IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 15, 2020, accepted June 14, 2020, date of publication June 17, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003056

Optimal Channel Selection Using Correlation Coefficient for CSP Based EEG Classification

YONGKOO PARK AND WONZOO CHUNG, (Senior Member, IEEE)
Division of Computer and Communications Engineering, Korea University, Seoul 02841, South Korea

Corresponding author: Wonzoo Chung (wchung@korea.ac.kr)

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00432, Development of non-invasive integrated BCI SW platform to control home appliances and external devices by user's thought via AR/VR interface) and Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00451, Development of BCI based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-00079, Artificial Intelligence Graduate School Program (Korea University)).

ABSTRACT In this paper, we present an optimal channel selection method to improve common spatial pattern (CSP) related features for motor imagery (MI) classification. In contrast to existing channel selection methods, in which channels significantly contributing to the classification in terms of the signal power are selected, *distinctive* channels in terms of correlation coefficient values are selected in the proposed method. The *distinctiveness* of a channel is quantified by the number of channels with which it yields large difference in correlation coefficient values for binary motor imagery (MI) tasks, rather than by the largeness of the difference itself. For each *distinctive* channel, a group of channels is formed by gathering strongly correlated channels and the Fisher score is computed using the feature output, based on the filter-bank CSP (FBCSP) exclusively applied to the channel group. Finally, the channel group with the highest Fisher score is chosen as the selected channels. The proposed method selects the fewest channels on average and outperforms existing channel selection approaches. The simulation results confirm performance improvement for two publicly available BCI datasets, BCI competition III dataset IVa and BCI competition IV dataset I, in comparison with existing methods.

INDEX TERMS Electroencephalography (EEG), brain-computer interfaces (BCIs), correlation coefficient, common spatial pattern (CSP), channel selection.

I. INTRODUCTION

Brain-computer interfaces (BCIs) provide non-muscular communication between humans and computer using brain signals. Electroencephalogram (EEG)-based BCIs, which directly translate the intent reflected by EEG signals into a control command, have been used due to its high temporal resolution and non-invasiveness [1], [2]. Motor imagery (MI) is an area of active research in EEG-based BCIs, as the power of motor-relevant cortex-related EEG signals is decreased or increased during imaging of body movements; these changes are known as event-related desynchronization (ERD) or event-related synchronization (ERS) [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan-Pin Lin.

Especially, the common spatial pattern (CSP) approach [1], [4]–[6] has been used successfully in MI classification by extracting ERD/ERS-related features. Recently, the various extensions of the CSP which overcome a frequency band dependency problem have been proposed such as filter-bank CSP (FBCSP) [7], sub-band regularized CSP (SBRCSP) [8], filter-bank regularized CSP (FBRCSP) [9], sparse filter band common spatial pattern (SFBCSP) [10], and filter band combined with Tikhonov regularization CSP (FB-TRCSP) [11]. Moreover, the temporally constrained sparse group spatial pattern (TSGCSP) [12] and sparse group representation model (SGRM) [13] is proposed to overcome a time period dependency and subject-dependency problem, respectively and shows improved performance for MI-classification.

In order to improve performance, EEG-based BCIs use EEG signals from multiple sites of scalp, thereby increasing the number of EEG channels [14]. However, the use of a large number of EEG channels does not always guarantee performance improvement [15]. Redundant task-irrelevant channels tend to introduce undesirable interference and boost noise level, as the activated brain regions for the same intent are relatively small and differ among subjects [16]. Therefore, several MI-relevant channel selection algorithms have been proposed [17]–[21].

The sparse common spatial pattern (SCSP) approach described in [17] selects the MI-relevant channels that correspond to the high sparse CSP filter coefficient, which is extracted by l_1/l_2 norm regularization. The CSP-rank for multiple frequency band (CSP-R-MF) approach described in [18] selects the MI-relevant channels for each frequency band and extracts features from selected channels of each frequency band using the least absolute shrinkage and selection operator (LASSO) algorithm [22]. The frequency-optimized local region CSP (LRFCSP) approach described in [19] selects MI-relevant channels using the variance ratio dispersion score (VRDS) and inter-class feature distance (ICFD) of small EEG channel groups and outperforms existing CSP-based channel selection MI classification algorithms. Recently, MI-relevant channel selection approaches have been proposed based on time domain parameters (TDPs) that utilize the frequency-band-insensitive characteristics of time domain signals [23]. TDP Fisher's discriminant analysis (TDP-FDA) described in [20] selects the MI-relevant channels with high Fisher ratios of TDPs. The feature compressing and channel ranking (FCCR) approach described in [21] attempts to reduce the TDP feature dimension by clustering and selects the MI-relevant channels based on a robust feature selection (RFS) algorithm [24].

These channel selection algorithms rely on temporal, spectral, or spatial EEG features, which have been used effectively for MI classification related to regionally well-separated ERD/ERS patterns. However, these features may exhibit limitations in terms of classifying subtle and mixed ERD/ERS pattern [25]. It is known that execution of even simple MI tasks require the participation of multiple brain regions that are mutually and subtly interconnected [26], [27]. Hence, features based on brain neural dynamic patterns (regarded as brain connectivity) have attracted considerable attention and may provide performance improvement for binary MI tasks. Recently, a correlation-based channel selection regularized CSP (CCS-RCSP) approach proposed in [28] selects the MI-related channels using correlation coefficient of EEG signal. CCS-RCSP performs the regularized-CSP [6] using the MI-related channels which are highly correlated with others and improves the classification performance. However, this approach still uses too many channels for classification.

In this paper, a novel MI-relevant channel selection method based on correlation coefficients that represent brain connectivity is proposed to improve MI classification performance. For a given channel, the differences in average correlation coefficients between two MI tasks, relative to other channels, is computed. The differences are normalized using the t-statistics and the number of channels with which a given channel has a p-value below the significant level is counted and referred to as the MI-score. The channels that have higher MI-scores than their average are selected as *distinctive* channels. For each *distinctive* channel, we construct a group of supporting channels that are strongly correlated with the channel (i.e., channels with which the given *distinctive* channel yields a correlation coefficient higher than a predetermined threshold). To identify the best supporting channel group for MI classification, FBCSP is performed to calculate the Fisher score, (i,e., the Fisher ratio of the resulting FBCSP features). Finally, we choose the supporting channel group with the highest Fisher score to determine the optimal channel set for MI classification. The FBCSP features of the optimal channel set are used as input for the support vector machine (SVM) classifier [29]. The performance of the proposed method is evaluated by simulation using the BCI competition III dataset IVa and BCI competition IV dataset I. The proposed algorithm selects the smallest number of channels (on average), but yields the best performance, in comparison with existing MI classification algorithms.

The paper is organized as follows. Section II explains the proposed channel selection method. Section III provides the data and experiments. Section IV analyzes the results. Finally, conclusion for this paper is made in Section V.

II. METHOD

A. SYSTEM MODEL

Let us consider the K channel binary MI EEG signals. The sampled EEG signal at the k-th channel is denoted as $x^{(k)}(n)$ for $k = 1, 2, \ldots, K$ and $n = 1, 2, \ldots, N$, where N is the number of samples per channel. We assume that M trials of training EEG signals are available and the i-th trial EEG signal for channel k is denoted as $\mathbf{x}_i^{(k)} = [x_i^{(k)}(1), x_i^{(k)}(2), \ldots, x_i^{(k)}(N)]^T$. Each trial of EEG signal consists of N time samples. As we consider binary MI classification, each trial belongs to index set I_1 or I_2 ($I_1 \cup I_2 = \{1, 2, \ldots, M\}$) corresponding to each MI task and $|I_c|$ denotes the number of training trials for the task $c (\in \{1, 2\})$.

In this paper, we consider the correlation coefficient to identify *distinctive* channels. The (sample) correlation coefficient of the i-th trial EEG channel signal pair for the channel k and the channel p, denoted by $\rho_i^{(k,p)}$, is defined as

$$\rho_i^{(k,p)} := \frac{C(\mathbf{x}_i^{(k)}, \mathbf{x}_i^{(p)})}{\sqrt{P(\mathbf{x}_i^{(k)})}\sqrt{P(\mathbf{x}_i^{(p)})}}, \quad k, p = 1, 2, \ldots, K \tag{1}$$

where $C(\mathbf{x}_i^{(k)}, \mathbf{x}_i^{(p)})$ is the sample covariance and $P(\mathbf{x}_i^{(k)})$ is the sample variance, defined as the following:

$$C(\mathbf{x}_i^{(k)}, \mathbf{x}_i^{(p)}) := \sum_{n=1}^{N}\left(x_i^{(k)}(n) - \frac{1}{N}\sum_{n=1}^{N} x_i^{(k)}(n)\right)\left(x_i^{(p)}(n) - \frac{1}{N}\sum_{n=1}^{N} x_i^{(p)}(n)\right) \tag{2}$$

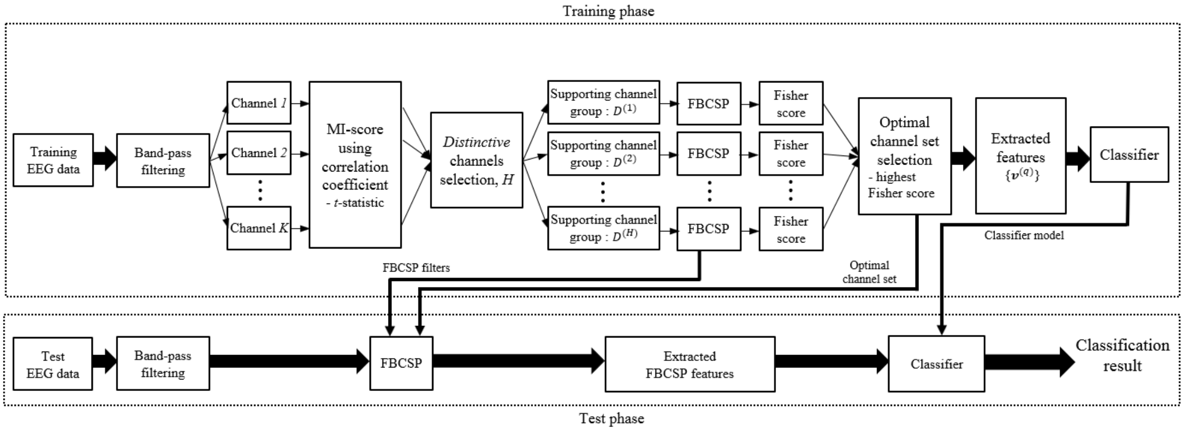

FIGURE 1. Process of the proposed method.

$$P(\mathbf{x}_i^{(k)}) := \sum_{n=1}^{N}\left(x_i^{(k)}(n) - \frac{1}{N}\sum_{n=1}^{N} x_i^{(k)}(n)\right)^2 \tag{3}$$

Figure 1 shows a block diagram of the proposed channel selection method. First, we select the distinctive channels based on correlation coefficient using t-statistic. Second, the strongly correlated supporting channel groups for each distinctive channel are constructed. Then, the Fisher score of FBCSP features for each supporting channel group is calculated to select the optimal channel set with the highest Fisher score. The FBCSP features of the optimal channel set is used as the final output features of the proposed method. In the following subsections, each step of the process is described in detail.

B. DISTINCTIVE CHANNEL SELECTION BASED ON THE CORRELATION COEFFICIENT

The correlation coefficient of the EEG channel pair represents brain connectivity and has been used for analysis of the relationship between two channel EEG signals [30]. The MI-relevant channel typically possesses a distinct signal property for different MI tasks (e.g., signal power). In terms of correlation coefficient, defining the *distinctiveness* of a channel with respect to different MI tasks is not straightforward, as the correlation coefficient cannot be computed for a single channel. Therefore, our measure of *distinctiveness* comprises the number of channels with which a channel has a significant difference in correlation coefficient with regard to MI tasks.

To statistically quantify the difference in correlation coefficient between two MI tasks for a number of different trials, we use the t-statistic [31]. The t-statistic of the correlation coefficient for channels k and p, denoted by $T^{(k,p)}$, given I trials is expressed as [31]:

$$T^{(k,p)} = \frac{\bar{\rho}_1^{(k,p)} - \bar{\rho}_2^{(k,p)}}{\sqrt{\frac{\sigma_1^2}{|I_1|} + \frac{\sigma_2^2}{|I_2|}}} \tag{4}$$

where

$$\bar{\rho}_c^{(k,p)} = \frac{1}{|I_c|}\sum_{i\in I_c} \rho_i^{(k,p)} \tag{5}$$

$$\sigma_c^2 = \frac{1}{|I_c - 1|}\left(\rho_i^{(k,p)} - \bar{\rho}_c^{(k,p)}\right)^2 \tag{6}$$

and $|I_c|$ denotes the number of training trials for the task $c(\in \{1,2\})$. It has known that $T^{(k,p)}$ has the Student's t-distribution with degree of freedom $|I_1| + |I_2| - 2$; the probability of $T^{(k,p)}$ occurring by chance from the given trials is designated as p-value and given by $P(X \geq |T^{(k,p)}|)$, where X is a random variable with Student's t-distribution of freedom $|I_1|+|I_2|-2$ ([32]). Let $P^{(k,p)}$ denote the p-value of t-statistic for channel k and p. To measure the *distinctiveness* of a channel k, we define the MI-score, denoted by $S^{(k)}$, which is the number of channels in which $P^{(k,p)}$ is below a given significant level (P_{thr}):

$$S^{(k)} = \left|\left\{p \in \{1,2,\ldots,K\} \,\middle|\, P^{(k,p)} < P_{thr}\right\}\right| \tag{7}$$

The significant level P_{thr} is typically set to 0.05; in this paper, it is set by cross-validation through training data. After MI-scores are calculated for each channel, we select the *distinctive* channels that have MI-score higher than the average MI-score and define an index set of *distinctive* channels as H:

$$H = \left\{k \in \{1,2,\ldots,K\} \,\middle|\, S^{(k)} > \frac{1}{|S^{(k)}|}\sum_{k=1}^{K} S^{(k)}\right\} \tag{8}$$

C. SUPPORTING CHANNEL GROUP OF A DISTINCTIVE CHANNEL

For each *distinctive* channel in set H, we construct a group of channels that are strongly correlated with the *distinctive* channel. Specifically, for a *distinctive* channel $h \in H$, the supporting channel group of h, denoted by $D^{(h)}$, is formed by the channels with which the channel h has an average correlation coefficient higher than a predetermined threshold, ρ_{thr}:

$$D^{(h)} = \left\{p \in H \,\middle|\, \bar{\rho}_1^{(h,p)} \geq \rho_{thr} \text{ and } \bar{\rho}_2^{(h,p)} \geq \rho_{thr}\right\} \tag{9}$$

In this paper, ρ_{thr} is set by cross-validation through training data. Note that some of the *distinctive* channels may share the same supporting channel group; consequently, the total number of supporting groups can be fewer than the number of *distinctive* channels H.

D. OPTIMAL CHANNEL SET SELECTION USING FISHER SCORE

Among the supporting channel groups, we select the best group for MI classification using the Fisher ratio [34], or Fisher score, of FBCSP features [7]. FBCSP approach extracts frequency optimized log-variance features of spatial filtered EEG signals with maximum variance difference between two MI tasks and consists of three stages: a filter bank of multiple bandpass filters, spatial filtering using the CSP algorithm, and frequency band selection based on mutual information [7]. In the first stage, a filter bank with M bandpass filters is used to generate M bandpass EEG signals for each supporting channel group. We denote the i-th trial of the m-th bandpass filtered EEG signals of $D^{(h)}$ as $X_i^{(h,m)} \in \mathbb{R}^{|D^{(h)}|\times N}$. The CSP algorithm is applied to each $X_i^{(h,m)}$ with two spatial filters to extract two CSP features, $\{v_{i,max}^{(h,m)}$ and $v_{i,min}^{(h,m)}\}$, where $v_{i,max}^{(h,m)}$ is the output power of the maximum CSP filter for $X_i^{(h,m)}$ and $v_{i,min}^{(h,m)}$ is the output power of the minimum CSP filter for $X_i^{(h,m)}$ ([19]). In the third stage, only the two best frequency bands for each supporting channel, denoted as $M_1^{(h)}$ and $M_2^{(h)}$, are selected using the mutual information based individual feature (MIBIF) algorithm [33]. The vector of four CSP features for the selected bands is called the FBCSP feature [19], and the FBCSP feature for the supporting group h and the i-th trial is given as

$$\boldsymbol{u}_i^{(h)} = \left[v_{i,max}^{(h,M_1^{(h)})}, v_{i,min}^{(h,M_1^{(h)})}, v_{i,max}^{(h,M_2^{(h)})}, v_{i,min}^{(h,M_2^{(h)})} \right] \tag{10}$$

After the FBCSP feature vectors are extracted for each supporting group, we use the Fisher score to measure the distinguishable power of extracted features [34]. The Fisher score of FBCSP features, denoted as $Z^{(h)}$, is given by:

$$Z^{(h)} = \frac{\| \frac{1}{|I_1|}\sum_{i\in I_1} \boldsymbol{u}_i^{(h)} - \frac{1}{|I_2|}\sum_{i\in I_2} \boldsymbol{u}_i^{(h)} \|}{\frac{1}{2}\sum_{c=1}^{2} \frac{1}{|I_c|}\sum_{i\in I_c} \|\boldsymbol{u}_i^{(h)} - \frac{1}{|I_c|}\sum_{i\in I_c} \boldsymbol{u}_i^{(h)}\|}, \quad h \in H \tag{11}$$

where $\|\cdot\|$ denotes the ℓ_2 norm. Finally, the supporting channel group with the highest Fisher score is selected as the optimal channel set for MI classification. The selected optimal channel set is denoted as $D^{(q)}$. The FBCSP feature of the $D^{(q)}$ is the final output feature of the proposed channel selection method and is used as the input value of the support vector machine (SVM) classifier [29].

III. EXPERIMENTAL STUDY

A. DATA DESCRIPTION

To evaluate the performance of the proposed method, we use two publicly available MI-EEG datasets: BCI competition III Dataset IVa [35] and BCI competition IV Dataset I [36]. The BCI competition III dataset IVa is widely used to evaluate binary MI classification performance with small training sets. EEG signals with 118 channels ($K = 118$) were recorded from five healthy subjects (‘*al*’, ‘*aa*’, ‘*av*’, ‘*aw*’, and ‘*ay*’), with a 100 Hz sampling rate. All subjects performed 280 trials, 140 per class.

TABLE 1. The number of trials for the training data and the test data for BCI Competition III dataset IVa.

Subject	Training data	Test data
al	224	56
aa	168	112
aw	84	196
av	56	224
ay	28	252

The BCI competition IV dataset I was recorded from seven subjects with 59 channels ($K = 59$) for binary MI tasks [36]. EEG data were recorded at a sampling rate of 1000Hz and bandpass filtered between 0.05Hz and 200Hz. The downsampled signal at 100Hz was used and each subject performed 200 trials (i.e., 100 trials per class).

B. DATA PROCESSING

For the experiments, the EEG signals from 0.5s to 3s after cue are used. The data are band-pass filtered using a filter bank consisted of the eight fourth-order Butterworth filter dividing the frequency range from 4 to 36Hz evenly with 4Hz for FBCSP. In the BCI competition III dataset IVa, 5 × 5 cross-validation is used to determine the P_{thr} and ρ_{thr} ($P_{thr} \in \{0.2, 0.15, 0.1, 0.05, 0.01\}$, $\rho_{thr} \in \{0.75, 0.8, 0.85, 0.9, 0.95\}$). The P_{thr} and ρ_{thr} are chosen as 0.05 and 0.9, respectively, shows the highest validation accuracy.

In the BCI competition IV dataset I, we use data from four healthy subjects (‘*a*’, ‘*b*’, ‘*f*’, and ‘*g*’) to evaluate performance. The parameters P_{thr} and ρ_{thr} are set by 5 × 5 cross-validation using training data for each subject ($P_{thr} \in \{0.2, 0.15, 0.1, 0.05, 0.01\}$, $\rho_{thr} \in \{0.75, 0.8, 0.85, 0.9, 0.95\}$). Since the statistical property of the subject ‘*b*’ is quite different from the rest; the correlation coefficient coefficients values are generally small and, hence, the differences are small too, we use different cross-validation parameter candidate sets for subject ‘*b*’ ($P_{thr} \in \{0.6, 0.5, 0.4, 0.3, 0.2\}$, $\rho_{thr} \in \{0.55, 0.6, 0.65, 0.7, 0.75\}$). After cross-validation, P_{thr} is chosen as 0.05 for subjects ‘*a*’, ‘*f*’, and ‘*g*’, and chosen as 0.4 for subject ‘*b*’. ρ_{thr} is chosen to 0.9, 0.6, 0.85, and 0.85 for subjects ‘*a*’, ‘*b*’, ‘*f*’, and ‘*g*’, respectively.

IV. RESULTS AND DISCUSSION

In this section, we present the experimental results for BCI competition III Dataset IVa and BCI competition IV Dataset I. For the BCI competition III dataset IVa, the performance measure is the number of correctly classified test trials divided by the total number of test trial data specified by BCI competition III. The number of trials for the training data and the test data of the five subjects are shown in Table 1. Note that the proposed method is evaluated using whole 118 channels. However, the conventional CSP-based methods are evaluated using only carefully chosen 18 channels as

TABLE 2. **Classification accuracies of the proposed method and CSP-based classification methods without channel selection for BCI Competition III dataset IVa.**

subject	CSP	FBCSP	SBRCSP	FBRCSP	Proposed method
al	91.07	94.64	**98.21**	94.64	**98.21**
aa	75.89	88.39	86.61	**91.07**	89.29
av	56.63	61.22	63.78	**75**	73.47
aw	66.51	80.22	89.05	76.78	**92.86**
ay	79.36	82.14	77.78	**93.65**	89.29
mean	73.89	81.28	83.09	86.23	**88.62**

presented in their original paper, since they perform better with the 18 channels than with 118 channels.

For BCI competition IV dataset I, 5×5 cross-validation is performed to evaluate performance. The trials are divided into five sets and one set is selected as test data while all others are selected as training data.

A. PERFORMANCE COMPARISON FOR BCI COMPETITION III DATASET IVa

1) PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH CSP-BASED CLASSIFICATION METHODS

In this experiment, we compared the performance of the proposed method with the performances of conventional CSP-based classification methods without channel selection, CSP [1], FBCSP [7], sub-band regularized CSP (SBRCSP) [8], and filter-bank regularized CSP (FBRCSP) [9]. Table 2 lists the classification accuracies of the proposed method and the conventional methods. Although the FBRCSP algorithm, which is the frequency optimized version of regularized CSP, outperforms the proposed method for some subjects, FBCSP with the proposed channel selection method shows the highest mean classification accuracy. The Figure 2 plots the distributions of the most significant two CSP features derived by FBRCSP and proposed method from subject '*aw*' of BCI Competition III dataset IVa. Clearly the features of the proposed method are more separable with respect to different tasks.

2) PERFORMANCE COMPARISON OF THE PROPOSED METHOD WITH CHANNEL SELECTION-BASED CLASSIFICATION METHODS

In this experiment, the performance of the proposed method is compared with the performances of existing channel selection methods, including filter bank-sparse CSP (FBSCSP) (a frequency optimized version of SCSP [17]), CSP-R-MF [18], LRFCSP [19], TDP-FDA [20], FCCR [21], and CCS-RCSP [28]. Table 3 shows the classification accuracies of each method for the BCI competition III dataset IVa. The numbers of selected channels are shown in parentheses. As shown in Table 3, our proposed method shows the best mean classification accuracy (88.62%) than the conventional ethods.

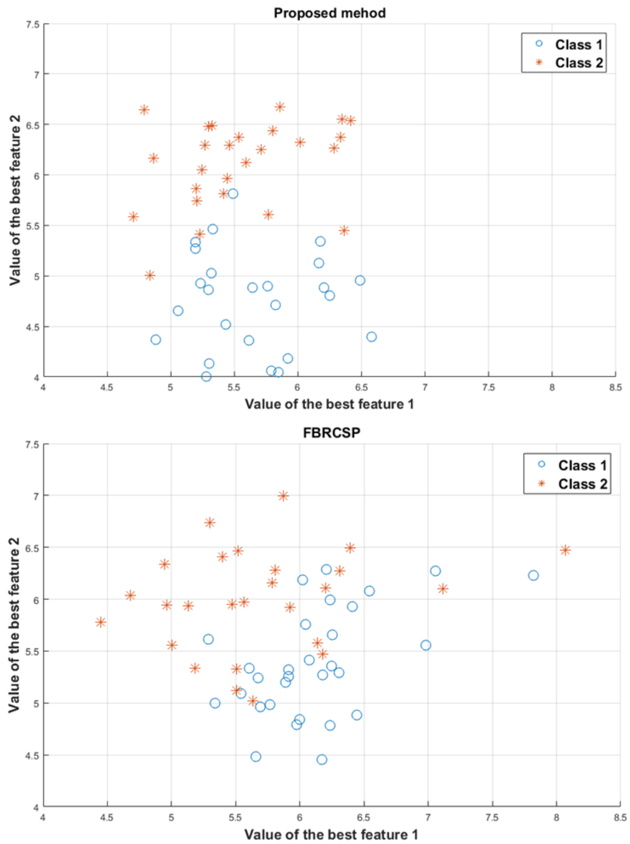

FIGURE 2. **Distributions of the most significant two features obtained by FBRCSP and proposed method from subject '*aw*' of BCI Competition III dataset IVa.**

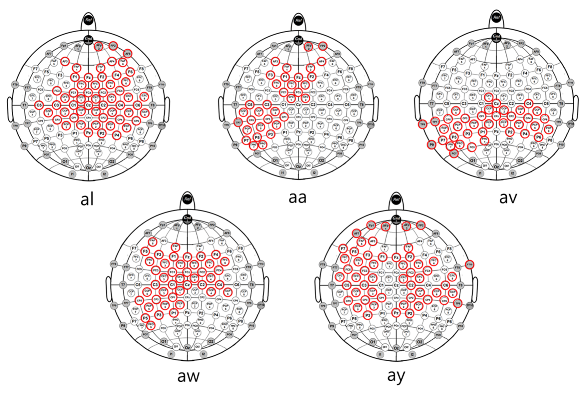

FIGURE 3. **Location of all *distinctive* channels for BCI Competition III dataset IVa.**

Table 3 also shows the number of selected channels for each subject. The average number of selected channels in the proposed method is 9, which is lower than the numbers in conventional methods. Our proposed method achieves the best classification performance using the smallest number of MI-relevant channels.

Similarly to the proposed methods, the CCS-RCSP method uses the correlation coefficients to select MI-related channels. The mean classification accuracy of CCS-RCSP tends to be

TABLE 3. Classification accuracies of the proposed method and channel selection-based classification methods for BCI Competition III dataset IVa.

subject	TDP-FDA	FBSCSP	CSP-R-MF	FCCR	CCS-RCSP	LRFCSP	Proposed method
al	87.50 (6)	96.42 (9)	96.42 (12)	**98.21** (10)	96.42(33)	96.42 (7)	**98.21** (7)
aa	67.86 (8)	83.93 (12)	82.14 (12)	78.57 (10)	83.03(42)	83.93 (22)	**89.29** (10)
av	61.22 (11)	69.39 (9)	72.14 (12)	72.45 (5)	70.91(52)	**74.49** (6)	73.47 (12)
aw	81.25 (10)	86.61 (12)	84.38 (12)	87.05 (15)	92.41(14)	88.84 (7)	**92.86** (9)
ay	92.06 (11)	90.87 (12)	**94.28** (12)	93.25 (9)	92.46(67)	89.29 (11)	89.29 (7)
mean	77.98 (9.2)	85.44 (10.8)	85.85 (12)	85.91 (9.8)	87.05(41.6)	86.59 (10.6)	**88.62** (9)

TABLE 4. Classification accuracies of the proposed method and channel selection based classification methods for BCI Competition IV dataset I.

subject	FBSCSP	TDP-FDA	CSP-R-MF	FCCR	CCS-RCSP	LRFCSP	Proposed method
a	81.5 (10)	82.5 (21)	81.5 (24)	83.5 (14)	85.5(46)	**87** (12.6)	86.5 (9.2)
b	62.5 (14)	69 (9)	63 (24)	**72.5** (9)	67.0(30)	69 (12.8)	69.5 (8.8)
f	77.5 (10)	79 (16)	79 (24)	81 (9)	79.5(10)	**87.5** (15.2)	**87.5** (8.2)
g	75 (18)	73.5 (19)	87.5 (24)	83.5 (18)	94.5(3)	92.5 (13.6)	**94** (6.7)
mean	74.13 (13)	76 (16.3)	77.75 (24)	80.13 (12.5)	81.6(22.2)	84 (13.6)	**84.4** (8.2)

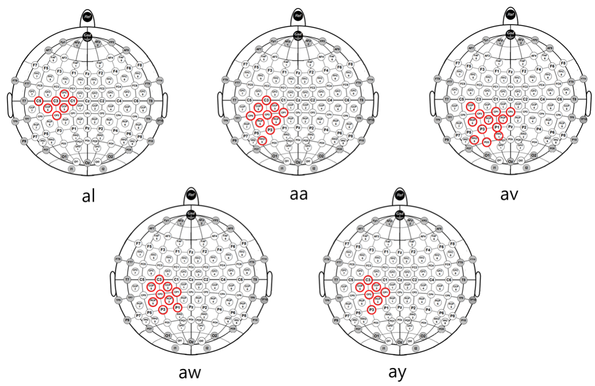

FIGURE 4. Location of optimally selected channels by the proposed method for BCI Competition III dataset IVa.

TABLE 5. Complexity of the proposed method and conventional methods.

Method	Complexity
CSP	$O(K^2(MN+K))$
FBSCSP	$O(K^2(MN+K))$
LRFCSP	$O(K^3(MN+K))$
CSP-R-MF	$O(K^2(MN+K))+O(P^2M)$
FCCR	$O(Q(KM)^3)$
Proposed method	$O((MK)^2)+O(H^3(MN+H))$

* K : the number of channels, M : the number of trials, N : the number of temporal samples per channel, P : the dimension of feature vector, Q : the number of iterations, H : the number of selected distinct channels

higher than conventional channel selection methods. This implies that the correlation coefficients are useful features for channel selection. Comparing CCS-RCSP with the proposed method, the proposed algorithm substantially outperforms CCS-RCSP. CCS-RCSP quantifies the distinctiveness of a channel using mean correlation coefficient values with other channels, while the proposed method quantifies the distinctiveness of a channel using the number of channels with which it yields large difference in correlation coefficient values between tasks. The proposed method exhibits higher mean classification accuracy with fewer channels.

Figure 3 shows the location of the *distinctive* channels, H. The sizes of H is 55, 34, 38, 42, and 54 for subjects '*al*', '*aa*', '*av*', '*aw*', and '*ay*', respectively. As shown in Figure 3, the locations of H varies among subjects, but are mostly distributed in the motor areas of the cerebral cortex. Figure 4 shows the finally selected supporting channel of the proposed channel selection method.

B. PERFORMANCE COMPARISON FOR BCI COMPETITION IV DATASET I

Table 4 compares the 5×5 classification accuracies of the proposed method and channel selection based classification methods, including FBSCSP [17], CSP-R-MF [18], LRFCSP [19], TDP-FDA [20], FCCR [21], and CCS-RCSP [28]. The numbers of channels selected by each algorithm are shown in parentheses. The proposed method achieves the highest mean classification accuracy.

The proposed method selects the smallest number of channels for each subject. This result shows that sufficiently high classification accuracy could be achieved when using a small number of channels, based on accurate channel selection.

The Figure 5 shows the optimal channels finally selected by the proposed method for each subject. The selected channels are generally located in the motor areas of the cerebral cortex for subjects '*b*', '*f*', and '*g*'. However, the selected channels of subjects '*a*' are mainly located at right motor area.

C. COMPLEXITY ANALYSIS

In this subsection, we analyze the computational complexity of six methods by comparing their complexity of the algorithm. Table 5 lists the complexities of the proposed method and the conventional channel selection methods. Our proposed method requires a relatively high computational complexity mainly due to the additional computation

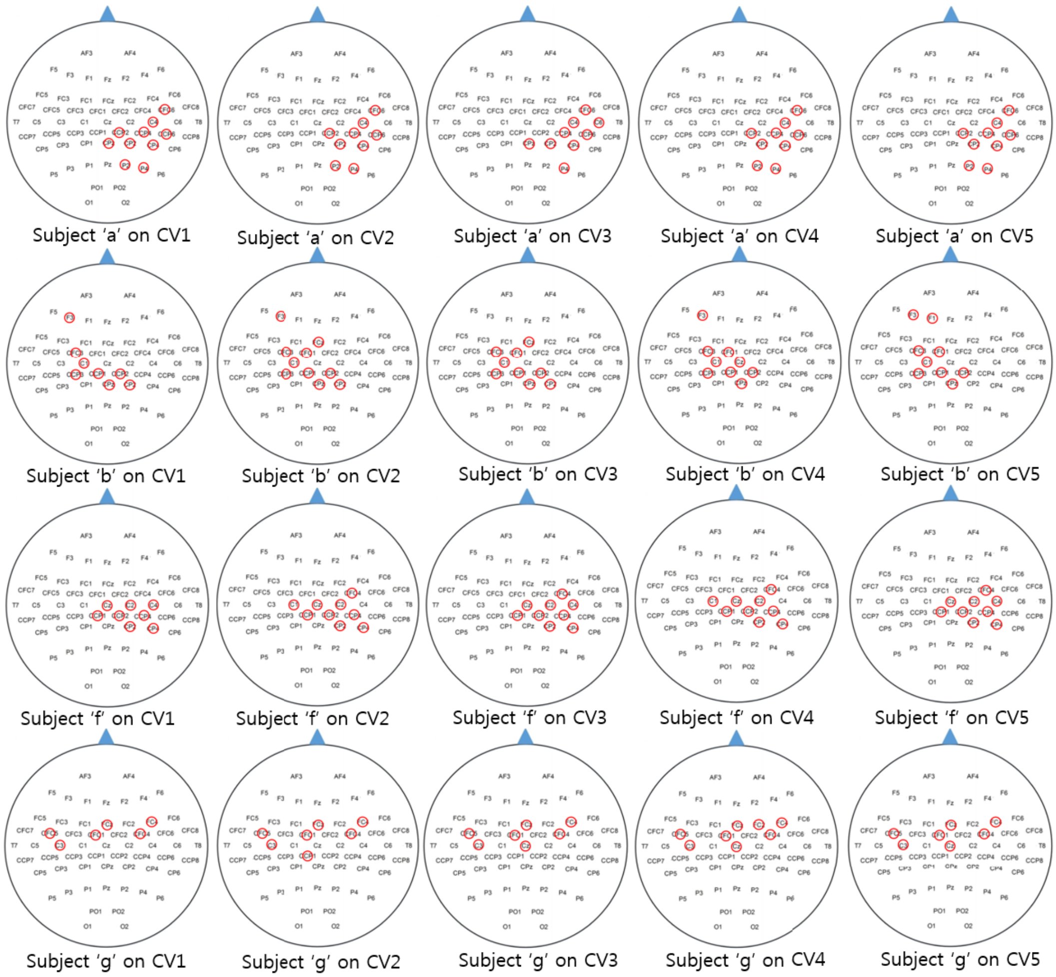

FIGURE 5. Location of optimal selected channels for BCI Competition IV dataset I. For each subject, five cross-validation results are displayed.

of correlation coefficients, $O((MK)^2)$, especially for a large channel size K unlike other methods. However, this increased computational complexity does not necessarily result in increased computational time, since a parallel computation of correlation coefficients can dramatically reduce the computational time.

V. CONCLUSION

In this paper, a novel MI-relevant channel selection method is considered based on correlation coefficients. First, our proposed method determines the *distinctive* channels, with respect to MI, which have significant correlation coefficient differences with as many other channels as possible. For each *distinct* channel, a supporting channel group is formed by gathering strongly correlated channels. By comparing the Fisher score of each supporting channel group, the channel set with the highest Fisher score is finally selected. Channel selection using two criteria, the correlation coefficient and the Fisher score of FBCSP features, effectively reduces the number of channels used and improves classification accuracy, compared to conventional channel selection methods.

REFERENCES

[1] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[2] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002.

[3] K. LaFleur, K. Cassady, A. Doud, K. Shades, E. Rogin, and B. He, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface," *J. Neural Eng.*, vol. 10, no. 4, Jun. 2013, Art. no. 046003.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[5] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787–798, May 1999.

[6] H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 653–662, Mar. 2012.

[7] K. Keng Ang, Z. Yang Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intelligence)*, Jun. 2008, pp. 2390–2397.

[8] S.-H. Park and S.-G. Lee, "Small sample setting and frequency band selection problem solving using subband regularized common spatial pattern," *IEEE Sensors J.*, vol. 17, no. 10, pp. 2977–2983, May 2017.

[9] S.-H. Park, D. Lee, and S.-G. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018.

[10] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *J. Neurosci. Methods*, vol. 255, pp. 85–91, Nov. 2015.

[11] R. Zhang, X. Xiao, Z. Liu, W. Jiang, J. Li, Y. Cao, J. Ren, D. Jiang, and L. Cui, "A new motor imagery EEG classification method FB-TRCSP+RF based on CSP and random forest," *IEEE Access*, vol. 6, pp. 44944–44950, 2018.

[12] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.

[13] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.

[14] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Muller, "The Berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2452–2462, Oct. 2008.

[15] J. Asensio-Cubero, J. Q. Gan, and R. Palaniappan, "Multiresolution analysis over simple graphs for brain computer interfaces," *J. Neural Eng.*, vol. 10, no. 4, Jul. 2013, Art. no. 046014.

[16] Y. Liu, H. Zhang, M. Chen, and L. Zhang, "A boosting-based spatial-spectral model for stroke Patients' EEG analysis in rehabilitation training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 169–179, Jan. 2016.

[17] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, Jun. 2011.

[18] 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," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–10, May 2019.

[19] Y. Park and W. Chung, "Frequency-optimized local region common spatial pattern approach for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1378–1388, Jul. 2019.

[20] Y. Yang, O. Kyrgyzov, J. Wiart, and I. Bloch, "Subject-specific channel selection for classification of motor imagery electroencephalographic data," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1277–1280.

[21] J. Han, Y. Zhao, H. Sun, J. Chen, A. Ke, G. Xu, H. Zhang, J. Zhou, and C. Wang, "A fast, open EEG classification framework based on feature compression and channel ranking," *Frontiers Neurosci.*, vol. 12, p. 217, Apr. 2018.

[22] K. Belwafi, O. Romain, S. Gannouni, F. Ghaffari, R. Djemal, and B. Ouni, "An embedded implementation based on adaptive filter bank for brain–computer interface systems," *J. Neurosci. Methods*, vol. 305, pp. 1–16, Jul. 2018.

[23] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for EEG-based brain–computer interfaces," *Neural Netw.*, vol. 22, no. 9, pp. 1313–1319, Nov. 2009.

[24] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell 2,1$-norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.

[25] B. Graimann, *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Berlin, Germany: Springer, 2010.

[26] F. Cona, M. Zavaglia, L. Astolfi, F. Babiloni, and M. Ursino, "Changes in EEG power spectral density and cortical connectivity in healthy and tetraplegic patients during a motor imagery task," *Comput. Intell. Neurosci.*, vol. 2009, pp. 1–12, Jun. 2009.

[27] M. Hamedi, S.-H. Salleh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: A review," *Neural Comput.*, vol. 28, no. 6, pp. 999–1041, Jun. 2016.

[28] 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 Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[29] J. Ye and T. Xiong, "SVM versus least squares SVM," in *Proc. 11th Int. Conf. Artif. Intell. Statist. (AISTATS)*, vol. 2, Mar. 2007, pp. 640–647.

[30] J. D. Bonita, L. C. C. Ambolode, B. M. Rosenberg, C. J. Cellucci, T. A. A. Watanabe, P. E. Rapp, and A. M. Albano, "Time domain measures of inter-channel EEG correlations: A comparison of linear, nonparametric and nonlinear measures," *Cognit. Neurodyn.*, vol. 8, no. 1, pp. 1–15, Feb. 2014.

[31] M. Hollander, D. A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[32] M. Krzywinski and N. Altman, "Points of significance: Significance, P values and t-tests," *Nature Methods*, vol. 10, no. 11, pp. 1041–1042, 2013.

[33] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.

[34] Y. Park and W. Chung, "Selective feature generation method based on time domain parameters and correlation coefficients for filter-bank-CSP BCI systems," *Sensors*, vol. 19, no. 17, p. 3769, Aug. 2019.

[35] (2005). *Data Set IVa for the BCI Competition III*. [Online]. Available: http://www.bbci.de/competition/iii/descIVa.html

[36] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, Aug. 2007.

YONGKOO PARK was born in Wonju, South Korea, in 1989. He received the B.E. degree in computer and radio communications engineering from Korea University, in 2013, where he is currently pursuing the Ph.D. degree. His current research interests include brain–computer interface, electroencephalography signals processing, and deep learning.

WONZOO CHUNG (Senior Member, IEEE) received the Ph.D. degree from Cornell University, Ithaca, NY, USA, in 2003. He is currently a Professor with Korea University, South Korea. His current research interest includes statistical machine learning for bio signals.