.00 ± 2.10
4 Classes	81.51 ± 1.79	72.78 ± 1.52	84.31 ± 1.61	79.34 ± 1.95

tures, which indicates that there is hardly any difference in performance between the four distance measures. This could be related to the proposed updated distance (Eq. (3)) that attempts to identify the true neighbours for each test window by incorporating the decision similarity of the classifiers.

In addition, we tried using 10-fold cross validation and found the performance to be comparable to that of the 4-fold cross-validation.

4.2. Comparison with static channel selection

In order for the evaluation to be meaningful, we implemented static channel selection and applied it to the three classification scenarios. We calculated the classification accuracy for static subsets of features with different sizes (range between 1 and 64). We used the differential evolution (DE) based feature selection algorithm described in Al-Ani, Alsukker, and Khushaba (2013), as it was shown to produce very competitive results in comparison to other feature selection algorithms. We used a population size of 50 and run the DE-based algorithm for 600 iterations, which was used as a stopping criterion. Each selected channel and its three neighbours were fed to the five classifiers of SVM, LDA, QDA, Bay and ELM. A weighted majority sum approach was used to combine the classification results and estimate the output class of each testing window.

Fig. 6 shows the classification accuracy of the static channel selection with different subset sizes and the proposed dynamic channel identification algorithm when considering the five EEG rhythm features. The three plots show that the best subset size for static channel selection for all three classification scenarios was around 10, which indicates that the inclusion of large number of channels in a static fashion could degrade the performance. This shows that if a relevant set of channels exists and can be identified, then the inclusion of additional channels may not be useful, as they would fail to provide complementary information to the classification task under investigation, and thus confuse the classifiers. In addition, the three plots of Fig. 6 clearly demonstrate the superior performance of the proposed method over static channel selection. This can be justified by the fact that the relevant set of channels for a given alertness state could be different from that of another state (as indicated in Durmer and Dinges (2005) and elaborated in Section 4.4). In addition, the effect of noise and artefact (will be discussed in the next section) makes it not possible for any channel, even though relevant, to accurately represent a given alertness state.

The three plots also show a noticeable difference in the classification accuracy between the scenario of four classes and the other

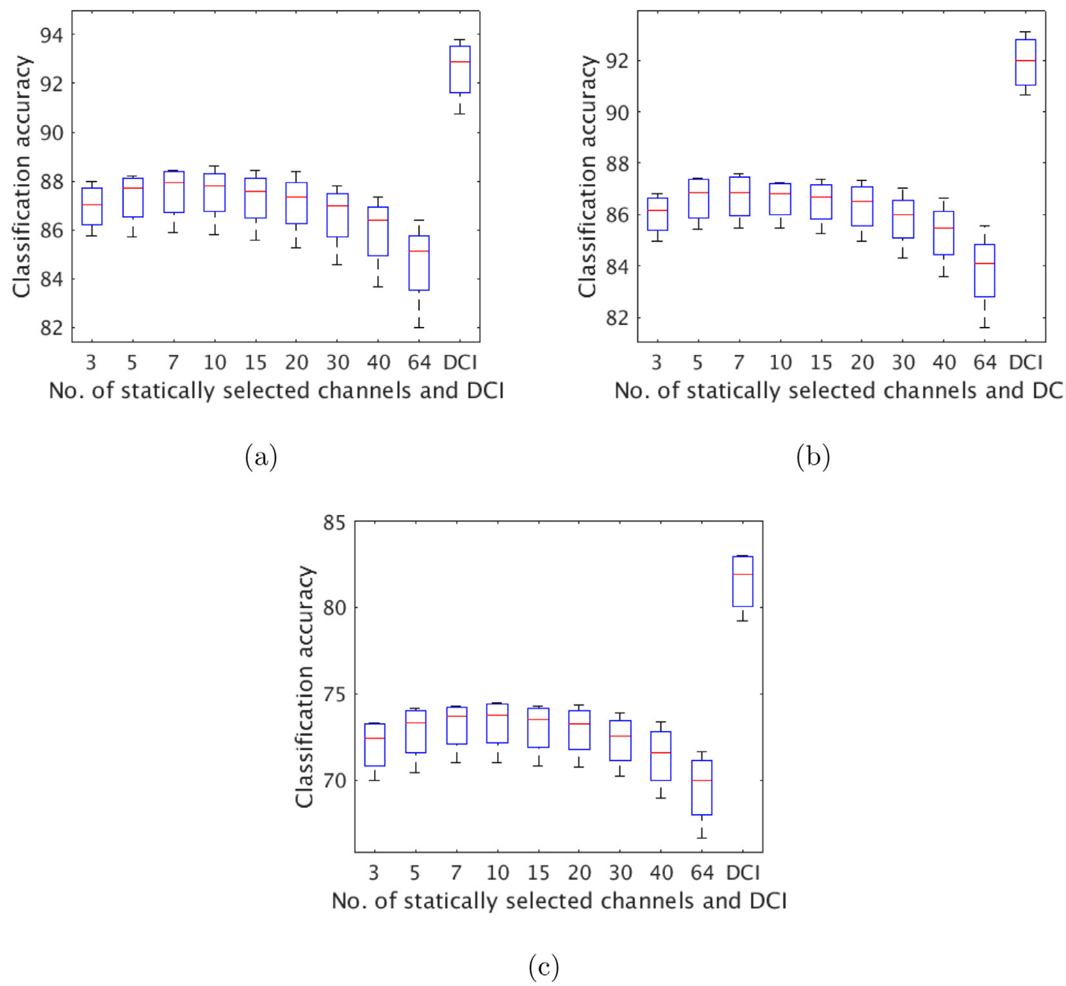

Fig. 6. Classification accuracy of static channel selection (first nine ticks) and the proposed dynamic channel identification (last tick) using EEG rhythm features for (a) 2 classes, (b) 3 classes, and (c) 4 classes.

two scenarios of two and three classes. This is mainly due to a large overlap between the *calm* and the *drowsy* states and also between the *calm* and *engaged* states. One possible explanation could be the different interpretation of the *calm* state by the different subjects.

The classification results using the wavelet features of static channel selection and our proposed dynamic channel identification algorithm for the three classification scenarios are shown in Fig. 7. When compared with the three plots of the five EEG rhythm features, it can be noticed that the wavelet features produced slightly better results for both static and dynamic channel selection (similar to the findings of Section 4.1). The results also indicate that static channel selection tends to favour smaller number of channels when using the wavelet features compared to the EEG rhythm features. This could be due to the fact that more relevant information could be extracted from individual channels, which reduces the need for incorporating additional channels. On the other hand, the two sets of figures clearly demonstrate the superiority of our proposed dynamic channel identification algorithm for all classification scenarios.

It is worth mentioning that empirical comparison with the two dynamic channel selection methods presented in Faul and Marnane (2012) and Kota et al. (2009) could not been carried out because the codes of those two methods have not been made available, and re-implementation of the two methods was found to be quite challenging.

4.3. Avoidance of noisy channel segments

EEG is a highly sensitive signal that is easily affected by different types of artefacts and noise, such as eye movement, muscle activities and poor electrode contact (LeVan, Urrestarazu, & Gotman, 2006). It is quite common for the different EEG channels not to be equally affected by any single artefact or noise source, as observed in Fig. 8 that shows the first 30 (out of 64) EEG channels of a 5 s noisy window. Since it is hard to extract relevant information from noisy EEG channels, it will be useful to avoid channels that are highly affected by artefacts and noise and to give more emphasis to less noisy ones, given that they contain relevant classification information. The figure also shows that noisy portions of data may occur for limited periods of time, before and after which clean data is observed. This makes dynamic channel selection more attractive than static selection of subsets of channels that are fixed throughout the recording session.

In order to get a rough estimation of the noise level of each channel in a given window, we calculated the energy of the individual channels in that window. We then calculated the number of testing windows in which the dynamically selected channel has lower energy compared to the average energy of all channels, and found them to make approximately 91% of the total number of testing windows. This gives a clear indication that the proposed method tends to inherently avoid noisy channels. A rationale justification for this behaviour is that noisy channels are not expected

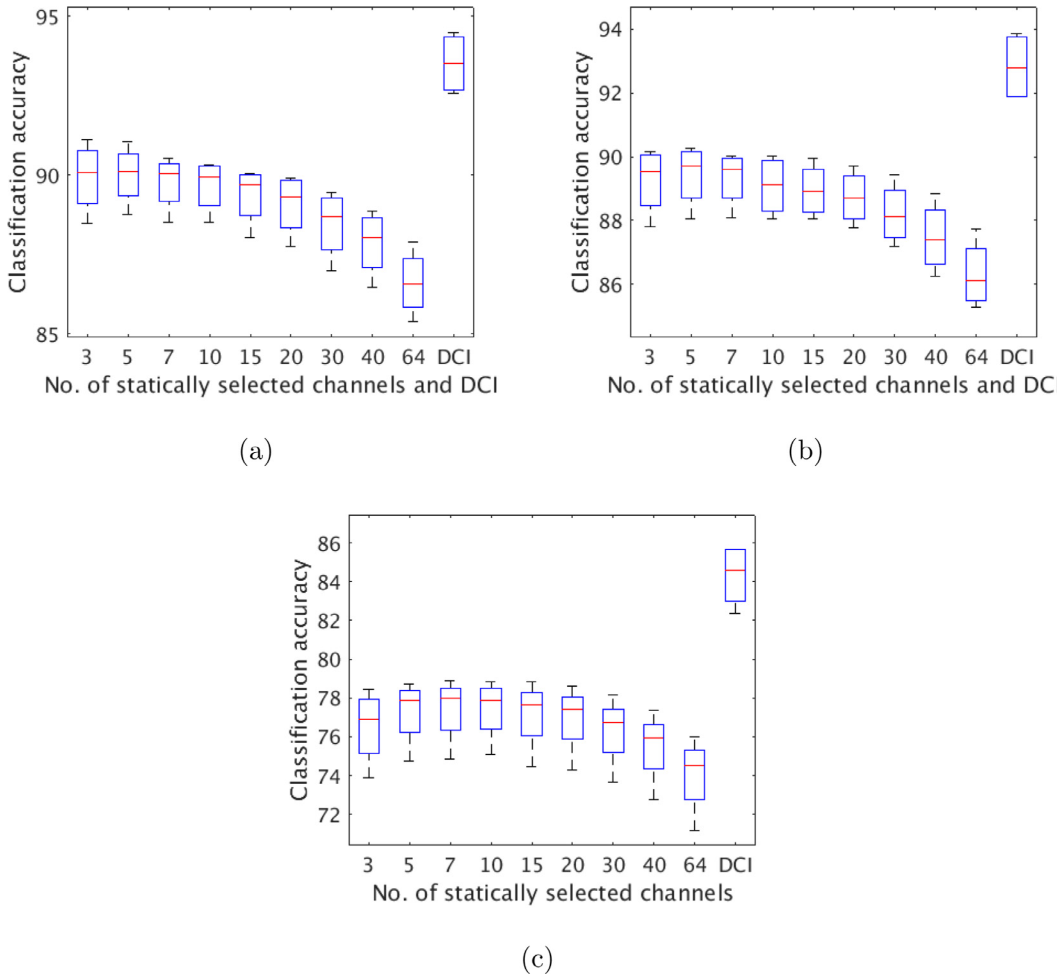

Fig. 7. Classification accuracy of static channel selection (first nine ticks) and the proposed dynamic channel identification (last tick) using wavelet features for (a) 2 classes, (b) 3 classes, and (c) 4 classes.

to have high similarity nor high accuracy values according to Eqs. (1) and (4) respectively.

4.4. Relevance of selected channels

EEG pattern identification, which involves the identification of relevant region(s) of the brain and relevant frequency rhythms, plays a very important role for many EEG applications. A number of previous studies have explored the alertness state identification problem. For example, Varri, Hirvonen, Hasan, Loula, and Haikkinen (1992) associated reduction in vigilance with a decrease in the amplitude, quantity and frequency of the posterior alpha rhythm and an increase in the slow wave components. It is reported in Nakamura, Sugi, Ikeda, Kakigi, and Shibasaki (1996) and Varri et al. (1992) that reduction in vigilance is associated with a decrease in the amplitude, quantity and frequency of the posterior alpha rhythm and an increase in the slow wave components. They also stated that reduction in vigilance can also be accompanied by a maximum of activity at the occipital or parieto-occipital region. Durmer and Dinges (2005) stated the following: (i) The dorsolateral prefrontal cortex (PFC) is one of the critical structures in a network of anterior and posterior ”attention control” areas; (ii) a posterior network including the superior parietal lobes, superior colliculus, and pulvinar is involved in switching attention from one target to another; (iii) increasing subjective sleepiness is associated with increases in spectral slow wave activity in the parietal and occipital regions and simultaneous decreases in α frequency waveforms. The wakefulness-sleep transition was studied in De Gennaro, Ferrara, Curcio, and Cristiani (2001), where it was stated that after the sleep onset, more EEG power was found in the range of slow frequencies at the centro-frontal scalp locations and a second peak of EEG activity was revealed within the range of the sigma frequency, higher at the centro-parietal scalp locations.

For the classification scenario of four classes, we counted the number of windows that each channel (and its three neighbours) has been selected for each of the four alertness states using both feature extraction methods (see Figs. 9 and 10, and Tables 2 and 3). We can notice that there is a good degree of similarity between the two figures, which indicates that despite using different feature sets, the proposed dynamic channel identification algorithm tends to have some degree of consistency in terms of selecting similar channels for each of the four classes.

For the *sleep* state, the selected channels are found to be mainly concentrated in the parietal region, with some activity in the central electrodes, which is consistent with De Gennaro et al. (2001). For the *drowsy* state, the selected channels are mainly from the parietal region, with some minor activities in other regions, including the parietal/occipital areas. These findings support to some extent the results reported in Durmer and Dinges (2005) and Nakamura et al. (1996). Although relatively more scattered, the selected channels for the *engaged* state show noticeable activity in the frontal area, which was not the case with the *sleep* and *drowsy*

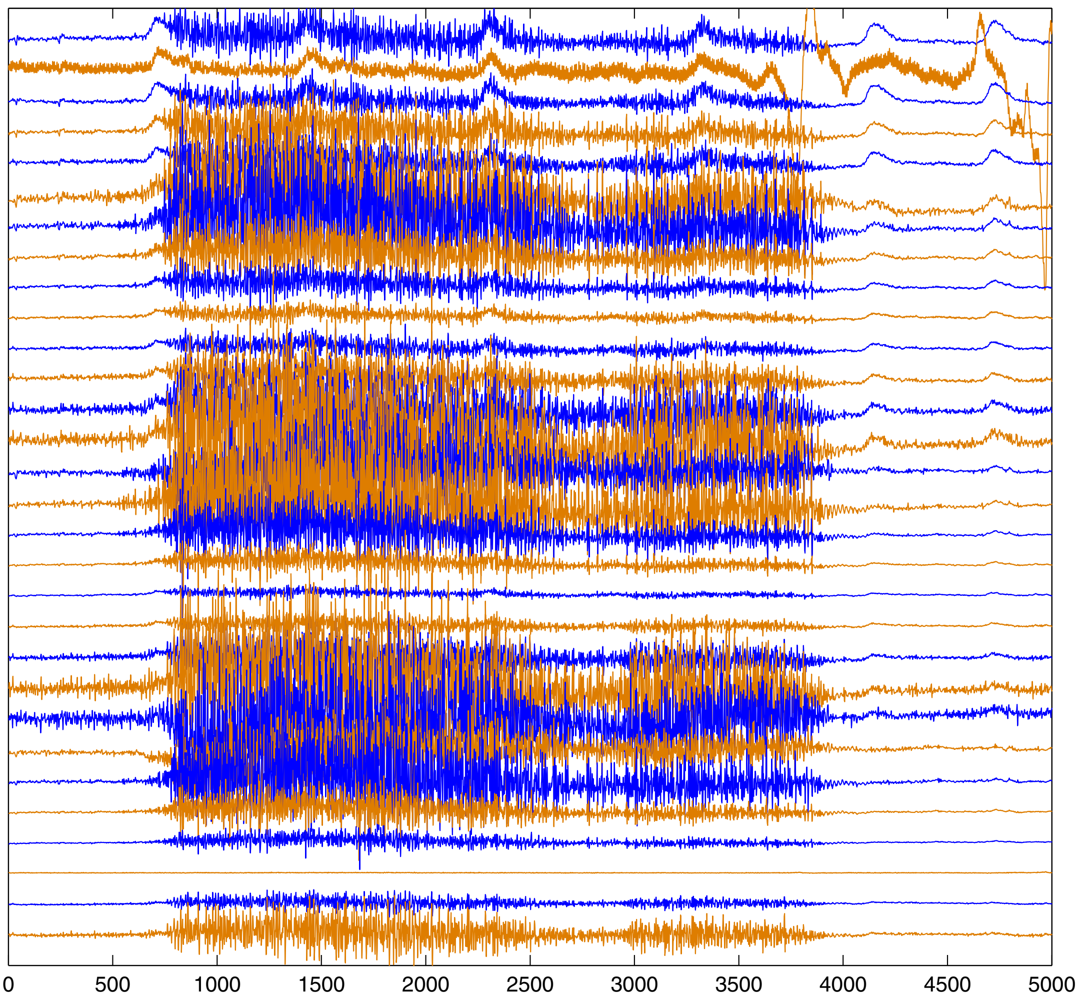

Fig. 8. Example of noisy EEG window.

states, and is also consistent with the "attention control" area defined in Durmer and Dinges (2005). As mentioned above, it was also reported in Durmer and Dinges (2005) that switching attention from one target to another involves a posterior network that includes the posterior parietal lobes. As there is no clear definition of the *calm* state in the literature, and the fact that the 10 subjects involved in the experiment made different interpretation of this state, which was reflected by their inconsistent labelling, it is quite hard to evaluate the selected channels of this particular state. We can however notice that the activities are mainly concentrated in the central and parietal regions. Activity in the parietal region demonstrate similarity (or overlap) with the *drowsy* state. Slight activity can also be noticed in the frontal region (similar to the *engaged* state). On the other hand, when looking at the statically selected channels, we noticed that they are also concentrated in the central/parietal region with some selections from frontal channels. Accordingly, those regions prove to be useful for both static and dynamic selection of channels.

It is important to mention that we are not claiming that our proposed method is capable of perfectly identifying the relevant channels of each class. However, the results reported in this paper show the potential of the proposed method in giving some indication of the importance of the different regions of the brain for the investigated classification task.

5. Conclusion

In this paper we proposed a novel dynamic channel identification algorithm for the classification of multichannel EEG data. The rationale behind proposing the method is to inherently facilitate the avoidance of noisy portions of data for the various channels, and the selection of relevant channel(s) for each time segment of the recording. The proposed method, which is inspired by the concept of multiple classifier behaviour, attempts to estimate the local accuracy of each channel for each time segment of the signal, which enables the dynamic identification of relevant channels. We showed that when the method was applied to the task of alertness state classification, it produced more accurate results than the traditional static channel selection, even when varying the channel subset sizes. Our evaluation was conducted using two different feature sets extracted to represent the signal segments – rhythmic and wavelet. Both feature sets performed well, with the wavelet slightly outperforming the rhythmic. We also validated the relevance of the selected channels, and the obtained results showed that our method tends to select channels from the brain regions that are known to be important for the given alertness states. For future research, we will investigate extending and applying the proposed method to other multichannel biomedical signals.

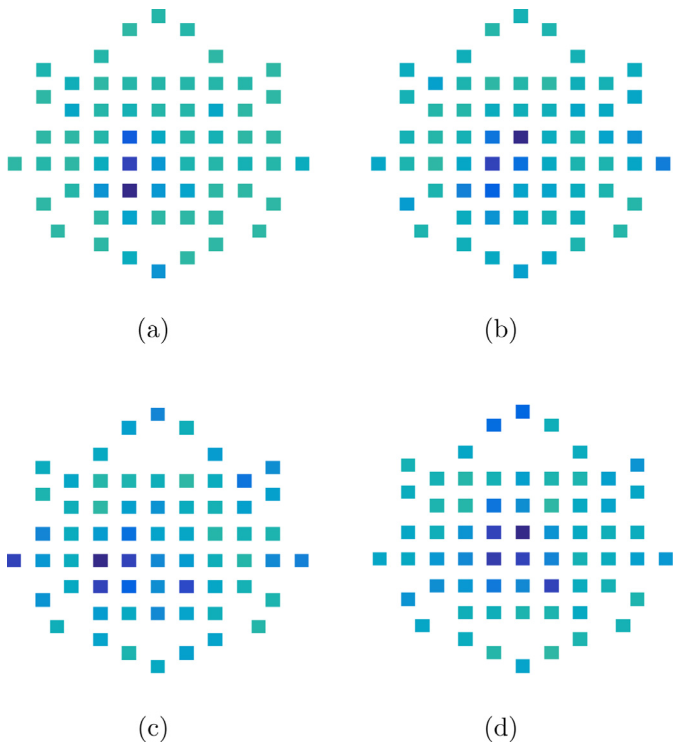

Fig. 9. EEG caps that reflect the number of times each of the 64 channels and their corresponding three neighbours are selected using the rhythm features for: (a) Sleep, (b) Drowsy, (c) Calm, and (d) Engaged. The darker the channel the more it is selected. Refer to Fig. 2 for channel locations.

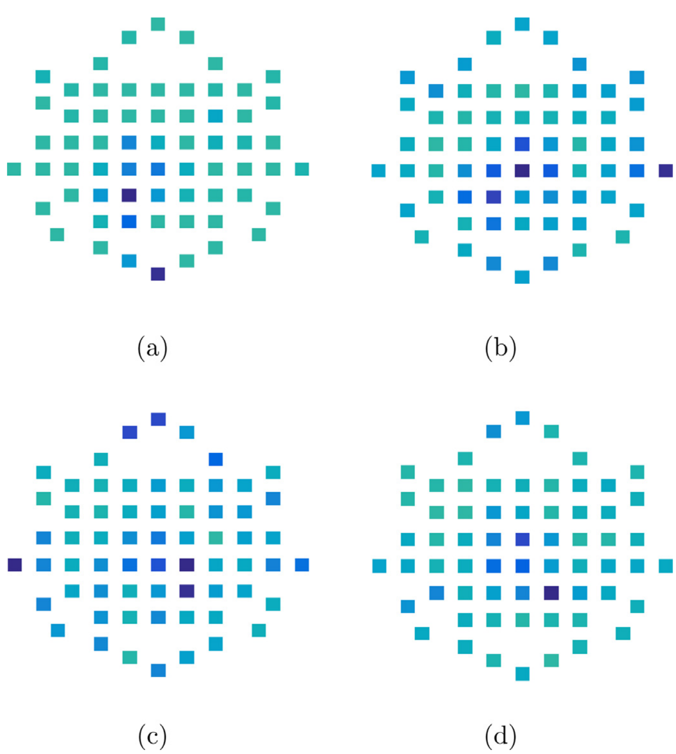

Fig. 10. EEG caps that reflect the number of times each of the 64 channels and their corresponding three neighbours are selected using the wavelet features for: (a) Sleep, (b) Drowsy, (c) Calm, and (d) Engaged. The darker the channel the more it is selected. Refer to Fig. 2 for channel locations.

Table 2
Number of times each of the 64 channels are selected for the four classes.

	EEG Rhythms				Wavelet features			
	Sleep	Drowsy	Calm	Engaged	Sleep	Drowsy	Calm	Engaged
FP1	2	13	57	253	3	59	137	225
FPZ	3	53	95	271	2	86	138	186
FP2	6	41	22	51	0	86	59	43
AF3	3	69	39	89	6	109	20	42
AF4	0	51	59	64	1	138	119	37
F7	9	91	30	41	14	123	26	17
F5	20	150	43	39	6	148	22	22
F3	1	35	7	10	2	44	18	2
F1	3	8	12	35	0	10	27	50
FZ	3	13	26	116	0	3	55	72
F2	6	6	2	21	0	18	25	21
F4	6	67	33	26	3	118	56	93
F6	5	51	119	114	0	96	41	100
F8	0	88	79	148	5	140	25	31
FT7	0	44	12	63	0	72	4	15
FC5	29	20	24	24	4	17	16	1
FC3	4	12	1	14	0	9	9	1
FC1	12	83	57	236	0	36	24	193
FCZ	5	65	71	162	1	32	50	128
FC2	4	72	29	6	0	35	8	13
FC4	31	114	19	73	30	58	63	63
FC6	0	102	54	104	2	56	57	79
FT8	0	119	57	81	4	110	80	66
T7	0	57	95	87	0	53	83	97
C5	1	9	30	44	0	2	21	10
C3	3	50	65	143	1	12	27	81
C1	109	309	123	333	66	84	67	263
CZ	26	514	49	382	21	287	97	457
C2	6	82	46	144	0	123	46	186
C4	13	28	7	75	15	13	3	11
C6	1	126	16	53	10	80	48	18
T8	1	203	13	63	0	133	44	65

Table 3
Number of times each of the 64 channels are selected for the four classes - cont.

	EEG Rhythms				Wavelet features			
	Sleep	Drowsy	Calm	Engaged	Sleep	Drowsy	Calm	Engaged
M1	0	110	168	90	0	89	166	139
TP7	2	38	67	78	0	63	80	119
CP5	1	8	34	131	0	14	25	45
CP3	20	117	191	174	21	145	68	81
CP1	121	454	169	348	69	270	105	357
CPZ	48	337	87	328	80	349	132	394
CP2	8	102	64	179	23	270	165	178
CP4	4	25	24	33	0	35	38	67
CP6	12	37	8	65	1	99	34	121
TP8	3	122	83	115	1	230	86	116
M2	21	275	99	138	8	340	105	106
P7	0	165	76	150	1	92	80	205
P5	3	28	40	165	1	31	37	257
P3	55	266	167	134	48	237	67	72
P1	140	383	142	227	135	308	24	164
PZ	51	140	79	174	48	116	57	256
P2	35	128	158	338	14	148	161	536
P4	1	89	43	58	5	116	55	176
P6	0	28	19	34	5	87	42	98
P8	0	15	11	28	0	92	27	66
PO7	0	35	26	83	2	34	60	96
PO5	1	82	65	52	14	32	63	82
PO3	36	192	36	52	96	219	16	49
POZ	0	87	84	21	2	64	75	26
PO4	0	42	48	25	1	105	33	30
PO6	0	71	24	78	0	89	30	40
PO8	0	9	8	42	0	27	20	61
CB1	0	64	51	59	1	62	75	91
O1	21	106	14	13	48	152	8	34
OZ	55	143	36	103	131	103	81	97
O2	0	119	57	1	0	156	46	8
CB2	0	27	51	22	0	27	45	46

References

Al-Ani, A., Alsukker, A., & Khushaba, R. N. (2013). Feature subset selection using differential evolution and a wheel based search strategy. *Swarm and Evolutionary Computation, 9*, 15–26.

Al-Ani, A., Mesbah, M., Van Dun, B., & Dillon, H. (2013). Fuzzy logic-based automatic alertness state classification using multi-channel eeg data. In *International conferenceon neural information processing (ICONIP)* (pp. 176–183).

Ansari, A., Karim, G. C., & Thierry, P. (2007). A channel selection method for eeg classification in emotion assessment based on synchronization likelihood. In *15th European signal processing conference (EUSIPCO)* (pp. 1241–1245).

Arvaneh, M., Guan, C., Ang, K. K., & Quek, C. (2011). Optimizing the channel selection and classification accuracy in eeg-based bci. *IEEE Transactions on Biomedical Engineering, 58*(6), 1865–1873.

De Gennaro, L., Ferrara, M., Curcio, G., & Cristiani, R. (2001). Antero-posterior eeg changes during the wakefulness-sleep transition. *Clinical Neurophysiology, 112*(10), 1901–1911.

Durmer, J., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology, 25*(1), 117–129.

Duun-Henriksen, J., Kjaer, T. W., Madsen, R. E., Remvig, L. S., Thomsen, C. E., & Sorensen, H. B. D. (2012). Channel selection for automatic seizure detection. *Clinical Neurophysiology, 123*(1), 84–92.

Faul, S., & Marnane, W. (2012). Dynamic, location-based channel selection for power consumption reduction in eeg analysis. *Computer Methods and Programs in Biomedicine, 108*(3), 1206–1215.

Giacinto, G., & Roli, F. (2001). Dynamic classifier selection based on multiple classifier behaviour. *Pattern Recognition, 34*(9), 1879–1–881.

Gianluca, B., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews, 44*, 58–75.

Gonzalez, A., Nambu, I., Hokari, H., & Wada, Y. (2014). Dynamic eeg channel selection using particle swarm optimization for the classification of auditory event-related potentials. *The Scientific World Journal, 2014*(1–12).

He, L., Hu, Y., Li, Y., & Li, D. (2013). Channel selection by rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery eeg. *Neurocomputing, 121*, 423–433.

Huang, Y. S., & Suen, C. Y. (1995). A method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(1), 90–94.

Koprinska, I., Pfurtscheller, G., & Flotzinger, D. (1996). Sleep classification in infants by decision tree-based neural network. *Artificial Intelligence in Medicine, 8*(4), 387–401.

Kota, S., Gupta, L., Molfese, D. L., & Vaidyanathan, R. (2009). A dynamic channel selection strategy for dense-array erp classification. *IEEE Transactions on Biomedical Engineering, 56*(4), 1040–1051.

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., & Scholkopf, B. (2004). Robust classification of eeg signal for brain-computer interface. *IEEE Transactions on Biomedical Engineering, 51*(6), 1003–1010.

Lan, T., Erdogmus, D., Adami, A., Mathan, S., & Pavel, M. (2007). Channel selection and feature projection for cognitive load estimation using ambulatory EEG. *Computational Intelligence and Neuroscience*, 8–20.

LeVan, P., Urrestarazu, E., & Gotman, J. (2006). A system for automatic artifact removal in ictal scalp eeg based on independent component analysis and bayesian classification. *Clinical Neurophysiology, 117*, 912–927.

Mirowski, P., Madhavan, D., LeCun, Y., & Kuzniecky, R. (2009). Classification of patterns of eeg synchronization for seizure prediction. *Clinical Neurophysiology, 120*(11), 1927–1940.

Nakamura, M., Sugi, T., Ikeda, A., Kakigi, R., & Shibasaki, H. (1996). Clinical application of automatic integrative interpretation of awake background. eeg: Quantitative interpretation, report making, and detection of artifacts and reduced vigilance level. *Electroen cephalography and Clinical Neurophysiology, 98*(2), 103–112.

Ning-Han, L., Chiang, C. Y., & Hsu, H. M. (2011). Improving driver alertness through music selection using a mobile eeg to detect brainwaves. *Sensors, 13*(7), 8199–8221.

Piryatinska, A., Woyczynski, W. A., Scher, M. S., & Loparo, K. A. (2012). Optimal channel selection for analysis of eeg-sleep patterns of neonates. *Computer Methods and Programs in Biomedicine, 106*(1), 14–26.

Platt, B., & Gernot, R. (2011). The cholinergic system, eeg and sleep. *Behavioural Brain Research, 221*(2), 499–504.

Schrder, M., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N., Birbaumer, N., Rosenstiel, W., & Schlkopf, B. (2005). Robust eeg channel selection across subjects for brain-computer interfaces. *EURASIP Journalon Advances in Signal Processing*, 3103–3112.

Siamac, F., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Mller, K., & Blankertz, B. (2012). Enhanced performance by a hybrid nirs-eeg brain computer interface. *Neuroimage, 59*(1), 519–529.

Thulasidas, M., Guan, C., & Wu, J. (2006). Robust classification of eeg signal for brain-computer interface. *IEEE Transactions on Neural Systemsand Rehabilitation Engineering, 14*(1), 24–29.

Torres-Garcia, A., Reyes-Garcia, C., Villasenor-Pineda, L., & Garcia-Aguilar, G. (2016). Implementing a fuzzy inference system in a multibjective eeg channel selection model for imagined speech classification. *Expert Systems With Applications, 59*(1), 1–12.

Varri, A., Hirvonen, K., Hasan, J., Loula, P., & Haikkinen, V. (1992). A computerized analysis system for vigilance studies. *Computer Methods and Programs in Biomedicine, 39*(1), 113–124.

Wang, Y., Shangkai, G., & Xiaorong, G. (2006). Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *27th annual international conference IEEE engineering in medicine and biology society* (pp. 5392–5395).

Woods, K. (1997). Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(4), 405–410.