P300 Detection with Brain–Computer Interface Application Using PCA and Ensemble of Weighted SVMs

Article *in* IETE Journal of Research · August 2017

DOI: 10.1080/03772063.2017.1355271

Sourav Kundu
C V Raman Global University, Bhubaneswar

P300 Detection with Brain-Computer Interface Application using PCA and Ensemble of Weighted SVMs

Sourav Kundu, Samit Ari
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela

June 23, 2017

ABSTRACT

Brain-computer interface (BCI) P300 speller can be used as a powerful aid for severely disabled people in their everyday life. The character recognition using P300 speller involves two stages for classification. First stage is to detect the P300 signal and second one is to determine the right character from the detected P300. Features are important for classification, but large feature dimension is a problem for P300 classification as computational complexity increase due to more number of features. In this work principal component analysis (PCA) based ensemble of weighted support vector machine (PCA-EWSVM) is used for character recognition. The proposed method includes PCA for feature extraction and an ensemble of weighted SVM (EWSVM) for classification. PCA is used to reduce the redundant features and ensemble of weighted classifier for minimizing the classifier variability. The proposed algorithm has been evaluated on data set of the BCI Competition II and data set II of the BCI Competition III.

Keywords:
Brain-computer interface (BCI), principal component analysis (PCA), ensemble of weighted support vector machine (EWSVM), electroencephalogram (EEG), P300.

1. INTRODUCTION

Brain-computer interface (BCI) might be the only medium of communication for individuals [1] who are not able to communicate through ordinary means because of severe motor disabilities like spinal cord injuries or amyotrophic lateral sclerosis (ALS) [2]. There are many alternative ways of communication for disabled people like voice or gesture based systems. However, these systems are not suitable for those individuals who suffer neuromuscular impairments. They are incapable of any muscular movement but have some cognitive abilities. The BCI system analyzes electroencephalogram (EEG) signal and sends the command to the outside world. Several types of EEG signals are used for BCI system like P300, steady-state visually evoked potential (SSVEP), event-related desynchronization/synchronization (ERD/ERS) produced by motor imageries [3], etc. There are several types of BCI systems like brain control wheelchair [4], brain control mobile application [5] or character recognition system [6] etc. The BCI framework for character recognition used in this work is based on P300 which is a typical response of the brain to some predefined stimulus.

A P300 signal appears in EEG data due to the infrequent auditory or visual stimuli. It is named as P300 because a positive peak appears after 300ms of stimuli. When the P300 has been detected, it has occurred for the stimulus that had been applied 300ms before the detection. From the detected P300 signal and flashing row-column information, the character information can be extracted. The row-column intersection gives the character position in the speller board.

Over the last few years, several EEG classification algorithms have been developed. Ensemble support vector machine (ESVM) as a classifier and a recursive channel elimination method for channel reduction are reported in [2]. The recursive channel elimination is a time consuming task. Wavelet based feature with ensemble of fisher's linear discriminant (FLD) classifier is used in [7]. In [8] a multi-resolution approximation based feature selection is applied and linear discriminant analysis (LDA) is used as a classifier. To classify the P300 signal convolutional neural networks (CNN) and temporal features are used in [9]. A semi-supervised classifier based on least squares support vector machine (LS-SVM) is reported in [10]. Binary de-based channel selection and ensemble support vector machines (ESVMs) [11] is used for P300 detection. A novel distance coupled hidden Markov models (HMM) classifier is proposed in [12]. Sparse bayesian classification and spatial-temporal discriminant analysis are introduced in [13–15] as EEG classifiers. In [16] aggregation of sparse linear discriminant analyses (ASLDA) is applied to overcome the curse of dimensionality and bias-variance trade-off. In BCI, the data set is highly imbalanced. To overcome this problem Twin SVM is proposed in [17] as it is insensitive to imbalance class sizes. Most of the above reported techniques use down-sampling or decimation for feature reduction. In down-sampling some important features are removed and as a result the classification accuracy is reduced.

In BCI system feature selection is an important task because all the features are not necessary for classification. In [18] spectral, wavelet and complexity based features are computed for diagnosis of alzheimer's disease. In [19] features like common spatial patterns (CSP), wavelength optimal spatial filter (WOSF) and approximate entropy are used for classification of motor imagery (MI) signal. Canonical correlation analysis (CCA) has been one of the most popular methods for frequency recognition in BCI system [20, 21]. Here a PCA based ensemble of weighted SVM (PCA-EWSVM) is proposed. The redundant features can reduce the classification accuracy and a reduced number of feature set decreases the classification complexity and computational time. Here PCA is used for feature selection. After PCA transformation, 99% variance of the total transform data is taken for classification as it is enough to represent the whole data with less number of features. Also, an ensemble of weighted SVM (EWSVM) classifier is proposed as ensemble of classifier reduces the classifier variability and a weight is assigned to the classifier so that better classifiers get more weight compared to other classifiers. As a result, when the classifier's outcomes are averaged out, best classifier provides more impact on the output for the weight assigned to it. We also used an ensemble of weighted LDA (EWLDA) and combination of SVM and LDA classifiers.

This paper is organized as follows: Section II briefly describes the data set which is provided by the BCI competition and the description of speller paradigm. In Section III, the details about PCA, SVM and LDA are mentioned. The proposed framework is explained in Section IV. Finally, Section V represents the experimental results and comparisons with earlier reported works and in Section VI conclusions of the work are given.

2. The Data Set

The P300 speller is based on the oddball paradigm which states that when a rarely expected stimulus occurs, a positive deflection is observed in EEG signal after about 300ms. The oddball paradigm is shown in Figure 1. The data set is provided by the organizer of BCI Competition II [22] and the BCI competition III [23]. A large number of research articles [2, 9, 10] [24–29] have been published using these benchmark data sets.

2.1 BCI Paradigm

The user interface for speller matrix contains 36 characters in 6×6 matrix. The user has to focus his attention on one character at one time. The rows and columns of the matrix are intensified randomly and successively. This speller matrix's flashing rate

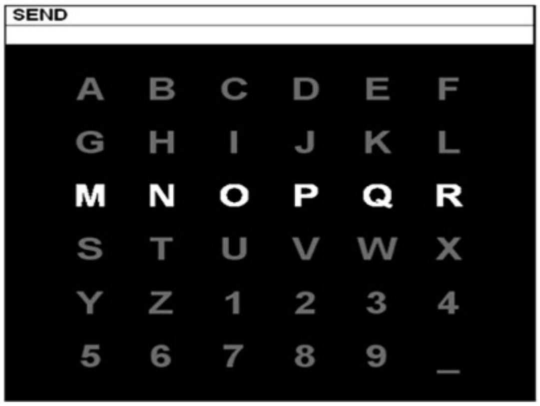

Figure 1: P300 speller Paradigm [22]

is 5.7 Hz when the above mention datasets are collected [22, 23]. Two flashes contained the desired character out of 12 intensifications of rows or columns i.e., one row out of 6 rows and other for out of 6 columns. The responses evoked by these rare stimuli are not the same those evoked by the stimuli that don't contain the desired character. These evoked signals are called P300 signal as previously reported by Farwell and Donchin [30].

2.2 Database Used

In the BCI Competition II [22] data set, the data is collected from a single subject. The data set consists of 11 training words which are composed of 42 symbols and 8 words or 31 test symbols. Out of 42 characters only 39 characters are used for training as the last set of data contained an error in event cue information. Another one is BCI competition III [23] data set II. Here two different subjects have participated in data collection and the database is composed of random 85 training and 100 test characters for each subject. The data is collected in five different sessions. Every session is made out of various runs and for every run, a subject is asked to spell a character. The character matrix is intensified for 100 ms and remains blank for 75 ms. For one round there are 12 flashing and for one character the sets of 12 intensifications are repeated 15 times (i.e., each row/column is intensified 15 times and thus there are $12 * 15 = 180$ total intensifications for a single character). Each repetition is called epoch. So, each character data consist of fifteen epochs. The EEG data is collected continuously from 64-channel. After bandpass-filtering from 0.1 - 60 Hz, the signal is digitized at a sampling rate of 240 Hz.

3. Theoretical Background

3.1 Principal Component Analysis (PCA)

In most of the signal processing application, PCA is used as a data dimension reduction technique. It extracts the most important feature from the high dimensional data set. PCA is an unsupervised linear projection which maximizes the scatter of all sample [31].

Let there be N samples of n dimensional vectors. Now a linear projection (PCA) is projecting n dimensional data to m dimensional data space where $m < n$. The input data set is $X = \{x_1, x_2, ...x_N\}$ where $x_k \in \mathbb{R}^n$. After linear transformation the new feature set $y_k \in \mathbb{R}^m$ is defined as

$$y_k = W_{PCA}^T x_k \qquad k = 1, 2, ..., N \tag{1}$$

where $W_{PCA} \in \mathbb{R}^{n \times m}$ is the matrix with orthonormal basis vector in the columns. These columns are the eigenvectors of m largest eigenvalues corresponding to the scatter matrix S_T, which is determined as

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \tag{2}$$

where μ is the mean features of all data sample.

In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$\begin{aligned} W_{opt} &= \arg \max_{W} \left| W^T S_T W \right| \\ &= [W_1 \, W_2 \, ...W_m] \end{aligned} \tag{3}$$

where $\{w_i |\ i = 1, 2, ...m\}$ is the set of n-dimensional eigenvectors of S_T corresponding to the m large eigenvalues.

3.2 Support Vector Machine (SVM)

SVM is an excellent tool for classification problems with a good generalization performance. Vapnik [32] designed this classifier for binary class problem. Let considered a training data set of N points $(x_i, y_i)_{i=1}^N$, where $x_i \in \mathbb{R}^m$ is ith input pattern and $y_i \in \{-1, 1\}$ is ith output pattern. To construct an optimal hyperplane which maximizes the margin boundary and minimizes the error (ξ). To solve this optimization problem quadratic programming (QP) problem is used.

$$\min_{w,\xi} \left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i \right] \tag{4}$$

where, w is weight vector and C is the regularization parameter. The regularization parameter C plays an important role in classification [33]. Smaller value of C ignores the points near to margin and increases the margin boundary, whereas the larger value of C considered all the points and to do so it is reduced the boundary. The Lagrangian representation of above function is

$$\begin{aligned}
&\max_{\alpha} \left[\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j \, y_i y_j k(x_i, x_j) \right] \\
&w = \sum_{i=1}^{N} y_i \alpha_i \Phi(x_i) \\
&\sum_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \ \forall i
\end{aligned} \tag{5}$$

where α_is are Lagrange multipliers related to each training point, $k(x_i, x)$ represent the kernel function. The constructed SVM decision function is

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b \tag{6}$$

where bias b is a real constant.

3.3 Linear Discriminant Analysis (LDA)

LDA [34] is a simple classifier which provides satisfactory result with low computational cost. In case of two class problem LDA assume that the data is normally distributed [3], so that they are linearly separable. LDA determines a linear discrimination function [35] which corresponds a hyperplane in the feature space in order to differentiate the classes. The hyperplane can be defined as

$$g(x) = w^T x + w_0 \tag{7}$$

where, w is the weight vector, x is input and w_0 is the threshold.
The weight vector w is calculated as follows

$$w = \Sigma_c^{-1}(\mu_2 - \mu_1) \tag{8}$$

where μ_1 and μ_2 are the mean of class one and two respectively and $\Sigma_c = \frac{1}{2}(\Sigma_1 + \Sigma_2)$ is the estimated common covariance matrix; Σ_1 and Σ_2 are the covariance of class one and two respectively.

4. Proposed framework

Three successive stages followed in the EEG based character recognition algorithm are preprocessing, feature extraction with reduction and classification as shown in the flowchart of Figure 2.

4.1 Preprocessing

The preprocessing stage involves the following substages: (i) From each channel a data of duration 0 to 667 ms is extracted after each flashing. As from the previous knowledge about P300 a positive peak will appear after 300 ms of stimulus. Therefore, it is postulated that a 667 ms window i.e. 160 samples are large enough to capture all necessary information for classification [2]. These windows are overlapping windows. (ii) Each extracted signal has been filtered by an 8^{th} order bandpass Chebyshev filter of Type I and cut-off frequency lies within 0.1 and 20 Hz. (iii) This post-stimulus signal means 160 samples from each channel has been transformed into a vector by concatenation of all 64 channels. Thus, BCI Competition II training data is consists of $7020 = 12*15*39$ and for BCI Competition III a single subject, the training set is composed of $15300 = 12*15*85$ post-stimulus vectors x_i of dimension $10240 = 160*64$.

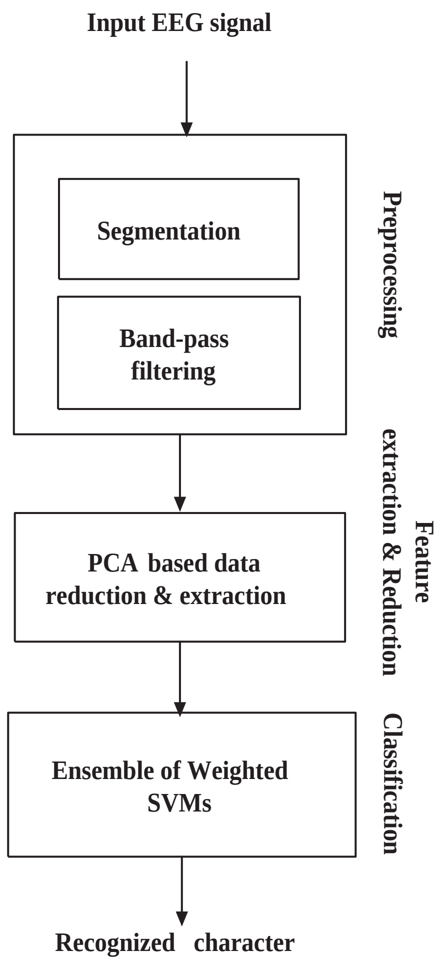

Figure 2: Flowchart of proposed EEG based character recognition algorithm

4.2 Feature Extraction and Reduction

All the samples in preprocessing stage are used as a feature set. To select the best features from extracted feature set PCA is applied in the work. PCA removes the less important features and few important features are kept for classification. The algorithm for input feature selection is as follows:

Step 1: Let $N \times n$ is the dimension of input feature set. The objective is to find a subset of dimension $N \times m\,(m < n)$, that contain important information of the feature set for classification of character recognition.

Step 2: Generate the scatter matrix by removing the features mean from the feature. The scatter matrix S_T is represented as

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \tag{9}$$

Step 3: Find the eigenvector of S_T and project the data onto the eigenvector.

Step 4: Calculate the data variance after projecting the data set.

Step 5: Choose the first few principal components in such a way that 99% of the total data variance contained in a small number of feature sets.

4.3 Model Selection

Here SVM is applied as a classifier. The regularization parameter C plays an important role in classifier performance. To select a proper C for SVM, a model selection procedure has been followed [2]. Training data of BCI Competition III is divided into 17 equal part which contains five characters in each part. Now each classifier is trained on one of the 17 partitions. These 17 partitions are divided into two subsets as mentioned in [2]. In the first subset, there are 8 partitions and in second subset there are 9 partitions. So, for the first subset, it is 8 fold cross-validation and for second part it is 9 fold cross-validation. The validation set for first subset is composed of $7 * 5 * 180 = 6300$ and second subset is composed of $8 * 5 * 180 = 7200$ post-stimulus signals. For BCI Competition II, 10 different partitions are made from 10 training words set. These 10 partitions are divided into two subsets. Each subset consists of 5 words. At the time of C parameter selection, we have used one of it as a training and rest are for testing. Before classification, the training data should be normalized to zero mean and unit variance. According to the normalized parameters obtain from the training dataset, the testing data is also normalized. The margin-error trade-off parameter for each SVM classifier has been selected by running the model selection procedure for different values of C. Then select the C value that maximizes the score C_{cs}.

$$C_{cs} = \frac{t_p}{t_p + f_p + f_n} \tag{10}$$

where f_p, f_n, t_p are the number of false positive, false negative and true positive respectively for the validation set. Here true negative value is ignored because the data is unbalanced and the target is to detect the positive responses which are fewer compared to negative response. In this case, different values of C are $[0.01, 0.05, 0.1, 0.5, 1.0]$.

4.4 Ensemble of Weighted Support Vector Machine (EWSVM)

ESVM is based on the averaging classifier output as its reduced the classifier variability [36].

The flow chart of the EWSVM is shown in Figure 3. Now if there are K number of classifiers and numbers

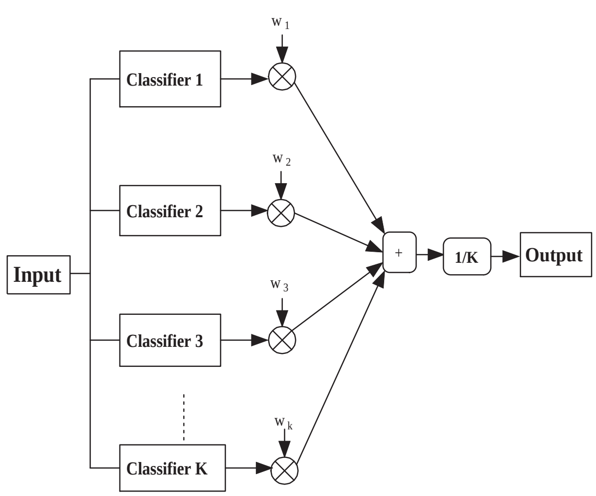

Figure 3: Flow chart of proposed EWSVM algorithm.

of sequences are J, then the ESVM decision function is written as

$$f_{avg}(x) = \frac{1}{K}\frac{1}{J}\sum_{k=1}^{K}\sum_{j=1}^{J} f_k(x) \tag{11}$$

Here all the classifiers provide same weight to determine the output but efficiency of all the classifiers are not same. For this reason EWSVM is proposed here. The algorithm for EWSVM is as follows:

Step 1: The score (C_{cs}) is calculated for each classifier according to section 4.3.

Step 2: The score is normalized and assigned weightage to the each SVM classifiers as follows:

$$\begin{aligned} W_k &= \frac{C_{cs,k}}{\sum_{k=1}^{K} C_{cs,k}} \\ where \quad & \sum_{k=1}^{K} W_k = 1 \end{aligned} \tag{12}$$

Step 3: The EWSVM can be written as:

$$\begin{aligned} f_{wavg}(x) &= \tfrac{1}{K}\tfrac{1}{J}\sum_{k=1}^{K}\sum_{j=1}^{J} W_k f_k(x) \\ &= \tfrac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{N} W_k \alpha_i y_i k(x_i, \tfrac{1}{J}\sum_{j=1}^{J} x) + b \end{aligned} \tag{13}$$

Here W_k represents the given weightage to the different classifiers. Weight is assigned more to the better classifier and less for other classifiers according to their performance at the time of cross-validation. In classifier equation the first averaging is applied in the data space: as sequences increase, a signal for each row and column are averaged out. The second averaging is done on different classifiers score. This last process headed towards a more robust classification arrangement. If a classifier gives an awful score for a test data, then it can be rectified by other classifiers.

5. Results and Discussion

This section presents the results of the proposed method for different datasets. PCA is applied on the whole training data set. Then EWSVM, EWLDA and combination of single SVM and single LDA are used as a classifier. In case of SVM linear kernel is used here. After PCA is applied on the data set the feature dimension reduced to 680 and 619 for subject A and B respectively for BCI Competition III data set and 988 for BCI Competition II data set from

Table 1: Number of correctly classified symbols for BCI competition III data set

		Epochs														
Method	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
EWSVM	A	17	25	52	54	65	69	74	78	81	82	88	94	96	97	99
	B	39	62	67	75	79	82	86	91	91	93	92	91	92	94	97
	Mean	28	43.5	59.5	64.5	72	75.5	80	84.5	86	87.5	90	92.5	94	95.5	98
EWLDA	A	16	23	30	42	47	55	61	62	67	76	76	80	84	82	86
	B	29	45	51	59	63	63	70	70	81	85	85	86	87	89	90
	Mean	22.5	34	40.5	50.5	55	59	65.5	66	74	80.5	80.5	83	85.5	85.5	88
SVM and LDA	A	16	40	53	56	65	70	76	74	79	87	87	95	97	96	95
	B	43	62	68	76	81	82	88	90	91	92	91	95	94	94	97
	Mean	29.5	51	60.5	66	73	76	82	82	85	89.5	89	95	95.5	95	96

10240 if 99% of the total variance of data is taken. In case of training cross-validation is used to calculate the weight and select the regularization parameter. For testing no cross-validation is performed. In BCI Competition III data set the training data set is divided into seventeen equal parts as the training data comprises of 85 characters. So it can only be divided into five equal parts consist of seventeen characters or vice versa. The second method is preferred for both EWSVM and EWLDA classifiers as it gives more number of classifiers and combining five characters give $900 = 5 * 180$ sample points for each of the training sets, which is more than the features size 680 and 619. The performance of character recognition accuracy is shown in Table 1 for BCI Competition III data set. For different epochs we have calculated the accuracy for each subject and also the average accuracy is calculated for different methods. From the Table 1, it is observed that the performance of EWSVM is better compare to EWLDA. LDA offered least overall performance, particularly when more training data was provided, whereas SVMs are commonly regarded as well suited for high-dimensional data, as their generalization error bounds do not explicitly depend on input dimensionality [37]. In case of BCI Competition II data set the training data is divided into 10 parts. Each part is consist of one word for EWSVM. The words predicted after $1^{st}, 2^{nd}, 3^{rd}, 4^{th}$ and 5^{th} epochs and the actual words of the test data are shown in Table 2. For EWLDA, the training data set is equally divided into three part as there are 39 characters for training. These 39 characters can be divided into 13 equal parts consist of 3 characters or vice versa. We choose the second one because after PCA the feature dimension is 988. For better result in LDA classifier, the feature dimension should be less than the number of samples. In Table 3 a comparison of the proposed method with earlier reported techniques for BCI Competition II data set is shown in term of number of correctly classified symbols. When LDA and SVM are combined together, SVM's regularization parameter (C) value is 0.01. From the Table 1, it is observed that as the number of epochs or number of sequences increases the percentage of character recognition accuracy increases. After 11^{th} epoch the increase rate is slow.

Table 2: The words predicted (with 1, 2, 3, 4 and 5 trials) and actual words for all the runs of session 12 for BCI Competition II data set

Run Number (Session 12)	Word Predicted After					Actual Word
	1 epochs	2 epochs	3 epochs	4 epochs	5 epochs	
1	FOO2	FOOD	FOOD	FOOD	FOOD	FOOD
2	MOON	MPOT	MOOT	MOOT	MOOT	MOOT
3	BAM	HAM	HAM	HAM	HAM	HAM
4	JIE	PIE	PIE	PIE	PIE	PIE
5	CAHE	CAIE	CAKE	CAKE	CAKE	CAKE
6	TUNA	TUNA	TUNA	TUNA	TUNA	TUNA
7	ZYSOT	ZMGON	ZSGON	ZYGOT	ZYGOT	ZYGOT
8	4567	4567	4567	4567	4567	4567

From Table 3, it is observed that after 4^{th} epoch we achieved 100% accuracy which is better than [27] and equal to [7] for BCI Competition II data set. After one epoch we correctly classify 25 characters which is almost 81% accurate. In Table 2 result are shown upto 5^{th} epoch because after 5^{th} epoch all are achieving 100% accuracy. The aim of the BCI competition is to report the classification result using all fifteen flash sequences (epochs) and additionally, only the

Table 3: Comparison of the proposed technique with earlier reported techniques for BCI Competition II data set in terms of number of correctly classified symbols

	Epochs					
	1	2	3	4	5	6
WT-EFLD [7]	17	25	28	31	31	31
Kaper *et al.* [27]	20	22	26	30	31	31
Chaurasiya *et al.* [24]	-	17	25	-	31	31
PCA-EWSVM	25	27	29	31	31	31
PCA-EWLDA	24	25	27	29	29	30
PCA-SVM and LDA	27	29	29	30	31	31

first five flash sequences. The number of characters correctly classified and the classification accuracy is same for BCI Competition III data set as the number of test characters are 100 for each subject. A comparison between proposed method and other earlier reported methods is shown in Table 4 using the first 5 flash sequences and all 15 sequences for BCI Competition III data set. The BCI III results have been received from the BCI competition website [6]. From the result, it is observed that the performance of proposed method is better compared to earlier reported techniques at 15^{th} epoch and at 5^{th} epoch the performance is comparable. The results referred in [2] is 73.5% after 5^{th} epoch and 96.5% after 15^{th} epoch whereas proposed method achieves 72.0% and 98.0% respectively.

Table 4: Performance Comparison of the proposed techniques with earlier reported techniques for BCI Competition III data set

Method	Epoch	
	5	15
Yandong [6]	55.0	90.5
LDA [38]	60.5	92.0
WT-EFLD [7]	71.5	95.0
ESVM [2]	73.5	96.5
MCNN-1 [9]	69.0	95.5
PCA-EWSVM	72.0	98.0

In Table 5, a comparison of feature dimensions is shown between proposed and other methods from literature for BCI Competition III data set. The proposed method chooses only 680 and 619 features for subject A and B respectively whereas other methods consider more features for classification.

Table 5: Comparison of the feature dimensions with other methods from literature for BCI Competition III data set

Method	Feature dimension	
	Sub A	Sub B
ESVM [2]	896	896
MCNN-1 [9]	4992	4992
WT-EFLD [7]	1280	1280
PCA-EWSVM	680	619

The performance of the proposed method is better compared to other techniques because only the significant features from the whole data set is considered, whereas other are down-sampled the data. At the time of down-sampling many significant samples are removed, as a result performance decreases. In classification more weight is assigned to the better classifier and less to the others. As a result, the impact of the good classifier is more on the output results. If the noise level is high, then proposed algorithm does not perform well.

6. CONCLUSION

In this paper, a PCA based feature selection and an efficient P300 classifier based on EWSVM is proposed to improve the character recognition performance in P300 speller. Feature extraction, feature selection and classification are the important steps in character recognition. PCA is used to reduce the redundant and irrelevant feature effect on the classification accuracy. The regularization parameter C of SVM is selected based on the model selection procedure. In training phase, according to the value of C the weight vector changes itself to minimize the error. The performance of the above algorithm is evaluated on BCI competition II and BCI competition III data set, which are benchmark data available online. For BCI competition III data set, the proposed algorithm outperforms the best result when the number of epochs is 15 and nearly equivalent when first 5 epochs are used. For BCI competition II data set, 81% accuracy is achieved after 1^{st} epoch, which is a significant improvement. The training time would be reduced by more efficient feature selection algorithm. Moreover for machine leaning and neuroscience community P300 detection is a challenging task.

Acknowledgment

The authors would like to thank A. Rakotomamonjy for his help towards the ESVM method. This work is supported by the Young Faculty Research Fellowship under Visvesvaraya Ph D scheme for Electronics & IT, DeitY, India.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

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

[3] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[4] X. Fan, L. Bi, T. Teng, H. Ding, and Y. Liu, "A brain–computer interface-based vehicle destination selection system using p300 and ssvep signals," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 274–283, 2015.

[5] S. Soman, S. Srivastava, S. Srivastava, and N. Rajput, "Brain Computer Interfaces for Mobile Apps: State-of-the Art and Future Directions," in *Proceedings of the 9th International Conference on Interfaces and Human Computer Interaction (IHCI 2015)*. Las Palmas de Gran Canaria, Spain, 2015.

[6] B. Blankertz, "BCI Competition III-Final Results," 2008. [Online]. Available: http://www.bbci.de/competition/iii/results/

[7] M. Salvaris and F. Sepulveda, "Wavelets and ensemble of FLDs for P300 classification," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2009, pp. 339–342.

[8] U. Celik and S. Anca, "Classification of the visual evoked eeg using multiresolution approximation based on excitatory post-synaptic potential waveform," in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of*. IEEE, 2010, pp. 929–933.

[9] H. Cecotti and A. Graser, "Convolutional neural networks for p300 detection with application to brain-computer interfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, 2011.

[10] Z. Gu, Z. Yu, Z. Shen, and Y. Li, "An online semi-supervised brain–computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2614–2623, 2013.

[11] R. K. Chaurasiya, N. D. Londhe, and S. Ghosh, "Binary de-based channel selection and weighted ensemble of svm classification for novel brain–computer interface using devanagari script-based p300 speller paradigm," *International Journal of Human–Computer Interaction*, pp. 1–17, 2016.

[12] S. Zhong and J. Ghosh, "Hmms and coupled hmms for multi-channel eeg classification," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, 2002, pp. 1254–1159.

[13] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse bayesian classification of eeg for brain–computer interface," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2256–2267, 2016.

[14] Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse bayesian learning for obtaining sparsity

of eeg frequency bands based feature vectors in motor imagery classification," *International Journal of Neural Systems*, p. 1650032, 2016.

[15] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for erp-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 233–243, 2013.

[16] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Aggregation of sparse linear discriminant analyses for event-related potential classification in brain-computer interface," *International journal of neural systems*, vol. 24, no. 01, p. 1450003, 2014.

[17] S. Soman and Jayadeva, "High performance eeg signal classification using classifiability and the twin svm," *Applied Soft Computing*, vol. 30, pp. 305–318, 2015.

[18] N. Kulkarni and V. K. Bairagi, "Extracting salient features for eeg-based diagnosis of alzheimer's disease using support vector machine classifier," *IETE Journal of Research*, vol. 63, no. 1, pp. 11–22, 2017.

[19] S. S. Gupta, S. Soman, P. G. Raj, and R. Prakash, "Improved classification of motor imagery datasets for BCI by using approximate entropy and WOSF features," in *Signal Processing and Integrated Networks (SPIN), 2014 International Conference on.* IEEE, 2014, pp. 90–94.

[20] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Frequency recognition in ssvep-based bci using multiset canonical correlation analysis," *International journal of neural systems*, vol. 24, no. 04, p. 1450013, 2014.

[21] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A speedy hybrid bci spelling approach combining p300 and ssvep," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 473–483, 2014.

[22] B. Blankertz, K.-R. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger *et al.*, "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 1044–1051, 2004.

[23] B. Blankertz, K.-R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[24] R. K. Chaurasiya, N. D. Londhe, and S. Ghosh, "An efficient p300 speller system for brain-computer interface," in *Signal Processing, Computing and Control (ISPCC), 2015 International Conference on.* IEEE, 2015, pp. 57–62.

[25] J. Long, Z. Gu, Y. Li, T. Yu, F. Li, and M. Fu, "Semi-supervised joint spatio-temporal feature selection for p300-based bci speller," *Cognitive neurodynamics*, vol. 5, no. 4, pp. 387–398, 2011.

[26] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI competition 2003-data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 1067–1072, 2004.

[27] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073–1076, 2004.

[28] 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," *IEEE Transactions on Biomedical engineering*, vol. 51, no. 6, pp. 1057–1061, 2004.

[29] P. J. Kindermans, H. Verschore, and B. Schrauwen, "A unified probabilistic approach to improve spelling in an event-related potential-based brain–computer interface,"

IEEE Transactions on Biomedical Engineering, vol. 60, no. 10, pp. 2696–2705, 2013.

[30] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[31] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[32] V. Vapnik, *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[33] B. Zhang, "Is the maximal margin hyperplane special in a feature space," *Hewlett-Packard Research Laboratories Palo Alto*, 2001.

[34] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[35] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of erp componentsâĂŤa tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[36] Y. Grandvalet, "Bagging equalizes influence," *Machine Learning*, vol. 55, no. 3, pp. 251–270, 2004.

[37] J. Lesniak, R. Hupse, R. Blanc, N. Karssemeijer, and G. Székely, "Comparative evaluation of support vector machine classification for computer aided detection of breast masses in mammography," *Physics in medicine and biology*, vol. 57, no. 16, p. 5295, 2012.

[38] N. Liang and L. Bougrain, "Averaging techniques for single-trial analysis of oddball event-related potentials," in *4th International Brain-Computer Interface Workshop*, 2008.

Authors

Sourav Kundu received his B. Tech degree in electronics and communication engineering in the year 2011. He received his M. Tech degree from IIEST, Shibpur in Mechatronics in the year 2014. He is currently pursuing his PhD degree in signal processing at National Institute of Technology, Rourkela, in the Department of Electronics and Communication Engineering.
E-mail: souravkundu90@gmail.com

Samit Ari received B. Tech in Electronics and Tele-communication Engineering degree from University of Kalyani and M. Tech degree in Instrumentation Engineering from Calcutta University. He completed his Ph.D in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur. He joined National Institute of Technology (NIT) Rourkela as a faculty member in 2009, where he presently holds the position of Assistant Professor in the department of Electronics and Communication Engineering.
E-mail: samit.ari@gmail.com