RANDOM FOREST CLASSIFICATION FOR P300 BASED BRAIN COMPUTER INTERFACE APPLICATIONS

Faisal Farooq and Preben Kidmose
Department of Engineering, Aarhus University, Denmark

ABSTRACT

One of the most successful types of brain computer interfaces (BCI) is based on the P300 evoked potential (EP) elicited by oddball type of paradigms. Given a particular paradigm the main challenge is to obtain an efficient and robust classification. This paper proposes the use of Random Forest (RF), a tree based ensemble learning method providing state-of-the-art generalization performance, for P300 BCI classification. The performance of the proposed method is compared to both the most commonly used classifiers for this problem: the support vector machine (SVM), and the step-wise linear discriminant analysis (SWLDA); and to two state-of-the-art methods: the multiple convolutional neural networks (MCNN) and the ensemble support vector machine (ESVM). The proposed method has been evaluated on two public available BCI datasets: the BCI competition dataset II for healthy subjects and the image driven paradigm dataset for disabled subjects. The proposed method demonstrated a significant improvement in classification accuracy on both datasets.

***Index Terms:* —** Brain Computer Interfaces (BCIs), evoked potential (EP), Random Forest (RF).

1. INTRODUCTION

A brain computer interface (BCI) allows a user to interact with a computer directly through the signals generated by the brain without relying on physical action. This is done by utilizing certain electrophysiological activities that reflect the function of the brain [1]. BCIs can be divided into two classes: endogenous, which are based solely on mental tasks; and exogenous, which are based on evoked potentials (EP). The usage of EP for BCIs is being extensively researched as they require a minimum of subject training and can be based on single/few EEG channels. One of the most explored EP for BCIs is the so-called P300, which is a positive EP that is elicited approximately 300ms after an attended external stimulus. As an example, one of the most popular BCI system based on the P300 is the *P300 speller* proposed by Farwell and Donchin [2] [3], in which the P300 is elicited by an oddball paradigm. The *P300 speller* system detects which character (target) in a matrix of characters the user is attending, display the detection result on the screen, and thereby enabling the user to communicate.

Previous works on P300 BCIs have utilized various binary classifiers for P300 classification. Lenhardt [4] and Bostanov [5] have utilized the linear discriminant analysis (LDA) method to maximize the separation between classes. Based on the LDA, a Bayesian LDA classifier was further proposed in [16] and has been used in the P300 speller paradigm [6]. Krusienski et. al. proposed the SWLDA and demonstrated superior performance compared to the previous mentioned methods [20]. Also SVM have shown satisfactory results in P300 BCI systems using channel selection method [7] [8], compared to classical neural networks classifiers [9]. From the literature it appears that the SVM and the SWLDA have been the most popular methods for the classification of P300 BCI.

In recent years ensemble methods have shown favorable performance over more classical methods in many practical machine learning problems. For instance recently the ensemble methods i.e. ESVM [10] and SWLDA [11] have shown to be better than their classical versions in [12] and [13] for P300 BCI. One of the most successful ensemble methods is the Random Forest (RF) [14]; and it has demonstrated generalization errors that compare favorable to the best statistical and machine learning methods, and has been used for many different applications e.g. [21] [22] and [23]. So far it has never been used for the P300-based BCI problem. Therefore, with the objective to improve the classification performance, we in this work propose the use of the RF classifier for P300-based BCIs. In addition to the excellent classification performance the RF classifiers have a relatively low computational complexity in the training and in the classification, which is a desirable characteristic in practical and real-time BCI-systems.

2. SIGNAL PROCESSING SCHEME

The signal processing scheme of the BCI system has a simple feed-forward structure comprising the following 4 steps:

1. **Preprocessing.** The raw EEG signals from all the channels were filtered using a third order Butterworth band pass filter with pass band between 0.1 and 10 Hz.
2. **Windsorizing.** To remove outliers originating from eye-blinks, muscle activity, motion artifacts etc. the

signal amplitude is limited by the 10 and 90% percentile; as mentioned in [16].

3. **Ensemble Averaging.** The feature vector for the classifier is the time-averaging of 1 sec. time-segments over a set of stimuli (this is an estimate of the EP). For the P300 speller for instance, each trial comprises 6 column and 6 row intensifications, and each trial is repeated 15 times. The feature vector is thus the time-aligned averaging over 2xN time segments, where N is the number of epochs used in the averaging.
4. **Classification.** The feature vector is feed to the RF classifier.

For the purpose of a comparative performance assessment the RF classifier has been compared to four other classifiers: the SVM, the SWLDA, the MCCN-1 and the ESVM. The signal processing step 1 through 3 are identical across all classifiers. The RF and the application to the P300 BCI classification problem are explained in the following section.

3. THE RANDOM FOREST CLASSIFIER

In this section, we propose a RF classifier for the classification of P300 signals. From the literature, we have found that there is no classical decision tree approach to increase both classification and generalization accuracy. For this purpose the RF was introduced by Tin Ho [15].

RF classifiers relies on an ensemble of single tree classifiers where each tree is trained on a random set of vectors from the training set, each such set is denoted a bootstrap. In the bootstrap methodology used in this work each bootstrap is generated as a random selection (with replacement) from the complete training set.

Consider a RF classifier of L trees. The lth tree $h_l(\boldsymbol{x}, \varphi_l)$, of the input vector $\boldsymbol{x}$, is based on a random vector, φ_l, generated from the training set. This vector is independent of past vectors $\varphi_1, \varphi_2 \dots \dots, \varphi_{l-1}$, but sampled from the same distribution of input data. In the construction of the tree, at each node *m* variables are selected at random from the random vector φ_l. The RF method here uses the Gini criterion to select the best splitting vector to split the node into child nodes. The Gini criterion is defined as follows:

$$Gini = \sum_{0,1} p_i(1-p_i) \qquad (1)$$

where, p_i is the proportions of data samples from the two classes. For an ensemble of tree classifiers, $\{h_1(\boldsymbol{x}), h_2(\boldsymbol{x}) \dots, h_L(\boldsymbol{x})\}$, the margin function is given as:

$$mg(\boldsymbol{X}, Y) = E\{I(h_l(X) = Y)\} - \max_{j \neq Y}(E\{I(h_l(X) = J)\}) \qquad (2)$$

where the expectation is taken over the training set $(\boldsymbol{X}, Y)$, and I is the indicator function. The generalization error using the above mentioned margin function is given as:

$$PE^* = P_{X,Y}(mg(\boldsymbol{X}, Y) < 0) \qquad (3)$$

where $P_{X,Y}$ indicates the probability over the $(\boldsymbol{X}, Y)$-space of input vectors **X** and class labels Y. Breiman [14] shows that the generalization error is asymptotically upper bounded by:

$$PE^* \leq \bar{\rho}\frac{(1-w^2)}{w^2} \qquad (4)$$

where $\bar{\rho}$ is the mean correlation between classifiers and w is the strength of the ensemble. Thus the generalization error depends upon two important factors: - the correlation between the random trees and the strength of the individual classifiers present in the forest. Thus, by increasing number of trees the generalization error converges to a limit, and therefor over-fitting is not a problem.

3.1. P300 Classification using Random Forest

A RF model, dedicated for P300 classification, is used to learn the mapping between features of averaged segments of input data and their corresponding classes, i.e. P300 and non-P300. The RF training model is constructed as follows:

- Associate each P300/non-P300 averaged feature vector $\boldsymbol{x}_i$ to their corresponding class label $y_i \in [0,1]$.
- Join all the feature vectors together to form the data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]$ and the corresponding class label vector $Y = [y_1, y_2, \cdots, y_N]$.
- Construct L independent bootstraps, $(\boldsymbol{X}_1, Y_1)$, $(\boldsymbol{X}_2, Y_2)$, $\cdots, (\boldsymbol{X}_L, Y_L)$, by random selection (with replacement) of N samples from the data matrix and the corresponding label vector.
- Construct the set of 'l' random trees $\{h_1, h_2, \dots., h_l\}$. Each tree h_l is trained based on the bootstrap sample $(\boldsymbol{X}_l, Y_l)$.
- At each node, m variables are selected randomly out of 'M' input variables i.e. m << M. Then GINI criterion was employed to select the best split among these m variables.

The classification of RF is based on a weighting (majority voting) of the L trees. This is illustrated in Fig.1.

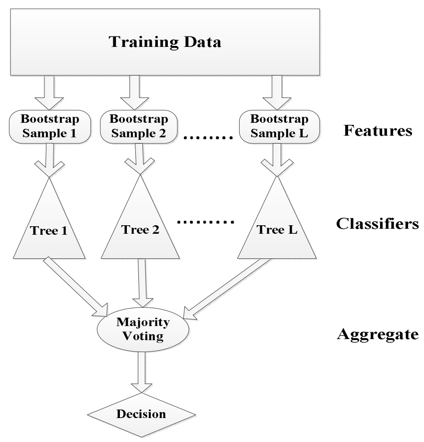

Figure.1: Architecture of the random forest classifier

4. DATASETS

We have tested the proposed technique on two public available BCI datasets: the BCI competition 2003 II of normal subjects [16] and a dataset based on an image driven paradigm recorded from disabled subjects provided by the EPFL BCI group [17].

4.1. Dataset I

As the first dataset we used the BCI competition 2003 dataset II provided by the Wadsworth center for our experiment. This competition has two datasets for Subject A and Subject B. In these datasets the users were presented with a 6x6 matrix of characters on the screen. The recordings consist of 64 EEG channels and the user's task was to attend one character prescribed by the investigator. For each character, the stimulus was as follows: the matrix was displayed for a 2.5 s period and during this time each character had the same intensity. Subsequently each row and column in the matrix was randomly intensified for 100ms. After intensification of a row or column, the matrix remains blank for 75ms. Row or column intensifications were block randomized in blocks of 12. For each character 12 intensifications were repeated 15 times. Each sequence of 15 sets of intensifications was followed by a 2.5 s period during which the matrix was blank. The training database contains 85 characters, while the test database contains 100 characters. For the training, the database consists of $85\times2\times15= 2550$ samples of targets and $85\times10\times15= 12750$ samples of non-targets; and for testing the database consists of $100\times 2 \times15= 3000$ samples of targets and $100\times 10 \times15 = 15000$ samples of non-targets.

4.2. Dataset II

This dataset was recorded from four disabled subjects. The subjects looked at computer screen images on which one out of six were shown; the images showed a telephone, a television, a lamp, a door, a window and a radio. These images were flashed in a random order and the subjects were asked to attend a target image and count the number of times the target was flashed. As a target image is a rare event (with probability one out of six) it induces a P300 response. Each image was flashed for 100 ms and there was a pause for 300 ms between the images. The images were flashed in blocks of six images, where each block consisted of random order of six images. The dataset contains data recorded from four sessions from each subject, and the EEG was recorded from 32 channels. Each session comprises six runs and in each run there were 23 iterations. The total number of targets is $6\times 1 \times 23= 138$, and the total number of non-targets is $6\times 5 \times 23= 690$.

For each dataset, balanced training data i.e. equal no. of targets and randomly selected non-targets were used for each character and image.

5. PERFORMANCE MEASURE

To evaluate the performance and compare the different methods, we have estimated the classification accuracy. The feature vector for the classifier is the estimated evoked response, obtained by averaging the EEG time-synchronously to the stimuli. In principle, as the number of averages is increased, the mean response converge towards the underlying evoked response, and therefore the classification accuracy will increase. Thus, the increase in classification accuracy comes at the expense of lower bandwidth of the BCI. To take this aspect in to consideration we have used the information transfer rate (ITR) as the performance metric as also used in e.g. [18] and [19]. The ITR is defined as:

$$ITR = \frac{60}{T}\left(\log_2(n)+p\log_2 p+(1-p)\log_2(\frac{1-p}{n-1}) \right) \qquad (5)$$

where 'T' is the time in seconds required to spell one character, 'n' is the number of different symbols a user can select and 'p' is the probability of correctness (i.e. classification accuracy). The ITR is measured in bits/minutes.

6. RESULTS AND DISCUSSION

For both dataset I and II the RF classifier was trained on the individual subjects. The RF classifier contained $l = 10$ trees, and the splitting function in each node of the trees was based on $m = 15$ variables. The classification accuracy of the RF classifier for 5, 10 and 15 epoch-averages and for all the subjects in dataset I and II are shown in Table 1 and Table 2 respectively. It is observed that a very similar performance is obtained across all the subjects, and that there is a consistent increase in the performance as the number of epoch-averages increases.

To assess the classification performance of the RF and to compare it against ESVM, MCCN-1, SVM, and the SWLDA classifiers we have estimated the classification accuracy as a function of the number of epoch-averages (from 1 to 15 averages). The results for dataset I are shown in Fig. 2. For all classifiers there was, as expected, a monotonically increase in classification accuracy performance with increasing number of epoch- averages. From Fig. 2 it is observed that the RF classifier has superior performance compared to the competing classifiers in most of the studied examples.

To compare the performance in a more rigorous statistical analysis, the standard deviation for the RF classifiers was estimated for 5, 10, and 15 epoch averages. This was obtained by pooling the training and test data, and randomly selecting new training and testing datasets. The standard deviation is shown in Fig. 2 as error bars around the mean curve for the RF classifier. As a measure of significance we have calculated how much of the probability mass of the RF classifier that is above the mean value of the competing classifiers (assuming the classification accuracy follows a Gaussian distribution). The results are summarized in Table 3 for 5, 10 and 15 epoch averages. Based on this analysis we conclude that the RF classifier has a performance that is significant better than both the ESVM and the MCNN-1.

The trade-off between classification accuracy and channel capacity of the BCI is assessed using the ITR measure. The result of this analysis is shown in Fig. 3. It is observed that the RF has a superior performance over all competing classifiers from 1 to 10 epoch averages. Above 10 the performance of the RF is very similar to the ESVM and MCNN classifiers.

Table 1: Classification performance in % of correctly recognized characters for two normal subject's w.r.t increasing no. of epochs. The values in the brackets show standard deviations.

Dataset I	**No .of epochs**		
	5	**10**	**15**
Subject A	73.80 (3.2)	89.0 (1.3)	97.30 (1.6)
Subject B	80.32 (1.2)	92.3 (1.5)	98.41 (0.8)
Average	77.06 (2.2)	90.65 (1.4)	97.85 (1.2)

Table 2: Classification performance in % of correctly recognized images for the four disable subjects w.r.t increasing number of sequences.

Dataset II	**No .of epochs**		
	5	**10**	**15**
Subject 1	74.0	85.34	94.13
Subject 2	78.32	89.56	97.72
Subject 3	69.51	80.23	95.28
Subject 4	75.23	88.12	98.43
Average	74.26	85.80	96.39

Table 3: Probability mass values of RF w.r.t state-of-art classifiers for dataset I on different number of epochs.

Methods	**No .of epochs**		
	5	**10**	**15**
RF vs. ESVM	0.947	0.995	0.87
RF vs. MCNN-1	0.9998	0.995	0.975

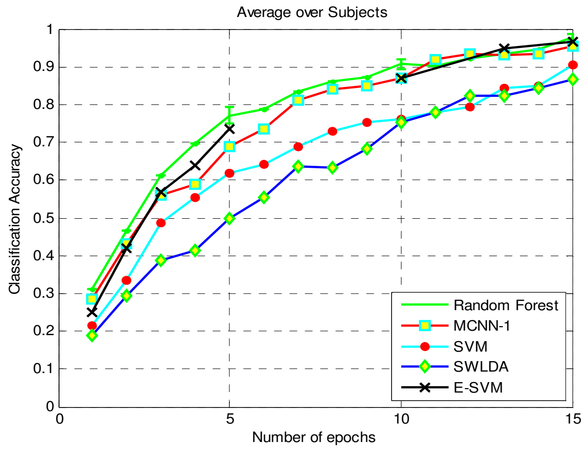

Figure 2: Classifiers performance verses the number of epochs for dataset I.

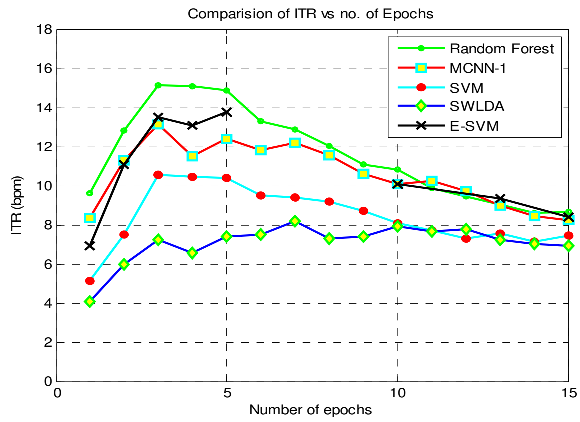

Figure 3: ITR for ESVM, MCCN-1, SWLDA, SVM, and RF number of epochs.

7. CONCLUSION

This paper has proposed a signal processing scheme, utilizing the random forest (RF) classifier for P300-based brain computer interfaces (BCI). The performance of the methodology has been evaluated over two types of P300-based BCIs using public available data sets. By means of a comparative study, the classification accuracies and the information transfer rates [bits/minutes] over number of epochs was evaluated. The study demonstrated that for these datasets the RF classifier had significantly better performance than state-of-the-art methods for most of the studied examples (i.e. 1 to 10 epochs). Furthermore we found that the overall classification accuracy of the RF classifier at 15 epoch-averages outperformed the state-of-the-art methods based on ESVM and MCNN. In addition to the excellent classification performance the RF classifier also have a relatively low computational complexity in the training and in the classification, and this is a desirable characteristic for practical and real-time BCI-systems.

8. REFERENCES

[1] J. R .Wolplaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control" in Clin.Neurophysiol., vol. 113, pp. 767–791,2002.

[2] L.A. Farwell, and E. Donchin, "Taking off the top of your head: Toward a mental prosthesis utilizing event related brain potentials" in Electroenceph. Clin. Neurophysiology. vol. 70, pp. 510-523, 1988.

[3] E. Donchin, K. M. Spencer, and R.Wijesinghe, "The mental prosthesis: Assessing the speed of a P300-based brain-computer interface," in IEEE Trans. Rehabilitation Engineering, vol. 8, no. 2, pp. 174-179, 2000.

[4] A. Lenhardt, M. Kaper, and H. J. Ritter, "An adaptive P300-based online brain-computer interface," in IEEE Trans. Neural Systems and Rehabilitation Engineering, vol. 16, no. 2, pp. 121-130, 2008.

[5] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," in IEEE Trans. Biomedical Eng, IEEE, vol.51, no.6, pp.1057-1061, June 2004.

[6] B.Rivet, A. Souloumiac, V. Attina and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain-computer interface," in IEEE Trans. Biomedical Engineering, vol. 56, no. 8, pp. 2035-2043, 2009.

[7] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner and H. Ritter,"BCI competition 2003-Dataset IIb: Support vector machines for the P300 speller paradigm," IEEE Trans. Biomedical Engineering, vol. 51, no. 6, pp. 1073-1076, 2004.

[8] V.Vapnik, Statistical learning theory, Springer, Berlin Heidelberg New York, 1998.

[9] C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.

[10] A. Rakotomamonjy and V Guigue, "BCI competition III: dataset II- ensemble of SVMs for BCI P300 speller," IEEE Trans. Biomedical Engineering, vol. 55, no. 3, pp. 1147-1154, 2008.

[11] G.D. Johnson, D.J. Krusienski; "Ensemble SWLDA Classifiers for the P300 Speller," Human-Computer Interaction. Novel Interaction Methods and Techniques, volume 5611, pp. 551-557, 2009.

[12] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw,"Toward Enhanced P300 Speller Performance", Neurosci Methods. January 15; 167(1): 15–21, 2008.

[13] M. Thulasidas, C. Guan and 1. Wu, "Robust Classification of EEG signal for brain-computer interface," IEEE Trans. Neural Systems and Rehabilitation Engineering, vol. 14, no. 1, pp. 24-29, 2006.

[14] L. Breiman., "Random forests", Mach. Learn., vol. 45, pp. 5–32, 2001.

[15] Tin Kam Ho, "Random decision forests," in ICDAR, pp. 278–282, 1995.

[16] B. Blankertz, BCI Competition III Webpage [online]. http://ida.first.fraunhofer.de/projects/bci/

[17] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," J. Neuroscience, Jan 15; 167(1):115-25, 2008.

[18] O.I.Khan, F.Farooq, F.Akram, M.T.Choi, S.M.Han, T.S.Kim, "Robust extraction of P300 using constrained ICA for BCI appplications", Medical & Biological Engineering & Computing., Vol. 50, No. 3, pp 231-24, 2012.

[19] H.Cecotti, A.Graser , "Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces," Pattern Analysis and Machine Intelligence, IEEE Transactions *on* , vol.33, no.3, pp.433-445, March, 2011.

[20] D.J.Krusienski, E.W.Seller, F.Cabestaing, S.bayoudh, D.J.Mcfarland, T.M.vaughan, J.R.Wolpaw,"A comparison of classification techniques for the P300 speller" Journal of Neural Engineering, vol. 3, no. 4, pp. 299, 2006.

[21] Cutler, D.R., T.C. Edwards, K.H. Beard, A. Cutler, K.T Hess, J.C. Gibson and J.J. Lawler. 2007. Random forests for classification in ecology. Ecology 88(11):2783-2792

[22] B.C. Ko, S.H. Kim, and J-Y Nam, "X-ray Image Classification Using Random Forests with Local Wavelet-Based CS-Local Binary patterns" Journal of Digital Imaging. 2011 Dec; 24(6)1141-1151.

[23] S.Bernard, S.Adam, L. Heutte, "Using Random Forests for Handwritten Digit Recognition", in ICDAR 2007. Ninth International Conference on, vol.2, no., pp.1043, 1047, 23-26 Sept. 2007.