A Novel EEG Correlation Coefficient Feature Extraction Approach Based on Demixing EEG Channel Pairs for Cognitive Task 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 and Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No. 2019-0-00079, Artificial Intelligence Graduate School Program (Korea University)] and Institute of 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).

ABSTRACT

This paper presents a novel feature extraction method for electroencephalogram (EEG)-based cognitive task classification based on the correlation coefficients of EEG channel pairs by introducing preprocessing of the EEG signals. The preprocessing attempts to optimally demix each pair of EEG channels using a two-dimensional rotation matrix in order to mitigate the interference between channel pairs and, consequently, to enhance the resulting correlation coefficient features for cognitive task classification. For the optimization, the following criteria are proposed with an optimal rotation angle approximated for each criterion: i) maximum inter-class correlation coefficient distance (ICCD); ii) minimum within-class correlation coefficient distance (WCCD); and iii) maximum Fisher ratio (FR), which is the ratio of ICCD to WCCD. Performance evaluation based on the cognitive task dataset, dataset IV and Ib in BCI competition II, and Keirn and Aunon’s dataset, shows that ICCD optimization with the ‘above the mean’ and 1.5 interquartile range (IQR) feature selection method yields the best classification performance in comparison with other existing cognitive task classification methods.

INDEX TERMS

electroencephalography (EEG), brain-computer interfaces (BCIs), correlation coefficient, cognitive task classification.

I. INTRODUCTION

Brain-computer interfaces enable direct communication between the computer and the human brain through neural signals [1]–[3]. Recently, electroencephalogram (EEG)-based BCIs have been studied in the field of brain science, neuroscience, and rehabilitation [4], [5]. EEG is one of the most studied neural signals for detecting a user’s intention. EEG signals are composed of electrical potentials arising from several sources and has the advantages of high temporal resolution and non-invasiveness [6], [7]. However, the direct use of EEG signals is problematic, as several sources can be activated simultaneously and the electrical potentials from task-irrelevant sources are superimposed on the measured EEG signal, which is known as the volume conduction effect [8]. Therefore, various methods to extract the task-related information buried in EEG signals have been developed in BCI systems.

In particular, motor imagery (MI) is widely used in EEG-based BCIs, as the power of EEG signals from the cerebral cortex is decreased or increased while imaging the movements of specific parts of the body, called event-related desynchronization (ERD)/event-related synchronization (ERS). The common spatial pattern (CSP) approach [9] is the most commonly used MI feature extraction method based on ERD/ERS. CSP produces output powers maximizing the difference between two MI-tasks by using optimal spatial filters that indirectly suppress the volume conduction effect [10]. Various extensions of the CSP algorithm, i.e., regularized version of CSP [11]–[13], and frequency optimized version of CSP [14], have been studied with successful results. Moreover, the CSP algorithms based on
MI-task related EEG channel selection have been proposed in [15], [16]. Contrast to the CSP, attempts have also been made to develop a direct method to mitigate the volume conduction effect. The adaptive spatial filtering (ASF) approach ([17]) partially enhances each EEG channel signal by mitigating the interference from only the non-significantly correlated channels using noise canceling adaptive filters.

Although MI-EEG classification approaches have been used successfully for BCI systems, several limitations have recently been noted [8]. For example, the extracted EEG features based on temporal, spectral, and spatial characteristics for MI may not provide enough information for BCI systems and no successful communication has been established for certain types of subjects [8]. Therefore, there is increasing interest in BCI systems based on brain connectivity (neuronal activation patterns of separated brain regions) triggered by general cognitive tasks.

In recent years, several studies have been conducted to measure brain connectivity to obtain informative features on BCI systems for cognitive tasks [18]–[25]. These studies developed various algorithms based on various features to characterize the EEG-based interactions of the brain region: Phase lag index (PLI), defined as the absolute mean value of the sign of the difference between the instantaneous phase for the EEG channel pairs, is used in [18]. Partial directed coherence (PDC) method ([19]) uses coherence, i.e., the magnitude of normalized cross-spectral density between the EEG channel pairs. Phase locking value (PLV), defined as the absolute value of the mean phase difference between EEG channel pairs, is used in [20] with the Fisher score, which plays a role in selecting good PLV features (hereafter referred to as the PLV-FS method) and in [22]. The method using PLV with the recursive feature elimination (RFE) is referred to as the PLV-RFE method [27]. Correlation coefficient, a normalized covariance of EEG channel pairs, is considered in [21], [22]. Finally, the amplitude envelope correlation (AEC), the correlation coefficient for the envelop of Hilbert transformed EEG channel pairs, is used in [23].

Among these various brain connectivity related features, the algorithms based on correlation coefficient and PLV features show superior performance compared to those based on other brain connectivity features ([24], [25]). Furthermore, correlation coefficient features have several advantages over PLV features, such as insensitivity to frequency band optimization and noise-robust characteristics, making them suitable for EEG signals with a low signal-to-noise ratio (SNR) [22], [26]. Despite these advantages, the large number of correlation coefficient features generated from all pairs of EEG channels requires a sophisticated feature selection method to yield satisfactory classification performance. The Gaussian curve (GC) approach described in [21] selects the EEG channel pairs that produce significant statistical differences between the two tasks by approximating the feature distribution as Gaussian. The r-statistic approach proposed previously [25] selects the EEG channel pairs with ‘good’ correlation coefficient features using r-score.

Correlation coefficient feature selection using RFE is proposed in [22] with various modified support vector machines (SVMs).

However, merely selecting good correlation coefficients has limitations with regard to improving performance. For example, connectivity/time-frequency mask (CM/TFM) algorithm demonstrate that integration of time-frequency power features and carefully selected correlation coefficient features can yield a better performance than that produced by existing correlation coefficient-based algorithms [25]. Therefore, fundamental approaches to enhance the quality of correlation coefficient features are required.

In this paper, preprocessing of raw EEG signals is proposed to enhance correlation coefficient features in contrast to conventional correlation coefficient-based methods. Ideally, a perfect source separation of the EEG signals that makes EEG channels statistically independent would dramatically enhance the correlation coefficients as features representing brain connectivity. However, such perfect source separation is not feasible and we attempt a two step sub-optimal signal processing. Adaptive spatial filtering (ASF, [17]) is firstly applied to all channels to mitigate mild interferences among the weakly correlated channels. Then, instead of global demixing of EEG signals, we consider feasible demixing of each channel pair EEG signals with a two-dimensional rotation matrix that has a single parameter, the rotation angle. Furthermore, the demixing matrix is adjusted to a tangible criteria to directly enhance classification accuracy. The following three demixing matrix optimization criteria are proposed and the optimal rotation angles are investigated: i) maximum distance of averaged correlation coefficients between two cognitive tasks, termed as inter-class correlation coefficient distance (ICCD), ii) minimum sum of the variances of correlation coefficients for each cognitive task, termed as within-class correlation coefficient distance (WCCD), and iii) maximum Fisher ratio (FR) of the correlation coefficients for two cognitive tasks i.e., the ICCD to WCCD ratio. The closed form expression for the approximated optimum demixing matrix for each criterion is derived in terms of the rotation angle. Furthermore, feature selection methods for the optimized correlation coefficients, i.e., selection of good correlation coefficient features, are thoroughly investigated. Four prevailing feature selection methods i.e., the ‘above the mean’ and 1.5 interquartile range (IQR) method [28], GC approach [21], r-statistic method [25], and RFE method [27] are considered with respect to each criterion. The classification performances of the proposed schemes are evaluated for three cognitive task EEG datasets, Ib and IV in BCI competition II [31] and Keirn and Aunon’s dataset [32]. The ICCD optimization with ‘above the mean’ and 1.5 IQR rule shows the best performance.

The paper is organized as follows. In Section II, we introduce the system model and the proposed method. In Section III, the experimental setting using the Ib and IV in BCI competition II and Keirn and Aunon’s dataset are...
explained. Section IV analyzes the experiments results. Finally, conclusion for this paper is made in Section V.

II. METHOD

A. SYSTEM MODEL

Let us consider an EEG signal dataset recording binary cognitive tasks using K channels. The EEG signal corresponding to channel k is denoted as the vector \( x^{(k)} = [x^{(k)}(1), x^{(k)}(2), \ldots, x^{(k)}(N)]^T \), where \( k = 1, 2, \ldots, K \) and \( N \) is the number of temporal samples per channel. We assume that \( M \) trials of EEG signals are available as a training dataset and each trial consists of \( N \) sample points, indexed by \( x^{(k)}_i \in \mathbb{R}^{N \times 1} \), where \( i = 1, \ldots, M \). Each trial is included in two index sets, \( I_1 \) and \( I_2 (I_1 \cup I_2 = \{1, 2, \ldots, M\}) \), corresponding to each cognitive task, respectively, as we consider binary cognitive task classification. We assume that \( x^{(k)}_i \) is already band-pass filtered.

For the cognitive task classification, we use sample correlation coefficients between selected EEG channel pairs after preprocessing. The (sample) correlation coefficient of the \( i \)-th trial EEG channel pair \( k \) and \( p \) is given by:

\[
\rho^{(k,p)}_i = \frac{C(x^{(k)}_i, x^{(p)}_i)}{\sqrt{P(x^{(k)}_i)} \sqrt{P(x^{(p)}_i)}}, \quad k, p = 1, 2, \ldots, K,
\]  

(1)

where \( C(x^{(k)}_i, x^{(p)}_i) \) is the sample covariance and \( P(x^{(k)}_i) \) is the sample variance,

\[
C(x^{(k)}_i, x^{(p)}_i) := \sum_{n=1}^N \left( x^{(k)}_i(n) - \bar{x}^{(k)}_i \right) \left( x^{(p)}_i(n) - \bar{x}^{(p)}_i \right)
\]

(2)

\[
P(x^{(k)}_i) := \frac{1}{N} \sum_{n=1}^N \left( x^{(k)}_i(n) - \bar{x}^{(k)}_i \right)^2
\]

(3)

and \( \bar{x}^{(k)}_i = \frac{1}{N} \sum_{n=1}^N x^{(k)}_i(n) \) denotes the sample mean of \( x^{(k)}_i \).

B. COST FUNCTIONS OF CORRELATION COEFFICIENTS

1) INTER-CLASS CORRELATION COEFFICIENT DISTANCE

The inter-class correlation coefficient distance (ICCD) for each EEG channel pair \( (k, p) \), denoted by \( D^{(k,p)} \), is the distance between the averaged correlation coefficients for two tasks [28]. The mean correlation coefficient for the task \( c \), denoted by \( \bar{\rho}^{(k,p)}_c \) \((c = 1, 2)\), is given by:

\[
\bar{\rho}^{(k,p)}_c = \frac{1}{|L_c|} \sum_{i \in L_c} \rho^{(k,p)}_i, \quad c = 1, 2
\]

(4)

where \( |L_c| \) denotes the number of training trials for the task \( c \).

For a training data set \( \{x^{(k)}_i\} \), the ICCD can be computed as following:

\[
D^{(k,p)} = \left( \bar{\rho}_1^{(k,p)} - \bar{\rho}_2^{(k,p)} \right)^2.
\]

(5)

2) WITHIN-CLASS CORRELATION COEFFICIENT DISTANCE

The within-class correlation coefficient distance (WCCD) for EEG channel pair \( (k, p) \), denoted by \( W^{(k,p)} \), is the sum of the variances of correlation coefficients for two tasks. Using the training dataset, WCCD is given by:

\[
W^{(k,p)} = \sum_{c=1}^2 \frac{1}{|L_c|} \sum_{i \in L_c} \left( \rho^{(k,p)}_i - \bar{\rho}^{(k,p)}_c \right)^2
\]

(6)

3) FISHER RATIO

The Fisher ratio (FR) for the correlation coefficient of EEG channel pair \( (k, p) \), denoted by \( F^{(k,p)} \), is defined as the ratio of ICCD to WCCD [29]:

\[
F^{(k,p)} = \frac{D^{(k,p)}}{W^{(k,p)}}.
\]

(7)

Fig. 1 depicts the block diagram of the proposed algorithm.

The EEG signals are preprocessed using the ASF method [17] for interference mitigation. For each pair of ASF-filtered EEG signals, a rotation matrix is applied to optimize one
of the above criteria, maximize ICCD, minimize WCCD, or maximize FR. Various feature selection schemes are examined with the resulting correlation coefficients.

C. ADAPTIVE SPATIAL FILTERING

Fig. 2 illustrates the block diagram of the adaptive spatial filtering algorithm for EEG channel signals. For each channel $k = 1, 2, \ldots, K$, the ASF method [17] first determines a reference set of EEG channels, denoted by $H^{(k)}$, with the channel $k$ computed using training data in both classes, the correlation coefficients of which are below a predetermined threshold, $\rho_{thr}$, as follows:

$$H^{(k)} = \left\{ p \in [1, 2, \ldots, K] \left| \rho_1^{(k,p)} < \rho_{thr} \wedge \rho_2^{(k,p)} < \rho_{thr} \right. \right\} \quad (8)$$

Once $H^{(k)}$ is set, ASF attempts to reduce the residual signal components of $H^{(k)}$ in $\mathbf{x}^{(k)}(n)$ by minimizing the power of the resulting signal, denoted by $\mathbf{y}^{(k)}(n)$ with adaptive interference canceling filters [17]:

$$\mathbf{y}^{(k)}(n) = \mathbf{x}^{(k)}(n) - \sum_{h \in H^{(k)}} [\mathbf{w}^{(k,h)}]^T(n)\mathbf{\tilde{x}}^{(h)}(n) \quad (9)$$

where $\mathbf{\tilde{x}}^{(h)}(n) = [x^{(h)}(n - D + 1), \ldots, x^{(h)}(n)]^T$ denotes a vector of the past $D$ samples of $x^{(h)}(n)$ at time index $n$ and $\mathbf{w}^{(k,h)}(n) = [w_0^{(k,h)}(n), \ldots, w_D^{(k,h)}(n)]^T$ denotes the canceling filter for the channel $h \in H^{(k)}$ initialized to zeros. The goal of ASF is to minimize the power of $\mathbf{y}^{(k)}(n)$ with optimal filter $\mathbf{w}^{(k,h)}$, hoping the power term contributed by interference is minimized. That is equivalent to find the filter minimizing the cost function,

$$J(\mathbf{w}^{(k,h)}) = E \left[ \| \mathbf{y}^{(k)}(n) \|^2 \right], \quad (10)$$

where $E(\cdot)$ denotes the expectation operation. To minimize the cost function, a stochastic descent algorithm [30] can be considered as follows:

$$\mathbf{w}^{(k,h)}(n+1) = \mathbf{w}^{(k,h)}(n) - \mu \frac{\partial}{\partial \mathbf{w}^{(k,h)}} J(\mathbf{w}^{(k,h)})$$

$$= \mathbf{w}^{(k,h)}(n) + \mu \mathbf{y}^{(k,h)}(n) \mathbf{\tilde{x}}^{(h)}(n) \quad (11)$$

where $\mu$ is a step size. The above adaptive algorithm is equivalent to the least mean square (LMS) algorithm [30]. In order to guarantee the stability of the algorithm, we apply the normalized LMS (NLMS) algorithm [30], the normalized version of LMS by the power of the input signal:

$$\mathbf{w}^{(k,h)}(n+1) = \mathbf{w}^{(k,h)}(n) + \mu \frac{y^{(k)}(n)x^{(h)}(n)}{||\mathbf{x}^{(h)}(n)||^2} \mathbf{x}^{(h)}(n), \quad h \in H^{(k)} \quad (12)$$

After applying the ASF method, we denote the ASF-filtered $\mathbf{x}^{(h)}_i$ as $\mathbf{y}^{(h)}_i = [y^{(h)}_i(1), y^{(h)}_i(2), \ldots, y^{(h)}_i(N)]$.

D. EEG CHANNEL PAIR OPTIMIZATION USING ROTATION MATRICES

After processing with ASF, we optimize the correlation coefficients of each EEG channel pair by following three criteria: maximize ICCD, minimize WCCD, and maximize FR. We apply a two-dimensional rotation matrix to the ASF-processed signals from each channel pair $(k, p)$ as follows:

$$G_i^{(k,p)}(\theta) = \frac{C(\mathbf{y}^{(k,p)}_i)}{\sqrt{P(\mathbf{y}^{(k,p)}_i)} \sqrt{P(\mathbf{y}^{(k,p)}_i)}}$$

$$= \frac{(P(\mathbf{y}^{(k,p)}_i)) \sin(2\theta) + 2C(\mathbf{y}^{(k,p)}_i) \cos(2\theta)}{P(\mathbf{y}^{(k,p)}_i) \cos(2\theta) + 2C(\mathbf{y}^{(k,p)}_i) \sin(2\theta)^2}$$

where $\theta$ denotes the rotation angle. The correlation coefficient of the rotated $\mathbf{y}^{(k,p)}$ and $\mathbf{y}^{(k,p)}_i$ can be expressed as the following function of the rotation angle $\theta$ with respect to the original $\mathbf{y}^{(k)}_i$ and $\mathbf{y}^{(k,p)}_i$ with some simplification:

$$\rho_i^{(k,p)}(\theta) = \frac{C(\mathbf{y}^{(k,p)}_i)}{\sqrt{P(\mathbf{y}^{(k,p)}_i)} \sqrt{P(\mathbf{y}^{(k,p)}_i)}}$$

$$= \frac{C(\mathbf{y}^{(k,p)}_i)}{\sqrt{P(\mathbf{y}^{(k,p)}_i)} \sqrt{P(\mathbf{y}^{(k,p)}_i)}}$$

For EEG signals, we can assume that the sum of powers, $P(\mathbf{y}^{(k,p)}_i)$, is significantly greater than the difference, $|P(\mathbf{y}^{(k,p)}_i)|$, and the covariance, $C(\mathbf{y}^{(k,p)}_i)$. Hence, $G_i^{(k,p)}(\theta)$ is dominated by the $P(\mathbf{y}^{(k,p)}_i)$ term and fluctuates due to the $\sin^2(\theta)$ term. Therefore, $G_i^{(k,p)}(\theta)$ is periodic with period $\pi/2$ and reaches its maximum value when $\sin^2(2\theta + \psi) = 1$ and its minimum value when $\sin^2(2\theta + \psi) = 0$. As such, we approximate $G_i^{(k,p)}(\theta)$ with its mean value, the $G_i^{(k,p)}(\theta)$
The ICCD of the demixed signals can be written as a function denoted with \( \hat{\rho} \). Rotation angle for the channel pair (C4, CP5) of BCI competition II dataset 87426 is shown in Fig. 3(a) shows the rotation angle \( \rho \) with respect to the original \( \theta \).

Fig. 3 plots an example of the approximation for the (C4, CP5) channel pair of the BCI competition II dataset IV. Fig. 3(a) shows \( G_i^{(k,p)}(\theta) \) and \( \bar{G}_i^{(k,p)}(\pi - \psi/2) \). True \( \rho_i^{(k,p)}(\theta) \) and approximated \( \hat{\rho}_i^{(k,p)}(\theta) \).

evaluated at \( \sin^2(2\theta + \psi) = 1/2 \) or \( \theta = \pi/8 - \psi/2 \), and denote it with \( \bar{G}_i^{(k,p)} \) as follows:

\[
\bar{G}_i^{(k,p)} := G_i^{(k,p)} \left( \frac{\pi}{8} - \frac{\psi}{2} \right)
= \sqrt{P^2(y_{i,+}^{(k,p)}) - C^2(y_{i,+}^{(k,p)}) + \frac{1}{2}P^2(y_{i,-}^{(k,p)})} \tag{17}
\]

Based on these assumptions, \( P(y_{i,+}^{(k,p)}) \gg |P(y_{i,-}^{(k,p)})| \) and \( C(y_{i,+}^{(k,p)}) \), \( \rho_i^{(k,p)}(\theta) \) can be approximated follows:

\[
\rho_i^{(k,p)}(\theta) \approx \frac{P(y_{i,-}^{(k,p)}) \sin(2\theta) + 2C(y_{i,-}^{(k,p)}) \cos(2\theta)}{\bar{G}_i^{(k,p)}}. \tag{18}
\]

Fig. 3 plots an example of the approximation for the (C4, CP5) channel pair of the BCI competition II dataset IV. Fig. 3(a) shows \( G_i^{(k,p)}(\theta) \) and \( \bar{G}_i^{(k,p)} \) for \( i = 1 \) and Fig. 3(b) shows the \( \rho_i^{(k,p)}(\theta) \) and approximated \( \hat{\rho}_i^{(k,p)}(\theta) \) for \( i = 1 \).

1) OPTIMAL ROTATION ANGLE FOR ICCD MAXIMIZATION

The ICCD of the demixed signals can be written as a function of \( \theta \) with respect to the original \( y_i^{(k,p)} \) and \( y_i^{(p)} \) as the following:

\[
D_{\text{opt}}^{(k,p)}(\theta) = \left( \rho_i^{(k,p)}(\theta) - \hat{\rho}_i^{(k,p)}(\theta) \right)^2. \tag{19}
\]

The optimal angle maximizing \( D_{\text{opt}}^{(k,p)}(\theta) \) is approximately given by (refer to Appendix A for the detailed derivation):

\[
\theta_{\text{opt}}^{(k,p)} = \arg \max_{\theta} D_{\text{opt}}^{(k,p)}(\theta)
= \frac{\pi}{4} - \frac{1}{2} \cos^{-1} \left( \frac{\bar{T}_1^{(k,p)} - \bar{T}_2^{(k,p)}}{\sqrt{\bar{T}_1^{(k,p)} - \bar{T}_2^{(k,p)} + (\bar{U}_1^{(k,p)} - \bar{U}_2^{(k,p)})^2}} \right) \tag{20}
\]

where

\[
T_i^{(k,p)} := \frac{P(y_{i,-}^{(k,p)})}{\bar{G}_i^{(k,p)}}, \quad \bar{T}_c := \frac{1}{|U_c|} \sum_{i \in U_c} T_i^{(k,p)} \tag{21}
\]

\[
U_i^{(k,p)} := 2C\left( \frac{y_{i,-}^{(k,p)}}{\bar{G}_i^{(k,p)}} \right), \quad \bar{U}_c := \frac{1}{|U_c|} \sum_{i \in U_c} U_i^{(k,p)} \tag{22}
\]

2) OPTIMAL ROTATION ANGLE FOR WCCD MINIMIZATION

The WCCD can be written as a function of \( \theta \) as follows:

\[
W_{\text{opt}}^{(k,p)}(\theta) = \sum_{c=1}^{2} \frac{1}{|U_c|} \sum_{i \in U_c} \left( \rho_i^{(k,p)}(\theta) - \hat{\rho}_i^{(k,p)}(\theta) \right)^2. \tag{23}
\]

The optimal angle minimizing \( W_{\text{opt}}^{(k,p)}(\theta) \) is given approximately by (refer to Appendix B for the detailed derivation):

\[
\theta_{\text{opt}}^{(k,p)} = \arg \min_{\theta} W_{\text{opt}}^{(k,p)}(\theta)
= -\frac{1}{4} \cos^{-1} \left( \frac{V^{(k,p)}}{\sqrt{[V^{(k,p)}]^2 + [U^{(k,p)}]^2}} \right) - \frac{\pi}{8}. \tag{24}
\]

where

\[
U_{\text{opt}}^{(k,p)} := \sum_{c=1}^{2} \frac{1}{|U_c|} \sum_{i \in U_c} \left( \rho_i^{(k,p)}(\theta) - \hat{\rho}_i^{(k,p)}(\theta) \right)^2. \tag{25}
\]

3) OPTIMAL ROTATION ANGLE FOR FR MAXIMIZATION

The FR is computed using ICCD and WCCD, i.e., \( F_{\text{opt}}^{(k,p)}(\theta) = \frac{D_{\text{opt}}^{(k,p)}(\theta)}{W_{\text{opt}}^{(k,p)}(\theta)} \), and the optimal angle maximizing \( F_{\text{opt}}^{(k,p)}(\theta) \) is given approximately by; see eq.28, as shown at the bottom of the next page (refer to Appendix C for the detailed derivation), where

\[
\theta_{\text{opt}}^{(k,p)} = \cos^{-1} \left( \frac{\gamma^{(k,p)}}{\sqrt{[\gamma^{(k,p)}]^2 + [\nu^{(k,p)}]^2}} \right) \tag{29}
\]

\[
\gamma^{(k,p)} := 4H^{(k,p)} \cos(2\psi_{i}^{(k,p)}) \sum_{c=1}^{2} \frac{1}{|U_c|} \sum_{i \in U_c} \left( T_i^{(k,p)} - \bar{T}_c \right)^2 + 8H^{(k,p)} \left( [V^{(k,p)}]^2 + [U^{(k,p)}]^2 \right) \sin(\psi_W^{(k,p)}) \tag{30}
\]
We have considered three optimization criteria for the correlation coefficients. The number of correlation coefficients generated from the $K$-channel EEG signals is $K(K - 1)/2$. From these overabundant correlation coefficients, a few 'good' EEG channel pairs are selected using training data during the training stage to determine the final output features. There are several feature selection methods for correlation coefficients. In this paper, four feature selection methods, the 'above the mean rule' and 1.5 IQR method, are tested with the optimized correlation coefficients with three different criteria. The simulation results presented in the next section show that ICCD maximized optimization with the 'above the mean rule' and 1.5 IQR method yield the best performance.

1) 'ABOVE THE MEAN' AND 1.5 IQR METHOD

The 'above the mean' rule selects the correlation coefficients that have the values above the mean in terms of the criterion for which the rotation matrix is optimized [28]. For example, when the rotation matrix is optimized for ICCD maximization, the channel pairs whose resulting ICCD is greater than the average of the ICCDs of all EEG channel pairs are selected. The selected set of EEG channel pairs, denoted by $Z$, is given by:

$$Z = \{(k, p) \in [1, \cdots, K]^2 \mid D^{k,p}(\theta_{opt}) \geq \bar{D}^{k,p}\}$$

where $\bar{D}^{k,p}(\theta_{opt}) = \frac{2}{K(K-1)} \sum_{k \neq p} D^{k,p}(\theta_{opt})$. When the rotation matrix is optimized with respect to WCCD minimization, we select the channel pairs with WCCDs that are lower than the averaged WCCD, as follows:

$$Z = \{(k, p) \in [1, \cdots, K]^2 \mid W^{k,p}(\theta_{opt}) \leq \bar{W}^{k,p}\}$$

where $\bar{W}^{k,p}(\theta_{opt}) = \frac{2}{K(K-1)} \sum_{k \neq p} W^{k,p}(\theta_{opt})$. For a small $K$, the 'above the mean' rule yields a suitable number of EEG channel pairs. However, for a larger $K$ the size of $Z$ may be still large. In such case, we further apply the 1.5 IQR rule [28] to select more discriminative EEG channel pairs, since the 1.5 IQR rule is effective feature selection method from the large number of features. The IQR thresholds for the optimization criteria, $D_{thr}$, $W_{thr}$, and $F_{thr}$ are defined as follows:

$$D_{thr} := Q_3 + 1.5(Q_3 - Q_1)$$
$$W_{thr} := Q_1 - 1.5(Q_3 - Q_1)$$
$$F_{thr} := Q_3 + 1.5(Q_3 - Q_1)$$

where $Q_1$ and $Q_3$ denote the middle value between the minimum and median and the middle value between the median and maximum, respectively, with respect to the metric of interest (ICCD, WCCD, or FR) among the pairs in the 'above the mean' set $Z$. We further refine the selected channel pairs with respect to the threshold. For example, for ICCD optimization

$$Z' = \{(k, p) \in Z \mid D^{k,p}(\theta_{opt}) \geq D_{thr}\}$$

and for WCCD optimization

$$Z' = \{(k, p) \in Z \mid W^{k,p}(\theta_{opt}) \leq W_{thr}\}$$

2) THE GAUSSIAN CURVE METHOD

The GC method approximates the distribution of correlation coefficients of an EEG channel pair for a given class as Gaussian with a sample mean and sample variance [21]. Hence, the probability of a misclassification error for a correlation coefficient for the channel $k$ and $p$, denoted by $A^{k,p}$, is given as the intersection area of the two approximated Gaussian probability density functions corresponding to each class, which can be easily computed by Gaussian probability density functions. The selected EEG channel pair set $Z$ consists of EEG channel pairs with lower misclassification errors than a threshold, denoted as $O_{thr}$, as follows:

$$Z = \{(k, p) \mid A^{k,p} \leq O_{thr}\}$$

$$\theta_{opt} = \arg \max_{\theta} F^{k,p}(\theta)$$

$$\approx \left\{ \begin{array}{ll}
\frac{1}{4} \sin^{-1} \left( \frac{-8H^{k,p} \sqrt{[V^{k,p}]^2 + [U^{k,p}]^2} \cos(2\psi^k_p - \psi^k_w)}{\sqrt{[\gamma^{k,p}]^2 + [\nu^{k,p}]^2}} \right) - \kappa^{k,p} & \text{if } 0 \leq \kappa^{k,p} < \frac{\pi}{4} \\
\frac{1}{4} \sin^{-1} \left( \frac{-8H^{k,p} \sqrt{[V^{k,p}]^2 + [U^{k,p}]^2} \cos(2\psi^k_p - \psi^k_w)}{\sqrt{[\gamma^{k,p}]^2 + [\nu^{k,p}]^2}} \right) + \frac{\pi}{4} - \kappa^{k,p} & \text{if } \frac{\pi}{4} \leq \kappa^{k,p} < \frac{\pi}{2}
\end{array} \right.$$
3) THE T-STATISTIC METHOD
The t-statistic method applies a statistical method based on the Student’s t-distribution to select discriminative EEG channel pairs [25]. The t-score for correlation coefficients, the normalized difference of the averaged correlation coefficients between two classes, approximately follows the t-distribution and the significance of the t-score, denoted by $P(k,p)$, is measured using the approximated t-distribution. The selected EEG channel pair set $Z$ consists of EEG channel pairs with $P(k,p)$ values higher than the threshold $P_{thr}$, as follows:

$$Z = \{(k,p) \mid P(k,p) \geq P_{thr}\}$$ (41)

4) RECURSIVE FEATURE ELIMINATION (RFE)
The RFE method is a feature selection method that determines the ranks of features by iteratively eliminating the features with the lowest contribution to the SVM classifier [27]. Then, a certain number of top features, $N_Z$, are used for the SVM. In our experiment we optimized $N_Z$ based on cross-validation.

III. DATA AND EXPERIMENTS
A. DATA DESCRIPTION
1) BCI COMPETITION II DATASET IV
The goal of public benchmark dataset IV from BCI competition II is to predict the laterality of upcoming finger movements [31]. This dataset is widely used to evaluate the performance of cognitive task classification algorithms. During the recording, the subject was requested to press the key using a self-chosen order and timing for predicting the movement of the left or right hand. This dataset was recorded without a feedback session. The EEG signal was recorded using 28 channels ($K = 28$) and bandpass filtered between 0.05 and 200 Hz. The sampling rate was 1000Hz. The dataset consisted of 416 trials (316 training trials and 100 test trials).

2) BCI COMPETITION II DATASET IB
The public benchmark dataset Ib from BCI competition II was taken from an artificially respirated amyotrophic lateral sclerosis (ALS) patient [31]. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were monitored. The EEG signal was recorded using seven channels: A1-Cz, A2-Cz, 2cm frontal of C3, 2cm parietal of C3, vEOG, 2cm frontal of C4, and 2cm parietal of C4. The sampling rate was 256Hz. The dataset consisted of 380 trials (200 training trials and 180 test trials).

3) KEIRN AND AUNON’S DATASET
This public benchmark dataset was originally reported by Keirn and Aunon [32]. The data were recorded from seven subject aged 21-48 years who were performing five distinct mental tasks [32]: i) the baseline task in which the subjects were asked to relax as much as possible; ii) the letter task in which the subjects were instructed to mentally compose a letter to a friend or relative without vocalizing; iii) the math task in which the subjects were asked to solve nontrivial multiplication problems without vocalizing or making any other physical movements; iv) the visual counting task in which the subjects were asked to imagine a blackboard and to visualize numbers being written on the board sequentially, v) the rotation task in which the subjects were asked to visualize a particular three-dimensional block figure being rotated about an axis. The EEG data were recorded for 10s at sampling rate of 250Hz using six channels (C3, C4, P3, P4, O1, and O2).

B. DATA PROCESSING
1) BCI COMPETITION II DATASET IV
For the experiments with dataset IV, the EEG data were band-pass filtered using a fourth-order Butterworth filter operating at 10-33Hz. The time segment of EEG data from 0.15 to 0.5s after the cue was extracted. We set $\rho_{thr}$ to 0.65 in the ASF algorithm. The step size ($\mu$) for NLMS algorithm is set to 0.0001 by 10 x 10 cross-validation and filter length ($D$) is set to 3. For feature selection, the ‘above the mean’ and 1.5 IQR rule, GC method, t-statistic method and RFE are used. For the GC method and RFM, we select 30 EEG channel pairs. For the t-statistic method, the probability threshold is set to 0.995.
TABLE 3. Classification accuracy (%) of proposed method for Keirn and Aunon’s dataset using 10 × 10 cross-validation.

| Method      | Accuracy |       |
|-------------|----------|-------|
|             | subject 1 | subject 2 | subject 3 | subject 4 | mean |
| ‘above the mean’/1.5 IQR rule | 94.4 | 76.8 | 77 | 77.9 | 81.5 |
| GC          | 94.7 | 72.4 | 73.9 | 76.4 | 79.4 |
| t-statistic | 95 | 74.8 | 73.7 | 81.6 | 81.3 |
| RFE         | 95.6 | 73.6 | 74.6 | 79.5 | 80.8 |

ICCD

WCCD

| Method      | Accuracy |       |
|-------------|----------|-------|
|             | subject 1 | subject 2 | subject 3 | subject 4 | mean |
| ‘above the mean’/1.5 IQR rule | 93.5 | 72.8 | 71.1 | 72 | 77.4 |
| GC          | 93.9 | 74.8 | 72.6 | 78.4 | 79.9 |
| t-statistic | 94 | 76.8 | 71.1 | 79.4 | 80.3 |
| RFE         | 93.9 | 72.8 | 73.3 | 75.9 | 79 |

FR

| Method      | Accuracy |       |
|-------------|----------|-------|
| ‘above the mean’/1.5 IQR rule | 89.2 | 62.4 | 70.8 | 72.5 | 73.7 |
| GC          | 83.3 | 61.2 | 71.2 | 68.1 | 71 |
| t-statistic | 91.9 | 60.4 | 72.6 | 58.3 | 70.8 |
| RFE         | 82.5 | 63.6 | 70.6 | 68.8 | 71.4 |

FIGURE 5. Topography of EEG channel correlation coefficients with C4 for BCI competition II dataset IV. (a) Class 1: Raw EEG signals (left) and ICCD optimized signals (right). (b) Class 2: Raw EEG signals (left) and ICCD optimized signals (right).

FIGURE 6. Selected EEG channel pairs using ICCD optimization with the ‘above the mean/1.5 IQR’ rule for BCI competition II dataset Ib.

TABLE 4. ICCD optimization and performance improvement for BCI competition II dataset IV.

| Channel Without preprocessing pairs | ICCD (angle) | Accuracy | ICCD optimization | ICCD (angle) | Accuracy |
|------------------------------------|--------------|----------|-------------------|--------------|----------|
| C4-C3                              | 0.002 (0°)   | 0.149 (56.21°) | 0.149 (56.21°) | 0.149 (56.21°) |
| C4-CP5                             | 0.003 (0°)   | 0.089 (39.33°) | 0.089 (39.33°) | 0.089 (39.33°) |
| C4-CP3                             | 0.002 (0°)   | 0.159 (41.19°) | 0.159 (41.19°) | 0.159 (41.19°) |
| C4-CP1                             | 0.017 (0°)   | 0.093 (44.99°) | 0.093 (44.99°) | 0.093 (44.99°) |

2) BCI COMPETITION II DATASET IB
For the experiment with dataset Ib, we use six channels excluding the vEOG channel (K = 6). A fourth-order bandpass filter operating at 0.5-7.5Hz was applied as the difference in EEG spectral power between the two classes of dataset Ib is concentrated in the low frequency band [25].

3) KEIRN AND AUNON’S DATASET
For the experiment with Keirn and Aunon’s dataset, we use four subjects as the other three subjects had fewer than 10 sessions or some errors in the recording [33]. As we consider the binary classification, we test our proposed method on two mental tasks, a letter task and a math task. The EEG signal for each mental task was segmented into 10 segments with length 1s. The \( \rho_{thr} \) for the ASF algorithm is set to 0.6. The step size (\( \mu \)) is set to 0.0001 by 10 × 10 cross-validation and filter length (\( D \)) is set equal to the experiments with dataset IV. For feature selection, the ‘above the mean’ rule, GC method, t-statistic method and RFE are used. Eight EEG channel pairs are selected using the GC method and RFE, and probability threshold is set to 0.9 for the t-statistic method.

The time segment of EEG data from 1 to 3s after the cue were extracted. The \( \rho_{thr} \) for the ASF algorithm is set to 0.6. The step size (\( \mu \)) is set to 0.0001 by 10 × 10 cross-validation and filter length (\( D \)) is set equal to the experiments with dataset IV. For feature selection, the ‘above the mean’ rule, GC method, t-statistic method and RFE are used. Eight EEG channel pairs are selected using the GC method and RFE, and probability threshold is set to 0.9 for the t-statistic method.

The time segment of EEG data from 1 to 3s after the cue were extracted. The \( \rho_{thr} \) for the ASF algorithm is set to 0.6. The step size (\( \mu \)) is set to 0.0001 by 10 × 10 cross-validation and filter length (\( D \)) is set equal to the experiments with dataset IV. For feature selection, the ‘above the mean’ rule, GC method, t-statistic method and RFE are used. Eight EEG channel pairs are selected using the GC method and RFE, and probability threshold is set to 0.9 for the t-statistic method.
TABLE 6. Classification performance (%) comparison of the proposed method with existing methods for BCI competition II dataset IV.

| CM/TFM | Correlation coefficient | PLV | Coherence | Proposed method |
|--------|-------------------------|-----|-----------|-----------------|
|        | GC                      | CC-RFE | PLV-FS | PLV-RFE | PDC | 90 |
|        | 85                      | 80    | 74       | 77       | 76  | 74 |

TABLE 7. Classification performance (%) comparison of the proposed method with existing methods for BCI competition II dataset Ib.

| CM/TFM | Correlation coefficient | PLV | Coherence | Proposed method |
|--------|-------------------------|-----|-----------|-----------------|
|        | GC                      | CC-RFE | PLV-FS | PLV-RFE | PDC | 62.78 |
|        | 60.56                   | 58.33 | 57.22    | 58.89     | 57.78 | 60 |

TABLE 8. Classification performance (%) comparison of the proposed method with existing methods for Keirn and Aunon’s dataset.

| CM/TFM | Correlation coefficient | PLV | Coherence | Proposed method |
|--------|-------------------------|-----|-----------|-----------------|
| subject 1 | GC                      | CC-RFE | PLV-FS | PLV-RFE | PDC | 94.2 |
| subject 2 | 86.7                    | 72.2  | 71.4     | 87.9     | 88.9 | 72.3 |
| subject 3 | 76.4                    | 70.3  | 70.8     | 61.6     | 62.8 | 68  |
| subject 4 | 74.5                    | 60.4  | 61.4     | 64.6     | 66.1 | 65.5 |
| mean    | 77                      | 67.5  | 69       | 72.4     | 74.6 | 68  |
| p value | 0.404                   | 0.0832 | 0.1528 | 0.1418 | 0.3552 | 0.0616 |

IV. RESULT AND DISCUSSION

A. PERFORMANCE COMPARISON OF PROPOSED THREE OPTIMIZATION SCHEMES

We compared the performance between three correlation coefficient optimization schemes (ICCD, CC-RFE, and FR optimization) with four feature selection methods (‘above the mean’/1.5 IQR rule, GC method, $t$-statistic method, and RFE) using the three EEG datasets mentioned above.

First, Table 1 presents the classification accuracies of BCI competition II dataset IV. The ICCD optimization with the ‘above the mean’/1.5 IQR rule outperforms the other optimization criterion with feature selection methods.

Table 2 shows the performance of three optimized schemes with feature selection methods for BCI competition II dataset Ib. For this dataset, the ICCD optimization with the ‘above the mean’/1.5 IQR rule also better than other schemes.

Table 3 shows the simulation results of Keirn and Aunon’s dataset using 10 × 10 cross-validation. ICCD optimization with the ‘above the mean’/1.5 IQR rule yields the highest classification accuracy in average, although the performance varies among subjects. The simulation results with the three datasets have shown that ICCD optimization with the ‘above the mean’/1.5 IQR rule perform the best in terms of cognitive task classification.

Fig. 4 shows the EEG channel pairs selected by ICCD optimization with the ‘above the mean’/1.5 IQR rule for BCI competition II dataset IV. Four EEG channel pairs, C4-C3, C4-CP5, C4-CP3, and C4-CP1, are selected, and channel C4 is included in all selected channel pairs.

Table 4 lists the ICCD values and the corresponding optimal rotation angles of the selected EEG channel pairs by the proposed algorithm based on ICCD optimization with the ‘above the mean’/1.5 IQR rule. The ICCD values for raw EEG pairs are also listed. The classification accuracies with/without the proposed preprocessing are also reported. The classification performance is enhanced by 25%.

Fig. 5 shows the correlation coefficient topography of channel C4 for BCI competition II dataset IV indicating the changes in correlation coefficient pattern due to ICCD optimization. Fig. 5(a) illustrates the topography of the raw EEG signals and fig. 5(b) shows the ICCD optimized EEG signals.

Fig. 6 shows the final EEG channel pairs selected using ICCD optimization with the ‘above the mean’/1.5 IQR rule for BCI competition II dataset Ib. Six EEG channel pairs are selected and ‘2cm frontal of C4’ is included in four of the six selected EEG channel pairs.

Table 5 shows the ICCD values and optimal rotation angles of the EEG channels selected using ICCD optimization with the ‘above the mean’/1.5 IQR rule. The ICCD values and classification accuracy for raw EEG pairs are compared with the optimized ICCD values, and the resulting performance improvement is 6.67%.

B. COMPARISON OF THE PERFORMANCE OF THE PROPOSED METHOD WITH EXISTING CONNECTIVITY BASED METHODS

To confirm the performance of proposed method, we compare the performance of proposed method with the various brain connectivity based methods, such as correlation coefficient based methods (GC [21], CC-RFE [22]), PLV feature based methods (PLV-FS [20], PLV-RFE [22]), coherence feature based method (PDC [19]) and CM/TFM (25) method for three EEG datasets. We use the ICCD optimization with the ‘above the mean’/1.5 IQR rule for proposed method that yields the best classification accuracy.
Table 8 shows the classification accuracy among the methods examined. IV as the former was recorded by a paralyzed patient with dataset Ib is lower than that for BCI competition II dataset and existing methods. The classification accuracy of connectivity based methods.

Table 6 lists the classification accuracies of the proposed method employing ICCD optimization with the ‘above the matrix that performed optimization according to one of three filtered EEG channel pairs are demixed using a rotation applied to select ‘good’ correlation coefficient features. The ASF-mitigates the mild interference of EEG signals using ASF. The ASF-features by preprocessing EEG signals. We roughly mitigate tive task classification using improved correlation coefficient and time-frequency information.

We describe a novel feature extraction method for cognitive task classification using improved correlation coefficient features by preprocessing EEG signals. We roughly mitigate the mild interference of EEG signals using ASF. The ASF-filtered EEG channel pairs are demixed using a rotation matrix that performed optimization according to one of three criteria, i.e., ICCD maximization, WCCD minimization, and FR maximization. The various feature selection methods are applied to select ‘good’ correlation coefficient features. The method employing ICCD optimization with the ‘above the mean’/1.5 IQR rule performs substantially better than existing brain connectivity based feature methods.

APPENDIX

A. ICCD MAXIMIZATION

With the approximation of $\rho_i^{(k,p)}$, the ICCD, $D_i^{(k,p)}(\theta)$, can be written as the follow:

$$D_i^{(k,p)}(\theta) = \left(\bar{\rho}_1^{(k,p)}(\theta) - \bar{\rho}_2^{(k,p)}(\theta)\right)^2$$

$$\approx \left\{\left(\frac{1}{|I_1|} \sum_{i \in I_1} \frac{P(Y_{i,\text{GSD}}^{(k,p)})}{G_i^{(k,p)}} - \frac{1}{|I_2|} \sum_{i \in I_2} G_i^{(k,p)}\right) \sin(2\theta) \right\}$$

$$+ \left\{\left(\frac{1}{|I_1|} \sum_{i \in I_1} 2C(Y_{i,\text{GSD}}^{(k,p)}) - \frac{1}{|I_2|} \sum_{i \in I_2} 2C(Y_{i,\text{GSD}}^{(k,p)})\right) \cos(2\theta) \right\}^2$$

B. WCCD MINIMIZATION

With approximation of $\rho_i^{(k,p)}$, the $W_i^{(k,p)}(\theta)$ can be written as the follow:

$$W_i^{(k,p)}(\theta) = \sum_{c=1}^{2} \frac{1}{|I_c|} \sum_{i \in I_c} \left(\rho_i^{(k,p)}(\theta) - \bar{\rho}_c^{(k,p)}(\theta)\right)^2$$

$$\approx \sum_{c=1}^{2} \frac{1}{|I_c|} \sum_{i \in I_c} L_{i,c}(k,p) \sin^2\left(2\theta + \cos^{-1}\left(\frac{T_i^{(k,p)} - \bar{T}_c^{(k,p)}}{\sqrt{T_i^{(k,p)} - \bar{T}_c^{(k,p)}}^2 + (\bar{T}_i^{(k,p)} - \bar{T}_c^{(k,p)})^2}\right)\right)$$

$$= \sqrt{V^{(k,p)}}^2 + [U^{(k,p)}]^2 \sin(4\theta + \psi_{WCCD}^{(k,p)}) + \sum_{c=1}^{2} \frac{1}{|I_c|} \sum_{i \in I_c} L_{i,c}(k,p)$$

Therefore, the optimal angle minimizing $V_i^{(k,p)}(\theta)$ is approximately give by:

$$\theta_{\text{opt}}^{(k,p)} \approx -\frac{1}{4} \cos^{-1}\left(\frac{V^{(k,p)}}{\sqrt{V^{(k,p)}}^2 + [U^{(k,p)}]^2}\right) - \frac{\pi}{8}.$$  (45)

C. FR MAXIMIZATION

Using the $D_i^{(k,p)}(\theta)$ and $V_i^{(k,p)}(\theta)$, the FR can be simplified as the follow:

$$F_i^{(k,p)}(\theta) = \frac{D_i^{(k,p)}(\theta)}{V_i^{(k,p)}(\theta)}$$

$$\approx \frac{H_i^{(k,p)} - H_i^{(k,p)} \cos[4\theta + 2\psi_{WCCD}^{(k,p)}]}{Y_i^{(k,p)} + Q_i^{(k,p)} \sin(4\theta + \psi_{WCCD}^{(k,p)})}.$$  (46)

where

$$Y_i^{(k,p)} = \sum_{c=1}^{2} \frac{1}{|I_c|} \sum_{i \in I_c} I_{i,c}(k,p)$$  (47)

$$Q_i^{(k,p)} = 2 \sqrt{V^{(k,p)}}^2 + [U^{(k,p)}]^2$$  (48)

The derivation of $F_{k,p}(\theta)$ is obtained as follows: (see eq.49) at the top of the next page.
The optimal angle maximizing $F^{(k,p)}$ is approximately given by: (see eq.50) at the top of this page.

REFERENCES

[1] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, “Brain-computer interface research at the wadsworth center,” IEEE Trans. Rehabil. Eng., vol. 8, no. 2, pp. 222–226, Jun. 2000.

[2] 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 Trans. Rehabil. Eng., vol. 8, no. 2, pp. 164–173, Jun. 2000.

[3] N. Birbaumer, “Brain-computer-interface research: Coming of age,” Clin. Neurophysiol., vol. 117, no. 3, pp. 479–483, Mar. 2006.

[4] W.-Y. Hsa, “EEG-based motor imagery classification using neuro-fuzzy prediction and wavelet fractal features,” J. Neurosci. Methods, vol. 189, no. 2, pp. 295–302, Jun. 2010.

[5] E. Curran, “Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems,” Brain Cognition, vol. 51, no. 3, pp. 326–336, Apr. 2003.

[6] H. Ramoser, J. Müller-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.

[7] C. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter bank common spatial pattern (FB CSP) in brain-computer interface,” in Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.), Jun. 2008, pp. 2390–2397.

[8] J. 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.

[9] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms,” IEEE Trans. Biomed. Eng., vol. 58, no. 2, pp. 355–362, Feb. 2011.

[10] 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.

[11] B. Blankertz, R. Tomioka, S. Lemm, M. kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” IEEE Signal Process. Mag., vol. 25, no. 1, pp. 41–56, Dec. 2008.

[12] 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.

[13] C. Wang, X. Wu, Z. Wang, and Y. Ma, “Implementation of a brain-computer interface on a lower-limb exoskeleton,” IEEE Access, vol. 6, pp. 38524–38534, 2018.

[14] J. A. Gaxiola-Tirado, R. Salazar-Varas, and D. Gutierrez, “Using the partial directed coherence to assess functional connectivity from motor imagery in brain-computer interface,” in Proc. Adv. Neural Inf. Process. Syst. (NIPS), vol. 21, pp. 10–22, Jan. 2013.

[15] M. Hamed, S.-H. Salleh, and A. M. Noor, “Electroencephalographic motor imagery brain connectivity analysis for BCI: A review,” Neural Computat., vol. 28, no. 6, pp. 999–1041, Jun. 2016.

[16] 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.

[17] B. Blankertz, R. Tomioka, S. Lemm, M. kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” IEEE Signal Process. Mag., vol. 25, no. 1, pp. 41–56, Dec. 2008.

[18] 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.

[19] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms,” IEEE Trans. Biomed. Eng., vol. 58, no. 2, pp. 355–362, Feb. 2011.
Y. Park, W. Chung: Novel EEG Correlation Coefficient Feature Extraction Approach

[25] J. Li, Y. Wang, L. Zhang, A. Cichocki, and T.-P. Jung, “Decoding EEG in cognitive tasks with time-frequency and connectivity masks,” IEEE Trans. Cognit. Develop. Syst., vol. 8, no. 4, pp. 298–308, Dec. 2016.

[26] 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.

[27] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature Extraction: Foundations and Applications. Berlin, Germany: Springer, 2006.

[28] 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.

[29] 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, 2019.

[30] S. Haykin and B. Widrow, Least-Mean-Square Adaptive Filters. Hoboken, NJ, USA: Wiley, 2003.

[31] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schroder, and N. Birbaumer, “The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials,” IEEE Trans. Biomed. Eng., vol. 51, no. 6, pp. 1044–1051, Jun. 2004.

[32] Z. A. Keirn and J. I. Aunon, “A new mode of communication between man and his surroundings,” IEEE Trans. Biomed. Eng., vol. 37, no. 12, pp. 1209–1214, Dec. 1990.

[33] R. Palaniappan, “Utilizing gamma band to improve mental task based brain-computer interface design,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 14, no. 3, pp. 299–303, Sep. 2006.

---

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, electroencephalogram 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.

* * *