



Multi-Frequent Band Collaborative EEG Emotion Classification Method Based on Optimal Projection and Shared Dictionary Learning

Jiaqun Zhu[†], Zongxuan Shen[†] and Tongguang Ni^{*}

School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

OPEN ACCESS

Edited by:

Yuanpeng Zhang,
Nantong University, China

Reviewed by:

Lu Chen,
Nanjing Institute of Technology (NJIT),
China

Runmin Liu,
Wuhan Sports University, China

*Correspondence:

Tongguang Ni
ntg@cczu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Alzheimer's Disease and Related
Dementias,
a section of the journal
Frontiers in Aging Neuroscience

Received: 04 January 2022

Accepted: 27 January 2022

Published: 17 February 2022

Citation:

Zhu J, Shen Z and Ni T (2022)
Multi-Frequent Band Collaborative
EEG Emotion Classification Method
Based on Optimal Projection
and Shared Dictionary Learning.
Front. Aging Neurosci. 14:848511.
doi: 10.3389/fnagi.2022.848511

Affective computing is concerned with simulating people's psychological cognitive processes, of which emotion classification is an important part. Electroencephalogram (EEG), as an electrophysiological indicator capable of recording brain activity, is portable and non-invasive. It has emerged as an essential measurement method in the study of emotion classification. EEG signals are typically split into different frequency bands based on rhythmic characteristics. Most of machine learning methods combine multiple frequency band features into a single feature vector. This strategy is incapable of utilizing the complementary and consistent information of each frequency band effectively. It does not always achieve the satisfactory results. To obtain the sparse and consistent representation of the multi-frequency band EEG signals for emotion classification, this paper propose a multi-frequent band collaborative classification method based on optimal projection and shared dictionary learning (called MBCC). The joint learning model of dictionary learning and subspace learning is introduced in this method. MBCC maps multi-frequent band data into the subspaces of the same dimension using projection matrices, which are composed of a common shared component and a band-specific component. This projection method can not only make full use of the relevant information across multiple frequency bands, but it can also maintain consistency across each frequency band. Based on dictionary learning, the subspace learns the correlation between frequency bands using Fisher criterion and principal component analysis (PCA)-like regularization term, resulting in a strong discriminative model. The objective function of MBCC is solved by an iterative optimization algorithm. Experiment results on public datasets SEED and DEAP verify the effectiveness of the proposed method.

Keywords: cognitive computing, EEG-based emotion classification, multi-frequency band EEG signals, subspace learning, dictionary learning

INTRODUCTION

Affective computing focuses on how to actively learn, reason, and perceive the surrounding world, as well as realize a certain level of brain-inspired cognitive intelligence by simulating people's psychological cognitive processes (Aranha et al., 2019; Samsonovich, 2020). Researchers in psychology and neurobiology investigate the changes and relationships in the human physiological

systems that occur during various emotional states and activities (Li et al., 2020). More and more evidences show that with the progress of neuroscience research, there is a connection between human emotional activity and the activity of specific areas of the brain, especially the cerebral cortex and central nervous system. For example, the amygdale is associated with emotions like fear and anxiety in the limbic system of the brain. Anger can activate the left frontal lobe of the brain (Davis and Whalen, 2001). Researchers have also studied the relationship between certain diseases and emotional activities, such as cancer, cardiovascular disease, and depression (Zhao et al., 2018). Wirkner et al. (2017) and Guil et al. (2020) studied the impact of emotional activity on the progression of breast cancer patients. Nurillaeva and Abdumalikova (2021) studied the pathways of communication between the heart and the brain, as well as the relationship between heart rhythm and cognitive and emotional functions. According to the study of Gianaros et al. (2014), there is a link between affective regulation and cardiovascular disease. The author discussed how intense emotional activity and the immune system interact, and how these close interactions affect the treatment of rheumatic cardiovascular disease. Tennant and McLean (2001) associated mood disorders such as anxiety, depression, and anger with coronary heart disease. Authors classified mood disorders as an important risk factor for coronary heart disease, and concluded that mood disorders are frequently associated with coronary heart disease events. Klatzkin et al. (2021) studied the food intake of emotional dieters during various emotional and stress responses. Researchers are also interested in the impact of emotional activities in the business field. According to research on the effect of emotion on commercial advertising, advertisements with emotional expression and influence are easier for consumers to remember, and publicity images with emotional color can influence consumers' access behavior (Shareef et al., 2018). It is clear that research on human emotional activities is important not only in the study and understanding of humanity, but also in medical health and commercial activities. As a result, the study of human emotions, including emotional activity intervention, can be regarded as scientific and practical.

Electroencephalogram activities are closely related to people's psychological attention consciousness and emotional experience. An emotional EEG signal is a physiological electrical signal collected by the human brain in a specific emotional state. EEG signals, as a window into brain thinking activities, cognitive processes, and mental states, are an important technical means for studying brain function and its neural mechanism. Wearable devices placed on the top of the head collect emotional EEG signals. The acquisition electrode's placement position is typically determined using the international standard 10–20 and other systems. Researchers in the field of artificial intelligence study the relationship between emotional activities caused by internal and external stimuli and the content of stimuli. Machine learning technology in artificial intelligence is widely used in EEG signals-based emotion classification. For example, Liu et al. (2020) developed a multi-level features guided capsule network to describe the internal relationship of multiple EEG signal channels. The advantage of this model is that different levels of feature mapping are integrated during

the process of forming the primary capsule, which can improve feature representation ability. Zhong et al. (2020) proposed a regularized graph neural network to mine both local and global relationships between various EEG channels. This method can alleviate the problem of time dependence in emotional process. Ni et al. (2021) developed a domain adaptation sparse representation classification model to alleviate the problem of insufficient training data in the new scene. This method employed the discriminative knowledge of historical data or related data to aid in establishing the classification model of the current scene.

According to intra-band correlation with a distinct psychological state, the EEG signals can be split into five frequency bands. Different frequency band EEG signals reflect the different states of brain state. **Table 1** briefly describes the information of five frequent bands of EEG signals (Gu et al., 2021a; Shen et al., 2021). Many scholars have studied EEG signals in different frequency bands. Mohammadi et al. (2017) used wavelet transform to decompose EEG signals into five sub-band signals, then extracted entropy and energy features from each sub-band signal and sent them to support vector machine and *k*-nearest neighbor, respectively. Li and Lu (2009) proposed a frequency band search method to find the best frequency band for emotion classification. According to their findings, the gamma frequency band is appropriate for EEG-based emotion classification using still images as stimuli. Zheng and Lu (2015) built a Multi-frequent band emotion recognition classifier using deep neural networks. This study had shown that the beta and gamma bands contained more discriminative information for emotion classification. Li et al. (2018) used the hierarchical deep learning model to train numerous classifiers on EEG signals. They verified that high-frequency bands played the most important role in emotion classification. Yang et al. (2018) developed a 3D representation of signal segment to extract representative features on bands. They integrated multiple frequency bands and used the constructed 3D signal cube as model input. Li et al. (2019) developed a sparse linear regression model using the technologies of graph regularization and sparse regularization. The authors compared the effects of different frequency band signals in emotion recognition on various EEG datasets.

Because there are internal relationships and differences between different frequency bands, a new learning method is required to make full use of the information in multi-frequency band data. Despite extensive research on the use of different frequency bands of EEG signals for emotion recognition, one traditional strategy is to directly concatenate features from

TABLE 1 | The basic information of five frequent bands of EEG signals.

Patterns	Frequency	Brain state
Delta (δ)	1–3 Hz	Slowest "sleep waves"
Theta (θ)	4–7 Hz	Light meditation and sleeping
Alpha (α)	8–13 Hz	Closing the eyes, relaxation
Beta (β)	14–30 Hz	Waking consciousness and reasoning waves
Gamma (γ)	30–100 Hz	Sensory and high-level information processing

multi-frequent bands in high dimensional space and consider this single feature vector as the model's input. Obviously, this strategy does not account for the complementarity and consistency of the data in each frequency band.

Our previous work named as optimized projection and Fisher discriminative dictionary learning (OPFDDL) (Gu et al., 2021a) extracted multi-frequent band EEG features in the optimal sparse representation subspace, and adopted the Fisher discrimination criterion to build a discriminative classifier. This method did not directly concatenate the features of each frequency band, but regarded each band signal as an independent feature. It incorporated the band-correlation knowledge into a dictionary learning model by learning independent projection matrices for each frequency band signal. Inspired of this work, we further use multi-frequent band shared information to exploit the intrinsic knowledge of EEG signals and achieve correlation modeling of multiple band data. Thus, in this study we propose a multi-frequent band collaborative EEG emotion classification method based on optimal projection and shared dictionary learning (MBCC). We construct a projection matrix for each frequency band. The projection matrix is composed of a common shared matrix (called shared component) and a frequency band-specific matrix (called specific component). The shared matrix well reflects the relationship between frequency bands. The EEG signal of each frequency band is projected to the subspace through the projection matrix, and the dictionary shared by each frequency band is learned in the subspace. The corresponding sparse representation is then obtained from the new data features using dictionary learning. According to Fisher's criterion, the MBCC method ensures that the coding reconstruction errors of the same class are as small as possible, while the coding reconstruction errors of different classes are as large as possible. Considering the information available in the original domain should not be lost in the projection space, we provide a regularization term similar to principal component analysis (PCA) that can retain discriminative knowledge to improve the discrimination ability of the model. An efficient alternating iterative optimization algorithm is designed to solve the proposed model. The experiment yielded good classification results on the public EEG emotion datasets SEED (Zheng and Lu, 2015) and DEAP (Koelstra et al., 2011).

The advantages of MBCC are as follows: (1) An effective discriminative dictionary is trained using the dictionary learning model framework by capturing common shared feature information from multi-frequency band data. The first correlation between data in multiple frequency bands is represented by the common shared dictionary. It creates a link between data from different frequency bands in order to obtain a new feature representation of EEG data. (2) Take into account the complementarity and difference of frequency band data, the projection matrix of each frequency band has the common shared and independent components. The common shared component reflects the second correlation between multiple frequency bands and can keep each frequency band consistent. (3) To assess the model's discriminative ability, the Fisher criterion based on coding error is introduced in the projection space. Furthermore, the PCA-like regularization term based on

the common shared projection component contributes to obtain more discriminative sparse coding.

BACKGROUND

Let $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_n] \in \mathbf{R}^{d \times n}$ be a set of d -dimensional n training signals. The traditional dictionary learning is to learn a dictionary matrix to sparsely represent the EEG signals \mathbf{Z} . The problem of dictionary learning (Jiang et al., 2013; Gu et al., 2021b) is,

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{A}} \|\mathbf{Z} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1, \\ \text{s.t. } \forall i, \|\mathbf{d}_i\|_0 = 1, \end{aligned} \tag{1}$$

where $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_k] \in \mathbf{R}^{d \times K}$ is the learned dictionary, K is the dictionary size. $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n] \in \mathbf{R}^{K \times n}$ is the sparse coding coefficient matrix. The first term in Eq. (1) is to minimize the reconstruction errors of \mathbf{Z} . The second term is the sparsity constraints.

In our previous work OPFDDL method (Gu et al., 2021a), $\mathbf{Z}^m = [\mathbf{z}_1^m, \dots, \mathbf{z}_{n_m}^m] \in \mathbf{R}^{d \times n_m}$ is the class m frequent band signal set, $m = 1, 2, \dots, M$, $n = \sum_m n_m$. By introducing the frequent band specific projection matrix $\mathbf{G}^m \in \mathbf{R}^{d \times p}$, each training signal \mathbf{z}_j^m is projected into a low-dimensional space, as $\mathbf{G}^m \mathbf{z}_j^m$. Suppose \mathbf{S}_w^m and \mathbf{S}_b^m are within-class and between-class reconstruction errors of the m -th frequent band signals, respectively. \mathbf{S}_w^m and \mathbf{S}_b^m are defined as,

$$\begin{aligned} \mathbf{S}_w^m &= \text{Tr}(\sum_{j=1}^{n_m} ((\mathbf{G}^m)^T \mathbf{z}_j^m - (\mathbf{G}^m)^T \mathbf{D} \delta(\mathbf{a}_j^m)) \times ((\mathbf{G}^m)^T \mathbf{z}_j^m \\ &\quad - \mathbf{G}^{mT} \mathbf{D} \delta(\mathbf{a}_j^m))^T) \\ &= \text{Tr}(\mathbf{G}^{mT} \mathbf{W}_w^m \mathbf{G}^m) \end{aligned} \tag{2}$$

where $\mathbf{W}_w^m = \sum_j (\mathbf{z}_j^m - \mathbf{D} \delta(\mathbf{a}_j^m)) \times (\mathbf{z}_j^m - \mathbf{D} \delta(\mathbf{a}_j^m))^T$. The function $\delta(\mathbf{a}_j^m)$ returns the coding coefficients consistent with the class of \mathbf{z}_j^m .

$$\begin{aligned} \mathbf{S}_b^m &= \text{Tr}(\sum_{j=1}^{n_m} ((\mathbf{G}^m)^T \mathbf{z}_j^m - (\mathbf{G}^m)^T \mathbf{D} \zeta(\mathbf{a}_j^m)) \times ((\mathbf{G}^m)^T \mathbf{x}_j^m \\ &\quad - \mathbf{G}^{mT} \mathbf{D} \zeta(\mathbf{a}_j^m))^T) \\ &= \text{Tr}((\mathbf{G}^m)^T \mathbf{W}_b^m \mathbf{G}^m) \end{aligned} \tag{3}$$

where $\mathbf{W}_b^m = \sum_j (\mathbf{z}_j^m - \mathbf{D} \zeta(\mathbf{a}_j^m)) \times (\mathbf{z}_j^m - \mathbf{D} \zeta(\mathbf{a}_j^m))^T$. The function $\zeta(\mathbf{a}_j^m)$ returns the coding coefficients not consistent with the class of \mathbf{z}_j^m .

According to the classification rule of Fisher criterion (Peng et al., 2020; Zhang et al., 2021), the OPFDDL method proposes the discriminative model on M frequent bands in the projection space,

$$\begin{aligned} \min_{\mathbf{G}^m, \mathbf{D}} \frac{\sum_m \text{Tr}(\mathbf{G}^{mT} \mathbf{W}_w^m \mathbf{G}^m)}{\sum_m \text{Tr}(\mathbf{G}^{mT} \mathbf{W}_b^m \mathbf{G}^m)}, \\ \text{s.t. } (\mathbf{G}^m)^T (\mathbf{G}^m) = \mathbf{I}, m = 1, 2, \dots, M \end{aligned} \tag{4}$$

Then the matrices $\tilde{\mathbf{G}}$, $\tilde{\mathbf{W}}_w$, and $\tilde{\mathbf{W}}_b$ are defined as $\tilde{\mathbf{G}} = [\mathbf{G}^1, \mathbf{G}^2, \dots, \mathbf{G}^M]$, $\tilde{\mathbf{W}}_w = \begin{bmatrix} \mathbf{W}_w^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W}_w^M \end{bmatrix}$, and

$$\tilde{\mathbf{W}}_b = \begin{bmatrix} \mathbf{W}_b^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W}_b^M \end{bmatrix}. \text{ With these definitions, the objective}$$

function of OPFDDL has the following form,

$$\begin{aligned} \min_{\tilde{\mathbf{G}}, \mathbf{D}, \lambda} & \lambda^2 \text{Tr}(\tilde{\mathbf{G}}^T \tilde{\mathbf{W}}_w \tilde{\mathbf{G}}) - \lambda \text{Tr}(\tilde{\mathbf{G}}^T \tilde{\mathbf{W}}_b \tilde{\mathbf{G}}), \\ \text{s.t.} & \tilde{\mathbf{G}}^T \tilde{\mathbf{G}} = \mathbf{I}, \end{aligned} \quad (5)$$

where λ is the adaptive weight parameter.

The training procedure of OPFDDL is given by **Algorithm 1**.

ALGORITHM 1 | The OPFDDL algorithm.

Repeat

1. Calculate the coding coefficients by solving the following problem:

$$\min_{\mathbf{A}} \|\tilde{\mathbf{G}}^T \mathbf{Z} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1 \quad (6)$$

2. Calculate the projection matrix by solving the following problem:

$$(\lambda^2 \tilde{\mathbf{W}}_w - \lambda \tilde{\mathbf{W}}_b) \tilde{\mathbf{G}} = \gamma \tilde{\mathbf{G}} \quad (7)$$

3. Calculate the dictionary \mathbf{D} by:

$$\mathbf{D} = \mathbf{D} - \frac{\lambda_D}{n} \sum_k \frac{\partial L(\tilde{\mathbf{Z}}_k)}{\partial \mathbf{D}}, k = 1, \dots, d \quad (8)$$

$$\frac{\partial L(\tilde{\mathbf{Z}}_k)}{\partial \mathbf{D}} = 2\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \mathbf{D}(\mathbf{A}\mathbf{A}^T + \mathbf{H}\mathbf{H}^T) - 2\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \tilde{\mathbf{Z}}_k(\mathbf{A}^T + \mathbf{H}^T) \quad (9)$$

$$\mathbf{A} = [\delta^1, \delta^2, \dots, \delta^M], \mathbf{H} = [\zeta^1, \zeta^2, \dots, \zeta^M] \quad (10)$$

4. Calculate the adaptive weight λ by:

$$\lambda = \frac{\text{tr}(\tilde{\mathbf{G}}^T \tilde{\mathbf{W}}_b \tilde{\mathbf{G}})}{2\text{tr}(\tilde{\mathbf{G}}^T \tilde{\mathbf{W}}_w \tilde{\mathbf{G}})} \quad (11)$$

Until convergence

MULTI-FREQUENT BAND COLLABORATIVE EEG EMOTION CLASSIFICATION METHOD BASED ON OPTIMAL PROJECTION AND SHARED DICTIONARY LEARNING

Objective Function of MBCC

The OPFDDL method can be regarded as the baseline algorithm of MBCC. The primary distinction between the MBCC method and OPFDDL is that, although OPFDDL also employs a projection matrix to project each frequency band to the subspace, the correlation between projection matrices is weak. The common shared component defined in MBCC is a key part of multi-frequent band collaborative learning. In addition, according to the consistency principle, the PCA-like regularization term in the shared potential space

further captures the discriminative information contained among multiple frequency bands. Thus, the MBCC method can balance discriminative knowledge and multi-frequent band correlation in the projection space.

We look for a projection matrix in the MBCC method to project the data from d -dimensional space to p -dimensional space. This study assumes that the projection matrix $\mathbf{G}^m \in \mathbf{R}^{d \times p}$ for each frequency band has two parts: the shared component $\mathbf{G}^0 \in \mathbf{R}^{d \times p}$, which is a common shared matrix that reflects the correlation between different frequency bands, and the band specific component $\tilde{\mathbf{G}}^m \in \mathbf{R}^{d \times p}$, which is the projection matrix for each frequency band. The matrix is equal to the sum of the shared component and the band specific component,

$$\mathbf{G}^m = (1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m, \quad (12)$$

where $\sigma \in [0, 1]$ is the balance parameter. When $\sigma = 1$, the projection matrix \mathbf{G}^m is degenerated into the band specific matrix $\tilde{\mathbf{G}}^m$, which is equivalent to the projection matrix in the OPFDDL method. When $\sigma = 0$, the model only learns the common shared matrix.

The projection of the signal in each frequency band is represented as,

$$(\mathbf{G}^m)^T \mathbf{z}_j^m = ((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{z}_j^m. \quad (13)$$

The within-class reconstruction error of the m -th frequent band in the projected space can be represented as

$$\begin{aligned} J_w^m &= \text{Tr}(\sum_j^{n_m} \sum_k^p [(1 - \sigma)\mathbf{G}^0(:, k)^T (\mathbf{z}_j^m - \mathbf{D}\delta(\mathbf{a}_j^m)) + \sigma\tilde{\mathbf{G}}^m(:, k)^T (\mathbf{z}_j^m - \mathbf{D}\delta(\mathbf{a}_j^m))]^2) \\ &= \text{Tr} \left(((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_w^m ((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m) \right). \end{aligned} \quad (14)$$

The between-class reconstruction error of the m -th frequent band in the projected subspace can be represented as

$$\begin{aligned} J_b^m &= \text{Tr}(\sum_j^{n_m} \sum_k^p [(1 - \sigma)\mathbf{G}^0(:, k)^T (\mathbf{z}_j^m - \mathbf{D}\zeta(\mathbf{a}_j^m)) + \sigma\tilde{\mathbf{G}}^m(:, k)^T (\mathbf{z}_j^m - \mathbf{D}\zeta(\mathbf{a}_j^m))]^2) \\ &= \text{Tr} \left(((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_b^m ((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m) \right). \end{aligned} \quad (15)$$

Thus, the Fisher criterion of all frequent bands is written as,

$$\min_{\mathbf{D}, \mathbf{G}^0, \tilde{\mathbf{G}}^m} \frac{\sum_m \text{Tr} \left(((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_w^m ((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m) \right)}{\sum_m \text{Tr} \left(((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_b^m ((1 - \sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m) \right)}. \quad (16)$$

Because different frequent band data describe the same object, there must be an internal connection between them. The model maximizes the commonality of multiple frequent band data in the shared projection space using the consistency principle. When projecting the data from multiple bands to the optimal subspace, we need to preserve the discriminative information available in the original space. To solve this problem, we use a PCA-like regularization term as follows,

$$J(\mathbf{G}^0) = \min_{\mathbf{G}^0} \sum_m \|\mathbf{Z}^m - \mathbf{G}^0(\mathbf{G}^0)^T \mathbf{Z}^m\|_F^2. \quad (17)$$

Ignoring the constant terms in $J(\mathbf{G}^0)$, Eq. (17) can be represented as,

$$J(\mathbf{G}^0) = - \min_{\mathbf{G}^0} \sum_m Tr(((\mathbf{G}^0)^T \mathbf{Z}^m)((\mathbf{G}^0)^T \mathbf{Z}^m)^T) \\ = - \min_{\mathbf{G}^0} \sum_m Tr((\mathbf{G}^0)^T \mathbf{Z}^m (\mathbf{Z}^m)^T \mathbf{G}^0). \quad (18)$$

Let $\Theta^m = \mathbf{Z}^m (\mathbf{Z}^m)^T$, Eq. (18) can be written as,

$$J(\mathbf{G}^0) = - \min_{\mathbf{G}^0} \sum_m Tr((\mathbf{G}^0)^T \Theta^m \mathbf{G}^0). \quad (19)$$

Combined the Fisher criterion and PCA-like regularization term, the objective function of MBCC is,

$$\min_{\mathbf{G}^0, \tilde{\mathbf{G}}^m, \mathbf{D}} \frac{\sum_m Tr(((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_w^m ((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)) - \alpha \sum_m Tr((\mathbf{G}^0)^T \Theta^m \mathbf{G}^0)}{\sum_m Tr(((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T \mathbf{W}_b^m ((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m))}, \\ s.t. \forall m, ((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m)^T ((1-\sigma)\mathbf{G}^0 + \sigma\tilde{\mathbf{G}}^m) = \mathbf{I}. \quad (20)$$

The projection matrix is orthogonal and it will result in an efficient procedure for optimization. We can see that the dictionary learned in the MBCC method have the stronger discriminative ability.

Define $\mathbf{G} = [\mathbf{G}^0; \mathbf{G}^1; \dots, \mathbf{G}^M] \in \mathbf{R}^{(M+1)d \times p}$, $\Omega^m = [(1-\sigma)\mathbf{I}_d, \sigma\mathbf{I}_d, \dots, \sigma\mathbf{I}_d] \in \mathbf{R}^{d \times (M+1)d}$, $\Delta^m = [\mathbf{I}_d, \mathbf{0}_{d \times d}, \dots, \mathbf{0}_{d \times d}] \in \mathbf{R}^{d \times (M+1)d}$, $\Lambda = \sum_m (\Omega^m)^T \mathbf{W}_w^m \Omega^m$, $\Theta = \sum_m (\Delta^m)^T \Theta^m \Delta^m$, $\mathbf{H} = \sum_m (\Omega^m)^T \mathbf{W}_b^m \Omega^m$, Eq. (20) is equivalent to,

$$\min_{\mathbf{G}, \mathbf{D}} \frac{Tr(\mathbf{G}^T \Lambda \mathbf{G}) - \alpha Tr(\mathbf{G}^T \Theta \mathbf{G})}{Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})}, \quad (21) \\ s.t. \mathbf{G}^T \mathbf{G} = \mathbf{I}.$$

By combining the two terms on the numerator, we can get,

$$\min_{\mathbf{G}, \mathbf{D}} \frac{Tr(\mathbf{G}^T (\Lambda - \alpha \Theta) \mathbf{G})}{Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})}, \quad (22) \\ s.t. \mathbf{G}^T \mathbf{G} = \mathbf{I}.$$

Optimization

It is not easy to directly solve the variables \mathbf{G} and \mathbf{D} in the objective function. Therefore, we will take the alternative iterative optimization scheme to decompose the original problem into two sub-optimization problems.

Update \mathbf{G} . For the given dictionary \mathbf{D} , there must be a minimum ρ , which satisfies the following formulation,

$$\frac{Tr(\mathbf{G}^T (\Lambda - \alpha \Theta) \mathbf{G})}{Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})} \geq \rho, \quad (23)$$

We have $F(\rho) = \min_{\mathbf{G}} Tr(\mathbf{G}^T (\Lambda - \alpha \Theta) \mathbf{G}) - \rho Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})$.

As a result, we can define the function of ρ by,

$$Tr(\mathbf{G}^T (\Lambda - \alpha \Theta) \mathbf{G}) - \rho Tr(\mathbf{G}^T \mathbf{H} \mathbf{G}) \geq 0, \quad (24)$$

According to Zhang et al. (2017), (1) $F(\rho)$ is a decreasing function of ρ . (2) $F(\rho) = 0$ if $\rho = \rho^*$. In addition, the minimum ρ always exists.

Then ρ can be updated by,

$$\rho = \rho + \lambda_\rho \frac{F(\rho)}{F'(\rho)}, \\ F'(\rho) = -Tr(\mathbf{G}^T \mathbf{H} \mathbf{G}), \quad (25)$$

where λ_ρ is the learning rate.

With the fixed ρ and \mathbf{D} , the objective function of \mathbf{G} is,

$$\min_{\mathbf{G}} Tr(\mathbf{G}^T (\Lambda - \alpha \Theta - \rho \mathbf{H}) \mathbf{G}), \quad (26) \\ s.t. \mathbf{G}^T \mathbf{G} = \mathbf{I},$$

The optimization of \mathbf{G} can be solved by the following eigenvalue decomposition,

$$(\Lambda - \alpha \Theta - \rho \mathbf{H}) \mathbf{G} = \gamma \mathbf{G}. \quad (27)$$

The columns of the matrix \mathbf{G} are the eigenvectors with respect to the first p minimum eigenvalues of Eq. (27).

Update \mathbf{D} . With the fixed \mathbf{G} , the objective function of \mathbf{D} is,

$$\min_{\mathbf{D}} \frac{Tr(\mathbf{G}^T \Lambda \mathbf{G})}{Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})}, \quad (28)$$

Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C]$ be the learned dictionary, and \mathbf{D}_j is the j -th class sub-dictionary. The Eq. (28) can be re-written as,

$$J(\mathbf{D}_j) = \min_{\mathbf{D}_j} \sum_{j=1}^C \frac{Tr(\mathbf{G}^T \Lambda_j \mathbf{G})}{Tr(\mathbf{G}^T \mathbf{H} \mathbf{G})}, \quad (29)$$

where $\Lambda_j = \sum_{m=1}^c \sum_{j \neq s} (\mathbf{Z}_j^m - \mathbf{D}_s \Gamma_{j,s}^m) \times (\mathbf{Z}_j^m - \mathbf{D}_s \Gamma_{j,s}^m)^T$, $\mathbf{H} = \sum_{m=1}^c \sum_{j=1}^c (\mathbf{Z}_j^m - \mathbf{D}_j \Gamma_{j,j}^m) \times (\mathbf{Z}_j^m - \mathbf{D}_j \Gamma_{j,j}^m)^T$. $\Gamma_{j,s}^m$ and $\Gamma_{j,j}^m$ are the coding coefficient matrices corresponding to classes s and j of the m -th frequent band, respectively, where $s \neq j$.

\mathbf{D}_j can be updated by gradient descent method, in which \mathbf{D}_j is computed as,

$$\mathbf{D}_j = \mathbf{D}_j + \eta \partial J(\mathbf{D}_j), \\ \partial J(\mathbf{D}_j) = \frac{\partial J(\mathbf{D}_j)}{\partial \Lambda_j} \frac{\partial \Lambda_j}{\partial \mathbf{D}_j} + \frac{\partial J(\mathbf{D}_j)}{\partial \mathbf{H}} \frac{\partial \mathbf{H}}{\partial \mathbf{D}_j}. \quad (30)$$

There is no connection between Λ_j and \mathbf{D}_j , i.e., $\frac{\partial \Lambda_j}{\partial \mathbf{D}_j} = 0$. Therefore, we only need compute $\frac{\partial J(\mathbf{D}_j)}{\partial \mathbf{H}} \frac{\partial \mathbf{H}}{\partial \mathbf{D}_j}$.

$$\frac{\partial J(\mathbf{D}_j)}{\partial \mathbf{H}} = \frac{-Tr(\mathbf{G}^T \Lambda_j \mathbf{G})(\mathbf{G})^T \mathbf{G}}{(Tr(\mathbf{G}^T \mathbf{H} \mathbf{G}))^2}, \quad (31)$$

$$\frac{\partial \mathbf{H}}{\partial \mathbf{D}_j} = (\Gamma_{j,j}^m)^T (\mathbf{D}_j \Gamma_{j,j}^m - \mathbf{Z}_j^m). \quad (32)$$

Update \mathbf{A} . With the fixed \mathbf{D} and \mathbf{G} , the sparse coding coefficient matrix \mathbf{A} can be computed as,

$$\min_{\mathbf{A}} \|\mathbf{G}^T \mathbf{Z} - \mathbf{D} \mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1, \quad (33)$$

Since \mathbf{A} is differentiable, it can be obtained by,

$$\mathbf{A} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} (\mathbf{D}^T \mathbf{G}^T \mathbf{Z}). \quad (34)$$

The alternating optimization procedure of MBCC is summarized in **Algorithm 2**.

ALGORITHM 2 | The MBCC algorithm.

- Repeat
1. Calculate the coding coefficient matrix \mathbf{A} by Eq. (34)
 2. Calculate the projection matrix \mathbf{G} by Eq. (27)
 3. Calculate the dictionary \mathbf{D} by Eq. (30)
- Until convergence

Testing

For the testing procedure, each frequency band feature of the signal \mathbf{z} is represented as \mathbf{z}^m . With the obtained $\{\mathbf{G}^m, \mathbf{D}\}$ by Algorithm 2, its label $l(\mathbf{z}^m)$ on the m -th frequency band can be computed by the following optimization problem,

$$l(\mathbf{z}^m) = \min_{z^m} \|(\mathbf{Q}^m)^T \mathbf{z}^m - \mathbf{D}_j (\mathbf{D}_j^T \mathbf{D}_j)^{-1} \mathbf{D}_j^T \mathbf{z}^m\|_2. \quad (35)$$

Then the majority voting strategy is used to determine the class label of signal \mathbf{z} ,

$$y = \arg \max_m l(\mathbf{z}^m). \quad (36)$$

EXPERIMENT

Datasets and Experimental Settings

Two EEG emotion recognition data sets used in the experiment, SEED and DEAP datasets, which are described as follows. The SEED dataset is an emotional EEG dataset collected and

provided by Shanghai Jiao Tong University’s BCMI Laboratory. The dataset is completed by requiring participants to wear EEG acquisition equipment and recording the emotional EEG signals produced by watching three different types of movie clips. Sixty-two channel electrodes are used in the SEED dataset. The dataset was compiled from 15 participants. With a total of 15 clips, the films are classified as positive, negative, or neutral in terms of their emotional impact. There are five clips of each type, and each movie clip lasts about 4 min. To ensure the experiment’s validity and accuracy, the playback sequence of the 15 videos is random, with no repeated clips. Every participant repeated the experiment three times. A few days were set aside in the middle of each experiment to allow participants to adjust their emotions so that they had a consistent emotional response to the same movie clip. In the experiment, EEG signals are divided into 5-s segments

TABLE 3 | The accuracy (standard deviations) of all methods on SEED dataset in session 2.

Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$	$\delta + \theta + \alpha + \beta + \gamma$
SVM	77.25 (8.53)	78.04 (7.81)	79.67 (7.96)	80.92 (9.92)
LC-KSVD	78.31 (7.74)	79.63 (7.23)	80.88 (9.65)	82.37 (9.62)
MvCVM	80.53 (8.28)	82.55 (8.33)	83.37 (8.24)	83.94 (9.54)
GLSRM	80.78 (7.09)	82.66 (8.75)	82.70 (8.56)	83.94 (9.11)
MVU	81.76 (8.88)	82.51 (8.89)	82.84 (9.57)	84.24 (9.82)
OPFDDL	81.30 (8.75)	83.22 (8.78)	84.76 (9.90)	86.21 (9.43)
MBCC	81.90 (8.52)	84.24 (7.99)	86.14 (8.85)	87.87 (8.07)

The best performance of each comparison is emphasized by the bold font.

TABLE 2 | The accuracy (standard deviations) of all methods on SEED dataset in session 1.

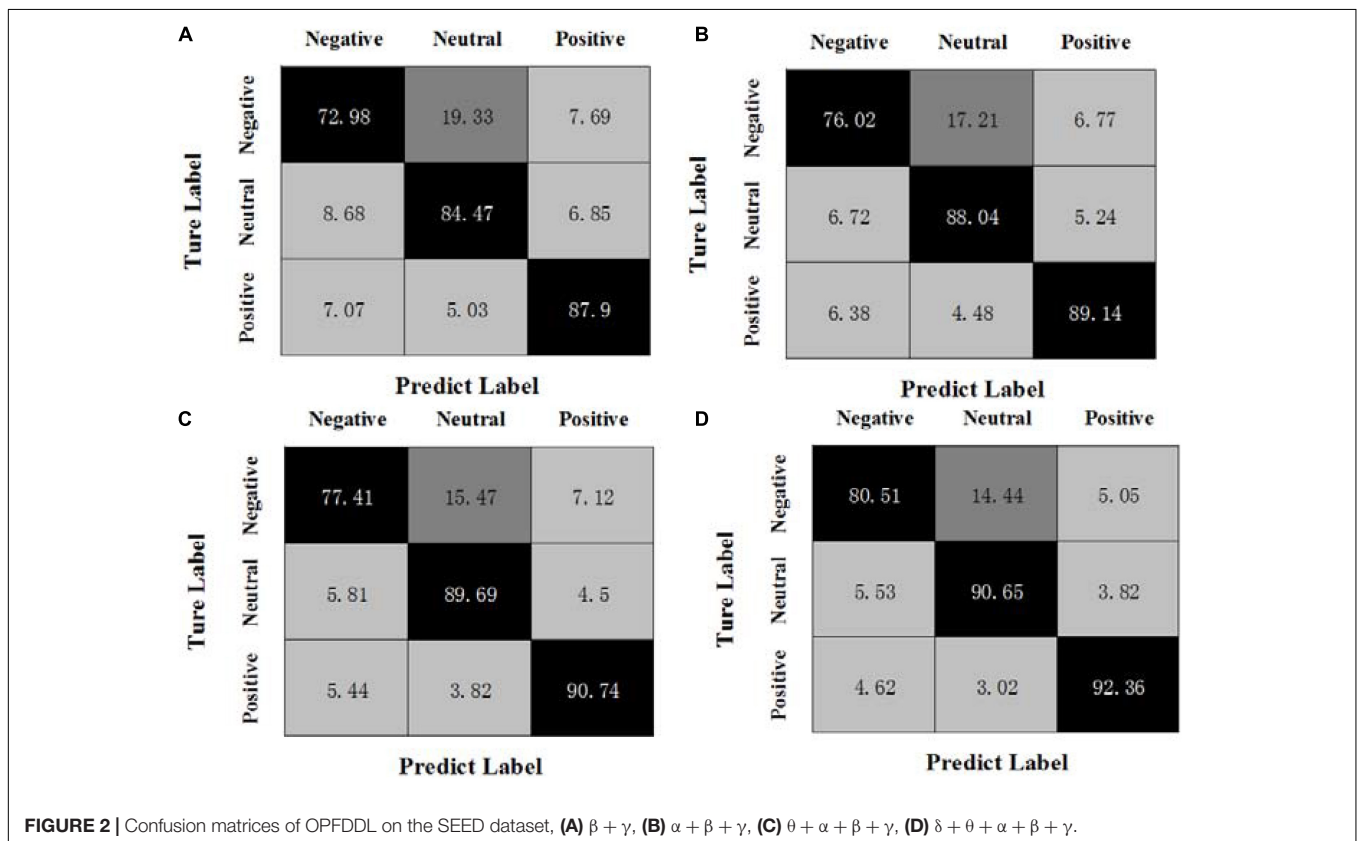
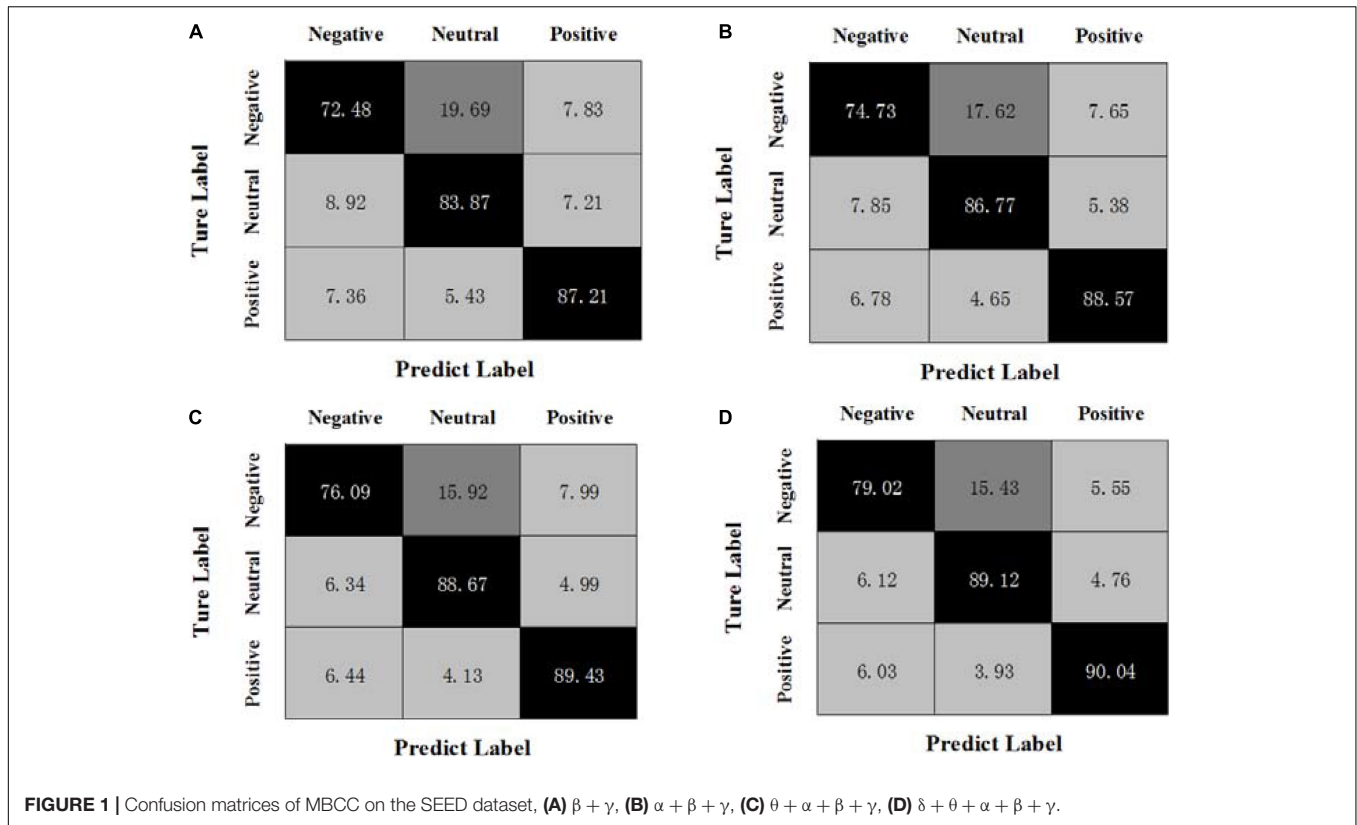
Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$	$\delta + \theta + \alpha + \beta + \gamma$
SVM	77.87 (8.69)	77.96 (9.42)	79.07 (9.96)	81.02 (8.26)
LC-KSVD	78.39 (9.81)	80.10 (8.13)	81.65 (9.19)	83.50 (8.38)
MvCVM	81.03 (9.84)	81.89 (8.72)	83.01 (9.67)	83.84 (9.93)
GLSRM	81.62 (9.66)	82.84 (8.95)	83.33 (8.11)	84.17 (9.39)
MVU	81.65 (9.93)	82.85 (9.50)	83.08 (8.67)	84.27 (9.04)
OPFDDL	81.18 (8.51)	83.67 (8.18)	84.81 (8.76)	86.52 (8.59)
MBCC	81.85 (7.98)	84.61 (8.69)	86.07 (8.82)	87.91 (8.26)

The best performance of each comparison is emphasized by the bold font.

TABLE 4 | The accuracy (standard deviations) of all methods on SEED dataset in session 3.

Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$	$\delta + \theta + \alpha + \beta + \gamma$
SVM	77.19 (9.23)	78.24 (9.65)	79.31 (9.41)	80.78 (9.50)
LC-KSVD	77.61 (8.12)	79.76 (7.09)	80.12 (7.32)	81.92 (9.87)
MvCVM	79.87 (8.67)	82.14 (8.12)	83.18 (8.02)	83.53 (9.13)
GLSRM	80.45 (9.09)	81.30 (9.59)	82.43 (9.34)	83.28 (8.84)
MVU	80.84 (9.27)	81.32 (8.13)	83.00 (9.72)	83.94 (9.28)
OPFDDL	81.05 (7.84)	83.18 (9.67)	84.61 (9.08)	86.43 (9.86)
MBCC	81.81 (8.04)	84.24 (8.98)	85.63 (8.22)	87.74 (8.90)

The best performance of each comparison is emphasized by the bold font.



and features are extracted every 0.5 s. Thus, the sequence length of each segment is 19.

The DEAP dataset is another open database for emotion recognition research that uses EEG and peripheral physiological signals. The dataset recorded the EEG data and 13 peripheral physiological signals of 32 participants using music videos as stimulus materials. The DEAP dataset employs 40 music videos, each of which is 1 min long, as stimulus materials. These music videos are labeled and screened using the general three-dimensional model of valence, arousal, and dominance.

To illustrate the effectiveness of the MBCC method, the comparison methods in the experiment are: SVM (Cortes and Vapnik, 1995), LC-KSVD (Jiang et al., 2013), multi-view CVM (MvCVM) (Huang et al., 2016), global and local structural risk minimization (GLSRM) (Zhu et al., 2016), multi-view learning

with universum (MVU) (Wang et al., 2014), and OPFDDL (Gu et al., 2021a). In detail, the Gaussian kernel is used in MvCVM. The kernel parameter and the weight parameter are searched in the grid $\{1/64, 1/32, \dots, 64\}$ and $\{1, 10^1, \dots, 10^3\}$, respectively. The weights and offsets in GLSRM are searched in the grid $\{0.1, 0.2, \dots, 1\}$, and its regularization parameters are searched in the grid $\{1, 10^1, \dots, 10^3\}$. In MVU, the learning rate is 0.99, and the relaxation of views is 10^{-6} . In OPFDDL and MBCC, the number of atoms in each class is selected in $\{5, 10, \dots, 35\}$. The dimension of matrix \mathbf{G} is set to be 90% of the dimension of the EEG signal features. The parameter α is searched in the grid $\{0.1, 0.2, \dots, 1\}$. The parameter σ is set as $\sigma = 1 - \alpha$. The regularization parameter in Eq. (2) was set as 0.01. All methods are implemented in MATLAB.

Experiments on the SEED Dataset

The commonly used power spectral density (PSD) features (Jenke et al., 2014) are adopted in δ , θ , α , β , and γ frequent bands. We obtain 62 dimensional features on each band. We divided the EEG signal data corresponding to the 15 movie clips collected and used 12 clips as training data and the remaining three clips as test data. In both the training and test sets, the proportion of three classes of EEG signals is the same. After the final preprocessing, the samples of three different classes of EEG signals in the training and test sets are balanced. The SEED dataset is divided into three sessions (sessions 1–3) according to the time interval of signal acquisition. The classification results of all methods in three sessions are shown in **Tables 2–4**. We can see that the MBCC method performs the best in terms of accuracy in all three sessions. In **Table 2**, the accuracies of the MBCC method are 0.67, 0.94, 1.26, and 1.39% better than the second best method OPFDDL in multi-frequent bands $\beta + \gamma$, $\alpha + \beta + \gamma$, $\theta + \alpha + \beta + \gamma$, $\delta + \theta + \alpha + \beta + \gamma$. The results in **Tables 3, 4** are similar to those in **Table 2**. Compared with the OPFDDL method, the proposed MBCC has the ability to take into account the complementarity and consistency between frequency bands while maintaining the PCA constraints of the data structure in the projection space, which is conducive to improving classification performance. Thus, the dictionary learned in the projection space by MBCC has good discriminative performance. The SVM and LC-KSVD methods merge all frequency band data into a vector for learning, and they cannot effectively find the internal connection between each frequency band. For joint learning of multiple perspectives, MvCVM, GLSRM, and MVU treat each frequency band as a learning view. Obviously, the MBCC method obtains a more discriminative model based on dictionary learning and subspace learning.

By calculating the average results of all experiments on three sessions, **Figures 1, 2** show the confusion matrices of MBCC and OPFDDL on the SEED dataset, respectively. The real label is represented by the ordinate of the confusion matrix, while the predicted label is represented by the abscissa.

It can be seen from **Figures 1, 2** that (1) the classification results of positive emotional EEG signals are relatively good on the SEED dataset, while the classification results of negative emotions are relatively poor. Positive emotion is easier to identify than negative and neutral ones. This shows that positive emotions

TABLE 5 | The accuracy (standard deviations) of all methods on the DEAP dataset in valence.

Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$
SVM	62.21 (8.30)	63.04 (8.09)	63.55 (8.76)
KSVD	62.63 (8.59)	63.51 (8.28)	63.94 (8.43)
MvCVM	63.87 (8.57)	64.20 (8.42)	64.65 (9.07)
GLSRM	64.15 (9.60)	66.26 (8.71)	66.84 (9.35)
MVU	64.14 (9.02)	66.18 (9.29)	66.79 (9.13)
OPFDDL	66.04 (8.74)	68.42 (8.40)	69.08 (8.95)
MBCC	66.64 (8.65)	68.85 (8.20)	69.97 (8.46)

The best performance of each comparison is emphasized by the bold font.

TABLE 6 | The accuracy (standard deviations) of all methods on the DEAP dataset in arousal.

Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$
SVM	64.77 (10.85)	65.37 (11.67)	65.85 (10.94)
KSVD	65.07 (10.46)	66.07 (11.09)	66.20 (11.86)
MvCVM	66.31 (11.01)	66.90 (11.48)	67.19 (11.24)
GLSRM	66.49 (10.33)	69.05 (10.27)	69.49 (10.48)
MVU	66.38 (10.79)	69.10 (10.75)	69.27 (11.12)
OPFDDL	68.46 (10.56)	70.35 (10.87)	70.59 (10.06)
MBCC	69.14 (10.39)	70.96 (10.88)	71.55 (10.70)

The best performance of each comparison is emphasized by the bold font.

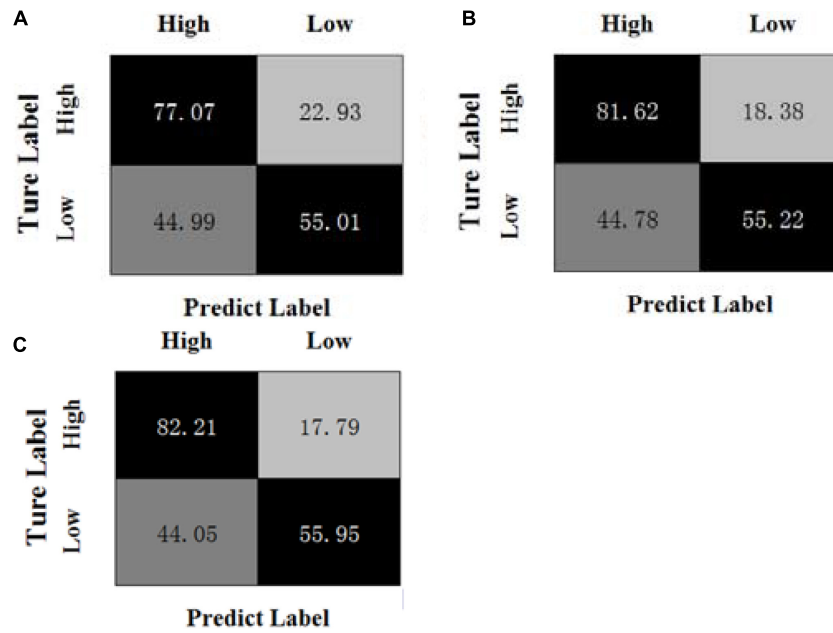


FIGURE 3 | Confusion matrices of OPFDDL on the DEAP dataset in valence, (A) $\beta + \gamma$, (B) $\alpha + \beta + \gamma$, (C) $\theta + \alpha + \beta + \gamma$.

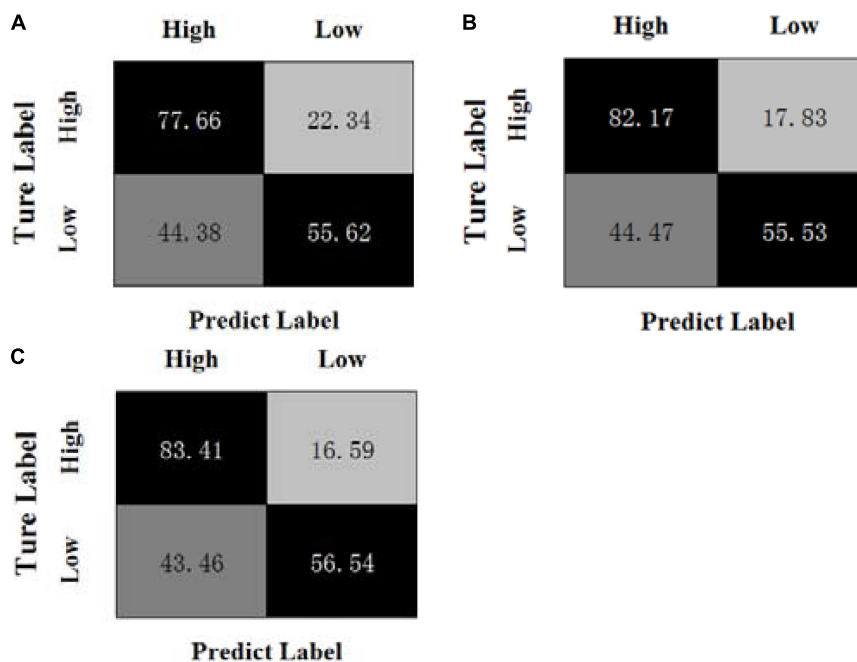


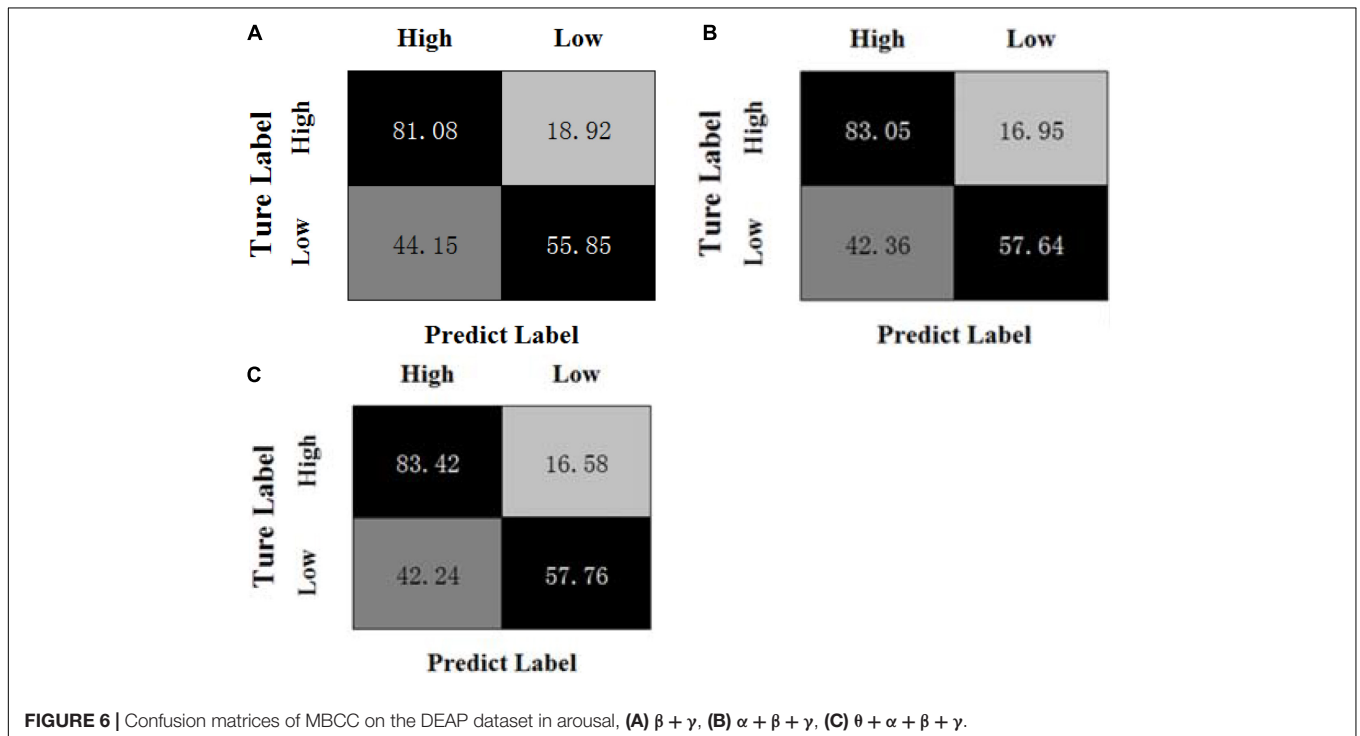
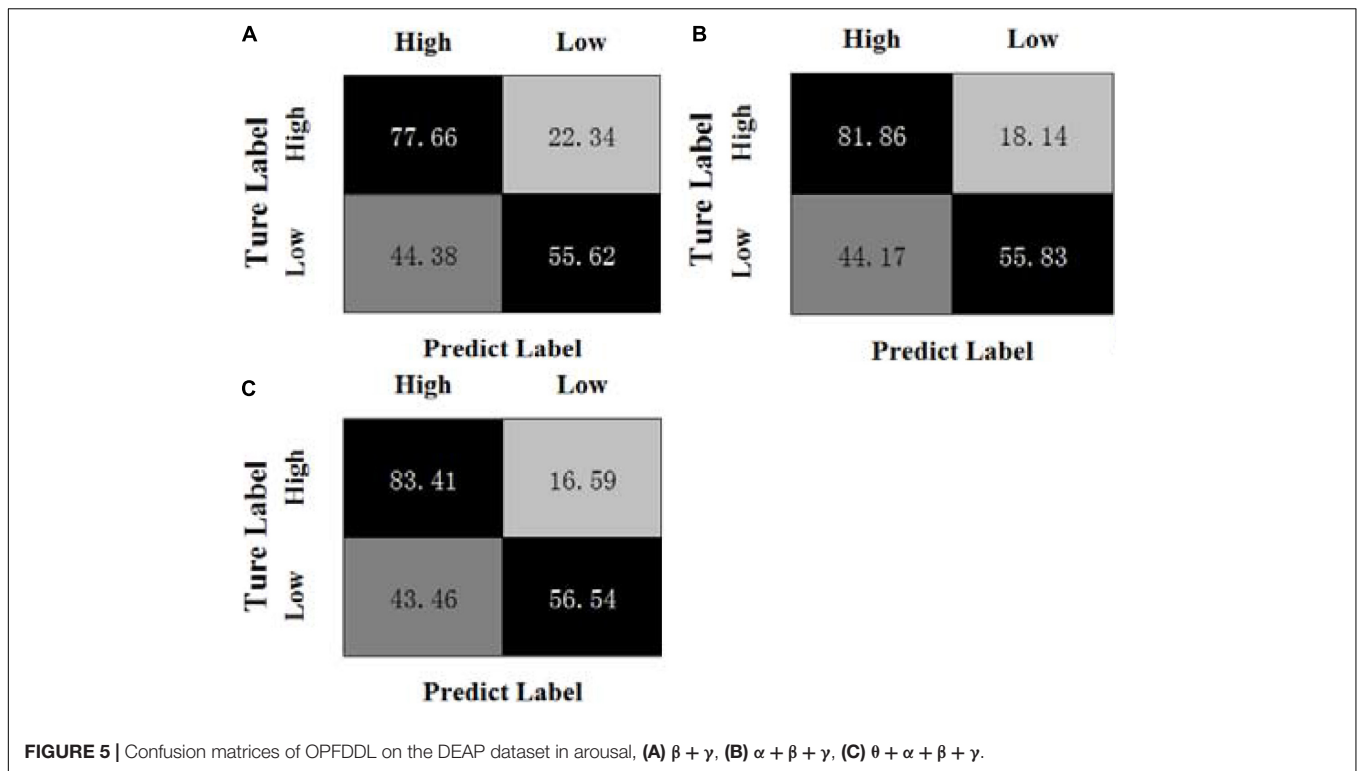
FIGURE 4 | Confusion matrices of MBCC on the DEAP dataset in valence, (A) $\beta + \gamma$, (B) $\alpha + \beta + \gamma$, (C) $\theta + \alpha + \beta + \gamma$.

can cause similar brain feedback between frequency bands more than neutral and negative emotions. (2) The data of different frequency bands are projected into subspaces, and the common shared component of the projection matrix represents the correlation between frequency bands. Obviously, the OPFDDL method does not have this characteristic. (3) In addition, the MBCC method use the PCA-like regularization term based on

shared projection matrix to make full use of the discriminative information of EEG data. Thus, the MBCC method achieves better classification accuracy on the SEED dataset.

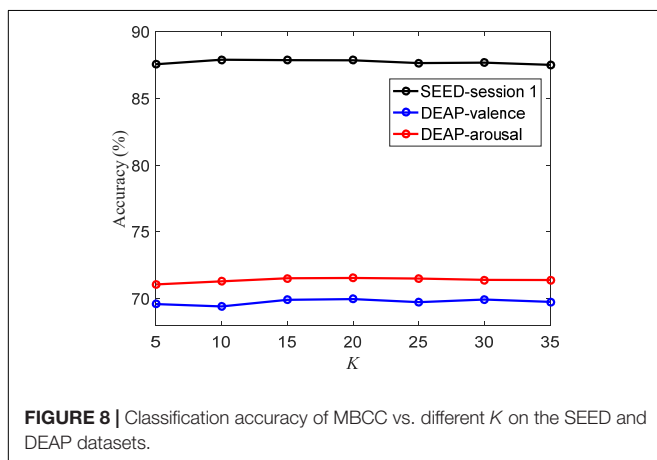
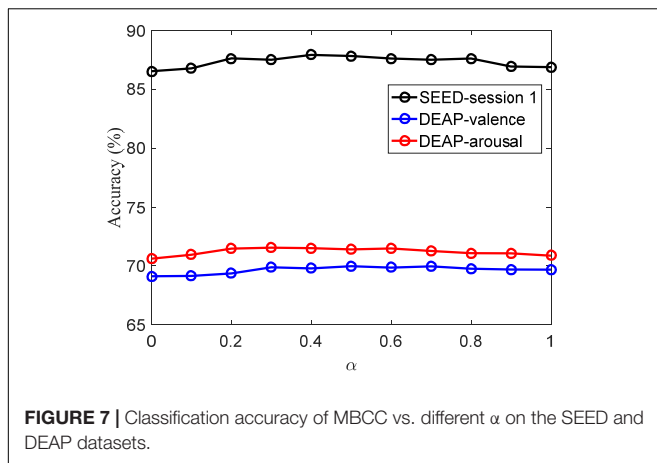
Experiments on the DEAP Dataset

In the DEAP dataset, music video stimulation is a three-dimensional emotion model based on valence, arousal, and



dominance. The valence and arousal of emotion are classified in this subsection. The binary valence-oriented classification refers to the classification of emotions according to high valence and low valence. Also, the binary arousal-oriented classification refers to the classification of emotions according to high arousal

and low arousal. The classification threshold is set to 5, the participant's score $\in [1, 5]$ for valence is low valence, and score $\in (5, 9]$ is high valence. Similarly, the participant's score $\in [1, 5]$ for arousal is low arousal, and score $\in (5, 9]$ is high arousal. The EEG signals are segmented by a 4-s time window with an overlap 2 s



for each frequency band. Similar to the feature extraction strategy in subsection “Experiments on the SEED Dataset,” PSD features are used in the DEAP dataset. Following (Shen et al., 2021), four frequency bands (θ , α , β , γ) are used in the experiment.

Tables 5, 6 compare the average recognition results on valence and arousal on the DEAP dataset, respectively. We can see that (1) all methods have achieved the better classification accuracy for the arousal than the valence on the DEAP dataset. The reason may be that arousal, as an indicator of physiological arousal, reflects the degree of activation of neurophysiological activities, which can be directly reflected in changes in physiological signals. The valence-oriented classification is a more complex task involving mental state, and PSD features may not fully reflect valence’s state. (2) Compared with the benchmark methods SVM and LC-KSVD, the MBCC method has achieved much better results. Compared with GLSRM, MVU, and OPFDDL methods, the classification performance of the MBCC method has further improved. The MBCC method has the accuracy rate of 69.97% for the valence-oriented classification, and 71.55% for the arousal-oriented classification using four frequency bands. The classification accuracies of the MBCC method are increased by 0.89 and 0.96%, respectively, when compared to the second best method. This is due to that the multi-frequent band data maintains the consistency between feature similarity and

semantic similarity in the learned subspace and can learn a more discriminative dictionary shared by frequency bands.

Figures 3, 4 show the confusion matrices of the OPFDDL method and the MBCC method in valence, respectively. **Figures 5, 6** show the confusion matrices of the OPFDDL method and the MBCC method in arousal, respectively. Compared with OPFDDL, MBCC has obvious advantages in valence-oriented and arousal-oriented classifications. When different band data describe the same object, there must be an internal connection between each band data. According to the consistency principle, the MBCC method maximizes the commonness of multiple frequent bands in the shared projection space. Furthermore, the Fisher criterion and PCA-like regularization term aids in learning more discriminative sparse representation and maintaining data structure.

Parameter Analysis

The parameter involved in the objective function of the MBCC method is α , and its impact on MBCC’s performance is analyzed here. The set value range specifies how to conduct experiments on the SEED session 1 and DEAP dataset, respectively. **Figure 7** depicts the accuracy values at various values of α . The figure shows that MBCC achieves the highest accuracy value when taking 0.4, 0.5, and 0.6 on the SEED session 1, DEAP in valence, and DEAP in arousal, respectively.

The atomic number K of the dictionary also directly determines the performance of the MBCC method. **Figure 8** shows the accuracy values under different K values. We can see that when K reaches 15 and 20 on the SEED session 1 and DEAP dataset, respectively, the accuracy rate tends to stabilize. It indicates that the learned dictionary well represents the data characteristics of the EEG data. Also it shows that the MBCC method can be well applied to the SEED and DEAP datasets using a small size of dictionary.

CONCLUSION

According to the consistent complementarity of multi-frequent band EEG signals and the internal correlation of data itself, this study proposes multi-frequency band collaborative EEG emotion classification method based on the idea of dictionary learning and subspace learning. Using a projection matrix, this method maps different frequency band data to the subspaces of the same dimension. Unlike most existing projection strategies, the projection matrix we designed is divided into two parts, a common shared component and a band-specific component. This strategy can fully use the relevance of different frequency bands and their shared semantics. In the subspace, the MBCC method learns the common shared dictionary between the frequency bands, which can represent the correlation and discrimination of the EEG data. Simultaneously, the incorporation of Fisher criterion and PCA-like regularization term into the subspace *via* dictionary learning makes the learned model more discriminative.

However, the time computation of MBCC is relatively high. It may be not suitable for real-time predicting emotional states

in applications of human-computer interaction. This is the problem we need to solve in the next stage. Furthermore, the work that can be studied further in the future includes: (1) Human emotions are susceptible to external influences. For example, the emotions of the subjects may change while watching a film. The overall emotions of watching the film may be consistent, but the emotions may be inconsistent with expectations at times. As a result, the collected EEG signals are mixed with abnormal samples. In practice, selecting the appropriate abnormal sample processing method is important. The use of the correct processing method can improve the accuracy of emotional EEG signal recognition. (2) EEG signals have the characteristics of randomness. That is, for the same individual subjects, EEG signals are different even in the same emotional state at different times. How to improve the robustness of emotion classification model in multiple domains still needs further research. In the future, we will continue to design and improve our method to be suitable in across time and individuals scenes.

REFERENCES

- Aranha, R. V., Corrêa, C. G., and Nunes, F. (2019). Adapting software with affective computing: a systematic review. *IEEE Trans. Affect. Comp.* 12, 883–899. doi: 10.1109/TAFFC.2019.2902379
- Cortes, C., and Vapnik, V. (1995). Support vector machine. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Davis, M., and Whalen, P. (2001). The amygdala: vigilance and emotion. *Mol. Psychiatry* 6, 13–34. doi: 10.1038/sj.mp.4000812
- Gianaros, P. J., Marsland, A. L., Kuan, C. H., Schirda, B. L., Jennings, J. R., Sheu, L. K., et al. (2014). An inflammatory pathway links atherosclerotic cardiovascular disease risk to neural activity evoked by the cognitive regulation of emotion. *Biol. Psychiatry* 75, 738–745. doi: 10.1016/j.biopsych.2013.10.012
- Gu, X., Fan, Y. Q., Zhou, J., and Zhu, J. Q. (2021a). Optimized projection and fisher discriminative dictionary learning for eeg emotion recognition. *Front. Psychol.* 12:705528. doi: 10.3389/fpsyg.2021.705528
- Gu, X., Zhang, C., and Ni, T. (2021b). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comp. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699
- Guil, R., Ruiz-González, P., Merchán-Clavellino, A., Morales-Sánchez, L., Zayas, A., and Gomez-Molinero, R. (2020). Breast cancer and resilience: the controversial role of perceived emotional intelligence. *Front. Psychol.* 11:595713. doi: 10.3389/fpsyg.2020.595713
- Huang, C., Chung, F., and Wang, S. (2016). Multi-view L2-SVM and its multi-view core vector machine. *Neural Networks* 75, 110–125. doi: 10.1016/j.neunet.2015.12.004
- Jenke, R., Peer, A., and Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* 5, 327–339. doi: 10.1109/TAFFC.2014.2339834
- Jiang, Z., Lin, Z., and Davis, L. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Patt. Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88
- Klatzkin, R. R., Nolan, L. J., and Kissileff, H. R. (2021). Self-reported emotional eaters consume more food under stress if they experience heightened stress reactivity and emotional relief from stress upon eating. *Physiol. Behav.* 243:113638. doi: 10.1016/j.physbeh.2021.113638
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: a database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15
- Li, J., Zhang, Z., and He, H. (2018). Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cogn. Comp.* 10, 368–380. doi: 10.1007/s12559-017-9533-x

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The SEED and DEAP datasets analyzed for this study can be found in the following links, respectively: <http://bcmi.sjtu.edu.cn/~seed/seed.html>, <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>.

AUTHOR CONTRIBUTIONS

JZ and TN conceived and designed the proposed model. ZS performed the experiment. All authors drafted and approved the manuscript.

FUNDING

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grants Bk202101333.

- Li, M., and Lu, B. L. (2009). “Emotion classification based on gamma-band EEG”. *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN. 1223–1226. doi: 10.1109/IEMBS.2009.5334139
- Li, W., Zhang, Z., and Song, A. (2020). Physiological-signal-based emotion recognition: an odyssey from methodology to philosophy. *Measurement* 172:108747.
- Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2019). EEG emotion recognition based on graph regularized sparse linear regression. *Neural Proc. Lett.* 49, 555–571. doi: 10.1007/s11063-018-9829-1
- Liu, Y., Ding, Y., Li, C., Cheng, J., Cheng, J., Song, R. C., et al. (2020). Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. *Comp. Biol. Med.* 123:103927. doi: 10.1016/j.combiomed.2020.103927
- Mohammadi, Z., Frounchi, J., and Amiri, M. (2017). Wavelet-based emotion recognition system using EEG signal. *Neural Comp. Appl.* 28, 1985–1990. doi: 10.1007/s00521-015-2149-8
- Ni, T. G., Ni, Y. Y., Xue, J., and Wang, S. H. (2021). A domain adaptation sparse representation classifier for cross-domain Electroencephalogram-based emotion classification. *Front. Psychol.* 12:721266. doi: 10.3389/fpsyg.2021.721266
- Nurillaeva, N. M., and Abdumalikova, F. B. (2021). Predictive importance of psycho-emotional syndrome of patients with coronary heart disease in the violation of platelet hemostatic system. *Atherosclerosis* 331:E204. doi: 10.1016/j.atherosclerosis.2021.06.625
- Peng, Y., Liu, S., Wang, X., and Wu, X. (2020). Joint local constraint and fisher discrimination based dictionary learning for image classification. *Neurocomputing* 398, 505–519. doi: 10.1016/j.neucom.2019.05.103
- Samsonovich, A. V. (2020). Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. *Cogn. Syst. Res.* 60, 57–76. doi: 10.1016/j.cogsys.2019.12.002
- Shareef, M. A., Mukerji, B., Alryalat, M., Wright, A., and Dwivedi, Y. K. (2018). Advertisements on Facebook: identifying the persuasive elements in the development of positive attitudes in consumers. *J. Retail. Cons. Serv.* 43, 258–268. doi: 10.1016/j.jretconser.2018.04.006
- Shen, F., Peng, Y., Kong, W., and Dai, G. (2021). Multi-scale frequency bands ensemble learning for EEG-based emotion recognition. *Sensors* 21:1262. doi: 10.3390/s21041262
- Tennant, C., and McLean, L. (2001). The impact of emotions on coronary heart disease risk. *J. Cardiovas. Risk* 8, 175–183. doi: 10.1177/174182670100800309

- Wang, Z., Zhu, Y., Liu, W., Chen, Z., and Gao, D. (2014). Multi-view learning with Universum. *Knowledge-Based Syst.* 70, 376–391. doi: 10.1016/j.knosys.2014.07.019
- Wirkner, J., Weymar, M., Lw, A., Hamm, C., Anne-Marie, S., Kirschbaum, C., et al. (2017). Cognitive functioning and emotion processing in breast cancer survivors and controls: an ERP pilot study. *Psychophysiology* 54, 1209–1222. doi: 10.1111/psyp.12874
- Yang, Y., Wu, Q., Fu, Y., and Chen, X. (2018). “Continuous Convolutional Neural Network with 3D Input for EEG-Based Emotion Recognition,” in *International Conference on Neural Information Processing*, Siem Reap, 433–443. doi: 10.1007/978-3-030-04239-4_39
- Zhang, G., Sun, H., Porikli, F., Liu, Y., and Sun, Q. (2017). Optimal couple projections for domain adaptive sparse representation-based classification. *IEEE Trans. Image Process.* 26, 5922–5935. doi: 10.1109/TIP.2017.2745684
- Zhang, G., Yang, J., Zheng, Y., Luo, Z., and Zhang, J. (2021). Optimal discriminative feature and dictionary learning for image set classification. *Inform. Sci.* 547, 498–513. doi: 10.1016/j.ins.2020.08.066
- Zhao, G., Zhang, Y., and Yan, G. (2018). Frontal EEG asymmetry and middle line power difference in discrete emotions. *Front. Behav. Neurosci.* 11:225. doi: 10.3389/fnbeh.2018.00225
- Zheng, W. L., and Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Mental Dev* 7, 162–175. doi: 10.1109/tamd.2015.2431497
- Zhong, P., Wang, D., and Miao, C. (2020). EEG-Based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comp. Early Access* 2020:2994159. doi: 10.1109/TAFFC.2020.2994159
- Zhu, C., Wang, Z., and Gao, D. (2016). New design goal of a classifier: global and local structural risk minimization. *Knowledge-Based Syst.* 100, 25–49. doi: 10.1016/j.knosys.2016.02.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhu, Shen and Ni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.