King's Research Portal

DOI:
10.1016/j.jneumeth.2021.109300

Document Version
Peer reviewed version

Link to publication record in King's Research Portal

Citation for published version (APA):
Li, S., Jin, J., Daly, I., Wang, X., Lam, H. K., & Cichocki, A. (2021). Enhancing P300 based character recognition performance using a combination of ensemble classifiers and a fuzzy fusion method. *Journal of Neuroscience Methods*, *362*, [109300]. https://doi.org/10.1016/j.jneumeth.2021.109300

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 06. Oct. 2023

Enhancing P300 based character recognition performance using a combination of ensemble classifiers and a fuzzy fusion method

Shurui Li，Jing Jin, Ian Daly, Xingyu Wang, Hak-Keung Lam, Andrzej Cichocki

IEEE Transactions on Fuzzy Systems

Abstract:

Background: P300-based Brain-computer interfaces (BCIs) provide communication pathways without the need for muscle activity by recognizing electrical signals from the brain. The P300 speller is one of the most commonly used BCI applications, as it is very simple and reliable, and it is capable of reaching satisfactory communication performance. However, as with other BCIs, it remains a challenge to improve the P300 speller's performance to increase its practical usability.

Methods: In this study, we propose a novel multi-feature subset fuzzy fusion (MSFF) framework for the P300 speller to recognize the users' spelling intention. This method includes two parts: 1) feature selection by the Lasso algorithm and feature division; 2) the construction of ensemble LDA classifiers and the fuzzy fusion of those classifiers to recognize user intention.

Results: The proposed framework is evaluated in three public datasets, i.e., BCI Competition II Dataset IIb, BCI Competition III dataset II, and the BNCI Horizon Dataset. The experimental results indicate that it yields better or comparable performance than previously reported machine learning algorithms.

Conclusion: The proposed method is able to improve the performance of P300-based BCIs.

Index Terms: P300 speller; MSFF framework; ensemble classifiers; fuzzy fusion

1. Introduction

A Brain-computer interface (BCI) is a control system that enables users to achieve direct communication with other people or with the external environment by brain activity alone and without needing to make any voluntary muscle movements [1]. It can be used for multiple applications such as moving a cursor on a computer screen, controlling external devices, or spelling out words. There are many potential user groups for BCI systems, including, but not limited to, individuals living with amyotrophic lateral sclerosis (ALS) or individuals who are in a locked-in state (LIS) [2, 3].

Brain-computer interfacing is also a multidisciplinary research topic that involves the fields of neuroscience, cognitive science, neural engineering, biomedical engineering, brain science, and artificial intelligence [4]. The brain activity used to control a BCI can be measured by using different signal acquisition approaches such as electroencephalogram (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), electrocorticogram (ECoG), or near-infrared spectroscopy (NIRS) [5, 6]. Since EEG signals are recorded via non-invasive electrodes placed on the surface of the scalp, EEG-based BCI systems are very commonly used [7]. Four key signal components of the EEG are frequently used for BCI control: slow cortical potential (SCP), event-related potentials (ERPs) [8], steady-state visual evoked potentials (SSVEP) [9], and motor imagery (MI) [10, 11].

The P300 is a type of ERP with a positive deflection, which usually occurs approximately 300 milliseconds after the onset of an unexpected, or out of sequence, stimulus [12]. The classical P300 speller system, called the oddball paradigm, was originally described by Farewell and Donchin in 1988 [13]. In the paradigm, a 6×6 matrix composed of 26 letters (A-Z) and 10 numbers (0-9) was presented on the computer screen and the rows and columns of characters were flashed randomly. During the experiment, subjects were requested to pay attention to the target character and count the number of times that the row and column that contained the target flashed. After signal processing, the predicted character would be presented on the screen. Since the subject does not need to be trained in order to produce a clear P300 ERP, in this study, we use it for character recognition. Due to the low signal-to-noise ratio (SNR) of the P300, multiple recordings of the P300 are typically required in order to enhance the quality of the ERP signal. The character recognition accuracy can be used to evaluate the performance of the P300 speller.

There are typically three steps used to analyze P300 ERPs: data preprocessing, feature extraction, and feature classification [14]. In order to improve the performance, several preprocessing methods may be used. These include frequency filtering [15], down-sampling [16], and channel selection (spatial filtering) [17]. Since the P300 is a time-locked response to an external stimulus, most BCIs make use of temporal information to analyze the signal and achieve good performance. Krusienski et al. [18] chose suitable temporal features for inclusion in the discriminant function and applied them to online P300 BCI processing. Lin et al. [19] designed a temporal matching filter to solve the time-alignment problem of the ERP data. Cecotti et al. [20] presented a method based on a convolutional neural network whose second hidden layer transformed the signal into the time domain. Moreover, other researchers employed different low-frequency characteristics of the P300 response [21, 22] or multiple domains including spatiotemporal features [23, 24], joint-domain time-space-frequency features [25], as the final features sent into the classifier. Kundu and Ari [26] proposed a sparse autoencoder (SAE) and stacked sparse autoencoder (SSAE) based on deep feature learning techniques to describe EEG signals. As a data preprocessing strategy, feature selection was usually introduced to reduce computational complexity.

Once the feature extraction is completed, it is important to select the classification algorithm. In the past few decades, several machine learning techniques have been effectively used with P300 based BCI systems, for instance, Linear Discriminant Analysis (LDA) [27], Bayesian Linear Discriminant Analysis (BLDA) [28], Support Vector Machines (SVMs) [29], and Convolutional Neural Networks (CNNs) [20] have been explored. In pursuit of higher performance, many researchers have also introduced ensemble classifiers. Ensemble classifiers combine the output of two or more individual classifiers in order to achieve higher classification performance. Rakotomamonjy et al. [30] used an ensemble of seventeen linear support vector machines (ESVM). Based on ESVM, Kundu and Ari [26] developed a novel data partitioning method to boost the diversity of ESVM for use with a P300 speller. Cavrini et al. [31] concentrated on the combination of different classifiers using the fuzzy integral.

Improved accuracies can be achieved by properly combining the outcomes from an ensemble of multiple classifiers, for example by weighting the output from individual classifiers. To date, there is a dearth of studies that apply ensemble classifiers within P300 based BCI systems. Additionally, the non-stationarity of the EEG is still a considerable challenge when developing classifiers for BCI systems. Non-stationarity in the EEG generally arises from a diverse range of sources. These include, but are not limited to, different psychological characteristics of the subjects, environmental noise, and trial-to-trial variability in mental activity on the part of the subject. Despite this empirical analysis of the EEG has revealed that signal averaging could boost the SNR of the EEG. Moreover, ensemble classifiers have been shown to lessen the influence of signal variability and improve classification performance. Compared with individual classifiers, appropriately designed ensemble classifiers can enhance prediction accuracy [32, 33]. Broadly speaking, ensemble classifiers are effective if the base classifiers are composed of increasing the diversity of the classification approach.

In this study, we propose a novel multi-feature subset fuzzy fusion (MSFF) framework for the P300 speller to recognize the users' spelling intention. This method extracts temporal features to construct an ensemble of multi-classifiers to improve classification performance. First, the feature selection method is applied to reduce the dimensionality of feature sets. Next, the selected temporal features are divided into two subsets to train based on the same classification algorithm. The classifiers are applied to estimate the offline classification accuracy of the whole training set. Then the offline accuracy is treated as the fuzzy density to calculate the fuzzy integral to contribute to the outcome of the feature vector. Figure 2 shows the framework of our proposed novel joint classification approach with fuzzy integral use in a P300 BCI system. It is composed of two parts: the first part represents the data processing structure, and the second part follows the MSFF framework embedded into the P300 based BCI system. We compare our proposed framework with several traditional classification algorithms including LDA, xDAWN, STDA, FC+LDA, and SWLDA.

The remainder of this work is listed as follows. Section 2 shows the materials and methods, including previous methods and our proposed MSFF framework. Section 3 illustrates the experimental setup and details the BCI competition datasets and applied framework. Section 4 and 5 describe the results and conclusions respectively.

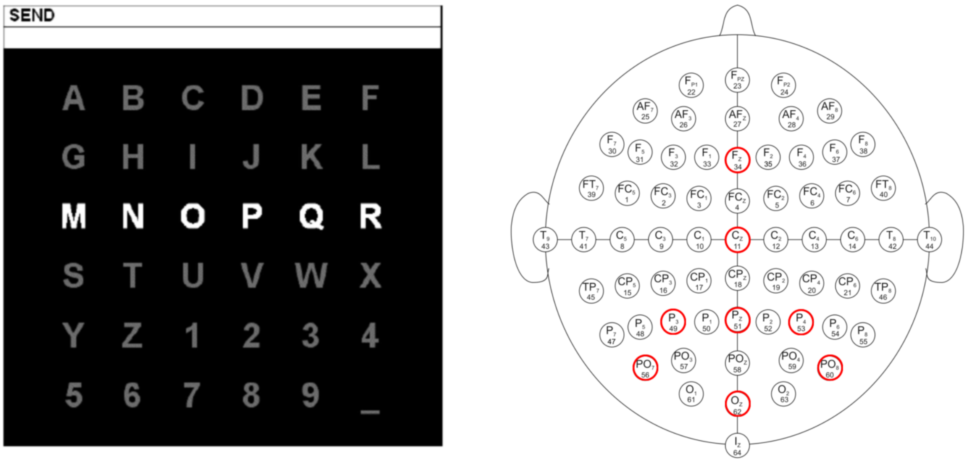

Fig.1 The P300 speller paradigm. (a) The 6×6 BCI paradigm for data collection in three datasets; (b) The electrode positions according to the International 10-20 System.

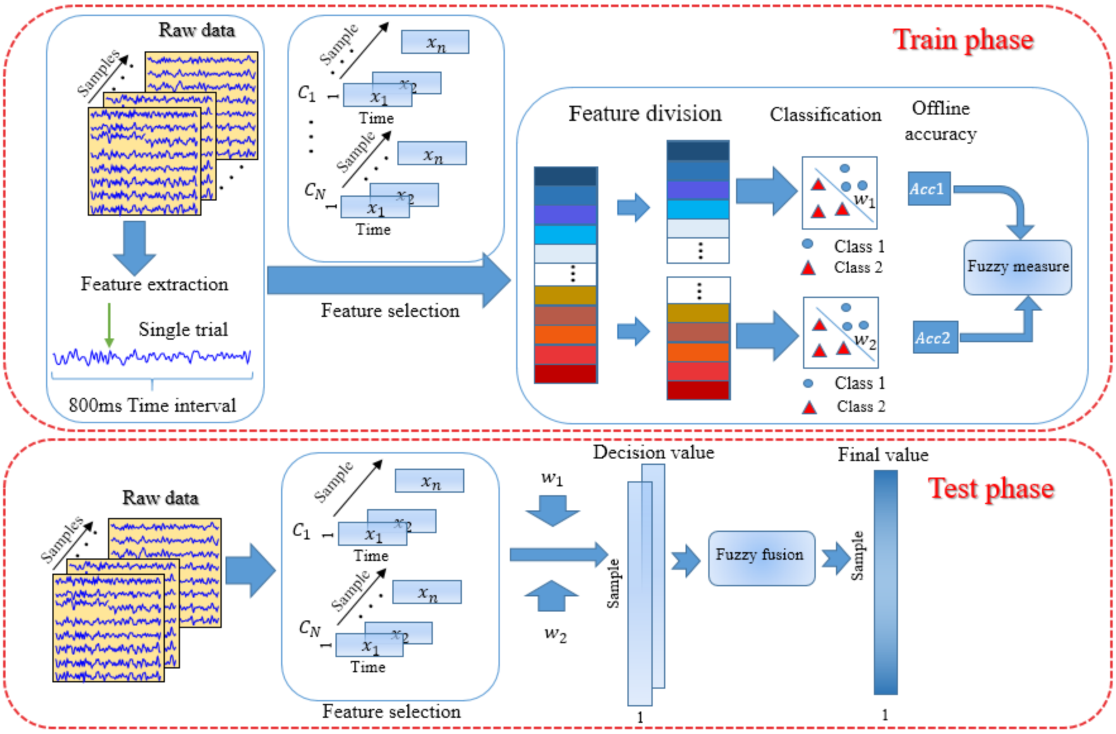

Fig. 2 Framework of our proposed novel multi-feature subset fuzzy fusion (MSFF) classification algorithm for the P300-based BCI systems.

2. Materials and methods

2.1 Background

2.1.1 Selection criteria of the comparison algorithm

This work compares our proposed MSFF framework with five other algorithms that are commonly used in the ERP-based BCI system: LDA, xDAWN, STDA, FC+LDA, and SWLDA. LDA is simple to use and generally provides good results, therefore it is a suitable algorithm to compare against our proposed method. The xDAWN algorithm has been shown to enhance the classification accuracy of P300. It has also been reported that the STDA algorithm is effective at reducing calibration time by using few training samples, which increases the classification accuracy. In addition, Fisher's criterion (FC)-based spatial filtering has been proposed to select features in combination with an LDA classifier (referred to as FC+LDA). Finally, the SWLDA method has been shown to be able to select suitable features automatically for constructing a multiple regression model. Therefore, these five machine learning approaches are compared to our proposed method.

2.1.2 Linear discriminant analysis (LDA)

As a simple and efficient approach, LDA has relatively frequently been employed in BCI systems [34]. The objective of LDA is to separate target and non-target data by calculating the optimal discriminant vector. Let $x_i \in R^D, i \in \{1,..,N\}$ be the input vector of the classifier, and $y_i \in \{-1,1\}$ be the corresponding label vector. The term N_1 denotes the number of samples in class C_1, and N_2 denotes the number of samples in class C_2. We can get the means m_k, k=1, 2 of two classes

$$m_k = \frac{1}{N_k} \sum_{i \in C_k} x_i \tag{1}$$

The between-class scatter matrix $S_b$and the within-class scatter matrix S_w can be defined as

$$S_b = (m_1 - m_2)(m_1 - m_2)^T \tag{2}$$

$$S_w = \sum_{k=1}^{2} \sum_{i \in C_k} (x_i - m_k)(x_i - m_k)^T \tag{3}$$

Thus, the optimal discriminant vector w of LDA can be expressed by

$$w = {S_w}^{-1}(m_1 - m_2) \tag{4}$$

2.1.3 xDAWN

The goal of the xDAWN algorithm is to enhance the P300 evoked potentials to aid classification. First, synchronous responses are estimated for each channel and then these responses are applied to estimate a spatial filter. The synchronous response model of the P300 is as follows.

$$X = DA + N \tag{5}$$

where X is the matrix of the collected EEG signals, D is the Toeplitz matrix whose first column is set to 1, A is the matrix of ERP signals or target stimuli signals, and N represents the noise and artifacts. In this study, we adopted the xDAWN algorithm for spatial filtering and the BLDA algorithm for classifying the features. More details can be seen in [35].

2.1.4 STDA

Instead of adopting traditional vectorized temporal features, spatial-temporal discriminant analysis (STDA) was introduced to learn two projection matrices from the spatial and temporal ERP feature subspaces. The trained projection matrices are then applied to convert the original spatial-temporal samples to new one-way samples. The STDA algorithm implements collaborative discriminant analysis and alternating spatial and temporal dimension optimization, which decreases feature dimensionalities, and therefore enhances the estimation of the covariance matrices in the discriminant analysis even when using a limited number of training samples. This contributes to improving the generalization performance of the model. Since the number of retained features could affect classification performance, we fixed the number of final features to 4 for each trial in our analysis. More details about the STDA method can be found in Zhang's study [36].

2.1.5 FC+LDA

The FC+LDA method combines the Fisher criterion (FC) and the LDA algorithm in order to identify the most relevant feature subset and then perform classification. The FC-based spatial filtering method has been widely used due to the performance of its de-noising and supervised dimensionality reduction stages. Therefore, the algorithm can lessen the effect of small training samples. In this paper, the Fisher criterion was carried out firstly as a pre-selector to remove the least relevant features [37], and then we applied LDA to build a classifier, which can obtain high classification performance.

2.1.6 SWLDA

Stepwise linear discriminant analysis (SWLDA) is considered as an extension of LDA, which incorporates feature selection [17]. SWLDA trains a linear discriminant model based on a combination of forward and backward stepwise regression starting with a set of empty features in a linear regression model. According to the statistical significance of the regression, this algorithm added the most informative input features (with p-values < 0.1) to the subset of selected features, and removed the least informative input features (with p-values > 0.5). This process is repeated until the model contains a predetermined number of variables, or until no additional variables meet the addition or removal criteria. In this paper, we set the number of final features to 60.

2.2 Proposed framework

Here, we propose a novel multi-feature subset fuzzy fusion (MSFF) framework that extracts temporal features to construct an ensemble of multiple classifiers to improve the classification performance. There are two parts: 1) feature selection and division; 2) the construction of ensemble classifiers and the fuzzy fusion of those classifiers.

First, we extract temporal features from all the trials and obtain the feature matrix $X \in R^{N_c \times N_t \times N_s}$ and label matrix $Y \in R^{N_s \times 1}$, where N_c is the number of channels, N_t is the number of features, and N_s is the number of trials. The acquired feature matrix is further improved by the least absolute shrinkage and selection operation (Lasso) [11]. With all the training trials, we partitioned it into two subsets according to the given signal, therefore the number of class labels for each subset is $N_s/2$. We obtain the feature matrix $\overline{X_1} \in R^{N_c \times N_t' \times N_s/2}$ and label matrix $\overline{Y_1} \in R^{N_s/2 \times 1}$, $\overline{X_2} \in R^{N_c \times N_t' \times N_s/2}$ and label matrix $\overline{Y_2} \in R^{N_s/2 \times 1}$, where N_t' is the number of sample points in each trial after feature selection. Then, the two feature subsets are classified by linear discriminant analysis separately. The fuzzy measure and fuzzy integral are adopted to incorporate different classifiers to achieve the final decision. Finally, the fuzzy integral aggregates the outputs of the two classifiers in the testing set to make the final decision.

In this study, the initial fuzzy density is determined using the offline accuracy of the classifiers. The 2nd-order fuzzy measure and the Choquet fuzzy integral are employed [38]. We have developed an effective fuzzy measure and fuzzy integral model fusion method for recognizing the P300 component. This measure is composed of a feature division schema, feature model fusion, and decision value fusion. At the same time, five classification algorithms are employed as base classifiers to compare with our proposed framework. Fig. 3(a) illustrates the proposed multiple classifier framework and Fig. 3(b) illustrates the individual classifier framework.

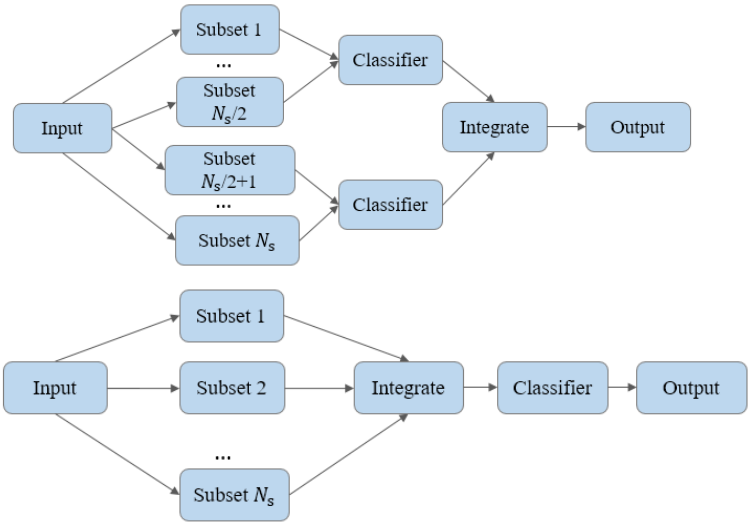

Fig. 3(a) Multiple classifiers framework; (b) Individual classifier framework

3. Experimental Result

3.1 Datasets

In this study, three public datasets involving the BCI Competition II Dataset IIb, BCI Competition III dataset II [39] and the BNCI Horizon Dataset are used to evaluate the performance of our proposed framework. All datasets were recorded with BCI2000 software using the paradigm developed by Donchin et al. [40], which includes the 6×6 matrix of characters shown in Fig. 1(a).

3.1.1 Dataset 1

BCI Competition II dataset IIb consists of one subject's EEG data, recorded during the presentation of 42 training characters and 31 test characters. According to the data analysis described in [41], the data file from experiment run 'AAS011R06' is inconsistent with the description of the dataset, so we removed one word and only retained 39 characters for training. More details about dataset 1 can be found at http://bbci.de/competition/ii/albany_desc/albany_desc_ii.html.

3.1.2 Dataset 2

BCI Competition III dataset II contains two subjects' data, subject A and subject B separately. Each subject's data contains EEG recorded during the presentation of 85 training characters and 100 test characters. In the experiment for Dataset 1 and Dataset 2, all the rows and columns are flashed at the rate of 5.7Hz, and the matrix is intensified in brightness for 100ms every 175ms. A block covers 12 trials in which every row and every column flashes once in random order where two flashes contain the desired character. The sets of 12 intensifications are repeated 15 times for each character and thus there are 180 total intensifications for each character epoch. The collected 64 channel EEG signals are filtered with a bandpass filter with a cut-off frequency of 0.1-60Hz and digitized at 240Hz. More details about dataset 2 can be found at http://www.bbci.de/competition/iii/desc_II.pdf.

3.1.3 Dataset 3

The BNCI Horizon dataset contains the data from 8 subjects (S1-S8) with ALS diagnosis [42]. Eight EEG channels, placed according to 10-10 standard, are used for collecting data. Specifically, EEG channels were placed at positions Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8. The scalp EEG signals were filtered with a bandpass filter of cut-off frequency 0.1-30Hz and digitized at 256Hz. In the experiment, all rows and columns are intensified for 125ms, with an inter stimulus interval (ISI) of 125 ms. For each character, all rows and columns are intensified 10 times. There are 35 predefined characters for each subject to spell. The EEG signals recorded from spellings of the first 20 characters are used for the training data, while the EEG data from the rest of the characters are used as the test data. More details about dataset 3 can be found at https://lampx.tugraz.at/~bci/database/008-2014/description.pdf.

All the above datasets share the same P300 paradigm, shown in Fig. 1(a), and Fig. 1(b), which illustrates the setup with a configuration of 64 electrode positions based on the international 10-20 system in terms of dataset 1 and dataset 2. The eight electrodes used to record dataset 3 are illustrated with red circles.

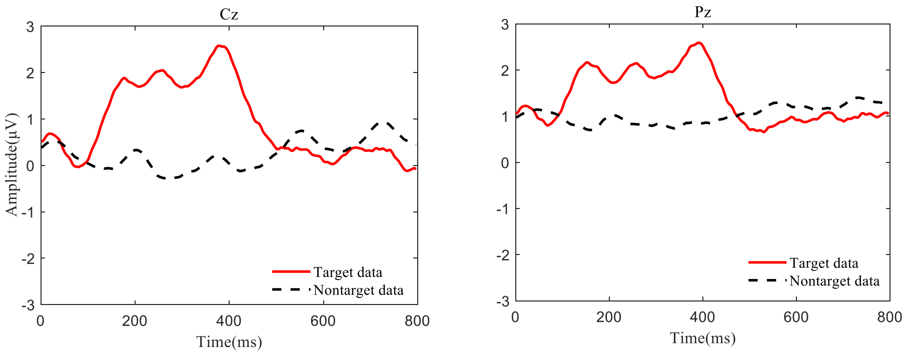

Fig. 4 Average ERPs for target trials and non-target trials for subject A in dataset 2 on EEG channels Cz (left) and Pz (right).

3.2 Implementation

3.2.1 Data preprocessing and feature extraction

Signal preprocessing is a crucial step to reduce dimensionality and increase the signal to noise ratio (SNR). For dataset 1 and dataset 2, the 64-channel EEG data were bandpass filtered from 0.5Hz to 20Hz and then down-sampled from 240Hz to 30Hz. The Parks-McClellan optimal equiripple finite impulse response (FIR) filter was used as the bandpass filter [43]. For dataset 3, the 8-channel EEG data were down-sampled by a factor of 8, and a 3rd order Butterworth filter between 0.1Hz and 30Hz was applied. Previous studies have shown that the P300 ERP appears about 300 ms after the target stimulus and the peak latency of the P300 signal differs across the subjects [44]. Based on this, we extract a range of EEG data from 0ms to 600ms after the stimulus from each channel as the temporal feature set. Therefore, the size of the feature vector for each stimulus is 64×18 where 64 is the number of channels and 18 is the number of samples per channel. For the label vector, we set the label to "+1" only if the row/column corresponding to the spelled character was flashed, indicating that the P300 component is reflected in this specific trial. Otherwise, we set the label to "+2". In particular, for dataset 1, we have a total of 7020 (39×12×15=7020) training feature vectors for each subject, while for dataset 2, we have a total of 15300 (85×12×15=15300) training feature vectors for each subject, and for dataset 3, there are 2400 (20×12×10=2400) training feature vectors for each subject.

3.2.2 Feature selection

The Least absolute shrinkage and selection operator (Lasso) is a kind of embedded method often used in feature selection, which has good stability. The algorithm has been demonstrated to be effective in feature selection. The main idea of this method is to minimize the sum of the predicted error. Lasso is a linear regressor with L1 regularization [45], which can be described in formula (6):

$$\frac{1}{2N}\sum_{i=1}^{N}(w_i - \alpha_0 - x_i^T \alpha)^2 + \lambda \sum_{j=1}^{n}|\alpha_j| \tag{6}$$

where N is the number of samples, w_i is the target of sample i, x_i is the input vector of sample i, λ is the regularization coefficient, α_0 and α are the regression coefficients in which α_0 is a constant and the size of α is $n \times 1$.

During the process of feature selection, the zero coefficient of the feature set after the shrinkage process is removed and the rest of the features are chosen to construct the model. As the value of λ grows, the number of zero components of α increases. In this study, we use the built-in Lasso function in Matlab 2017b, and use the default parameter values for this method.

3.2.3 Fuzzy fusion technology

Fuzzy measures are defined as subjective scales to evaluate the degree of fuzziness, which reflects the importance of attributes of the data and the relationships among those attributes [46]. A fuzzy measure μ can be described as:

Definition 1 Let X be an arbitrary set, μ over the power set of X can satisfy the theorem $\mu: P(X) \to [0,1]$, that is to say,

1) $\mu(\emptyset) = 0,\ \mu(X) = 1$
2) $\mu(A) \leq \mu(B),\ \forall A,B \in P(X),\ A \subset B$
3) $\lim_{n\to\infty} \mu(F_n) = \mu(\lim_{n\to\infty} F_n)$

where $F_n \in P(X)$ is a monotone set sequence. The fuzzy measure of different attributes is calculated by using fuzzy densities, which reveal the importance of different attributes. As mentioned above, if there are a set of n elements, the fuzzy measure has $2^n - 1$ parameters, which lead to complicated computation. The 2^{nd}-order additive fuzzy measure proposed by Grabisch has $n(n+1)/2$ fuzzy measure coefficients to be identified.

The fuzzy measures are used to denote the fuzzy integral, which is used to build a subjective evaluation model of the fuzzy object related to various attributes. Therefore, once we have evaluated the fuzzy measures of the different feature sets, they are applied as the fuzzy integral expression to further compute the fuzzy integral values.

To date, many different kinds of fuzzy integral methods, including the Sugeno integral and the Choquet integral, have been introduced [46]. Here we choose the Choquet integral. On the one hand, it provides a broad generalization of the ordinary integral. On the other hand, the learning process generated in this study can be treated as a convex quadratic program, which is the main reason for choosing the Choquet integral. The Choquet integral can be defined as:

Definition 2 Let f be a function on X: $X \to [0,\infty)$ and let μ be a fuzzy measure on X.

$$\int_X f(x) \circ \mu(\cdot) = \int_0^{\infty} \mu\,(F_\alpha) d\alpha \tag{7}$$

where $F_\alpha = \{x |\, f(x) \geq \alpha,\ x \in X\}$. Assuming that $X = \{x_1, x_2, \cdots x_n\}$ is a discrete set, and $f(x_1) \leq f(x_1) \leq \cdots \leq f(x_n)$. Then the discrete Choquet integral in relation to fuzzy measure μ can be transformed into

$$(c)\int f d\mu = \int_x f(x) \circ \mu(\cdot) = \sum_{j=1}^{n} [f(x_j) - f(x_{j-1})] \cdot \mu(A_j) \tag{8}$$

where $A_j = \{x_j, x_{j+1}, \cdots, x_n\}$, $f(x_0) = 0$.

3.3 Results

In this study, the proposed MSFF framework is tested on dataset 1, dataset 2, and dataset 3. In addition, five classifiers are then trained with the training data of three datasets and compared across the LDA, xDAWN, STDA, FC+LDA, and SWLDA classification methods. Testing data are processed similarly to the training data and then are sent to the classifier. Afterward, we use paired-samples *t*-tests with Bonferroni correction to estimate the significance of classification accuracy differences between our MSFF method and each of the other classification methods.

3.3.1 Comparison of classification methods for dataset 1

Table I represents the predicted words and errors for our proposed method for BCI competition II dataset II. The errors refer to the case when the row or column of the predicted character does not correspond to the target character. From Table I, we can easily see that the MSFF method obtains 100% accuracy after 4 epochs, which is better than the previously reported work in reference [26]. Table II describes the number of epochs that achieve an accuracy of 100% between our proposed method and the other methods for BCI competition II dataset II. In particular, the results achieved with our proposed MSFF method are compared with other conventional methods.

Table I Predicted words and error (%) of our proposed MSFF method for BCI competition II dataset II

Epoch	Predicted words	Error
1	XOOD NQON HAM PIA CANE S7HA Z5XOT 3567	38.71%
2	FOOD MOOT HAM PIE CAKE TUNA Z5AOM 4567	9.77%
3	FOOD MOON HAM PIE CAKE TUNA Z5SON 4567	12.99%
4	FOOD MOOT HAM PIE CAKE TUNA ZYGOT 4567	0.0%
5	FOOD MOOT HAM PIE CAKE TUNA ZYGOT 4567	0.0%

Table II The number of epochs achieving an accuracy of 100% with our proposed method and with other conventional methods for BCI competition II dataset II

Methods	The number of epochs when achieving an accuracy of 100%
SVM [29]	5 epochs
ICA [47]	5 epochs
SSAE-ESVM [26]	5 epochs
Proposed MSFF method	4 epochs

3.3.2 Comparison of classification methods for dataset 2

Fig. 4 shows the average ERPs for target trials and non-target trials for subject A in dataset 2. Fig. 5 describes the classification performance using different methods for different epochs. Fig. 6 shows the character recognition accuracy of the proposed method after 1, 5, 10, 13, and 15 epochs. It is apparent that the proposed framework achieved better or comparable averaged classification performance than previously reported methods after training for many number of epochs. Specifically, our proposed method obtained average accuracies of 31%, 72%, 88% and 96% after 1, 5, 10, 15 epochs, respectively. However, the LDA method achieved average accuracies of 25%, 64.5%, 86%, and 93.5% after 1, 5, 10, 15 epochs, respectively. In reference [35], the xDAWN was applied for classification and obtained average accuracies of 26.5%, 68.5%, 90% and 95% after 1, 5, 10, 15 epochs, respectively. The work of Zhang's proposed STDA technology learned spatial and temporal features achieved average accuracies of 28.5%, 67%, 89%, and 94.5% after 1, 5, 10, 15 epochs, respectively. The method of combining the fisher criterion and LDA obtained average accuracies of 22.5%, 60.5%, 84% and 90% after 1, 5, 10, 15 epochs, respectively. In addition, the SWLDA based feature selection method obtains average accuracies of 27.5%, 60.5%, 89% and 90.5% after 1, 5, 10, 15 epochs, respectively.

Researches have shown that the character recognition performance that is equal to or more than 70% can be used for effective BCI speller applications [26]. Our proposed work yielded 72% accuracy only using 5 epochs, which is suitable for use in BCI speller applications.

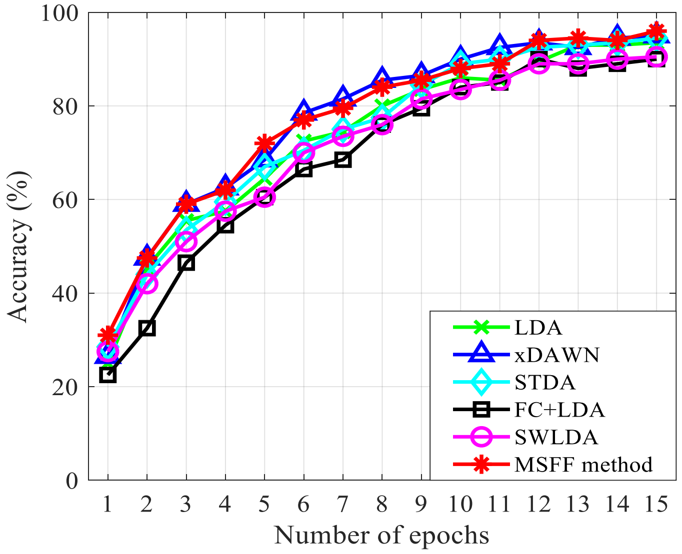

Fig. 5 Character recognition accuracy of the proposed method and traditional methods with different numbers of epochs for dataset 2

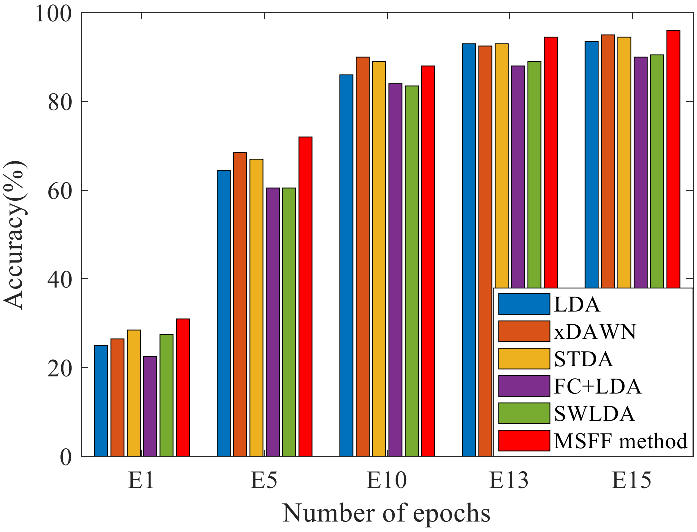

Fig. 6 Character recognition accuracy of the proposed method after 1, 5, 10, 13, and 15 epochs.

To compare our proposed method with the previous methods, a statistical analysis based on paired-samples *t*-test method is performed for different number of epochs. The paired *t*-test results reveal that our proposed method's classification performance is significantly than the LDA, STDA, FC+LDA, and SWLDA techniques ($p < 0.05$). However, there is no significant difference between our proposed MSFF method and xDAWN method. Nonetheless, our proposed MSFF method can achieve superior accuracy after 1 epoch, 5 epochs, 13 epochs, and 15 epochs, as shown in Fig. 6.

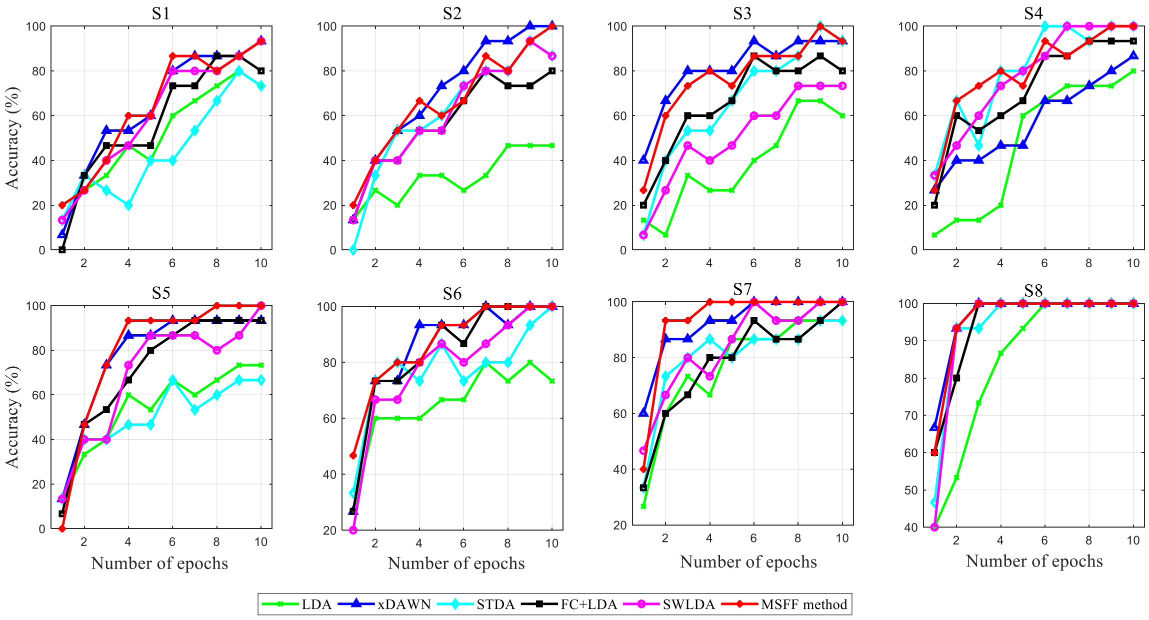

Fig. 7 The character detection accuracy obtained by the LDA, xDAWN, STDA, FC+LDA, SWLDA, and our proposed MSFF methods using one to ten epochs averaged for the eight subjects.

3.3.3 Comparison of classification methods for dataset 3

Figure 7 depicts the character detection accuracy obtained by the LDA, xDAWN, STDA, FC+LDA, SWLDA, and MSFF methods using one to ten epochs for each of the eight subjects. Table III shows the character recognition performance for dataset 3 using different methods. The result reflects the fact that the MSFF method yields significantly higher character detection average accuracies than the other methods across different numbers of epochs. Such superiority is more prominent for S5, S7, and S8 who can reach better accuracies compared to the other subjects. At the same time, Figure 8 depicts the character detection averaged accuracy obtained by our proposed MSFF method and the traditional methods using one to ten epochs for all subjects.

Table III Character recognition accuracy (%) for dataset 3 using different methods

	Epochs									
Method	1	2	3	4	5	6	7	8	9	10
LDA	20	35	43.3	50	57.5	64.2	68.3	74.2	76.7	80
xDAWN	31.7	60	70	76.7	79.2	88.3	90.8	91.7	94.2	95.8
STDA	22.5	56.7	59.2	64.2	70	77.5	79.2	81.7	90.8	89.2
FC+LDA	22.5	54.2	61.7	68.3	73.3	85	87.5	89.2	90.8	90.8
SWLDA	23.3	50.8	59.2	67.5	75.0	83.3	85.8	87.5	92.5	94.2
MSFF method	**30.0**	**62.5**	**73.3**	**82.5**	**81.7**	**90**	**92.5**	**92.5**	**97.5**	**98.3**

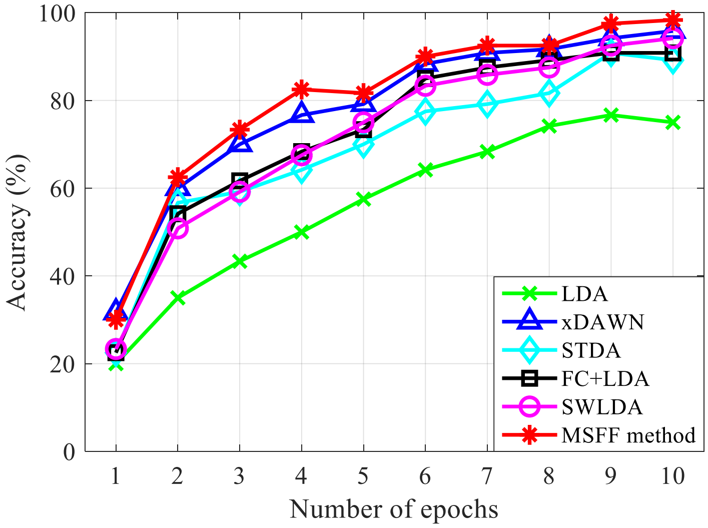

Fig. 8 Character recognition accuracy of the proposed method and and traditional methods with different numbers epochs for all subjects.

Table IV the results of the statistical analysis.

Proposed MSFF method paired with	*p*-value
LDA	$p < 0.001$
xDAWN	$p = 0.005$
STDA	$p < 0.001$
FC+LDA	$p < 0.001$
SWLDA	$p < 0.001$

Table IV describes the results of statistical analysis of the comparison. Paired t-tests reveal that our proposed MSFF method achieved significantly higher accuracies for character detection than those of the other methods with insufficient training samples (MSFF > LDA: $p < 0.001$, MSFF > xDAWN : $p < 0.01$, MSFF > STDA: $p < 0.001$, MSFF > FC+LDA: $p < 0.001$, MSFF > SWLDA: $p < 0.001$).

4. Discussion

P300-based BCI systems have been generally explored in the last few decades and have begun to be applied to clinical practice [2]. Although considerable progress has been made in the development of P300 BCI systems in recent years, it is still a challenge to improve system performance. For P300 spellers a crucial challenge is to enhance the performance in terms of character recognition accuracy.

In general, the character recognition performance mainly depends on good quality EEG signals and optimal signal processing algorithms. Preliminary research has indicated that the impact of signal variability reduces P300 speller performance, requiring the use of advanced machine learning techniques to overcome this difficulty [48]. Given the nonstationary effects of EEG on BCI applications, we proposed a novel framework that integrates multiple classifiers based on feature selection and fuzzy fusion. Although several works focus on the combination of classifiers using ensemble methods [30, 31], to our knowledge, this is the first approach to integrate the same classifiers to improve system performance.

In the feature extraction stage, the temporal features from each epoch were extracted and a feature selection method was adopted to solve the dimensionality reduction problem. In the feature classification stage, two feature subsets were sent to the same classifier and fuzzy fusion was applied to incorporate the classifier outputs. The proposed framework, based on EEG data from all available channels for each of the three datasets, yields performance that is higher than or equal to the performance of traditional classifiers.

5. Conclusion

This work focuses on the operation of the combination of multiple classifiers based on fuzzy measures and fuzzy integrals to the ERP-based BCI. Especially, the Choquet integral was used. Additionally, the Lasso algorithm was used to select features, and the LDA method was used to construct a classification model. To evaluate the effectiveness of the framework, we compared our proposed MSFF method with five classification algorithms: LDA, xDAWN, STDA, FC+LDA, and SWLDA. We applied the methods to attempt to recognize the P300 ERP component from three different datasets based on the visual ERP BCI system. The experimental results show that our proposed MSFF method can outperform the other methods.

Acknowledgments

The authors would like to thank all the public datasets and algorithms described in this paper.

References

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, “Brain-computer interface technology: a review of the first international meeting,” *IEEE transactions on rehabilitation engineering,* vol. 8, no. 2, pp. 164-173, 2000.

[2] J. J. Daly, and J. R. Wolpaw, “Brain–computer interfaces in neurological rehabilitation,” *The Lancet Neurology,* vol. 7, no. 11, pp. 1032-1043, 2008.

[3] J. Katona, and A. Kovari, “Examining the learning efficiency by a brain-computer interface system,” *Acta Polytechnica Hungarica,* vol. 15, no. 3, pp. 251-280, 2018.

[4] S. Bulárka, and A. Gontean, "Brain-Computer Interface review." pp. 219-222.

[5] J. J. Vidal, “Toward direct brain-computer communication,” *Annual review of Biophysics and Bioengineering,* vol. 2, no. 1, pp. 157-180, 1973.

[6] J. J. Vidal, “Real-time detection of brain events in EEG,” *Proceedings of the IEEE,* vol. 65, no. 5, pp. 633-641, 1977.

[7] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, “Correlation-based channel selection and regularized feature optimization for MI-based BCI,” *Neural Networks,* vol. 118, pp. 262-270, 2019.

[8] M. G. Coles, and M. D. Rugg, *Event-related brain potentials: An introduction*: Oxford University Press, 1995.

[9] S. Sutton, M. Braren, J. Zubin, and E. John, “Evoked-potential correlates of stimulus uncertainty,” *Science,* vol. 150, no. 3700, pp. 1187-1188, 1965.

[10] J. Jin, C. Liu, I. Daly, Y. Miao, S. Li, X. Wang, and A. Cichocki, “Bispectrum-based channel selection for motor imagery based brain-computer interfacing,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 10, pp. 2153-2163, 2020.

[11] J. K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, and A. Cichocki, “An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system,” *Computational Intelligence and Neuroscience,* vol. 2019, 2019.

[12] R. Ramele, A. J. Villar, and J. M. Santos, “EEG waveform analysis of P300 ERP with applications to brain computer interfaces,” *Brain Sciences,* vol. 8, no. 11, pp. 199, 2018.

[13] 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.

[14] J. Mak, Y. Arbel, J. W. Minett, L. M. McCane, B. Yuksel, D. Ryan, D. Thompson, L. Bianchi, and D. Erdogmus, “Optimizing the P300-based brain–computer interface: current status, limitations and future directions,” *Journal of neural engineering,* vol. 8, no. 2, pp. 025003, 2011.

[15] L. Bougrain, C. Saavedra, and R. Ranta, "Finally, what is the best filter for P300 detection?."

[16] J. Jin, Z. Chen, R. Xu, Y. Miao, X. yu Wang, and T.-P. Jung, “Developing a Novel Tactile P300

Brain-Computer Interface with a Cheeks-Stim Paradigm," *IEEE Transactions on Biomedical Engineering*, 2020.

[17] K. Colwell, D. Ryan, C. Throckmorton, E. Sellers, and L. Collins, "Channel selection methods for the P300 Speller," *Journal of neuroscience methods,* vol. 232, pp. 6-15, 2014.

[18] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of neuroscience methods,* vol. 167, no. 1, pp. 15-21, 2008.

[19] C.-T. Lin, I.-F. Chung, L.-W. Ko, Y.-C. Chen, S.-F. Liang, and J.-R. Duann, "EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment," *IEEE Transactions on Biomedical Engineering,* vol. 54, no. 7, pp. 1349-1352, 2007.

[20] 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, 2010.

[21] C. Başar-Eroglu, E. Başar, T. Demiralp, and M. Schürmann, "P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review," *International journal of psychophysiology,* vol. 13, no. 2, pp. 161-179, 1992.

[22] T. Demiralp, A. Ademoglu, M. Schürmann, C. Basar-Eroglu, and E. Basar, "Detection of P300 waves in single trials by the wavelet transform (WT)," *Brain and language,* vol. 66, no. 1, pp. 108-128, 1999.

[23] A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology,* vol. 46, no. 3, pp. 617-625, 2009.

[24] D. B. Ryan, G. Frye, G. Townsend, D. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain–computer interface: increasing the rate of communication," *Intl. Journal of Human–Computer Interaction,* vol. 27, no. 1, pp. 69-84, 2010.

[25] C. S. Syan, and R. E. Harnarinesingh, "Comparison of pre-processing and classification techniques for single-trial and multi-trial P300-based brain computer interfaces," *American Journal of Applied Sciences,* vol. 7, no. 9, pp. 1219, 2010.

[26] S. Kundu, and S. Ari, "P300 based character recognition using sparse autoencoder with ensemble of SVMs," *Biocybernetics and Biomedical Engineering,* vol. 39, no. 4, pp. 956-966, 2019.

[27] N. Liang, and L. Bougrain, "Averaging techniques for single-trial analysis of oddball event-related potentials."

[28] J. Jin, E. W. Sellers, and X. Wang, "Targeting an efficient target-to-target interval for P300 speller brain–computer interfaces," *Medical & biological engineering & computing,* vol. 50, no. 3, pp. 289-296, 2012.

[29] 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.

[30] 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.

[31] F. Cavrini, L. Bianchi, L. R. Quitadamo, and G. Saggio, "A fuzzy integral ensemble method in visual P300 brain-computer interface," *Computational intelligence and neuroscience,* vol. 2016,

2016.

[32] Y. Xu, X.-B. Wang, J. Ding, L.-Y. Wu, and N.-Y. Deng, “Lysine acetylation sites prediction using an ensemble of support vector machine classifiers,” *Journal of Theoretical Biology,* vol. 264, no. 1, pp. 130-135, 2010.

[33] M. Congedo, A. Barachant, and A. Andreev, “A new generation of brain-computer interface based on Riemannian geometry,” *arXiv preprint arXiv:1310.8115*, 2013.

[34] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, “An efficient P300-based brain–computer interface for disabled subjects,” *Journal of Neuroscience methods,* vol. 167, no. 1, pp. 115-125, 2008.

[35] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, “xDAWN algorithm to enhance evoked potentials: application to brain–computer interface,” *IEEE Transactions on Biomedical Engineering,* vol. 56, no. 8, pp. 2035-2043, 2009.

[36] 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.

[37] G. Pires, U. Nunes, and M. Castelo-Branco, “Statistical spatial filtering for a P300-based BCI: tests in able-bodied, and patients with cerebral palsy and amyotrophic lateral sclerosis,” *Journal of neuroscience methods,* vol. 195, no. 2, pp. 270-281, 2011.

[38] M. Grabisch, “K-order additive discrete fuzzy measures and their representation,” *Fuzzy sets and systems,* vol. 92, no. 2, pp. 167-189, 1997.

[39] D. Krusienski, and G. Schalk, “BCI Competition III Challenge 2004,” 2004.

[40] E. Donchin, K. M. Spencer, and R. Wijesinghe, “The mental prosthesis: assessing the speed of a P300-based brain-computer interface,” *IEEE transactions on rehabilitation engineering,* vol. 8, no. 2, pp. 174-179, 2000.

[41] L. Yang, J. Li, Y. Yao, and G. Li, "An algorithm to detect P300 potentials based on F-score channel selection and support vector machines." pp. 280-284.

[42] M. Sahu, N. Nagwani, and S. Verma, “Applying auto regression techniques on amyotrophic lateral sclerosis patients EEG dataset with P300 speller,” *Indian Journal of Science and Technology,* vol. 9, no. 48, pp. 1-8, 2016.

[43] G. V. Tcheslavski, “Coherence and phase synchrony analysis of electroencephalogram,” Virginia Tech, 2005.

[44] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichocki, “The study of generic model set for reducing calibration time in P300-based brain–computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 28, no. 1, pp. 3-12, 2019.

[45] A. M. Soleh, A. H. Wigena, A. Djuraidah, and A. Saefuddin, “Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO),” *Applied Mathematical Sciences,* vol. 9, no. 108, pp. 5361-5369, 2015.

[46] F. Cavrini, L. R. Quitadamo, L. Bianchi, and G. Saggio, "Combination of Classifiers using the Fuzzy Integral for Uncertainty Identification and Subject Specific Optimization." pp. 14-24.

[47] 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.

[48] J. Mladenović, J. Mattout, and F. Lotte, “A generic framework for adaptive EEG-based BCI training and operation,” *arXiv preprint arXiv:1707.07935*, 2017.