Progress in Artificial Intelligence manuscript No.
(will be inserted by the editor)

Interval Feature Extraction for Classification of Event-Related Potentials (ERP) in EEG Data Analysis

Ludmila I. Kuncheva · Juan J. Rodríguez

Received: date / Accepted: date

Abstract Event-related potential data can be used to index perceptual and cognitive operations. However, they are typically high-dimensional and noisy. This study examines the original raw data and 6 feature extraction methods as a pre-processing step before classification. Four traditionally used feature extraction methods were considered: principal component analysis, independent component analysis, auto-regression, and wavelets. We add to these a less well-known method called Interval Feature Extraction. It overproduces features from the ERP signal and then eliminates irrelevant and redundant features by the fast correlation based filter. To make the comparisons fair, the other feature extraction methods were also run with the filter. An experiment on two EEG data sets (4 classification scenarios) was carried out to examine the classification accuracy of four classifiers on the extracted features: support vector machines with linear and perceptron kernel, the nearest neighbour classifier and the random forest ensemble method. The interval features led to the best classification accuracy in most of the configurations, specifically when used with the Random Forest classifier ensemble.

Keywords Pattern recognition, EEG, feature extraction, event-related potentials (ERP), interval features, Fast Correlation-Based Filter (FCBF)

Ludmila I. Kuncheva
School of Computer Science, Bangor University, Bangor Gwynedd LL57 1UT, UK
E-mail: l.i.kuncheva@bangor.ac.uk

Juan J. Rodríguez
Departamento de Ingeniería Civil, Universidad de Burgos, Spain
E-mail: jjrodriguez@ubu.es

1 Introduction

Event-Related Potentials (ERP) are stereotyped electrophysiological responses to a volitional activity or an external stimulus. ERPs are derived by averaging the spontaneous electroencephalogram (EEG) across several occurrences of the same event, for example repeated presentations of the same stimulus. Assuming that the onset of the event is known, the ERP is a time series containing the ensuing brain reaction, termed an *epoch*. ERP can be used to distinguish between the brain states of the subject in response to the different stimuli within the particular examination context. This makes ERP analysis an ideal, long-standing tool in brain-computer interface (BCI) [2], ranging from predicting the laterality of imminent left-versus right-hand finger movements [1] to distinguishing between musical segments [22].

Classification of ERP data is challenging due to the high dimensionality and noise contamination of the raw data. Here we study the classification accuracy of the original data and four feature extraction methods used as a preprocessing step, and argue that ERP data benefits from Interval Feature Extraction.

ERP are associated to different electrodes. For each electrode a different classification problem is considered. The ERP time series is regarded as a collection of measurements, $x_1, \ldots, x_T$, each measurement is considered to be a feature. Thus each ERP series becomes a point in a T-dimensional space, $\mathbf{x} \in \Re^T$. A training data set is a collection of points $\mathbf{x}_1, \ldots \mathbf{x}_N$ and the corresponding labels $y_1 \ldots y_N$.

Section 2 explains the feature extraction methods, Section 3 lists the four classifiers, and Section 4 details the data, the experimental protocol and the results.

2 Feature extraction methods

2.1 The original (raw) data

A classification method is applied directly to the given training data set: N data points $\mathbf{x}_1, \ldots \mathbf{x}_N$ with $\mathbf{x}_i \in \Re^T$ and labels $y_1 \ldots y_N$.

To illustrate the difficulty in classifying ERP data, Figure 1 shows 44 individual ERPs coming from an experiment where 22 participants viewed two categories of objects (called here classes or conditions): faces and cars. The grand-average ERPs for the two classes are plotted with thick lines. There is no clear pattern to guide the classifications.

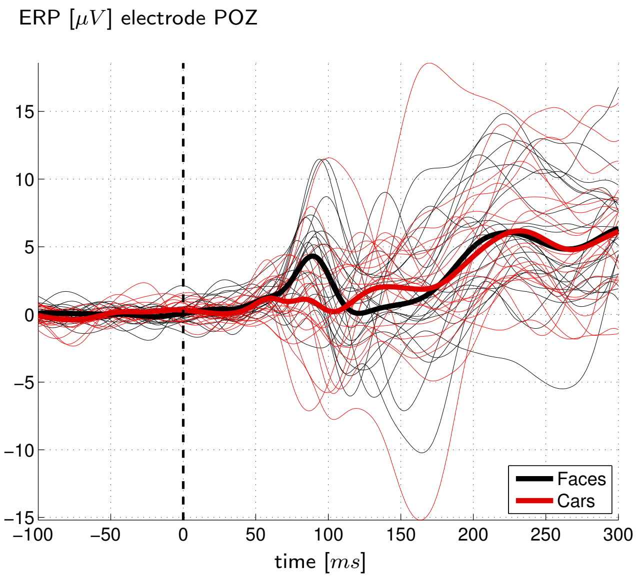


Fig. 1 Individual and grand-average ERP curves at electrode location POZ. The onset of the event is marked with a dashed line.

2.2 PCA

Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Auto-regression models (AR) are widely used feature extraction methods in EEG/ERP data analysis [4,7,16,15,25].

PCA calculates a transformation matrix W such that the original features are replaced with extracted features, obtained as $\mathbf{a} = W\mathbf{x}$. This effectively rotates the space so that the fist component of $\mathbf{a}$ is orientated along the dimension with the largest variance in the data, the second component is orthogonal to the first one, and is orientated along the largest remaining variance, and so on. The number of retained components is determined so that a chosen percentage of the variance of the data is preserved after the transformation (we used 95% in the experiments).

2.3 ICA

ICA also looks for a linear transformation of the original space in the form $\mathbf{a} = W\mathbf{x}$ but the derivation of W is governed by a different criterion: the components of $\mathbf{a}$ should be as independent as possible. Traditionally, ICA has been applied for the task of "blind source separation", where a signal is decomposed into a weighted sum of independent source signals [4,10], but there is no reason why we should stop here. Any feature set can be processed so as to derive a new smaller set of independent features. We used the FastICA algorithm due to Hyvärinen [9][1].

2.4 Auto-Regression (AR)

Auto-regressive (AR) models regard the ERP as a dynamic system of order K where the i-th value of the signal is predicted by the previous K values $\hat{x}_i = a_1\, x_{i-K+1} + a_2\, x_{i-K+2} + \cdots + a_K\, x_{i-1}$. The coefficients a_j are fitted to minimise the discrepancy between the prediction $\hat{x}_i$ and the true value x_i, and treated as the extracted features. Models of order up to $K = 50$ have been considered [16].[2]

2.5 Wavelets

A family of wavelets are scaled and shifted, so that the ERP signal is approximated in the best possible way. Wavelets capture both finer and coarser resolution features of the signal. The features which represent the ERP signal are the coefficients of the wavelets. Wavelet feature extraction has been widely applied to EEG analysis, opening up the perspectives of obtaining highly informative spatial-spectral-temporal descriptions [12]. The wavelet method has found its use in ERP analysis and classification too [17,14,18,21,28].

2.6 Interval features

Features derived from the signal over time segments of various lengths, called "interval features", have been shown to lead to high classification accuracy [19,20].

[1] `http://sourceforge.net/projects/fastica/`

[2] The Signal Processing Toolbox of Matlab was used for the fitting of the AR model.

Denote by x_i the value of the signal at time i, where $i = 1, \ldots, T$ spans the ERP epoch. Nested time intervals are formed for each i using the remaining series $x_i, \ldots, x_T$. The interval length is varied as the power of 2, while the starting point is kept at i. Thus the first interval at i is the point set $\{x_i, x_{i+1}\}$, the second interval is $\{x_i, x_{i+1}, x_{i+2}, x_{i+3}\}$, and so on until the interval length exceeds the remainder of the epoch. The last interval will therefore be $\{x_{T-1}, x_T\}$. All the intervals are pooled, and the following features are extracted from each interval: (1) the average amplitude of the point set μ, (2) the standard deviation σ, and (3) the covariance with the time variable. Denote the beginning of the interval by i_b, the end of the interval by i_e and the length by $l = i_e - i_b + 1$. The mean of the time variable is therefore $\bar{t} = \frac{(i_b+i_e)}{2}$. Denoting the sample estimate of the expectation by $E(.)$, the covariance is calculated as[3]

$$\gamma \;=\; E(tx) - E(x)E(t) \;=\; \frac{1}{l}\sum_{k=i_b}^{i_e} k\; x_k - \mu\;\frac{(i_b+i_e)}{2}.$$

Since the interval feature extraction is less well known compared to the remaining five feature extraction methods, we give an example below. Consider a hypothetical ERP time series spanning 10 time points tagged 1, 2, ..., 10. In this case $i_b = 1$, $i_e = 10$, and the length is $l = i_e - i_b + 1 = 10$. The complete set of 19 time intervals is shown in Table 1. These 19 intervals are also diagrammatically presented as horizontal line segments in Figure 2 (a). They have only an x-dimension, and are offset on the y-axis for visualisation purposes.

Subplots (b) and (c) show the amplitude features for intervals of length 2 and 4, respectively. The amplitudes are calculated as the mean value of the ERP for the respective interval.

In essence, the 'average' interval feature extraction produces a collection of outputs of a series of smoothing filters and can be interpreted as a wavelet-type feature. However, the remaining interval features are different by design. The total number of interval features for this data set is $19 \times 3 = 57$. The 10 original features are added to the collection, giving a total of 67 features. Typically, there is likely to be substantial redundancy among the interval features, which suggests that classification will benefit from further filtering.

2.7 FCBF

Fast Correlation-Based Filter (FCBF) is a feature selection method originally proposed for microarray data analysis [30] and adopted for other domains with very high-dimensional data [11]. The idea of FCBF is that the features that are worth keeping should be correlated with the class variable but not correlated among themselves. If they are mutually correlated, one feature from the correlated group will suffice. By design, the interval feature set will contain many redundant features, which makes FCBF particularly suitable.

FCBF starts by ranking the features according to their mutual information with the class variable and removing all features whose mutual information is below a chosen threshold value. Subsequently, features are removed from the remaining list by considering simultaneously their predictive value and their relationship with other features. Feature x is removed from the list if there exists feature y in the list, which is a better predictor of the class label than x, *and* feature x is more similar to feature y than to the class label variable. The 'similarity' is measured through mutual information. The process runs until there are no more features that can be removed. The number of selected features depends on the data and partly on the threshold value used to cut off the initial tail of features with low relevance.

3 Classification methods

We chose the nearest neighbour classifier (1-nn) which has been praised for being simple, accurate and intuitive. Next, the Support Vector Machine classifier (SVM) has been the classifier of choice in an overwhelming amount of neuroscience studies. In spite of the scepticism towards using classifiers of increased complexity for EEG data [7], classifier ensembles have made a successful début in the past decade [26,17,23,29]. Here we chose the Random Forest ensemble [3] (RF) because of its accuracy and ability to work with high-dimensional data. Each RF consisted of 100 individual trees.

If not specified otherwise, all feature extraction and classification methods were used within Weka[4] [8], with the default set-up. Two kernels were considered for the SVM: linear $K(x, y) = \langle x, y\rangle$ and perceptron $K(x, y) = ||x - y||_2$ [13]. The latter is less known than the RBF kernels but has the advantage of not requiring any parameter tuning.

For some classification methods the results could be very different if the parameters where adjusted instead of using default values. The focus of this work is the comparison of feature extraction methods on different electrodes, not the comparison among different classification methods. For each considered classification method, the feature extraction methods are com-

[3] The biased estimate is used without loss of generality.

[4] `http://www.cs.waikato.ac.nz/ml/weka/`

Table 1 The 19 intervals for a time series with 10 time points

Starting at	1	2	3	4	5	6	7	8	9
length 2	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10
length 4	1–4	2–5	3–6	4–7	5–8	6–9	7–10		
length 8	1–8	2–9	3-10						

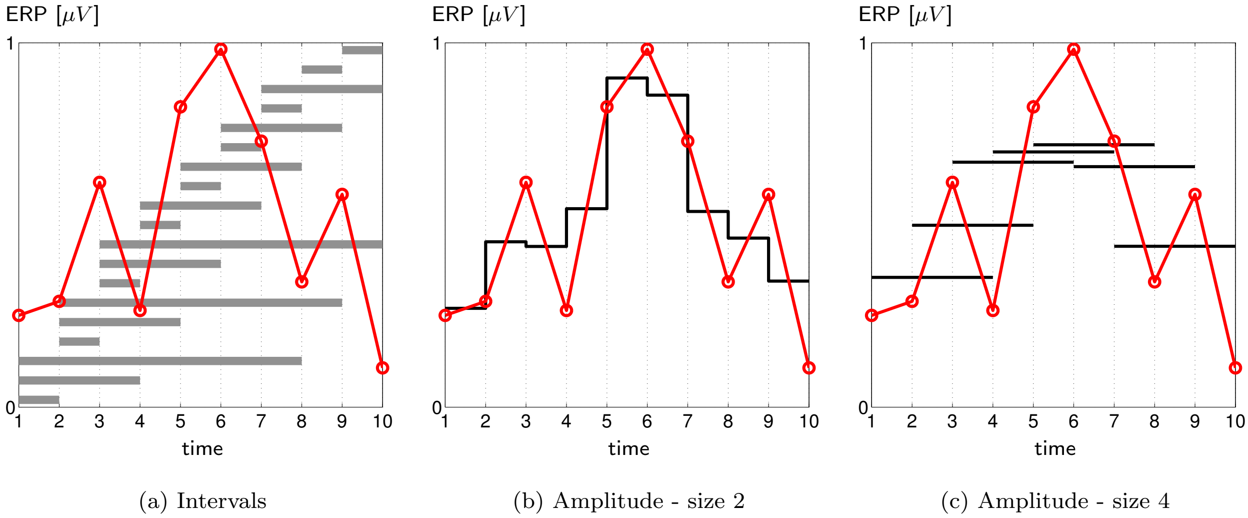


Fig. 2 An Interval Features example. (a) The grey horizontal line segments represents the intervals. They have only an x-dimension, and are offset on the y-axis for visualisation purposes. The curve is the ERP; (b) and (c) show the amplitude features for intervals of length 2 and 4, respectively. The amplitudes are calculated as the mean value of the ERP for the respective interval.

pared. Nevertheless, it must be noted that the reported results could be probably improved with a thoroughly adjustment of the parameters.

4 Experiments

The method that we recommend in this study is interval feature extraction followed by FCBF. To the best of the authors' knowledge interval features have not been applied previously to ERP classification, either as a full set or after FCBF filtering. The purpose of this experiment is to compare the Interval Features + FCBF with the traditional feature extraction methods for preprocessing of ERP data for classification. To make the comparisons fair, we applied FCBF to all competing methods as well.

4.1 Data sets

• *Face/Car data.* The first data set is about distinguishing between images of faces and of cars. A negative peak segment of the ERP, labelled N170, is commonly acknowledged to be larger in amplitude for face stimuli compared with any other visual object. Recently, the face selectivity of the N170 has been challenged [27]. The data set used in this study came from [5]. Stimuli were 400 grey-scale pictures of faces and cars, all presented in full-front views, scaled to produce a similar size and width/height ratio, and centred on the screen. The responses of each participant to the same group of stimuli were averaged to obtain the ERPs. The original data consisted of 64 electrodes and 44 ERPs for each electrode: 22 from class 'Face' and 22 from class 'Car'. Each ERP spanned exactly 300 ms starting at the onset of the event.

• *Alcoholism data.* The second data set arises from a study to examine EEG correlates of genetic predisposition to alcoholism. The data set is included in the UCI Machine Learning Repository collection [6]. It contains measurements from 64 electrodes placed on standard locations on the subject's scalps which were sampled at 256 Hz for 1 second [31]. There were two groups of subjects: alcoholic and control, which are the two class labels to discriminate between. Each subject was exposed to either a single stimulus (S1) or to two stimuli (S1 and S2) which were pictures of objects chosen from the 1980 Snodgrass and Vanderwart picture set [24].

The three experimental scenarios were as follows (A) Single stimulus S1; (B) Two matching stimuli (S1 and S1); and (C) Two different stimuli (S1 and S2). The chosen data set contains data for 10 alcoholic and 10 control subjects, with 10 runs per subject per scenario.

4.2 Experimental protocol

Ten times 10-fold cross-validation was carried out for each classifier model, thus giving 100 testing accuracies from which the average accuracies were calculated. The repeat of the cross-validation was done to reduce the effect of the split of the folds.

The feature extraction was done entirely on the *training* folds. The FCBF was applied to the original features and to the 4 feature extraction methods. Each electrode data was processed separately. There are 4 sets of results: one with the Face/Car data, and three with the A, B and C scenarios for the Alcoholism data. Each set contains results for 64 electrodes. For each electrode, we carried out 44 experiments = 4 (classifiers) $\times$ [5 methods (original+PCA+ICA+AR+WAV) $\times$ 2 (FCBF) + 1 (Interval Features FCBF)]. Thus the results from each of the 4 experiments are two 64-by-44 matrices - one with the average classification accuracies, and one with the standard deviations calculated from the 100 cross-validation testing folds.

4.3 Results

We grouped the results by classifiers and show in Table 2 the maximum achieved accuracies along with the respective electrode labels for the methods *without FCBF*. Table 3 has the same format but this time FCBF has been applied after the selection. The Interval method is the same because FCBF is an integral part of it. Wilcoxon signed rank tests were run to determine whether the difference with the Interval Features accuracies are statistically significant at level 0.05. A mark of $\circ$ in the tables means that the respective methods is significantly better than the interval features method, $\bullet$ means that it is significantly worse and $-$ means that the difference is not statistically significant.

The tables show that the top accuracies with the interval features are often larger than these with the other feature extraction methods, and in many experimental configurations these differences are statistically significant (level 0.05). The Interval method was found significantly worse only in four cases (two in Table 2 and two in Table 3). It is interesting to note that the differences in favour of the Interval methods are most pronounced in the bottom block-row of both tables. These rows correspond to the Random Forest ensembles, which give the highest classification accuracies across the classification models.

We chose the Face/Car data set to illustrate the results further. Figure 3 shows the distribution of the *classification accuracy* across the electrodes on the scalp for two combinations of a feature extraction method and a classifier. The anatomical reference of the electrode with the best accuracy is also shown.[5] Subplot (a) shows the combination with the highest maximal accuracy and subplot (b), the combination with the highest average accuracy across all electrodes.

The maximum of the classification accuracy for this data set (Experiment 1), achieved by the interval features with FCBF and the Random Forest classifier ensembles, falls in the parietooccipital electrode site, which is the expected brain area for distinguishing face from non-face objects [5].

5 Conclusions

Feature extraction methods are likely to enhance the classification accuracy in ERP classification. We showed that interval features followed by a Fast Correlation Based Filter (FCBF) are a suitable tool for this task. The first step over-produces interval-based features, a large part of which are redundant, irrelevant or both. The second step picks the most relevant and uncorrelated features. The resultant feature set works well with various classifier models, including SVM and classifier ensembles, represented here by Random Forest.

Compared to the other feature extraction methods, the Interval method is simple and intuitive. The simplicity is important in view of the large variability of the raw data, as illustrated in Figure 1. A more intricate and versatile model such as wavelets may overfit the noise. Unlike PCA and ICA, the Interval extraction method is non-linear, which gives it additional flexibility. Finally, the method does not rely on a parameter of significant importance as does auto-regression extraction method.

An interesting future direction is joining the electrode data and extracting interval features from the multidimensional ERP. This reflects the fact that brain activity is typically not localised but involves spatially coordinated brain activation patterns.

Acknowledgements The authors are grateful to Prof Guillaume Thierry and Dr Benjamin Dering from the School of

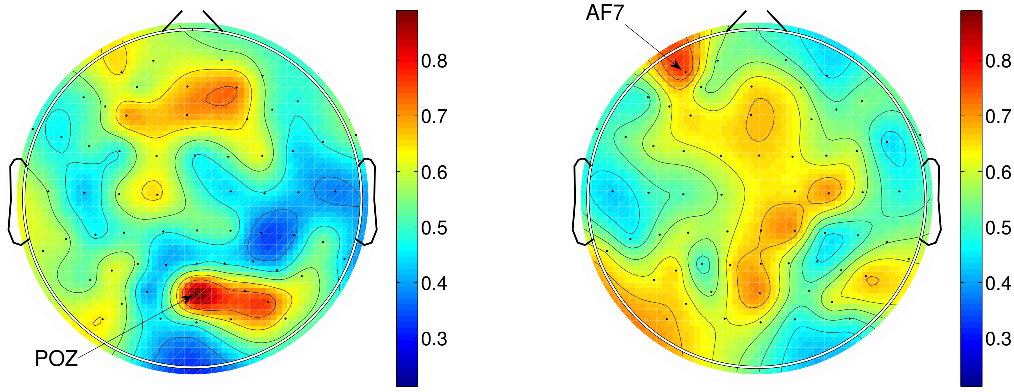
[5] The plots were produced with the EEGLAB toolbox for Matlab.

Table 2 Maximum accuracy across the 64 electrodes *without FCBF*. A mark of ○ means that the respective methods is significantly better than the interval features method, • means that the interval method is significantly better, and – means that the difference is not statistically significant (level 0.05).

		Experiment 1		Experiment 2		Experiment 3		Experiment 4	
SVM (lin)	Original	– 75.8	AF7	• 71.0	CPZ	○ 94.5	P3	– 81.0	CP1
	PCA	• 70.7	FCZ	– 77.0	FZ	– 79.5	P3	• 77.5	CP1
	ICA	• 66.8	FCZ	– 79.5	AFZ	– 68.5	P1	– 79.0	CP1
	AR	• 73.0	CP4	– 85.0	P1	– 75.5	CP6	– 87.0	CP6
	WAV	• 71.5	O1	○ 90.5	CP2	– 79.5	C4	– 83.0	OZ
	Interval	79.9	POZ	82.0	CP2	76.5	CP3	86.0	C6
SVM (perc)	Original	• 67.5	POZ	– 83.5	F5	– 80.5	P3	• 86.5	CP1
	PCA	• 68.4	C2	• 77.5	FZ	• 78.0	CP3	• 81.0	CP4
	ICA	• 64.5	F3	– 80.5	F5	• 71.0	P3	• 80.0	CP1
	AR	• 69.9	CP4	– 83.0	P1	– 82.5	CP6	• 80.5	CP4
	WAV	• 70.8	AF7	– 85.0	CP2	– 78.0	C4	• 76.5	PZ
	Interval	80.2	POZ	85.5	nd	85.0	CP3	97.0	F7
NN	Original	• 61.3	FP1	• 80.0	AF7	• 76.0	CP5	• 80.0	FCZ
	PCA	• 54.6	POZ	• 75.5	AFZ	• 77.5	P4	• 86.0	P2
	ICA	– 73.0	FZ	• 74.5	F5	• 73.0	C1	• 82.5	P1
	AR	– 67.5	CP2	– 81.5	P3	• 73.0	P2	• 71.5	F6
	WAV	• 57.5	P5	• 74.5	AF7	• 71.5	CP6	• 71.0	F7
	Interval	69.8	PO8	88.0	nd	90.5	CP3	97.0	F7
RF	Original	• 69.0	P8	• 76.0	F5	• 75.5	O1	• 81.0	PO2
	PCA	• 66.6	P7	– 84.5	F5	• 82.5	O2	• 82.5	CP1
	ICA	• 74.4	FZ	• 77.5	AFZ	• 69.0	P1	• 75.0	PZ
	AR	• 58.5	FT7	– 88.0	F7	• 75.5	POZ	• 80.0	CP6
	WAV	• 75.0	POZ	• 78.0	FP2	• 70.0	FCZ	• 73.5	P7
	Interval	89.0	POZ	88.0	nd	91.0	CP3	97.0	F7

Table 3 Maximum accuracy across the 64 electrodes *with FCBF*. A mark of ○ means that the respective methods is significantly better than the interval features method, • means that the interval method is significantly better, and – means that the difference is not statistically significant (level 0.05).

		Experiment 1		Experiment 2		Experiment 3		Experiment 4	
SVM (lin)	Original	• 69.3	C1	– 75.5	F5	– 80.0	PO1	– 80.5	PO7
	PCA	• 65.5	FC5	– 79.0	F5	• 70.5	P8	– 81.0	CPZ
	ICA	• 66.7	P8	– 76.0	AFZ	• 64.5	AFZ	• 65.0	F6
	AR	• 62.0	FT7	○ 90.5	CP1	– 80.0	AF1	– 80.0	AF2
	WAV	– 76.0	POZ	– 82.5	PZ	– 71.5	P8	– 86.0	CPZ
	Interval	79.9	POZ	82.0	CP2	76.5	CP3	86.0	C6
SVM (perc)	Original	• 71.1	P5	– 79.5	AFZ	– 85.5	PO1	• 80.5	CP1
	PCA	• 63.6	FC5	– 82.5	F5	– 81.5	O2	• 88.0	CP1
	ICA	• 63.7	C1	• 77.0	AF7	• 61.5	POZ	• 67.0	CP1
	AR	• 64.3	FT10	– 90.5	CP1	– 84.5	AFZ	• 81.0	FT7
	WAV	– 76.7	POZ	– 84.0	PZ	• 71.0	P8	• 82.5	CPZ
	Interval	80.2	POZ	85.5	nd	85.0	CP3	97.0	F7
NN	Original	– 68.8	F4	• 74.5	AFZ	• 80.5	PO2	• 82.5	CP3
	PCA	• 61.1	CPZ	• 74.5	PZ	• 79.0	FC1	• 86.0	CPZ
	ICA	– 73.3	FZ	• 66.5	AF7	• 64.0	C1	• 78.0	AF8
	AR	• 61.6	F4	– 88.0	CP1	• 80.5	FC1	• 77.0	AF2
	WAV	– 70.0	POZ	– 81.0	FZ	• 74.0	FCZ	• 81.0	CP2
	Interval	69.8	PO8	88.0	nd	90.5	CP3	97.0	F7
RF	Original	• 68.3	F4	• 77.5	F2	• 83.5	P2	• 90.0	CP3
	PCA	• 60.9	CPZ	• 73.5	F5	• 77.0	O2	• 86.5	CPZ
	ICA	• 72.5	FZ	• 70.0	AF7	• 61.0	O2	• 68.0	AF8
	AR	• 64.6	F4	○ 94.5	C3	– 90.0	FC2	• 82.0	AF2
	WAV	• 79.8	POZ	– 82.5	PO8	• 79.5	FCZ	• 83.5	P6
	Interval	89.0	POZ	88.0	nd	91.0	CP3	97.0	F7



(a) Interval features FCBF (b) Original features

Fig. 3 Approximation of the *classification accuracy* across the electrodes on the scalp for the Face/Car data (Experiment 1)

Psychology, Bangor University, UK, for providing the first data set used in this study and for the useful discussion and comments on an earlier version of this manuscript.

References

1. Blankertz, B., Dornhege, G., Schafer, C., Krepki, R., Kohlmorgen, J., Muller, K.R., Kunzmann, V., Losch, F., Curio, G.: Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial eeg analysis. IEEE Transactions on Neural and Rehabilitation Engineering **11**(2), 127–131 (2003)
2. Blankertz, B., Muller, K.R., Curio, G., Vaughan, T., Schalk, G., Wolpaw, J., Schlogl, A., Neuper, C., Pfurtscheller G .and Hinterberger, T., Schroder, M., Birbaumer, N.: The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering **51**(6), 1044–1051 (2004)
3. Breiman, L.: Random forests. Machine Learning **45**, 5–32 (2001)
4. Delorme, A., Sejnowski, T., Makeig, S.: Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. NeuroImage **34**, 1443–1449 (2007)
5. Dering, B., Martin, C.D., Thierry, G.: Is the N170 peak of visual event-related brain potentials car-selective? NeuroReport **20**, 902–906 (2009)
6. Frank, A., Asuncion, A.: UCI machine learning repository (2010)
7. Garrett, D., Peterson, D.A., Anderson, C.W., Thaut, M.H.: Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **11**, 141–144 (2003)
8. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. SIGKDD Explorations **11** (2009)
9. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neural Computation **9**(7), 1483–1492 (1997)
10. Jung, T., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.: Analysis and visualization of single-trial event-related potentials. Human Brain Mapping **14**(3), 166–185 (2001)
11. Lai, C., Reinders, M.J.T., Wessels, L.: Random subspace method for multivariate feature selection. Pattern Recognition Letters **27**(10), 1067 – 1076 (2006)
12. Li, J., Zhang, L.: Regularized tensor discriminant analysis for single trial EEG classification in BCI. Pattern Recognition Letters **31**, 619–628 (2010)
13. Lin, H.T., Li, L.: Novel distance-based SVM kernels for infinite ensemble learning. In: In Proceedings of the 12th International Conference on Neural Information Processing, ICONIP'05, pp. 761–766 (2005)
14. Lu, B.L., Shin, J., Ichikawa, M.: Massively parallel classification of single-trial EEG signals using a min-max modular neural network. IEEE Transactions on Biomedical Engineering **51**, 551–558 (2004)
15. Palaniappan, R., Ravi, K.: Improving visual evoked potential feature classification for person recognition using pca and normalization. Pattern Recognition Letters **27**(7), 726 – 733 (2006)
16. Peters, B.O., Pfurtscheller, G., Flyvbjerg, H.: Automatic differentiation of multichannel EEG signals. IEEE Transactions on Biomedical Engineering **48**, 111–116 (2001)
17. Polikar, R., Topalis, A., Green, D., Kounios, J., Clark, C.M.: Comparative multiresolutionwavelet analysis of erp spectral bands using an ensemble of classifiers approach for early diagnosis of Alzheimer's disease. Computers in Biology and Medicine **37**, 542–558 (2007)
18. Quian Quiroga, R., Garcia, H.: Single-trial event-related potentials with wavelet denoising. Clinical Neurophysiology **114**, 376–390 (2003)
19. Rodríguez, J.J., Alonso, C.J., Boström, H.: Boosting interval based literals. Intelligent Data Analysis **5**, 245–262 (2001)
20. Rodríguez, J.J., Alonso, C.J., Maestro, J.A.: Support vector machines of interval-based features for time series classification. Knowledge-Based Systems **18**, 171–178 (2005)
21. Samar, V.J., Swartz, K.P., Raghuveer, M.R.: Multiresolution analysis of event-related potentials by wavelet decomposition. Brain and Cognition **27**, 398–438 (1995)

22. Schaefer, R.S., Farquhar, J., Blokland, Y., Sadakata, M., Desain, P.: Name that tune: Decoding music from the listening brain. Neuroimage **56**(2, Sp. Iss. SI), 843–849 (2011)
23. Shen, K.Q., Ong, C.J., Li, X.P., Hui, Z., Wilder-Smith, E.P.V.: A feature selection method for multilevel mental fatigue EEG classification. IEEE Transactions on Biomedical Ednigneering **54**, 1231–1237 (2007)
24. Snodgrass, J.G., Vanderwart, M.: A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory **6**, 174–215 (1980)
25. Sun, S., Zhang, C., Lu, Y.: The random electrode selection ensemble for EEG signal classification. Pattern Recognition **41**, 1663–1675 (2008)
26. Sun, S., Zhang, C., Zhang, D.: An experimental evaluation of ensemble methods for eeg signal classification. Pattern Recognition Letters **28**(15), 2157 – 2163 (2007)
27. Thierry, G., Martin, C.D., Downing, P., Pegna, A.J.: Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nature Neuroscience **10**, 505–511 (2007)
28. Trejo, L.J., Shensa, M.J.: Linear and neural network models for predicting human signal detection performance from event-related potentials: A comparison of the wavelet transform with other feature extraction methods. In: Society for Computer Simulation, pp. 153–161 (1993)
29. Übeyli, E.D., Gler, I.: Features extracted by eigenvector methods for detecting variability of eeg signals. Pattern Recognition Letters **28**(5), 592 – 603 (2007)
30. Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: In Proceedings of the 20th International Conference on Machine Learning (ICML2003). Washington, DC (2003)
31. Zhang, X.L., Begleiter, H., Porjesz, B., Wang, W., Litke, A.: Event related potentials during object recognition tasks. Brain Research Bulletin **38**, 531–538 (1995)