Discriminative Canonical Pattern Matching for Single-Trial Classification of ERP Components

Xiaolin Xiao , *Member, IEEE*, Minpeng Xu, *Member, IEEE*, Jing Jin , *Member, IEEE*,
Yijun Wang , *Member, IEEE*, Tzyy-Ping Jung , *Fellow, IEEE*, and Dong Ming , *Senior Member, IEEE*

Abstract—**Event-related potentials (ERPs) are one of the most popular control signals for brain–computer interfaces (BCIs). However, they are very weak and sensitive to the experimental settings including paradigms, stimulation parameters and even surrounding environments, resulting in a diversity of ERP patterns across different BCI experiments. It's still a challenge to develop a general decoding algorithm that can adapt to the ERP diversities of different BCI datasets with small training sets. This study compared a recently developed algorithm, i.e., discriminative canonical pattern matching (DCPM), with seven ERP-BCI classification methods, i.e., linear discriminant analysis (LDA), stepwise LDA, bayesian LDA, shrinkage LDA, spatial-temporal discriminant analysis (STDA), xDAWN and EEGNet for the single-trial classification of two private EEG datasets and three public EEG datasets with small training sets. The feature ERPs of the five datasets included P300, motion visual evoked potential (mVEP), and miniature asymmetric visual evoked potential (aVEP). Study results showed that the DCPM outperformed other classifiers for all of the tested datasets, suggesting the DCPM is a robust classification algorithm for assessing a wide range of ERP components.**

Index Terms—**Brain–computer interface (BCI), electroencephalogram (EEG), single-trial classification, event-related potential (ERP), discriminative canonical pattern matching (DCPM).**

X. Xiao is with the Laboratory of Neural Engineering & Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University.

M. Xu is with the Laboratory of Neural Engineering & Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, and also with the Tianjin International Joint Research Center for Neural Engineering, Academy of Medical Engineering and Translational Medicine, Tianjin University.

J. Jin is with the Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology.

Y. Wang is with the State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, 100083 Beijing, China (e-mail: wangyj@semi.ac.cn).

T.-P. Jung is with the Swartz Center for Computational Neuroscience, University of California, and also with the Laboratory of Neural Engineering & Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University.

D. Ming is with the Laboratory of Neural Engineering & Rehabilitation, Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, 300072 Tianjin, China, and also with the Tianjin International Joint Research Center for Neural Engineering, Academy of Medical Engineering and Translational Medicine, Tianjin University, 300072 Tianjin, China (e-mail: richardming@tju.edu.cn).

Digital Object Identifier 10.1109/TBME.2019.2958641

I. INTRODUCTION

BRAIN-COMPUTER interfaces (BCIs) directly link the brain and the external world without the involvement of muscles and peripheral nerves, which allow users to communicate with their environments and control external devices [1]–[4]. Moreover, BCIs have multiple applications in attention test [5], mental workload assessment [6], glaucoma detection [7] and so on. Most BCI systems are based on the measurements of electroencephalography (EEG) because of its advantageous of non-invasiveness, high temporal resolution, low cost and mobility over other brain recording/imaging techniques such as electrocorticography (ECoG), functional magnetic resonance imaging (fMRI), functional near infrared reflectance spectroscopy (fNIRS) and magnetoencephalogram (MEG). Among all EEG features, event-related potentials (ERPs) are one of the most important brain control signals used for BCIs, which includes P300 [8]–[10], N170 [11], N200 [12], [13], motion visual evoked potential (mVEP) [14] and miniature asymmetric visual evoked potential (aVEP) [15], etc. ERP-based BCIs have many practical applications in both clinical and non-clinical fields.

A key component of an ERP-based BCI is to discriminate ERPs from the noisy background EEG. However, as the background EEG signals are non-linear, non-stationary and often many times larger than ERPs, it's difficult to recognize the single-trial ERPs and needs to collect multiple samples. Furthermore, the ERP profiles are sensitive to experimental settings. A change of experimental parameters would result in changes in the latencies and amplitudes of the ERPs. But the mechanism of the ERP generation remains unclear [16], making it impossible to construct a mathematical model to link the ERP patterns with experimental parameters. As a result, the traditional decoding algorithms often have variable success across different datasets. Therefore, it's desirable to develop a robust classification method that can adapt to a wide range of ERP patterns for practical applications of ERP-based BCIs.

Previous studies have developed a few algorithms for the single-trial ERP classification. In 2008, Krusienski *et al.* used

a stepwise linear discriminant analysis (SWLDA) to recognize the single-trial P300 potential and got an average accuracy of about 35% of character recognition [17]. In the same year, Hoffmann *et al.* introduced a bayesian version of regularized LDA, i.e., bayesian LDA (BLDA) for the single-trial P300 classification, which achieved an average accuracy of about 60% across eight subjects including four disabled patients [18]. In 2009, Rivet *et al.* proposed xDAWN to enhance the P300 features for P300-speller by projecting raw EEG on the estimated P300 subspace and the average character accuracy was nearly 30% for only one repetition [19]. In 2011, Blankertz *et al.* applied a shrinkage technique to LDA, i.e., SKLDA. It could achieve as high as 70% in the single-trial P300 classification [20]. In 2013, Zhang *et al.* designed a spatial-temporal discriminant analysis (STDA) algorithm and got an average accuracy about 61% for classifying P300 in single trials [21]. In 2018, Lawhern *et al.* proposed EEGNet which was regarded as a compact CNN architecture for EEG-based BCIs. It had been proved effective for single-trial P300 classification and the AUC was higher than 0.9 using about 1500 training samples [22]. Besides, researchers had tested a variety of other methods, such as support vector machine (SVM) [23], linearly constrained minimum variance (LCMV) [24] and independent component analysis (ICA) [25] *et al.*, for the single-trial ERP classification.

Recently, we have developed a new algorithm, i.e., discriminative canonical pattern matching (DCPM) [15], to recognize the miniature ERPs. It can first suppress the common-mode noise of the background EEG and then recognize the canonical patterns of ERPs. DCPM performs well in classifying the miniature aVEPs with even as small as 0.5 microvolts in amplitude. However, it's still unclear what the advantage of DCPM is over other traditional ERP classification methods, and whether DCPM works for other ERP components. To this end, this study compared the single-trial classification results obtained by DCPM and other methods including LDA, SWLDA, BLDA, STDA, SKLDA, xDAWN and EEGNet for five different BCI paradigms including a miniature aVEP-based speller, two different kinds of P300-spellers, a rapid serial visual presentation (RSVP)-speller and a motion center-speller, with small training sets as few as 30 samples.

II. MATERIALS AND METHODS

A. Introduction of Four BCI Datasets

1) Dataset 1: This study used a miniature aVEP dataset collected in our previous study (please find the online video from the URL: https://www.youtube.com/watch?v = kC7btB3mvGY) [15]. In the offline aVEP speller experiment, twelve subjects were asked to focus at the center of the target character indicated by a visual cue but not the stimuli around the character. The lateral visual stimuli with 25% duty cycle (i.e., 25 ms on and 75 ms off) would randomly appear at the bottom left or right area, which consisted of six tiny dots gathered within 0.5° of visual angle and located at an eccentricity of 2.1°, as shown in Fig. 1(a). Left stimuli evoked right aVEPs while right stimuli evoked left aVEPs.

EEG data were recorded using a Neuroscan Synamps2 system with 64 electrodes placed at the locations following the 10/20
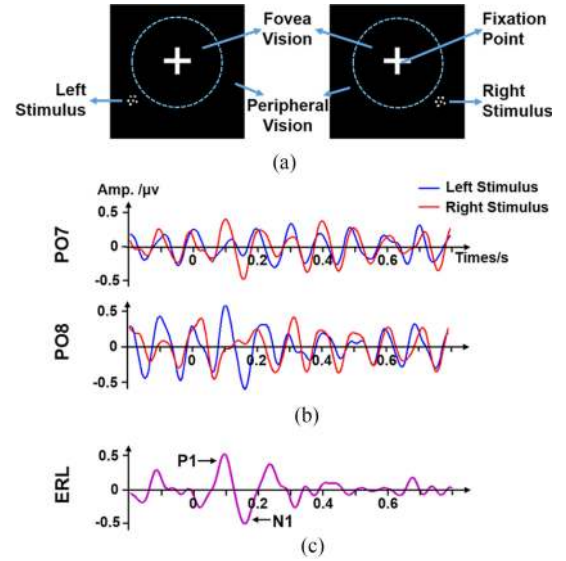

Fig. 1. Miniature aVEPs induced by repetitive peripheral vision stimuli. (a) The blue dash line lying at 2° away from the cross center, which was not displayed in the experiment, indicates the boundary between the fovea and peripheral visions. (b) Grand average miniature aVEPs of 90 trials across subjects were obtained for the left and right stimulus, respectively, which were recorded from channels PO7 and PO8. The zero point in the time axis indicates the stimulus onset. (c) Grand average event-related lateralization (ERL) across channels PO7 and PO8.

system. The reference electrode was put at the central area near Cz, while the ground electrode was put at the frontal lobe. The stored EEG data were first filtered at 2–70 Hz using a band-pass Chebyshev filter, and then downsampled to 200 Hz. After that, the signals were segmented from 0.05 s to 0.25 s after stimulus onset, extracted from 21 channels (P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, CB1, O1, Oz, O2 and CB2) and then used for classification. The sampling rate, filter band and selected channels are the same to our previous study. A total of 720 segments, i.e., trials, consisting of 360 left aVEPs and 360 right aVEPs were obtained for each subject.

2) Dataset 2: The P300 dataset was from the EPFL BCI group (https://mmspg.epfl.ch/research/page-58317-en-html/bci-2/bci_datasets/) [18]. EEG data were recorded from eight subjects, including four disabled and four healthy persons using a Biosemi Active Two amplifier (BioSemi, Netherland) at a sampling rate of 2048 Hz from 32 electrodes placed at the standard positions of the 10–20 international system. Visual stimuli consisted of six images and one of them was cued as the target for each selection. All images were highlighted one by one in a random sequence. Each flash lasted for 100 ms with an ISI of 300 ms. The subjects were asked to focus on the cued target stimulus during the experiment.

The EEG signals were first band-pass filtered from 1 Hz to 12 Hz by a sixth-order forward–backward Butterworth bandpass filter and then downsampled from 2048 Hz to 32 Hz. The same parameters were used with the original research in which the BLDA method was proposed [18]. The EEG trials used for classification were extracted from the data of 32 channels (Fp1, 2, AF3, 4, Fz, F3, 4, 7, 8, FC1, 2, 5, 6, T7, 8, Cz, C3, 4, CP1, 2, 5, 6, Pz, P3, 4, 7, 8, PO3, 4, Oz and O1, 2) after re-referenced

to the average of M1 and M2. The time window was from 0 to 1 s after stimulus onset. Least of 3240 trials consisting of 540 targets and 2700 non-targets were obtained for each subject.

3) Dataset 3: In order to further analyze the classification performance of P300 s in the classical paradigm, we implemented a row/column (RC) speller featuring a 6×6 character matrix. Each row and column would be intensified in succession according to the random sequence with an inter-stimulus interval (ISI) of 175 ms. Each round comprised 12 intensifications (six rows and six columns) and there were six rounds for each character. Each of 12 subjects randomly spelled 30 characters and was asked to concentrate on the specific character and silently count the number of intensifications. A total of 2160 trials, consisting of 360 targets and 1800 non-targets were collected for each subject.

EEG data were recorded using a Neuroscan Synamps2 system with 64 electrodes placed according to the 10/20 system. The reference electrode was put at the central area and the ground electrode was put at the frontal area. The stored EEG data were first re-referenced to the average of M1 and M2, and then filtered at 1–20 Hz using a band-pass Chebyshev I filter, and then downsampled to 200 Hz. When classifying, we used 0.05 s to 0.75 s after stimulus onset for each trial with 16 channels according to previous study [26] (F3, F4, Fz, T7, T8, C3, C4, Cz, CP3, CP4, P3, P4, Pz, PO7, PO8, and Oz).

4) Dataset 4: Dataset 4 was from an experiment of RSVP speller (http://bnci-horizon-2020.eu/database/data-sets) conducted by Acqualagna and Blankertz [27]. In the RSVP paradigm, all symbols with different colors are presented one-by-one in random order at the center of the screen. Participants were asked to concentrate on the target letter and silently count the number of times the target characters appeared on the screen.

EEG data were recorded at a sample rate of 1000 Hz with 63 channels (Fp1/2, AF3/4, Fz, F1-10, FCz, FC1-6, FT7/8, T7/8, Cz, C1-6, TP7/8, CPz, CP1-6, Pz, P1-10, POz, PO3/4/7-10, Oz and O1/2). All the electrodes were referenced to the left mastoid, and grounded to the forehead area. It should be noted that only eleven out of twelve subjects' data were used in this study, as the remaining one had two channels (PO6 and O2) lost. A total of 7200 training trials including 240 targets and 6960 non-targets and 12300 testing trials including 410 targets and 11890 non-targets were collected for each subject. Each trial was first filtered between 1–20 Hz using a band-pass Chebyshev filter, and then downsampled to 200 Hz. The trials for classification were extracted from the EEG data of 55 electrodes as the original study (all except for Fp1/2, AF3/4, F9/10, FT7/8) from 0.05 s to 0.75 s after stimulus presentation.

5) Dataset 5: This dataset was about the motion center speller developed by Schaeff *et al.* in Technical University Berlin [28] (http://bnci-horizon-2020.eu/database/data-sets). In the paradigm, a moving pattern in the middle of the hexagon is used to evoke an mVEP. The grid pattern consists of arrowheads with unique colors pointing alternately to one of the pieces. Participants have to fixate on a central fixation point in the middle of the moving pattern that was meant to elicit an mVEP.

EEG data from eleven subjects were recorded at 1000 Hz with 63 electrodes placed according to the international 10–10 system. The right mastoid was chosen as a reference site and

the ground electrode was placed at forehead. A total of 2160 training trials consisting of 360 targets and 1800 non-targets were collected for ten subjects, but one subject only has 2100 training trials. Each subject had a different number of testing samples ranging from 3960 to 5640 trials. Each EEG trial was directly filtered between 1–20 Hz using a band-pass Chebyshev filter, and then down sampled to 100 Hz. The trials for classification were extracted from the EEG data of 57 electrodes as the original study (Fz, F1-10, FCz, FC1-6, FT7/8, T7/8, Cz, C1-6, TP7/8, CPz, CP1-6, Pz, P1-10, POz, PO3/4/7/8, Oz and O1/2) from 0.05 s to 0.75 s after stimulus presentation.

B. Inclusion Criteria of Algorithms for Comparison

This study compared DCPM with seven other algorithms which were popular in the studies of ERP-BCIs, i.e., LDA, BLDA, SKLDA, SWLDA, STDA, xDAWN and EEGNet. As LDA is the easiest and most commonly used method for ERP classification [29], it was selected as a default method for comparison in this study. BLDA was selected because it was first proposed for addressing dataset 2 and got very good performance. Likewise, SKLDA performed well in datasets 4 and 5. SWLDA was selected in this study because it's also a common method for ERP-BCIs and demonstrated to be better than SVM [30]. The aforementioned methods concatenated temporal points and spatial channels, but DCPM adopted spatial-temporal samples and implemented collaboratively discriminant analysis. Therefore, we selected STDA as a comparison, which is also a collaboratively discriminant analysis method. xDAWN was demonstrated to have good performance on feature enhancement which was also claimed as an advantage of DCPM, so it was selected for comparison. At last, deep neural network is an emerging technique in the field of computer science, and shows great potential on addressing EEG signals. Therefore, this study selected a compact CNN method, i.e., EEGNet, for comparison.

C. Linear Discriminant Analysis

As one of the most popular algorithms in BCI applications, LDA uses hyperplanes to separate the data representing different classes (e.g., two classes: target and non-target). Let $X_k \in R^{N_D \times N_k} (N_D = N_t \cdot N_c)$ are the training samples of pattern k = 1, 2. N_t represents temporal points, N_c represents the number of channels and N_k represents the number of samples in class X_k. Each sample x_i is the concatenation of N_t and N_c. The means and empirical covariance matrices of two classes are computed from

$$\mu_k = \frac{1}{N_k} \sum_{i \in X_k} x_i, \; k = 1, 2 \tag{1}$$

$$\Sigma_k = \frac{1}{N_k - 1} \sum_{i \in X_k} (x_i - \mu_k)(x_i - \mu_k)^T, k = 1, 2 \tag{2}$$

$$\Sigma = \frac{N_1}{N_1 + N_2} \Sigma_1 + \frac{N_2}{N_1 + N_2} \Sigma_2 \tag{3}$$

where Σ is the common covariance matrix and the hyperplane, namely the projection vector w of LDA is written as

$$w = \Sigma^{-1} (\mu_1 - \mu_2) \tag{4}$$

Although LDA has been widely applied in BCI systems, it performs poorly in single-trial classification with few training samples. Therefore, some advanced versions of LDA have been proposed to address the problem.

D. Advanced Versions of LDA

1) Stepwise LDA (SWLDA): SWLDA has been employed in the case of a small sample size because of its effectiveness in feature dimension reduction [17]. SWLDA implements a combination of forward and backward stepwise analysis to select suitable features in the discriminant model. The input features are weighted using ordinary least squares regression to predict the target class labels. Starting with no initial features in the discriminant function, the most statistically significant input feature for predicting the target label (with p-value < 0.1) is added to the discriminant function. After each new entry to the discriminant function, a backward stepwise analysis is performed to remove the least significant input features, with p-values > 0.15. This process is repeated until the discriminant function includes a predetermined number of features or until no additional features satisfy the entry/removal criteria [30].

2) Bayesian LDA (BLDA): In BLDA, firstly, the neurophysiological and experimental priors are employed by modeling the trial-level covariance, and then the weight vector covariance of LDA is expressed explicitly as linearly separable components, meanwhile, the relative contribution of each component is controlled by the hyperparameters estimated using Restricted Maximum Likelihood (ReML) [31]. BLDA is a probabilistic method that based on Bayesian regression, and it performs better than LDA with a small number of training sets or strong noise contamination [18]. The Bayesian regression framework offers a more elegant and less time-consuming solution for the problem of choosing the hyperparameters. Details of the algorithm can be found in [31].

3) Shrinkage LDA (SKLDA): Shrinkage is a common remedy for compensating the systematic bias of the estimated covariance matrices and shrinkage parameter for high-dimensional feature spaces [20]. The SKLDA improves the traditional LDA by adjusting the extreme eigenvalues of the covariance matrix towards the average eigenvalue and has shown its superior performance when using insufficient training samples [21]. The poorly estimated covariance matrix Σ_c can be remedied as

$$\tilde{\Sigma}_c = (1 - \lambda)\Sigma_c + \lambda_c v_c \mathbf{I}, \ \lambda_c \in [0, 1] \tag{5}$$

where λ_c is a shrinkage parameter, $v_c = (tr(\Sigma_c))/(D)$ is defined as the average eigenvalue of Σ_c with D being the dimensionality of the feature space, and $\mathbf{I}$ is an identity matrix. The method to calculate the optimal shrinkage parameter can be found in [32].

4) Spatial-Temporal Discriminant Analysis (STDA): Unlike concatenating temporal points and spatial channels that adopted in the traditional classification methods, STDA adopts each spatial-temporal sample to a new one-way sample after projecting by the matrices that calculated from spatial and temporal feature subspaces, instead of the vectorized feature sample. The STDA method implements collaboratively discriminant analysis and performs alternatingly optimization in the spatial and temporal dimensions of samples, which reduces feature dimensionalities, and hence improves the estimation of covariance matrices in the discriminant analysis even using limited number of training samples. This assists in enhancing the generalization capability of the classifier. The details of the algorithm can be found in Zhang's study [21]. The classification performance depends on the selected number of retained eigenvectors. This study used the number of 6 after optimization.

E. xDAWN Algorithm

The main idea of xDAWN is to construct a spatial filter that can enhance the synchronous response induced by the target stimuli. In the xDAWN method, the synchronous responses are first estimated for each channel, and then these responses are used to estimate the spatial filters which will enhance the evoked P300 potentials. xDAWN unfolds the response pattern as:

$$X = D\,A + N \tag{6}$$

where A is the synchronized response with target stimuli, and D is a Toeplitz matrix whose first column is defined by 1 at stimuli onset. N are the noise and artifacts.

We used BLDA for classification after xDAWN spatial filtering, and a full description of xDAWN and BLDA can be found in [19] and [31].

F. EEGNet

CNNs were first used in computer vision and has been applied to BCIs recently. EEGNet is a compact CNN architecture for EEG-based BCIs which was introduced in 2018 [22]. EEGNet uses depthwise and separable convolutions to construct an EEG-specific model. The network has two blocks, and there are two convolutional steps in block 1. The first step is a temporal convolution that learns frequency filters, and the second is a depthwise convolution that learns frequency-specific spatial filters. In block 2, there is a separable convolution combined with a depthwise convolution, which is followed by a pointwise convolution. EEGNet has been generalized across four BCI paradigms including P300 s, error-related negativity responses, movement-related cortical potentials, and sensory motor rhythms. Another feature of EEGNet is the ability of being trained with limited data. More details about EEGNet can be found in [22].

G. Discriminative Canonical Pattern Matching (DCPM)

The DCPM method consists of three major parts: (1) the construction of discriminative spatial patterns (DSPs); (2) the construction of CCA patterns; (3) pattern matching (see Fig. 2).

Given a parameter $k = 1, 2$ denotes the number of patterns, $X_k \in R^{N_c \times N_t \times N_s}$ are different training sets, $Y \in R^{N_c \times N_t}$ is the testing sample, where N_c is the number of channels, N_t is the number of time points, N_s is the trial number. They are both zero mean across time. $\hat{X}_k \in R^{N_c \times N_t}$ is the template of pattern k, that is derived from the average of training trials. The
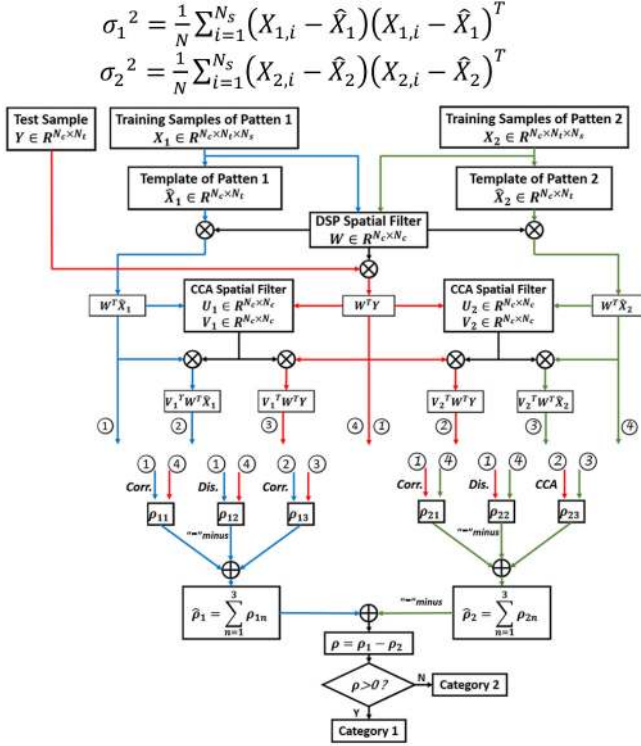
$$\sigma_1{}^2 = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{1,i} - \hat{X}_1\right)\left(X_{1,i} - \hat{X}_1\right)^T$$

$$\sigma_2{}^2 = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{2,i} - \hat{X}_2\right)\left(X_{2,i} - \hat{X}_2\right)^T$$

Fig. 2.   Flowchart of DCPM classification algorithm.

covariance matrix of $\begin{bmatrix}\hat{X}_1 \\ \hat{X}_2\end{bmatrix}$. is written as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \hat{X}_1\hat{X}_1^T & \hat{X}_1\hat{X}_2^T \\ \hat{X}_2\hat{X}_1^T & \hat{X}_2\hat{X}_2^T \end{bmatrix} \tag{7}$$

The variances of X_1 and X_2 are

$$\sigma_1{}^2 = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{1,i} - \hat{X}_1\right)\left(X_{1,i} - \hat{X}_1\right)^T \tag{8}$$

$$\sigma_2{}^2 = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{2,i} - \hat{X}_2\right)\left(X_{2,i} - \hat{X}_2\right)^T \tag{9}$$

Then a projection matrix W is constructed by DSP, λ_i is the eigenvalue of the ith column of W and W could be regarded as a set of spatial filters to make the two patterns more discriminative after transformation,

$$S_w{}^{-1}S_B * W = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{N_c} \end{bmatrix} * W \tag{10}$$

$$S_B = \Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{21} \tag{11}$$

$$S_w = \sigma_1{}^2 + \sigma_2{}^2 \tag{12}$$

After removing the common mode noise by W, the CCA algorithm is used to reveal the underlying correlation between $W^T\hat{X}_k$ and W^TY by finding two projection matrixes, U_k, V_k,

which equals to solve

$$\mathrm{CCA}(W^T\hat{X}_k, W^TY) =$$

$$\max_{U_k,V_k} \frac{\mathcal{E}[U_k{}^TW^T\hat{X}_kY^TWV_k]}{\sqrt{\mathcal{E}[U_k{}^TW^T\hat{X}_k\hat{X}_k^TWU_k] \cdot \mathcal{E}[V_k{}^TW^TYY^TWV_k]}} \tag{13}$$

where $\mathcal{E}[\cdot]$ is the expectation. In pattern matching, the similarity between the training template and the testing signal is represented as a vector

$$\rho_k = \begin{bmatrix} \rho_{k1} \\ \rho_{k2} \\ \rho_{k3} \end{bmatrix} = \begin{bmatrix} \mathrm{corr}\left(W^T\hat{X}_k, W^TY\right) \\ -\mathrm{dist}\left(W^T\hat{X}_k, W^TY\right) \\ \mathrm{corr}\left(V_k{}^TW^T\hat{X}_k, V_k{}^TW^TY\right) \end{bmatrix}, k=1,2 \tag{14}$$

where $\mathrm{corr}(*)$ refers to the Pearson's correlation, $\mathrm{dist}(*)$ refers to the Euclidean distance. The more similar it is between Y and $\hat{X}_k$, the larger the $\rho_{k,1}$, $\rho_{k,2}$, and $\rho_{k,3}$ will be. The final feature value can be the sum of all coefficients,

$$\tilde{\rho}_k = \sum_{i=1,2,3} \rho_{k,i} \tag{15}$$

Then the predicted code pattern of Y is

$$\hat{k} = \underset{k}{\mathrm{argmax}}\,\tilde{\rho}_k \tag{16}$$

III. RESULTS

A. Characteristics of ERPs

The characteristics of ERPs were first analyzed and compared for all datasets. Fig. 3 shows the temporal and spatial differences between two kinds of ERPs for each dataset, i.e., left vs. right ERPs for Dataset 1 and target vs. non-target ERPs for Datasets 2, 3, 4 and 5. It was obvious that the ERP's morphologies and topographic patterns were very different from each other for the five datasets. Specifically, the discriminative ERP features were restricted to the occipital areas within a short time window for Dataset 1, while spread over the whole scalp and lasted for a relatively long time for the other four datasets. Although some typical ERP components, such as P200, N200 and P300, could be found in Datasets 2, 3, 4 and 5, they showed evident differences among conditions. Therefore, it requires a robust and adaptable classification algorithm to address such diverse ERPs.

B. Comparison of Classification Methods for Dataset 1

Fig. 4 and Table I show the single-trial classification of miniature-aVEP-speller was conducted using LDA, BLDA, SWLDA, SKLDA, STDA, xDAWN, EEGNet and DCPM, respectively, with different numbers of training samples. For each subject, 30 to 360 samples with a step of 30 samples (50% left and 50% right aVEPs) were randomly selected for training the classifiers while remaining 360 samples (50% left and 50% right aVEPs) were selected for testing the trained classifiers. To ensure fairness, all classification methods used the same training and testing samples. This procedure repeated for 10 times, and the
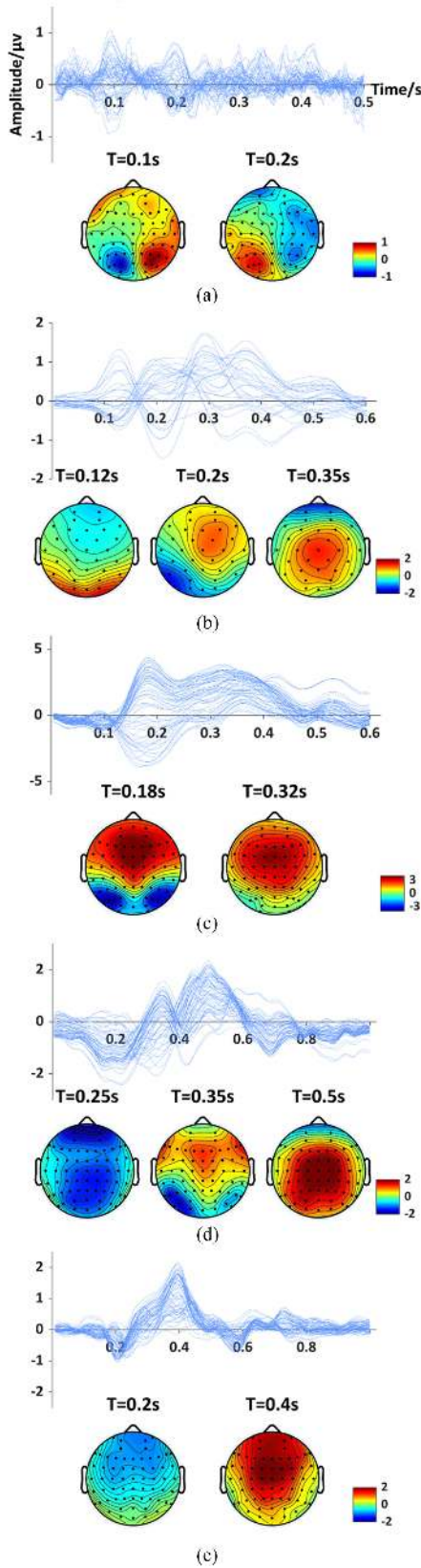
Fig. 3. Distributions of different ERP characteristics. (a) Left vs. right aVEPs for dataset 1. (b) Target vs. non-target ERPs for dataset 2. (c) Target vs. non-target ERPs for dataset 3. (d) Target RSVPs vs. non-target signals for dataset 4. (e) Target mVEPs vs. non-target signals for dataset 5.
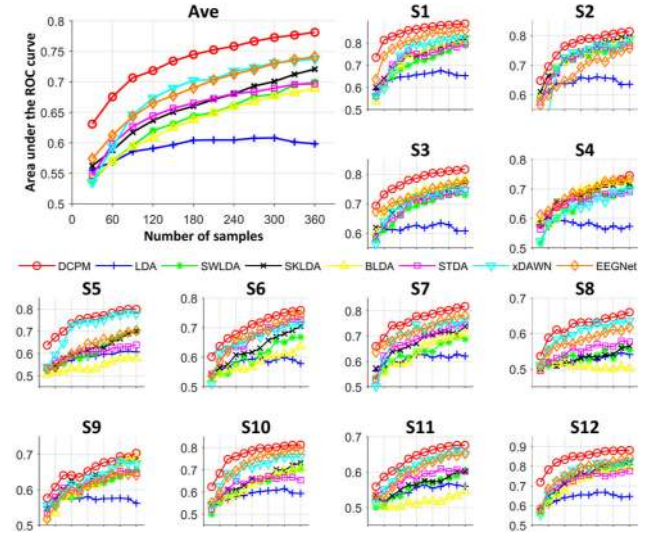

Fig. 4. Comparison of AUCs for dataset 1 under different numbers of training samples.

TABLE I
STATISTICAL ANALYSES BETWEEN DCPM AND THE OTHER
METHODS FOR DATASET 1

Method Comparison	Number of Training Samples											
	30	60	90	120	150	180	210	240	270	300	330	360
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SKLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. BLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. STDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. xDAWN	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. EEGNet	†	‡	‡	†	‡	‡	‡	‡	‡	‡	‡	‡

Note: ~nonsignificant, *p < 0.05, #p < 0.01, †p < 0.005, ‡p < 0.001

average area under the receiver operating characteristic (ROC) curves (AUCs) was computed. The AUC was adopted to evaluate the classification performance for all methods.

As shown in Fig. 4, the proposed DCPM method achieved higher AUCs than the other methods for across different numbers of training samples. DCPM possessed a stable advantage over the other methods for all numbers of training samples, while xDAWN and EEGNet performed better than other traditional classification methods. We used paired-samples T test to get the significances for all datasets. Individual analyses showed that DCPM achieved statistically significantly higher AUCs than other methods (see Table I).

C. Comparison of Classification Methods for Dataset 2

The single-trial classification of the P300-speller in EPFL dataset was conducted using LDA, BLDA, SWLDA, SKLDA, STDA, xDAWN, EEGNet and DCPM, respectively. 30 to 540 samples with a step of 30 samples (50% target and 50% non-target trials) were randomly selected for training classifiers while other 540 samples were selected for testing them. The classification procedure was as the same as used for Dataset 1 and the results are shown in Fig. 5. It's found that DCPM outperformed the other methods, and BLDA was the second-best
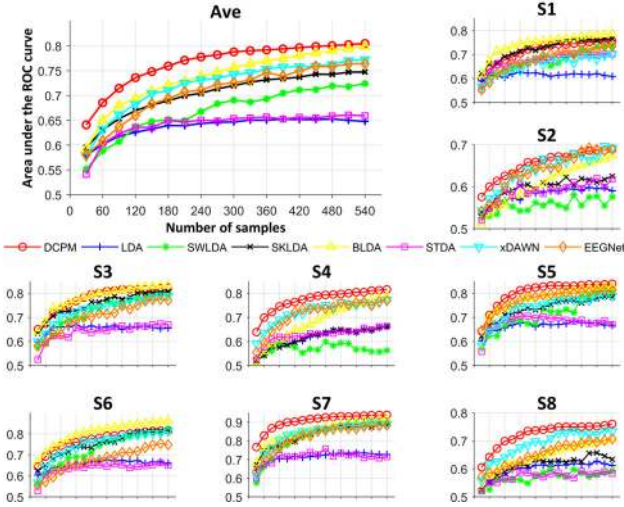
Fig. 5. Comparison of AUCs for dataset 2 under different number of training samples.

Fig. 6. Comparison of AUCs for dataset 3 under different number of training samples.

TABLE II
STATISTICAL ANALYSES BETWEEN DCPM AND THE OTHER
METHODS FOR DATASET 2

Method Comparison	Number of Training Samples								
	30	60	90	120	150	180	210	240	270
DCPM vs. LDA	†	†	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	†	‡	‡	‡	‡	‡	‡	†	†
DCPM vs. SKLDA	*	*	*	*	*	*	#	†	#
DCPM vs. BLDA	*	~	~	~	~	~	~	~	~
DCPM vs. STDA	‡	‡	†	‡	‡	‡	‡	‡	‡
DCPM vs. xDAWN	†	‡	‡	‡	‡	‡	‡	‡	†
DCPM vs. EEGNet	†	‡	‡	‡	†	‡	‡	‡	‡
	300	330	360	390	420	450	480	510	540
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	†	†	†	#	*	*	*	*	*
DCPM vs. SKLDA	#	#	*	*	*	*	*	*	*
DCPM vs. BLDA	~	~	~	~	~	~	~	~	~
DCPM vs. STDA	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. xDAWN	‡	†	†	#	‡	‡	†	#	†
DCPM vs. EEGNet	‡	‡	†	‡	†	†	†	†	†

Note: $\sim$ nonsignificant, $^*p < 0.05$, $^\#p < 0.01$, $^\dagger p < 0.005$, $^\ddagger p < 0.001$

TABLE III
STATISTICAL ANALYSES BETWEEN DCPM AND THE OTHER
METHODS FOR DATASET 3

Method Comparison	Number of Training Samples											
	30	60	90	120	150	180	210	240	270	300	330	360
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡	#	#	#	#
DCPM vs. SKLDA	‡	‡	‡	‡	†	*	*	~	~	~	~	~
DCPM vs. BLDA	†	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. STDA	‡	†	*	*	~	~	~	~	~	~	*	*
DCPM vs. xDAWN	†	*	*	~	~	~	~	~	~	~	~	~
DCPM vs. EEGNet	‡	†	†	†	#	~	~	~	~	~	~	~

Note: $\sim$ nonsignificant, $^*p < 0.05$, $^\#p < 0.01$, $^\dagger p < 0.005$, $^\ddagger p < 0.001$

randomly selected to test the classifiers. The classification procedure was as the same as addressed for Dataset 1. Fig. 6 and Table III show the results of an AUC analysis. The advantage of DCPM was more evident for different fewer training samples. Specifically, when using only 30 training samples, DCPM achieved 0.72 that was 0.07 higher than that of the second-best method, i.e., EEGNet. However, the AUC advantage reduced and even disappeared with the growth of training samples. For STDA, SKLDA, xDAWN and EEGNet, there was no significant difference between DCPM and them in different conditions of training samples, nevertheless, DCPM had higher AUCs than other methods and had significant differences with 30, 60 and 90 training samples. Therefore, DCPM had more advantages over the other methods with few training samples.

E. Comparison of Classification Methods for Dataset 4

Fig. 7 shows the single-trial classification of RSVP-speller was conducted using LDA, BLDA, SWLDA, SKLDA, STDA, xDAWN, EEGNet and DCPM, respectively, with different numbers of training samples. The classification procedure was as the same as described for Dataset 1. Here, 30 to 480 training samples were selected from the offline data (50% target trials),

algorithm for this dataset. From the individual performance, DCPM had a higher accuracy for most subjects except subjects 1 and 6 where BLDA was the best. Table II shows the statistical analyses between DCPM and the other methods. It's found that the advantage of DCPM over LDA, SWLDA, SKLDA, STDA, xDAWN and EEGNet achieved significant level. However, although DCPM had slightly higher AUCs than BLDA for the overall performance, no significant difference could be found between them except with 30 training samples, which might be due to the small number of subjects in this dataset.

D. Comparison of Classification Methods for Dataset 3

For each subject in Dataset 3, 30 to 360 samples with a step of 30 samples (50% are target responses) were randomly selected from the data set to train the classifiers, while the remaining 180 target samples and additional 180 non-target samples were
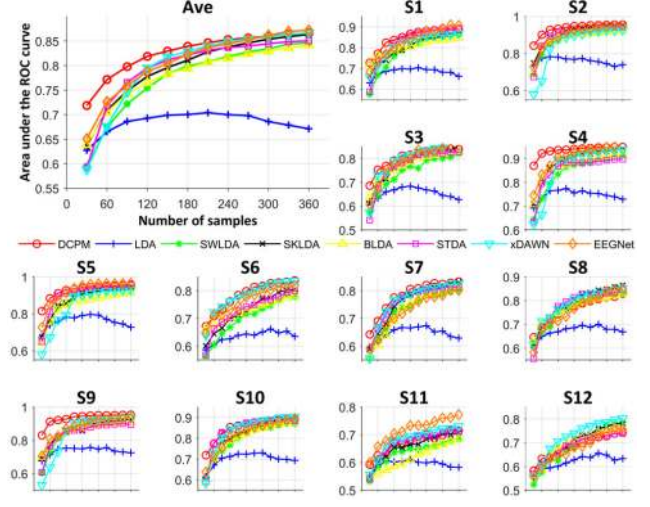
Fig. 7. Comparison of AUCs for dataset 4 under different numbers of training samples.
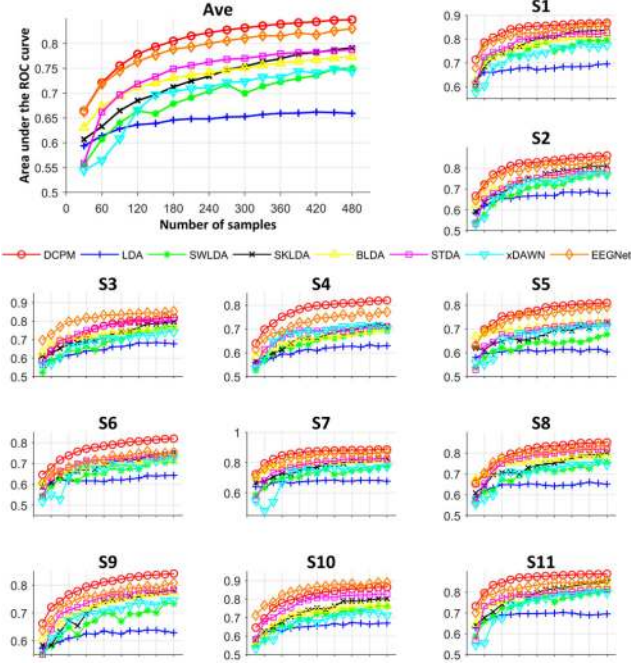

Fig. 8. Comparison of AUCs for dataset 5 under different numbers of training samples.

TABLE IV
STATISTICAL ANALYSES BETWEEN DCPM AND THE OTHER METHODS FOR DATASET4

Method Comparison	Number of Training Samples							
	30	60	90	120	150	180	210	240
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SKLDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. BLDA	*	†	‡	‡	‡	‡	‡	‡
DCPM vs. STDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. xDAWN	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. EEGNet	~	~	~	~	~	~	~	~
	270	300	330	360	390	420	450	480
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SKLDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. BLDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. STDA	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. xDAWN	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. EEGNet	~	~	~	*	*	*	*	~

Note: ~nonsignificant, *p < 0.05, #p < 0.01, †p < 0.005, ‡p < 0.001

TABLE V
STATISTICAL ANALYSES BETWEEN DCPM AND THE OTHER METHODS FOR DATASET 5

Method Comparison	Number of Training Samples											
	30	60	90	120	150	180	210	240	270	300	330	360
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. SKLDA	†	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. BLDA	†	‡	†	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. STDA	‡	‡	†	†	†	†	†	†	†	‡	†	‡
DCPM vs. xDAWN	†	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡
DCPM vs. EEGNet	†	‡	‡	‡	‡	‡	‡	†	†	†	‡	‡
	390	420	450	480	510	540	570	600	630	660	690	
DCPM vs. LDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. SWLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. SKLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. BLDA	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. STDA	†	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. xDAWN	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	‡	
DCPM vs. EEGNet	†	‡	‡	‡	†	†	†	†	†	‡	†	

Note: ~nonsignificant, *p < 0.05, #p < 0.01, †p < 0.005, ‡p < 0.001

while 480 testing samples were selected from the online data. It was obviously that DCPM had better performance than the other classification methods for most conditions of training samples. As shown in Table IV, the superiority of DCPM to the other methods was statistically significant. There was statistically significant difference between DCPM and EEGNet only with 360, 390, 420 and 450 training samples, although the AUC of DCPM was higher than EEGNet all along. It should be noted that EEGNet was firstly applied for recognizing P300 in the RSVP-speller [22], and the parameters about the networks maybe more applicable to RSVP-speller.
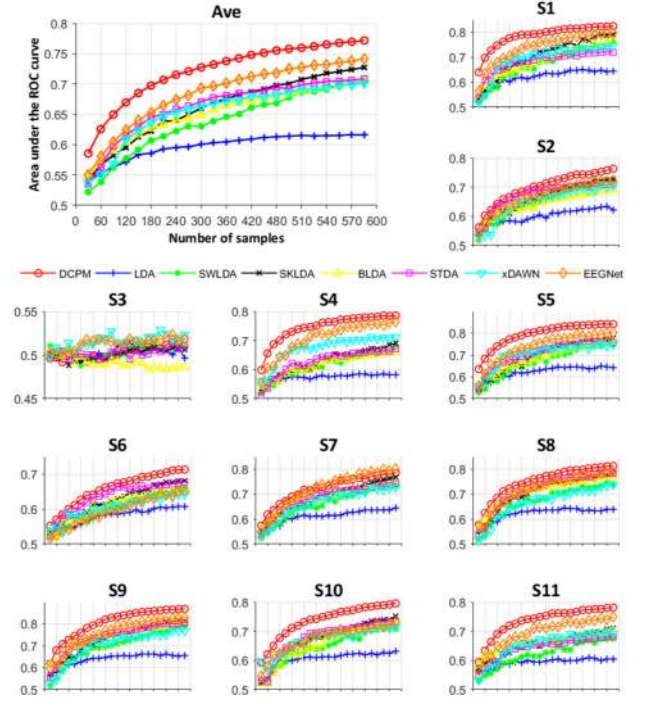
F. Comparison of Classification Methods for Dataset 5

Fig. 8 shows the single-trial classification of motion center speller was conducted using LDA, BLDA, SWLDA, SKLDA, STDA, xDAWN, EEGNet and DCPM, respectively, with different numbers of training samples. The classification procedure was as the same as described for Dataset 1. Here, the training samples were selected from the offline data, while 690 testing samples were selected from the online data. DCPM possessed an apparent advantage over the other methods for all numbers of training samples, which could achieve around 5.9% higher AUC than the second-best one, EEGNet. Table VI demonstrated that

TABLE VI
AVERAGE AUCs FOR LDA AND RANKINGS OF CLASSIFICATION METHODS

Dataset		1	2	3	4	5	Overall
Baseline AUC		0.59	0.64	0.68	0.64	0.60	0.63
Gain (%) and rankings	DCPM	23.75 / 1	20.08 / 1	20.85 / 1	24.75 / 1	20.88 / 1	22.06 / 1
	LDA	0 / 8	0 / 8	0 / 8	0 / 8	0 / 8	0 / 8
	SWLDA	7.50 / 6	4.92 / 6	13.36 / 7	6.79 / 7	6.80 / 7	7.87 / 7
	SKLDA	10.76 / 4	10.35 / 5	16.53 / 4	12.74 / 5	10.83 / 4	12.24 / 4
	BLDA	6.63 / 7	15.69 / 2	14.30 / 6	13.69 / 4	9.28 / 6	11.92 / 5
	STDA	10.25 / 5	0.54 / 7	16.39 / 5	15.03 / 3	11.08 / 3	10.66 / 6
	xDAWN	14.89 / 3	13.07 / 3	16.74 / 3	7.50 / 6	9.62 / 5	12.37 / 3
	EEGNet	15.15 / 2	10.98 / 4	18.01 / 2	22.06 / 2	14.66 / 2	16.17 / 2

performance of DCPM was significantly better than the other methods.

G. Overall Comparison for All Datasets

This study conducted an overall comparison of these classification methods for all datasets to have a comprehensive evaluation of their performance. Table VI shows the grand average AUCs of LDA across subjects and the training sample numbers for each dataset, which were regarded as the baseline. The performance improvements are described by percentage changes compared to the AUC gotten by the default LDA method. All classification methods were ranked according to the gain of AUC for all datasets. It's obvious that DCPM performed the best among all the datasets, and the total improvement of DCPM is nearly 6% higher than the total improvement of the second method, EEGNet. DCPM also nearly 10% higher than the third method, xDAWN, in terms of the total improvement.

IV. DISCUSSION

The performance of ERP-based BCIs depends heavily on the classification algorithm. To find out the most powerful algorithm for ERP classification, researchers have conducted a few comparison studies in the past. Krusienski et al. compared five methods for classifying P300 s, including Pearson's correlation method (PCM), LDA, SWLDA, linear support vector machine (LSVM) and Gaussian kernel support vector machine (GSVM), and proved that SWLDA and LDA performed better than other methods [30]. Contrary to Krusienski's findings, Aloise et al. compared these five methods for a gaze-independent P300-BCI, and found no statistically significant differences between the considered classifier's performance [33]. Later, Blankertz et al. proposed SKLDA for single-trial classification of ERP components and demonstrated its superiority over LDA and SWLDA [20]. In 2013, Zhang et al. developed STDA for ERP-BCIs and compared it with LDA, SWLDA and SKLDA. The results showed STDA performed best among all the methods. Lawhern et al. introduced EEGNet for the P300 dataset and it performed well in both within-subject classification and cross-subject classification [22]. It's interesting to find that different studies draw different conclusions, which might indicate that the best classification algorithm varied across ERP datasets. As we know, the ERP characteristics are sensitive to the BCI

paradigm and the experimental environment. Changes in experimental parameters would result in different ERP patterns. For example, the five datasets used in this study showed quite different ERP profiles in both temporal and spatial domains. For that reason, it's difficult for researchers to draw a consistent conclusion with their selected datasets. It also indicates that traditional algorithms cannot adapt to the ERP diversities in different datasets. Therefore, it's imperative to develop a robust classification algorithm that can deal with a wide range of ERP patterns.

To address the above problem, this study used five different ERP datasets collected in different paradigms and from different labs to test the performance of DCPM and seven other traditional classification methods. As a result, DCPM performed the best for all datasets. It demonstrated that DCPM had robust performance and was adaptable to the ERP patterns in different datasets especially with limited training samples. As to the other algorithms, their performance varied with datasets. Specifically, EEGNet had stable and good performance in all datasets and overall ranked the second. xDAWN ranked in the third place for the overall performance. In particular, it got the third place for dataset 1, 2 and 3 but had a poor performance in dataset 4 and 5. xDAWN will perform better with more training samples [19]. SKLDA and BLDA ranked the fourth and fifth for the overall performance, respectively. Notably, BLDA was in the second place for datasets 2 although it performed poorly for other datasets. STDA ranked sixth by getting third place in dataset 4 and 5. SWLDA and LDA were in the seventh and eighth place, respectively, for the overall performance. LDA was said to be stable and had a low complexity as linear discriminant analysis method [29]. SWLDA, SKLDA, and BLDA were regularized versions of LDA and all performed better than it. The results indicated that the traditional algorithms might have good performance for a certain dataset, but could hardly adapt to the ERP diversities across datasets.

ERPs were too weak and variable to be extracted stably. Traditional methods always collected multiple trials and carried out direct classification using time domain waveforms. As one of the most classical algorithm, LDA could achieve a satisfactory classification result with enough training samples. In previous, researchers paid most efforts on the balance between systematic error and bias, and developed a lot of advanced versions of LDA. Blankertz et al. indicated that SKLDA was effective with

insufficient training samples and got a target/non-target P300 binary validation error of 30% with 50 training samples [20]. EEGNet was a compact CNN method which could be trained with limited data. The AUC of single-trial ERP classification was about 0.85 using more than 250 training samples [22]. However, really few studies focused on ERP feature extraction and enhancement. xDAWN was proposed to enhance the ERP features by maximizing the signal-to-signal-plus-noise-ratio, and it had shown to be effective for the detection of P300 s using more than 500 training samples [19], [34]. The DSP filter could enlarge the difference between two patterns and the DCPM could enhance the SNR of ERPs [15], furthermore, the process of canonical pattern matching could improve over-fit effect. As a result, DCPM achieved an average of about 7.4% higher than the second best method in AUC across all the datasets with only 30 training samples. Meanwhile, all the methods in this study were used in the online BCI system. The computing time of each method could guarantee the real-time and precision of the system. Therefore, the outperformance of DCPM with small training set may contribute to reduce the BCIs' calibration time.

V. CONCLUSION

This study compared DCPM with seven other classification methods, i.e., LDA, SWLDA, BLDA, SKLDA, STDA, xDAWN and EEGNet on addressing different ERP patterns from five different datasets. The single-trial classification AUC was used to estimate their performance. As a result, DCPM outperformed the other methods for all datasets, while the other algorithms had a variation of performance for different datasets. The study results demonstrate that DCPM is a robust classification algorithm that has outstanding performance in a wide range of ERP patterns.

ACKNOWLEDGMENTS

REFERENCES

[1] J. R. Wolpaw *et al.*, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002.
[2] M. A. Lebedev *et al.*, "Brain-machine interfaces: Past, present and future," *Trends Neurosci.*, vol. 29, no. 9, pp. 536–546, Sep. 2006.
[3] H. Cecotti, "Spelling with non-invasive brain-computer interfaces–current and future trends," *J. Physiol. Paris*, vol. 105, no. 1/3, pp. 106–114, Jan.–Jun. 2011.
[4] J. R. Wolpaw *et al.*, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.
[5] Y. C. Lee *et al.*, "A visual attention monitor based on steady-state visual evoked potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 399–408, Mar 2016.
[6] C. L. Baldwin *et al.*, "Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48–56, Jan. 2012.
[7] M. Nakanishi *et al.*, "Detecting glaucoma with a portable Brain-Computer interface for objective assessment of visual function loss," *Jama Ophthalmol.*, vol. 135, no. 6, pp. 550–557, Jun. 2017.

[8] G. Townsend *et al.*, "Pushing the P300-based brain-computer interface beyond 100 bpm: Extending performance guided constraints into the temporal domain," *J. Neural Eng.*, vol. 13, no. 2, Apr. 2016, Art. no. 026024.
[9] E. Yin *et al.*, "A hybrid brain-computer interface based on the fusion of P300 and SSVEP scores," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 4, pp. 693–701, Jul. 2015.
[10] A. Pinegger *et al.*, "Control or non-control state: That is the question! An asynchronous visual P300-based BCI approach," *J. Neural Eng.*, vol. 12, no. 1, Feb. 2015, Art. no. 014001.
[11] Y. Zhang *et al.*, "A novel BCI based on ERP components sensitive to configural processing of human faces," *J. Neural Eng.*, vol. 9, no. 2, Apr. 2012, Art. no. 026018.
[12] B. Hong *et al.*, "N200-speller using motion-onset visual response," *Clin. Neurophysiol.*, vol. 120, no. 9, pp. 1658–1666, Sep. 2009.
[13] M. Li *et al.*, "Increasing N200 potentials via visual stimulus depicting humanoid robot behavior," *Int. J. Neural Syst.*, vol. 26, no. 1, Feb. 2016, Art. no. 550039.
[14] F. Guo *et al.*, "A brain-computer interface using motion-onset visual evoked potential," *J. Neural Eng.*, vol. 5, no. 4, pp. 477–485, Dec. 2008.
[15] M. Xu *et al.*, "A brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1166–1175, May 2018.
[16] M. Xu *et al.*, "Use of a steady-state baseline to address evoked vs. oscillation models of visual evoked potential origin," *NeuroImage*, vol. 134, pp. 204–212, Jul. 2016.
[17] D. J. Krusienski *et al.*, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008.
[18] U. Hoffmann *et al.*, "An efficient P300-based brain-computer interface for disabled subjects," *J. Neural Eng.*, vol. 167, no. 1, pp. 115–125, 2008.
[19] B. Rivet *et al.*, "xDAWN algorithm to enhance evoked potentials: Application to brain-computer interface," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2035–2043, Aug. 2009.
[20] B. Blankertz *et al.*, "Single-trial analysis and classification of ERP components–a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011.
[21] Y. Zhang *et al.*, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar 2013.
[22] V. J. Lawhern *et al.*, "EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
[23] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147–1154, Mar. 2008.
[24] M. van Vliet *et al.*, "Single-trial ERP component analysis using a spatiotemporal LCMV beamformer," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 55–66, Jan. 2016.
[25] W. L. Lee *et al.*, "Single-trial event-related potential extraction through one-unit ICA-with-reference," *J. Neural Eng.*, vol. 13, no. 6, Dec. 2016, Art. no. 066010.
[26] R. E. Alcaide-Aguirre *et al.*, "Novel hold-release functionality in a P300 brain-computer interface," *J. Neural Eng.*, vol. 11, no. 6, Dec. 2014, Art. no. 066010.
[27] L. Acqualagna *et al.*, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clin. Neurophysiol.*, vol. 124, no. 5, pp. 901–908, May 2013.
[28] S. Schaeff *et al.*, "Exploring motion VEPs for gaze-independent communication," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 045006.
[29] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.*, vol. 4, no. 2, pp. R1–R13, Jun. 2007.
[30] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 Speller," *J. Neural Eng.*, vol. 3, no. 4, pp. 299–305, Dec. 2006.
[31] X. Lei, P. Yang, and D. Yao, "An empirical Bayesian framework for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 6, pp. 521–529, Dec. 2009.
[32] J. Schafer *et al.*, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statistical Appl. Genetics Mol. Biol.*, vol. 4, 2005, Art. no. 32.
[33] F. Aloise *et al.*, "A comparison of classification techniques for a gaze-independent P300-based brain-computer interface," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 045012.
[34] H. Woehrle *et al.*, "An adaptive spatial filter for User-Independent single trial detection of event-related potentials," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1696–1705, Jul. 2015.