Ternary ECOC Classifiers Coupled with Optimized Spatio-Spectral Patterns for Multiclass Motor Imagery Classification

Soroosh Shahtalebi, Student Member, IEEE and Arash Mohammadi, Member, IEEE
Concordia Institute for Information System Engineering (CIISE), Concordia University, Montreal, QC, Canada H3G 1M8 Emails: {s_shahta, arashmoh}@encs.concordia.ca

Abstract—Modeling and representation of multiple tasks from brain signals is a crucial task in Electroencephalogram (EEG) based Brain-Computer Interfaces (BCIs). The motivation of this work comes from the need for a BCI system, intended to operate in real world scenarios, to discriminate multiple tasks and activities simultaneously. In this regard, the paper proposes a novel multi-class EEG-based BCI system via utilization of error correcting output coding (ECOC) classifiers. To best of our knowledge, the ECOC classifiers have not ever been applied to the EEG classification problems. In the ECOC method, the classification problem is modeled as communication over a noisy channel where the miss-classification error is corrected by error correcting techniques borrowed from information theory. In this work, we propose to utilize a modified version of the ECOC classifiers adopted to EEG classification problems which deploys ternary class codewords. Therefore, we analyze more combinations of the classes and greater number of classifiers vote for the final result. The proposed classifier is coupled with a Bayesian framework to compute the optimized spatio-spectral filters to extract the most discriminative feature sets of different classes. The proposed framework is applied to a motor imagery classification problem and evaluated over BCI Competition IV-2a dataset where the results indicate a noticeable enhancement over other methods developed for multi-class EEG classification.

Index Terms: Brain-computer interface (BCI), Common spatial patterns, Electroencephalogram (EEG), Error correction output coding, Motor Imagery.

I. INTRODUCTION

Recently, there has been a surge of interest in development of brain-computer interfaces (BCIs) to provide a non-muscular connection between brain and an external device such as a rehabilitation robot [1], [2]. BCI is a key member of human-in-the-loop cyber-physical systems (CPSs) [3], a new class of systems promoting innovative research that further augment human’s interaction with physical world. The BCI systems have several practical applications of engineering importance such as rehabilitation/assistive systems [4]–[7], and controlling a wheelchair or neuro-prosthesis for disabled individuals [8]. In particular, the BCIs are being used more rigorously in therapeutic applications where brain plasticity properties are deployed for better rehabilitation [4], [5]. Rehabilitation-based BCIs are becoming pervasive in practical rehabilitation applications including neuro-feedback (NFB), therapy for autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), schizophrenia, and motor rehabilitation for post-stroke patients to name a few.

BCI systems typically consist of two main modules, the brain imaging module and the processing unit. Different techniques have been developed for brain imaging such as: (i) Electroencephalogram (EEG), (ii) Magnetoencephalogram (MEG), (iii) Near-Infrared spectroscopy (NIRs), (iv) functional near-infrared spectroscopy (fNIRs), and any hybrid combination of the aforementioned techniques. Since the EEG approach is more affordable than any other technique and also provides higher temporal resolution, it is considered as the prime choice for any type of practical BCI system. Processing of EEG signals typically consists of two main steps: feature generation and feature translation (classification). While the latter tries to make sense of the previously extracted features, the former aims at extracting relevant information from raw signals and avoid the so-called “curse of dimensionality”. Several motor-related EEG signal modalities have been investigated in the literature among which typically sensorimotor activities [9] is considered as the leading developed modality. Frontal and parietal cortices exhibit rhythmic activity in the 8-12 Hz and 13-30 Hz ranges, respectively called μ and β rhythms. In case of an imagery movement some channels in specific frequency bands show increase in power, a phenomenon referred to as event-related synchronization (ERS), while some other frequency bands show decrease in power, referred to as event-related desynchronization (ERD).

The Common Spatial Patterns (CSP) [10] is an effective tool for discriminating imagery movements from EEG recordings and is widely used to detect abnormalities in EEG signals. The CSP algorithm introduces spatial filters for multi-channel EEG recordings to better locate and extract ERD and ERS waveforms. Consequently, the CSP methodology enhances EEG channels containing higher weights for the ERD and the ERS. Several recent works have shown that performance of the CSP for discriminating motor imagery (MI) tasks is superior in comparison to its counterparts. Therefore, the CSP has been extended and improved from different aspects to enhance its classification performance, e.g., filter bank common spatial patterns (FBCSP) [11], regularized common Spatial patterns (RCSP) [12]–[14], Bayesian spatio-spectral filter optimization (BSSFO) [16], and separable common spatio-spectral patterns
The performance of such artificial interfaces, however, rarely matches that of humans rendering their practical applications considerably limited. This calls for more in-depth research on alternative solutions to further enhance the performance of BCIs and increase their applicability which is the focus of this paper.

The classification step in any type of BCI system is an uncompromising step and the quality of the classifier may affect the overall performance of the system. The need for an appropriate classifier raises when it comes to multi-class classification problems. Although several techniques for multi-class problems are developed in the literature, yet their performance cannot outperform the binary classifiers. The error correcting output coding (ECOC) is one of well regarded methods for multi-class problems and is widely used for different multiclass classification problems such as text-classification. To the best of our knowledge, however, the ECOC has not yet been applied to motor imagery classification problems. The paper addresses this gap.

The ECOC technique splits the problem of multi-class classification to a number of binary sub-classification problems. The action of how to split the problem, requires generating proper codewords. Up until now, the binary codewords for classes have always been used, but in this paper, we propose an extension to ECOC approach for its adoption to EEG signal processing. More specifically, we propose to deploy a modified version of ECOC classifiers which is based on “ternary class codeword”, instead of binary ones. Ternary class codewords provide us with a scheme to analyze all possible combinations of the classes and reach higher performances in comparison to existing approaches. In brief, the proposed ternary codewords offer the following three distinct benefits: (i) Increased performance over conventional ECOC method which utilizes binary codewords; (ii) Increased performance over other multi-class classification approaches, and; (iii) Increased hamming distance between class codewords which reduces the misclassification error. Referred to as the ternary ECOC-Bayesian spatio-spectral filtering (TECOC-BSSF), in the proposed framework the TECOC classifier is coupled with a Bayesian framework to derive the optimized spatio-spectral filters [16], to extract the discriminative features from EEG recordings. The results show that the proposed TECOC-BSSF significantly improves the classification performance.

The rest of the paper is organized as follows: Section II formulates the problem. The proposed TECOC-BSSF is developed in Section III. Simulation results are provided in Section IV. Finally Section V concludes the paper.

II. PROBLEM FORMULATION

Throughout the paper, the following notations are used: non-bold letter $x$ denotes a scalar variable, lowercase bold letter $x$ represents a vector, and capital bold letter $X$ denotes a matrix. The real domain is represented by $\mathbb{R}$. The transpose and trace of a matrix $X$ are, respectively, denoted by $X^T$, and $\text{Tr}(X)$.

We consider supervised learning from EEG signals based on the available set of EEG epochs (trials) denoted by $X_i \in \mathbb{R}^{N_a \times N_t}$, for $(1 \leq i \leq N_{\text{trial}})$, where $N_{\text{trial}}$ is the total number of trials used for processing; $N_{\text{ch}}$ is the number of EEG channels (electrodes); and $N_t$ is the number of time samples collected from each electrode in one trial. The training dataset is denoted by $\{(X_i, \Omega_i)\}$, for $(1 \leq i \leq N_{\text{trial}})$, where $\Omega_i$ represents the label corresponding to the $i$th trial, e.g., $\Omega_i$ could be “MI of right hand”, “MI of left foot”, or “MI of left hand”. Before processing EEG signals for classifying MI tasks, typically, a pre-processing step is applied. In this stage, initially the power line interference is removed by applying a notch filter followed by bandpass filtering to extract required frequency contents. The pre-processing step is commonly followed by constructing second-order statistics of the EEG epochs, i.e., computing the sample covariance matrix corresponding to each trial $X_i$. Since $X_i$ is obtained from bandpass filtering of an EEG signal, all classes have zero mean, therefore, the discriminant information contained in the second-order statistics of the data can be represented, instead, by the normalized spatial covariance matrix given by

$$C_i = \frac{X_i X_i^T}{\text{Tr}(X_i X_i^T)}.$$

Then properly designed features are extracted from the normalized spatial covariance matrices and provided as input to the classifier for performing the MI classification task. In this paper, we use CSP features [10] as the CSP exploits the information contained in labels in a supervised manner. Intuitively speaking, the CSP methodology uses a linear transformation matrix to project multi-channel EEG signals into a lower-dimensional spatial subspace. The projection matrix is used to maximize the variance of two-class signals by simultaneous diagonalization of the normalized spatial covariance matrix of EEG signals corresponding to the two class. This completes a brief presentation of the problem at hand. Next, we present the proposed TECOC-BSSF framework.

III. THE TECOC-BSSF FRAMEWORK

In the following sub-sections, we present the proposed TECOC-BSSF which is a novel framework to discriminate $N_c > 2$ different imagery movement classes from EEG recordings. The TECOC-BSSF is designed by coupling Bayesian spatio-spectral filtering optimization [16] with a modified version of the ECOC classifiers.

A. The TECOC-based EEG Modeling

Prior to introducing the proposed modified version of the ECOC classification method, we briefly review the core idea behind the conventional one. The conventional ECOC typically uses a binary coding matrix consisting of $N_c$ bit-vectors of length $N_{\text{Bits}}$. The set of all bit-vectors (the coding matrix) is denoted by $C$, where the $i$th row $C_i$ is called “codeword” which is a unique bit-vector corresponding to class $i$, for $(1 \leq i \leq N_c)$. Each codeword denoted by $\Lambda = \{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N_{\text{Bits}})}\}$ is of a pre-defined length $N_{\text{Bits}}$. The ECOC algorithm constructs an individual binary classifier for each column of the coding matrix $C$. Classifier $\gamma^{(j)}$, for $(1 \leq j \leq N_{\text{Bits}})$, has positive
Fig. 1. All possible 4 bit binary codes. The possible classifiers to discriminate 4 classes in conventional ECOC method are numbered in red color from 1 to 7.

instances for each class $i$ when $C_{i(j)} = 1$. In other words, each classifier decides on whether their corresponding bit is zero or one resulting into two super-sets $S^{(0)}$ and $S^{(1)}$ for each column of the coding matrix. The conventional approach is to consider a multi-class setting consisting of ($3 \leq N_c \leq 7$) different classes, and construct codes of length $N_{Bts} = 2^{N_c-1} - 1$ based on the following procedure: (i) All the entries in the first row are ones and; (ii) The $i$th row consists of alternating runs of $2^{N_c-1}$ zeros and ones.

Now let us intuitively elaborate on the idea behind utilization of code words of length $2^{N_c-1} - 1$ in the conventional ECOC algorithm with an example. Consider the scenario when $N_c = 4$ classes, first we try to find all the possible ways that labels 0 and 1 can be assigned to any of the classes. In practice, we are defining all the binary numbers that could be formed with 4 bits. Hence, we have 16 binary numbers between 0000 and 1111. Fig. 1 depicts all the 16 possible codewords. As it is labeled in Fig. 1, we are interested in the cases that both $S^{(0)}$ and $S^{(1)}$ super sets could be formed, therefore, the first (all zero) and the last (all 1) cases are not suitable for the classification problem. Among the rest of the cases, it is observed that each case has its complement among the others. Since the 0s and 1s are just the labels for classes, there is no difference between a binary number and its complement (e.g., the 2nd column from left and right in Fig. 1 are the same). Hence, we only need to use half of the remaining binary numbers, i.e., out of 16 possibilities, 14 are acceptable, and these 14 cases reduce to 7 complement pairs. Therefore, in total we have $2^3 = 8$ exclusive codewords.

We believe that the drawback of the ECOC method, introduced above, is that in each of the cases that we form the super sets ($S^{(0)}$ and $S^{(1)}$), all the classes are involved and there is no chance to examine the classifier $\gamma^{(j)}$ in absence of features of a certain class. Therefore, we propose utilization of the ECOC classifiers which are designed in “Ternary System” [17], [18]. Referred in this paper as TECOC, the possible labels are now 0, 1 and 2 (ternary system). In this work, we take 0 and 1 as labels to form the super sets, and the classes with label 2, are not involved in the classifier design. More specifically, in the ternary system, for 4 classes of data, we can form 81 different numbers. By performing the similar procedure which was described for the conventional ECOC method for removing the complements and meaningless cases, and taking this fact into account that label "2" means that the class is not involved into the classifier training, we derive 25 exclusive cases. Any 4 digit ternary number which does not consist of either 0 or 1, is defined as meaningless.

Training the TECOC classifier comprises of learning a set $\Lambda = \{\gamma^{(1)}, \ldots, \gamma^{(N_{Bts})}\}$ of independent binary classifiers. Based on the learned $\Lambda$, the correct class of an unlabeled trial $X_i$ is hypothesized as follows: Evaluate each independent classifier based on $X_i$ resulting in generation of $\Lambda(X_i) = \{\gamma^{(1)}(X_i), \ldots, \gamma^{(N_{Bts})}(X_i)\}$. Most likely, the generated bit-vector $\Lambda(X_i)$ will not be a row of $C$, but it will certainly be closer (in Hamming distance $\Delta$) to some rows than to others. Trial $X_i$ is categorized as follows

$$\Phi(X_i) = \text{argmin}_\Delta\{\Delta(C_i, \Lambda(X_i))\},$$

where $\Delta(a, b)$ is the number of bits in which vectors $a$ and $b$ differ. In other words, when classifying a new signal $X_i$, we compute the Hamming distance between $\Lambda(X_i)$ and all available codewords $C_i$. We assign the new signal to class $i$ if it has the minimum distance among other classes. It is worth mentioning that the proposed TECOC also benefits from introducing extra Hamming distance between class codewords which eliminates the effect of misclassification for a certain number of classes. Features required for classificatin/training of $N_{Bts}$ binary classifiers are extracted from EEG epochs as described in the next sub-section.

B. CSP-based Feature Extraction/Classification

In the proposed TECOC-BSS framework, each classifier $\gamma^{(j)}$, for ($1 \leq j \leq N_{Bts}$), uses its specific features, i.e., features are classifier specific. In this regard, a set of class discriminative frequency bands and their associated spatial filters are derived within the Bayesian framework for each classifier. The goal is to jointly address couple of critical drawbacks of the conventional CSP approach and its extensions. The TECOC-BSS iteratively performs spectral and spatial filtering as outlined below.

Spectral Filtering: Similar to the BSSFO algorithm [16], we model the uncertainty in the cut-off frequencies of spectral filters with a prior probability denoted by $p(B)$ over random variable $B$. The prior density $p(B)$ describes relative probabilities of different states (frequency bands) in which a single-trial EEG recording is correctly discriminated. The posterior probability distribution (PDF) denoted by $p(B|X_i, \Omega_i)$ is then computed given a single-trial EEG recording $X_i$, for ($1 \leq i \leq N_{Trial}$), and its corresponding label denoted by $\Omega_i$ based on the Bayes rule as follows

$$p(B|X_i, \Omega_i) = \frac{p(X_i, \Omega_i|B)p(B)}{p(X_i, \Omega_i)}.$$  

(3)

Given the frequency band $B$ and a raw EEG signal $X_i$, the bandpass-filtered signal $Z^{(j)}_i$ is deterministically obtained. Therefore, the likelihood and the evidence are equal to $p(Z^{(j)}_i|\Omega_i, B)$ and $p(Z^{(j)}_i|\Omega_i)$, respectively. Hence, we can rewrite Eq. (3) by replacing the raw EEG signal $X_i$ with its bandpass-filtered version $Z^{(j)}_i$ as

$$p(B|Z^{(j)}_i, \Omega_i) = \frac{p(Z^{(j)}_i, \Omega_i|B)p(B)}{p(Z^{(j)}_i, \Omega_i)}.$$  

(4)
Algorithm 1 TECOC-BSSF IN TRAINING PHASE.

Input: EEG Trials: \( \{ X_i \}^{N_{\text{Trials}}} \); Labels: \( \{ \Omega_i \}^{N_{\text{Trials}}} \) with \( N_c \) distinct MI classes, and; Number of classifiers: \( N_{\text{Bits}} \).

Output: The coding matrix: \( \mathbf{C} \); \( N_{\text{Bits}} \) set of optimized spatio-spectral filters \( \{ h(j), W(j) \} \), and; \( N_{\text{Bits}} \) trained binary classifiers \( \{ \gamma^{(1)}, \ldots, \gamma^{(N_{\text{Bits}})} \} \).

S1. **Codebook Generation:** Generate the \(( N_c \times N_{\text{Bits}})\) binary coding matrix \( \mathbf{C} \).

S2. **Classifier Design:**
1. for \( 1 \leq j \leq N_{\text{Bits}} \) do
2. Construct two super-sets, \( S^{(j,0)} \) and \( S^{(j,1)} \) where \( S^{(j,1)} \) consists of all trials for which \( C_{ij} = 1 \), and \( S^{(j,0)} \) consists of all trials for which \( C_{ij} = 0 \). Trials with \( C_{ij} = 2 \) are dismissed.
3. Initialize all the particles \( \{ Z^{(j)}(k), p^{(j)}(k) \}^{N_p}_{k=1} \), where \( p_k = \frac{1}{N_p} \).
4. Perform the Bayesian spatio-spectral filtering the \( S^{(j,0)} \) and \( S^{(j,1)} \) super sets resulting is set of selected particles: \( \Pi^{(j)} = \cup_k (\pi_k > \tau) ; k \in 1, 2, \ldots, N_p \), updated particles \( Z^{(j)} \), and associated optimized spatial filters \( W^{(j)} \).
5. Extract features from the spectral and spatial filtered signals.
6. Construct a binary classifier \( \gamma^{(j)} \) to distinguish \( S^{(j,0)} \) from \( S^{(j,1)} \).

7: end for

The posterior \( p(\mathbf{B} | Z^{(j)}_i, \Omega_i) \) provides all the required information regarding \( \mathbf{B} \) which can be obtained from the bandpass-filtered signal \( Z^{(j)}_i \) and its corresponding class label \( \Omega_i \).

**Spatial Filtering:** A spatial filter \( W^{(j)} \) is computed from \( Z^{(j)}_i \) by applying standard CSP approach. More specifically, for trial \( X_i \), classifier \( \gamma^{(j)} \), for \( 1 \leq j \leq N_{\text{Bits}} \), first derives the normalized spatial covariance matrix denoted by \( C^{(j)} \) based on Eq. (1). As the goal of the CSP is to discriminate two classes of data, we define \( C^{(j,0)} \) and \( C^{(j,1)} \) as the average of spatial covariance matrices of different trials belonging to each super-set \( S^{(j,0)} \) and \( S^{(j,1)} \). Based on the computed average covariance matrices \( C^{(j,0)} \) and \( C^{(j,1)} \), the composite spatial covariance matrix denoted by \( C^{(j,c)} \) is computed as \( C^{(j,c)} = C^{(j,0)} + C^{(j,1)} \).

Next, eigenvalue decomposition is performed as \( C^{(j,c)} = U^{(j,c)} \lambda^{(j,c)} (U^{(j,c)})^T \), where \( U^{(j,c)} \) is the matrix of eigenvectors associated with the composite covariance, and \( \lambda^{(j,c)} \) is the diagonal matrix of its corresponding eigenvalues. In the next step a whitening transform is applied on \( U^{(j,c)} \) as \( P^{(j)} = \sqrt{\frac{\lambda^{(j,c)}}{\lambda^{(j,c)}}} [U^{(j,c)}]^T \).

Intuitively speaking, the whitening operator scales the variance in the space spanned by \( U^{(j,c)} \), i.e., all the eigenvalues of \( P^{(j)} C^{(j,c)} (P^{(j)})^T \) are equal to one. Using the whitening matrix, the average covariance matrices \( (C^{(j,0)} \) and \( C^{(j,1)} \) are transformed as\( S^{(j,0)} = P^{(j)} C^{(j,0)} (P^{(j)})^T \) and \( S^{(j,1)} = P^{(j)} C^{(j,1)} (P^{(j)})^T \), therefore, \( S^{(j,0)} \) and \( S^{(j,1)} \) share common eigenvectors denoted by \( B^{(j)} \), i.e., \( S^{(j,0)} = B^{(j)} \lambda^{(j,0)} (B^{(j)})^T \) and \( S^{(j,1)} = B^{(j)} \lambda^{(j,1)} (B^{(j)})^T \), with \( \lambda^{(j,0)} + \lambda^{(j,1)} = I \), where \( I \) denotes an identity matrix of appropriate dimension. The TECOC-BSSF projection matrix corresponding to classifier \( \gamma^{(j)} \) for \( 1 \leq j \leq N_{\text{Bits}} \), is then given by \( W^{(j)} = [P^{(j)}]^T B^{(j)} \), which is used to form the decomposition (mapping) of each trial \( X_i \), for \( 1 \leq i \leq N_f \) as follows

\[
W^{(j)}_i = [P^{(j)}]^T Z^{(j)}_i.
\]

As the variances of only a small number \( m \) of signals are suitable for discrimination analysis, only the first and last \( m \) rows of \( W^{(j)}_i \) are used for the construction of the classifier. In other words, matrix \( W^{(j)}_i \) is constructed from the first and last \( m \) rows of matrix \( Z^{(j)}_i \) which represents rows of \( W^{(j)}_i \) associated with the largest eigenvalues that maximizes the difference of variance between two super-sets

\[
f^{(j)}_{i,p} = \log \left( \frac{\text{var}(W^{(j)}_{i,p})}{\sum_{k=1}^{2m} \text{var}(W^{(j)}_{k,p})} \right),
\]

where \( \text{var}(\cdot) \) denotes the variance operator. Note that, the log-transformation in Eq. (6) is included to approximate normal distribution of the data.

In brief, a feature vector is extracted by multiplying \( Z^{(j)}_i \) with the spatial filter \( W^{(j)} \) which is computed from its second-order statistics followed by applying a monotonically increasing logarithmic function. Therefore, the posterior in Eq. (4) can be rewritten as follows

\[
p(\mathbf{B} | Z^{(j)}_i, \Omega_i) \propto p(\mathbf{B} | f^{(j)}_{i,p}, \Omega_i) = \frac{p(f^{(j)}_{i,p} | \Omega_i, \mathbf{B})p(\mathbf{B})}{p(f^{(j)}_{i,p} | \Omega_i)}.
\]

The goal to find the optimal spatio-spectral filter for discriminative feature extraction, therefore, can be defined as estimation of the posterior distribution in Eq. (7). However, term \( p(f^{(j)}_{i,p} | \Omega_i, \mathbf{B}) \) on the right hand side (RHS) of Eq. (7) is too complex in nature resulting in complex \( p(\mathbf{B} | f^{(j)}_{i,p}, \Omega_i) \) eliminating the possibility of direct evaluation in a closed-form. Alternatively, particle-based approximation techniques [19], [20] are utilized to address this issue. In brief, a set of \( N_p \) particles \( \{\mathbb{B}^{(j)}(k)\}^{N_p}_{k=1} \) generated from the prior density \( p(\mathbf{B}) \) are utilized where \( \mathbb{B} \) denotes a particle representing a single frequency band. The weight of each particle is computed as

\[
\pi_{i,p}^{(j)} (k) = \frac{p(f^{(j)}_{i,p} | \Omega_i, \mathbb{B}^{(j)}(k))}{\sum_{i=1}^{N_p} p(f^{(j)}_{i,p} | \Omega_i, \mathbb{B}^{(j)}(k))}
\]
Algorithm 2 TECOC-BSSF IN EVALUATION PHASE

Input: Unlabeled EEG Trial $X_{Test}$; The coding matrix: $C$; $N_{Bits}$ trained binary classifiers $\{\gamma^{(1)}, \ldots, \gamma^{(N_{Bits})}\}$, and $N_{Bits}$ set of optimized spatio-spectral filters.

Output: The assigned class label: $\Phi(X_{Test})$.

S1. Spatial filtering of $X_{Test}$ based on the optimized frequency bands for each set of particles $Z^{(j)} = h^{(j)} * X_{Test}$, where $*$ denotes convolution operator.

1: for $1 \leq j \leq N_{Bits}$ do
2: Spatial filtering of $Z^{(j)}$ with optimized spatial filters $W^{(j)}(Y^{(j)}) = (W^{(j)})^{T} \times Z^{(j)}$.
3: Extract classifier-specific features (transforming matrices) using Eq. (6).
4: end for

S2. Compute the $N_{Bits}$-bit codeword ($\Lambda(X_{Test})$).

S3. Compute the class of the test epoch using Eq. (2).

where $f_{i,p}^{(j)}(k)$ denotes a feature vector set extracted from the spectrally (using $B^{(j)}(k)$) and spatially (using $(W^{(j)}(k))$ filtered signal. Consequently, the set of particles and their associated weights $\{B_{k}, \pi_{k}\}_{k=1}^{N_{p}}$ approximate the distribution of the posterior and converges to the true density as the number of particles increases. For further details on particle filter please refer to [16], [19], [20]. The output is the particle set $\{B^{(j)}(k), \pi_{k}^{(j)}(k)\}_{k=1}^{N_{p}}$ and the spatial filter set $\{W^{(j)}(k)\}_{k=1}^{N_{p}}$. In other words, the Bayesian step forms the class-discriminative frequency bands, and the spatial patterns, one for each band. The weights $\{\pi^{(j)}(k)\}_{k=1}^{N_{p}}$ are used in constructing a classifier, i.e., the filter bank is composed as

$$\Pi^{(j)} = \bigcup_{k} \pi^{(j)}(k) > \tau, \quad $$ (9)

where $(1 \leq k \leq N_{p})$ and $\tau$ denotes a threshold parameter that is determined empirically. This completes development of the proposed TECOC-BSSF which is summarized in Algorithm 1.

IV. SIMULATIONS

The performance of the proposed TECOC-BSSF method is evaluated on the database form BCI competition IV-2a [21]. The dataset consists of four classes of motor imagery EEG measurements (Right hand MI, Left hand MI, Feet MI, and Tongue MI) obtained from nine subjects. Signals are recorded at sampling rate of 250Hz, using 22 EEG channels and 3 monopolar electrooculogram (EOG) channels (with left mastoid serving as the reference). The original EEG signal recordings are already bandpass filtered (0.5-100Hz) and notch filtered. For each subject, two sessions are recorded (one for training purposes and the other one for evaluation). Each session consists of six segments and each segment consists of 48 trials of length 3 seconds. In total and for each subject, 288 trials for training and 288 trials for evaluation are available. In order to measure the performance of the proposed framework and based on the recommendation from BCI competition [21], kappa coefficient $\kappa$ is used, i.e., $\kappa = \frac{CCR - P_{rand}}{1 - P_{rand}}$, where CCR represents the correct classification rate, and the value of $P_{rand}$ for this dataset is equal to 0.25. For both training and evaluation sessions, segments of 2 seconds (500 samples) from each trial are selected as the input to the TECOC-BSSF algorithm. The segment starts 0.5 seconds after the onset cue of each trial. In the training stage, ternary codewords are assigned to the four MI classes. According to the procedure described in section III-A, the generated codewords for the 4 class classification problem is as introduced in Table I.

In order to evaluate the algorithm over unseen data, each trial of the dataset which is labeled as “unknown” is selected and similar to the training step, a segment of 2 seconds length, from 0.5 to 2.5 seconds after the cue is selected. Likewise the training step, the mean of the signal for 2 seconds prior to the cue onset is computed and subtracted from the segment. Then, the spatio-spectral filters which were derived in the training step, are applied on the segment and the corresponding features are extracted. Each set of features is then fed into its corresponding classifier and the output of all classifiers are put together to form a 25 bit codeword $\Lambda(X_{Test})$ which is served as decision of the proposed algorithm over the segment. By deploying Eq. (2), we determine the class of the unknown trial, which is the one that has minimum distance with the derived codeword, i.e., $\Lambda(X_{Test})$. The performance of the proposed algorithm is evaluated on BCI-$IV_{2a}$ dataset and the results are provided in Table II. In order to compare the performance of our proposed approach with successful and state of the art algorithms, the results of five other approaches is also provided in Table II. The FBCSP method [11] is the winner of the BCI competition $IV$ and achieved the highest performance for this dataset. The SCSSP method [15] is one of the newest works on this problem which was published on 2016.

Since the novelty of this work is mainly in the classification step and we are proposing the TECOC classifiers as a new classification approach, we also have compared our results with conventional ECOC method which is based on binary class codewords. In addition, we have provided the simulation results for the two techniques which are served as extensions to binary classifiers, to fit them for multi-class classification problems. The two techniques are called ”One-Versus-One” (OVO) and ”One-Versus-All” (OVA) and they are well regarded in most of the multi-class classification problems. The two methods are believed to yield compromising results in extension of binary classifiers to multi-class classifiers.

In OVO approach, for any possible pair of classes, a classifier is trained, and then for an unknown trial, the final decision is made based on the most frequent class label in
the results of the classifiers. In a generalized example, for $N_c$ number of classes, $(\frac{N_c}{2})$ number of classifiers should be trained.

In OVA, each class out of $N_c$ classes should be discriminated from a superset of all other classes. So $N_c$ number of classifiers should be trained to form the extension of the binary classifiers to the multi-class classifiers. In our current problem of classifying 4 different classes, we have to train 4 classifiers for the following sets: $[1], [2, 3, 4], [2, 1, 3, 4], [3, 1, 2, 4], [4, 1, 2, 3]$. The final decision could be based on two criteria: (i) the class label which is reported more frequently than others be selected as the assigned label to the unknown trial or, (ii) the label which is reported with higher score be assigned to the unknown trial.

V. Conclusion

A successful BCI system needs to discriminate different patterns of signals accurately. This fact motivates us to design a reliable and efficient signal processing framework. Typically the processing unit of a BCI system consists of the following major steps: pre-processing, feature extraction, feature selection, and classification. One of the methods which is well regarded in feature extraction step for motor imagery classification, is CSP approach. In this work we deployed a Bayesian framework to optimize the spatio-spectral filters based on a CSP approach, to extract the most discriminative features. In addition, we proposed and utilized the ternary codewords for ECOC classifiers, referred to as TECOC, to enhance the performance of multi-class classification step. The proposed approach is evaluated over BCI Competition IV-2a dataset which is a four class classification problem of MI movements. The results indicate a noticeable performance improvement in comparison to its counterparts.

REFERENCES

[4] J. J. Daly, J. R. Wolpaw, “Brain-Computer Interfaces in Neurological Reha-

![Table II: Performance comparison for different approaches, tested on BCIC – IV2a dataset. Performance measure is in kappa (κ) value.](image)

![Table II: Performance comparison for different approaches, tested on BCIC – IV2a dataset. Performance measure is in kappa (κ) value.](image)

![Table II: Performance comparison for different approaches, tested on BCIC – IV2a dataset. Performance measure is in kappa (κ) value.](image)