



Optimized Projection and Fisher Discriminative Dictionary Learning for EEG Emotion Recognition

Xiaoqing Gu¹, Yiqing Fan², Jie Zhou^{3*} and Jiaqun Zhu^{1*}

¹ School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China, ² Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, ³ School of Electrical and Mechanical Engineering, Shaoxing University, Shaoxing, China

OPEN ACCESS

Edited by:

Yaoru Sun,
Tongji University, China

Reviewed by:

Yanhui Zhang,
Hebei University of Chinese
Medicine, China

Yufeng Yao,
Changshu Institute of
Technology, China

*Correspondence:

Jie Zhou
sxuj_zhou@163.com
Jiaqun Zhu
angelsoar@cczu.edu.cn

Specialty section:

This article was submitted to
Emotion Science,
a section of the journal
Frontiers in Psychology

Received: 05 May 2021

Accepted: 31 May 2021

Published: 28 June 2021

Citation:

Gu X, Fan Y, Zhou J and Zhu J (2021)
Optimized Projection and Fisher
Discriminative Dictionary Learning for
EEG Emotion Recognition.
Front. Psychol. 12:705528.
doi: 10.3389/fpsyg.2021.705528

Electroencephalogram (EEG)-based emotion recognition (ER) has drawn increasing attention in the brain–computer interface (BCI) due to its great potentials in human–machine interaction applications. According to the characteristics of rhythms, EEG signals usually can be divided into several different frequency bands. Most existing methods concatenate multiple frequency band features together and treat them as a single feature vector. However, it is often difficult to utilize band-specific information in this way. In this study, an optimized projection and Fisher discriminative dictionary learning (OPFDDL) model is proposed to efficiently exploit the specific discriminative information of each frequency band. Using subspace projection technology, EEG signals of all frequency bands are projected into a subspace. The shared dictionary is learned in the projection subspace such that the specific discriminative information of each frequency band can be utilized efficiently, and simultaneously, the shared discriminative information among multiple bands can be preserved. In particular, the Fisher discrimination criterion is imposed on the atoms to minimize within-class sparse reconstruction error and maximize between-class sparse reconstruction error. Then, an alternating optimization algorithm is developed to obtain the optimal solution for the projection matrix and the dictionary. Experimental results on two EEG-based ER datasets show that this model can achieve remarkable results and demonstrate its effectiveness.

Keywords: EEG signal, emotion recognition, dictionary learning, fisher discrimination criterion, brain computer interface

INTRODUCTION

Brain–computer interface (BCI) has been one of the research hotspots in recent years in health monitoring and biomedicine (Edgar et al., 2020; Ni et al., 2020b). The BCI does not rely on muscles and the peripheral nervous system. It establishes a direct information transmission channel between the brain and the outside world. The electroencephalography (EEG) signals captured by the BCI system are a powerful tool to analyze neural activities and brain conditions. EEG has the advantages of convenience (i.e. non-invasive, non-destructive and simple) and validity (i.e. sensitivity, validity and compatibility) (Sreeja and Himanshu, 2020). EEG signal is an important tool for revealing the emotional state of human beings. It has been shown that when people are in different thinking and emotional states, the rhythm components of EEG signals are different from their waveform. In BCI, the operation of emotion recognition (ER) starts from

external stimuli to subjects, which induce specific emotions such as happiness, sadness, and anger. These stimuli may be videos, images, music, and so on. During the session, EEG data are recorded by EEG devices. Subsequently, the first step is to extract and preprocess useful features obtained from the recorded EEG. The next step is to train the classifier and optimize the parameters. The final step is to test the training model with new EEG data that are not used in the training process.

Traditional machine learning classifiers have been widely used in EEG-based ER, such as support vector machine (SVM) (Zheng et al., 2019), deep learning (Hwang et al., 2020; Song et al., 2020), nearest neighbor classifier (Li et al., 2019), random forest (Fraivan et al., 2012), and probabilistic neural networks (Nakisa et al., 2018). In recent years, dictionary learning-based methods have achieved great success in EEG-based recognition tasks for BCI (Ameri et al., 2016; Gu et al., 2020; Ni et al., 2020a). In general, dictionary learning-based classification methods often learn the discriminative and robust dictionaries from training samples. The test sample is sparsely represented as a sparse linear combination of atoms by the learned dictionary, and then, the classification task can be carried out according to the reconstruction error and/or the sparse coefficients. Dictionary learning works well-even with noisy EEG signals. Barthélemy et al. (2013) developed an efficient method to represent EEG signals based on the adapted Gabor dictionary and demonstrated on real data that the learned multivariate model is flexible and the learned representation is informative and interpretable. Abolghasemi and Ferdowsi (2015) developed a dictionary learning framework to remove ballistocardiogram (BCG) artifacts from EEG. Given the advantage of the noise-robust sparse dictionary, a new cost function was proposed, which can model BCG artifacts and then remove them from the original EEG signals. Kashefpoor et al. (2019) developed a correlational label consistent K-SVD dictionary learning method applied to EEG-based screening tool. This method was applied to speckle extraction of EEG signals and extracted spectral features in both time and frequency domains. Aiming at the problem that eye movement and blinking can cause artifacts, Kanoga et al. (2019) proposed a multi-scale dictionary learning method to eliminate eye artifacts from single-channel measurement. Specifically, the time-domain waveforms related to repetitive phase events in EEG signals were learned within the framework of dictionary learning. And the proposed multi-scale dictionary learning method was used to represent the signal components on different timescales. To achieve the highly accurate classification of EEG in BCI, Huang et al. (2020) developed a signal identification model using sparse representation and fast compressed residual convolutional neural networks (CNNs). The authors used the common spatial patterns to extract EEG signal features and build a redundant dictionary using these features. Then, the proposed deep model as a classifier recognized the input EEG signals.

Although machine learning has achieved good classification performance in some application scenarios, the accuracy and applicability of the classification do not go far enough. Since EEG data provide comprehensive information across different frequency bands to characterize emotions, it was expected to

design an ER method, which utilizes the specific discriminative information of each frequency band and preserves the common discrimination information shared by multiple band signals. After the success of dictionary learning, in this study, we propose optimized projection and Fisher discriminative dictionary learning (DDL) for EEG-based ER. According to the Fisher discrimination criterion of minimum within-class sparse reconstruction error and maximum between-class sparse reconstruction error, we learn the discriminative projection to map the multiple band signals into a shared subspace and simultaneously build a shared dictionary that establishes the connection between different bands and represents the characteristics of signals well. Therefore, the joint learning of projection and dictionary ensures the common internal structure of multiple frequency bands of signals to be mined in the subspace.

The main contributions of this study are as follows:

- (1) A multiple frequency band collaborative learning is introduced in dictionary learning for the EEG-based ER. This learning mechanism can efficiently integrate the band-independent information and inter-band correlation information.
- (2) Through the feature projection matrix, the data of multiple frequency bands are projected into a common projection subspace to keep the latent manifold of EEG signals. Meanwhile, the discriminative dictionary is learned by enforcing the classification criterion so that the learned sparse code has a strong representation and discrimination ability.
- (3) This joint optimization method has some benefits. Learning independent projection matrices makes this model easily extensible; meanwhile, learning a dictionary in a subspace allows abandoning extraneous information in the original features. In addition, the alternating optimization procedure ensures the dictionary and projection are optimized at the same time.
- (4) These extensive experiments on the SEED and DREAMER datasets demonstrate that the multiple band collaborative learning is effective, and this method can improve the discrimination ability of sparse coding in EEG-based ER.

BACKGROUND

Datasets

The experimental data in this study are taken from two public EEG emotion datasets: SEED (Zheng and Lu, 2015) and DREAMER (Katsigiannis and Ramzan, 2018). **Table 1** briefly describes the information of the two datasets. Both SEED and DREAMER datasets are collected when subjects watched emotion-eliciting movies. In the SEED dataset, each subject participated in three experiments, which were separated into three time periods, corresponding to three sessions, and each session corresponds to 15 EEG data trials. Thus, a total of $15 \times 3 = 45$ trials are formed per subject. The SEED provided five frequency bands: δ band (1–3 Hz), θ band (4–7 Hz), α band (8–13 Hz), β band (14–30 Hz), and γ band (31–50 Hz). For the DREAMER dataset, the data recorded by each

subject contain three parts: 18 experimental signal segments, 18 baseline signal segments corresponding to relaxation state, and 18 corresponding labels. The DREAMER data provided EEG features with frequency bands θ , α , and β .

Machine Learning-Based EEG Signal Processing Program

For machine learning-based EEG-based ER, feature extraction and emotion classification are the critical procedures. Considering the SEED dataset as an example, the process of constructing five frequency band sequences is described in **Figure 1**. Firstly, EEG signals collected by BCI are preprocessed by filtering. Then, according to the characteristics of different rhythms of EEG signals, EEG signals usually can be divided into several rhythmic signal components ranging from 0 to 50 Hz. Secondly, EEG features can be extracted by various strategies.

Time-domain, frequency-domain, and non-linear analysis methods are the three types of most commonly used EEG feature extraction methods. The time-domain features aim to capture the temporal information of EEG signals, such as higher-order crossings (HOC) (Petranonakis and Hadjileontiadis, 2010), Hjorth features (Petranonakis and Hadjileontiadis, 2010), and event-related potential (ERP) (Brouwer et al., 2015). The frequency-domain features aim to capture primarily the EEG emotion information from a frequency perspective. Then, EEG features can be extracted by various methods, such as rhythm (Bhatti et al., 2016), wavelet packet decomposition (WPD) (Wu et al., 2008), and approximate entropy (AE) (Ko et al., 2009). Non-linear features are extracted from the transformed phase space. Non-linear features contain quantitative measures that represent the complex dynamic characteristics of the EEG signals, such as Lyapunov exponent (Lyap) (Kutepov et al., 2020) and correlation dimension (CorrDim) (Geng et al., 2011). Finally, many machine learning methods are established to handle EEG emotion classification on extracted feature sets.

A common approach to deal with multiple bands of EEG data using traditional dictionary learning methods is to directly concatenate features of multiple bands together in the high-dimensional space and treat this single feature vector as the input to the model. However, dictionary learning may not perform well because different band features usually carry different characteristics of EEG emotion.

TABLE 1 | The basic information of SEED and DREAMER datasets.

SEED dataset	DREAMER dataset
15 subjects	23 subjects
15 trials using Chinese movie clips (length of film clips about 4 min) to evoke emotions as negative, positive, and neutral	18 trials using movie clips (length of film clips 65–393 s) to evoke 9 emotions
In one session, 5 s hint before each clip, 45 s self-assessment, and 15 s rest after each clip	At least 61 s of pretrial baseline data available
Emotion rating metric: negative, positive, and neutral	Emotion rating scale: valence, arousal, and dominance on a continuous scale from 1 to 5
62-channel electrode cap	14-channel electrode cap
Sampling rate 1,000 Hz	Sampling rate 128 Hz
Frequency band from 0 to 75 Hz	Frequency band from 4 to 30 Hz

Dictionary Learning

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$ be a set of m -dimensional n training signals. To minimize the reconstruction error and satisfy the sparsity constraints, the sparse representation and dictionary learning of \mathbf{X} can be accomplished by

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{S}} \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2, \\ \text{s.t. } \|\mathbf{s}_i\|_0 \leq T, \end{aligned} \tag{1}$$

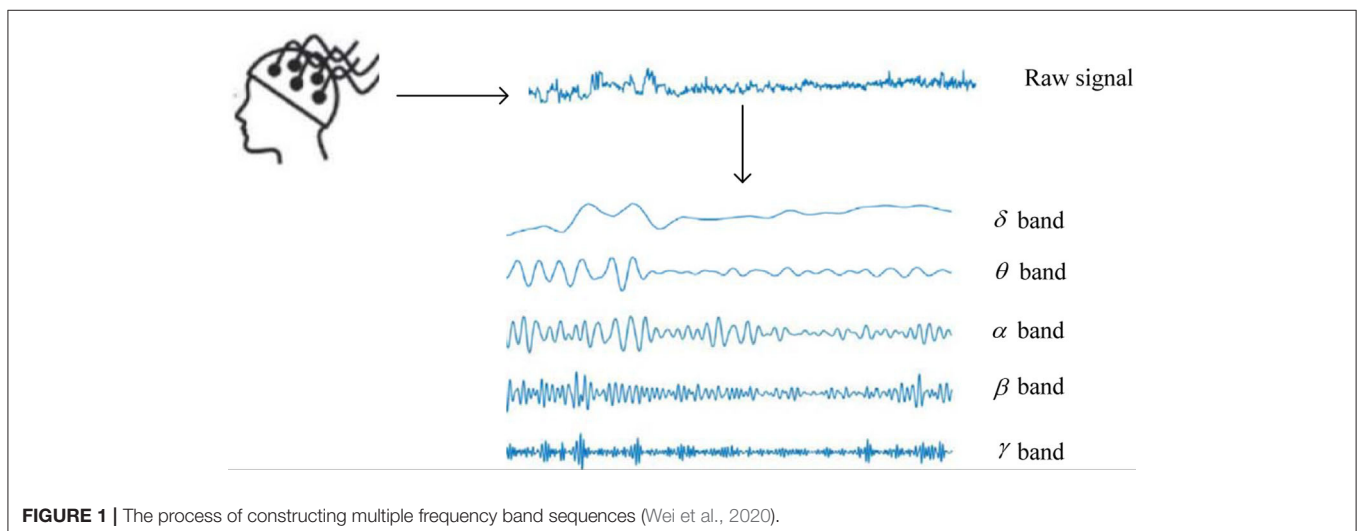


FIGURE 1 | The process of constructing multiple frequency band sequences (Wei et al., 2020).

where $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_K] \in \mathbf{R}^{m \times K}$ is a dictionary with K atoms $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_n] \in \mathbf{R}^{m \times K}$ is the sparse coefficient matrix of signals \mathbf{X} , and \mathbf{s}_i is the sparse coefficient vector of \mathbf{x}_i over \mathbf{D} . T is the sparse constraint factor. The $\|\mathbf{s}_i\|_0 \leq T$ term requires the signal \mathbf{x}_i to have fewer than T non-zero items in its decomposition.

It is not easy to find the optimal sparse solution using ℓ_0 -norm regularization term; thus, an alternative formulation of Equation (1) is to replace it with ℓ_1 -norm regularization as

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{S}} \|\mathbf{X} - \mathbf{D}\mathbf{S}\|_F^2 + \lambda \|\mathbf{S}\|_1, \\ \text{s.t. } \|\mathbf{d}_i\|_2^2 = 1, \forall i \end{aligned} \quad (2)$$

Equation (2) can be optimized by many efficient ℓ_1 optimization methods, such as the famous K-SVD algorithm (Aharon et al., 2006; Jiang et al., 2013). However, Equation (2) is an unsupervised learning framework. To learn a discriminative dictionary for classification tasks, different kinds of loss functions or Fisher discrimination criterion are considered in the dictionary learning. Fisher discrimination constraints on atoms of the dictionary (Peng et al., 2020) or sparse coefficient \mathbf{S} (Li et al., 2013) or reconstruction error of \mathbf{X} (Zheng and Sun, 2019; Zhang et al., 2021) strive to preserve the class distribution and geometric structure of data.

Suppose data matrix \mathbf{X} consists of samples from C different classes, from \mathbf{X} , both the sub-dictionary \mathbf{D}_i and sub-sparse coefficient matrix \mathbf{S}_i are learned for the i -th class data ($i = 1, 2, \dots, C$). The whole dictionary \mathbf{D} is represented as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_C]$. Let \mathbf{W}_w and \mathbf{W}_b denote the within-class scatter and between-class reconstruction error of \mathbf{X} , respectively, then

$$\mathbf{W}_w = \sum_j (\mathbf{x}_j - \mathbf{D}\delta_{l_j}(\mathbf{s}_j)) \times (\mathbf{x}_j - \mathbf{D}\delta_{l_j}(\mathbf{s}_j))^T, \quad (3)$$

and

$$\mathbf{W}_b = \sum_j (\mathbf{x}_j - \mathbf{D}\zeta_{l_j}(\mathbf{s}_j)) \times (\mathbf{x}_j - \mathbf{D}\zeta_{l_j}(\mathbf{s}_j))^T, \quad (4)$$

where $\delta_{l_j}(\cdot)$ function returns the sparse codes consistent with the class of \mathbf{x}_j ($j = 1, 2, \dots, n$), and $\zeta_{l_j}(\cdot)$ function returns the sparse codes not consistent with the class of \mathbf{x}_j .

Then, a discriminative dictionary can be learned by reducing the within-class diversity and by increase between-class separation using Equations (3) and (4) (Zheng and Sun, 2019; Zhang et al., 2021).

OPTIMIZED PROJECTION AND FISHER DISCRIMINATIVE DICTIONARY LEARNING

Objective Function

Here, we describe in detail the optimized projection and Fisher DDL model for collaborative learning of multiple frequency band EEG signals. The training framework of the OPFDDL model is illustrated in **Figure 2**. We learn a discriminative projection to map multiple frequency band EEG signals into a common subspace; simultaneously, we learn a common discriminative

dictionary to encode the band-invariant information of multiple frequency bands. In particular, to promote the discrimination ability of the model, we utilize the model according to the Fisher discrimination criterion (Gong et al., 2019) under the structure of dictionary learning.

Let $\mathbf{X}^r = \{\mathbf{x}_j^r\}$ denote the signal set \mathbf{X} of the frequency band r , where R is the number of frequency bands ($r = 1, \dots, R$). \mathbf{x}_j^r is the j th sample in \mathbf{X}^r . To build the connection between different frequency bands and exploit the specific characteristic of each representation, we project \mathbf{x}_j^r into a feature subspace as $\mathbf{z}_j^r = \mathbf{Q}^r \mathbf{x}_j^r$ by using a transformation matrix $\mathbf{Q}^r \in \mathbf{R}^{m \times d^r}$. Therefore, we obtain $\{\mathbf{z}_j^r\}_{r=1}^R$ by $\{\mathbf{Q}^r \mathbf{x}_j^r\}_{r=1}^R$ as the feature representations for R frequency bands. Then, we denote the within-class reconstruction error J_w^r and between-class reconstruction error J_b^r of the r th frequency band in the projection subspace

$$\begin{aligned} J_w^r &= \text{tr} \left(\sum_{j=1}^{n_r} (\mathbf{Q}^{rT} \mathbf{x}_j^r - \mathbf{Q}^{rT} \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{Q}^{rT} \mathbf{x}_j^r - \mathbf{Q}^{rT} \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r))^T \right) \\ &= \text{tr} (\mathbf{Q}^{rT} \sum_{j=1}^{n_r} (\mathbf{x}_j^r - \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r - \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r)) \mathbf{Q}^r) \\ &= \text{tr} (\mathbf{Q}^{rT} \mathbf{W}_w^r \mathbf{Q}^r) \end{aligned} \quad (5)$$

and

$$\begin{aligned} J_b^r &= \text{tr} \left(\sum_{j=1}^{n_r} (\mathbf{Q}^{rT} \mathbf{x}_j^r - \mathbf{Q}^{rT} \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{Q}^{rT} \mathbf{x}_j^r - \mathbf{Q}^{rT} \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r))^T \right) \\ &= \text{tr} (\mathbf{Q}^{rT} \sum_{j=1}^{n_r} (\mathbf{x}_j^r - \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r - \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r)) \mathbf{Q}^r) \\ &= \text{tr} (\mathbf{Q}^{rT} \mathbf{W}_b^r \mathbf{Q}^r) \end{aligned} \quad (6)$$

where $\mathbf{W}_w^r = \sum_{j=1}^{n_r} (\mathbf{x}_j^r - \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r - \mathbf{D} \delta_{l_j}(\mathbf{s}_j^r))^T$ is the within-class scatter matrix for sparse coding of the r th frequency band, and $\mathbf{W}_b^r = \sum_{j=1}^{n_r} (\mathbf{x}_j^r - \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r)) \times (\mathbf{x}_j^r - \mathbf{D} \zeta_{l_j}(\mathbf{s}_j^r))^T$ is the between-class scatter matrix for sparse coding of the r th frequency band.

From the classification point of view, minimizing within-class scatter and maximizing between-class scatter in the dictionary learning-based classifier can be represented as

$$\begin{aligned} \min_{\mathbf{Q}^1, \dots, \mathbf{Q}^R, \mathbf{D}} \frac{\sum_{r=1}^R \text{tr}(\mathbf{Q}^{rT} \mathbf{W}_w^r \mathbf{Q}^r)}{\sum_{r=1}^R \text{tr}(\mathbf{Q}^{rT} \mathbf{W}_b^r \mathbf{Q}^r)}, \\ \text{s.t. } (\mathbf{Q}^r)^T (\mathbf{Q}^r) = \mathbf{I}, \forall r \end{aligned} \quad (7)$$

With definition as $\tilde{\mathbf{Q}} = [\mathbf{Q}^1, \mathbf{Q}^2, \dots, \mathbf{Q}^R]$, $\tilde{\mathbf{W}}_w = \begin{bmatrix} \mathbf{W}_w^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W}_w^R \end{bmatrix}$, and $\tilde{\mathbf{W}}_b = \begin{bmatrix} \mathbf{W}_b^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{W}_b^R \end{bmatrix}$, Equation (7) can be written as follows:

$$\begin{aligned} \min_{\tilde{\mathbf{Q}}, \mathbf{D}} \frac{\text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_w \tilde{\mathbf{Q}})}{\text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_b \tilde{\mathbf{Q}})}, \\ \text{s.t. } (\tilde{\mathbf{Q}})^T (\tilde{\mathbf{Q}}) = \mathbf{I}, \end{aligned} \quad (8)$$

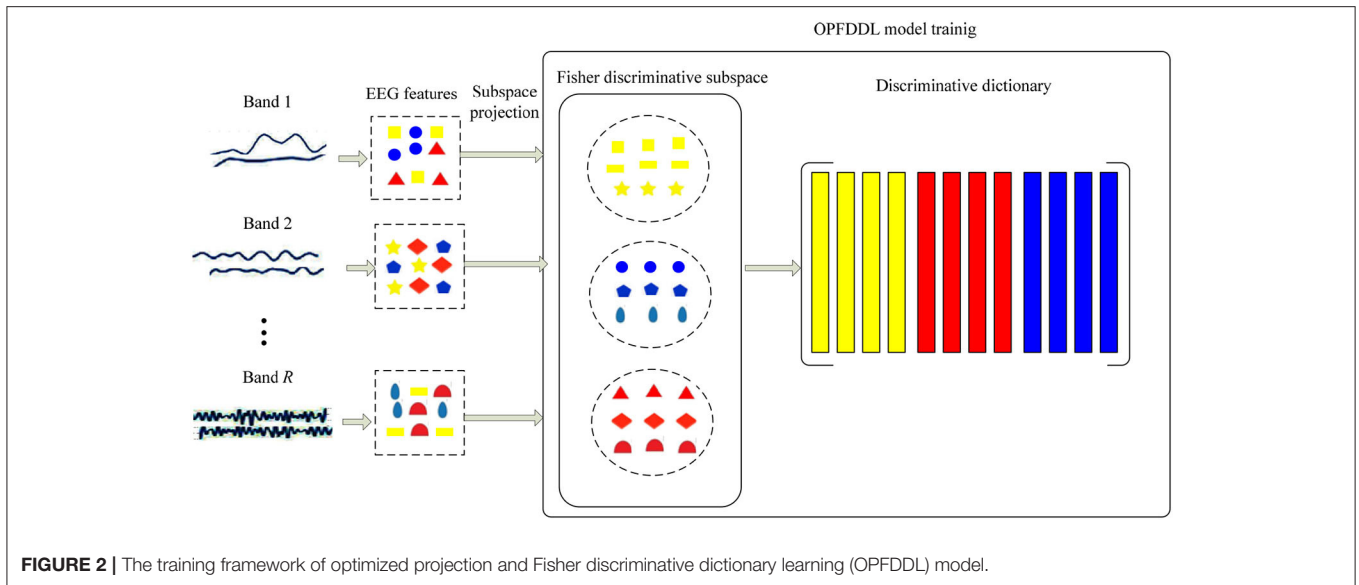


FIGURE 2 | The training framework of optimized projection and Fisher discriminative dictionary learning (OPFDDL) model.

The projection matrix $\tilde{\mathbf{Q}}$ is limited to be orthogonal, which is highly effective in the optimization process. The solution of Equation (8) refers to the complex inverse operation and is computationally intensive. Thus, we translate it into the following quadratic weighted optimization (QWO) problem and obtain the objective function of OPFDDL

$$\begin{aligned} \min_{\tilde{\mathbf{Q}}, \mathbf{D}, \mu} & \mu^2 \text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_w \tilde{\mathbf{Q}}) - \mu \text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_b \tilde{\mathbf{Q}}), \\ \text{s.t. } & \tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}} = \mathbf{I}, \end{aligned} \quad (9)$$

It is noted that the parameter μ is an adaptive weight that can be obtained by a closed-form solution but is not a manually adjusted parameter.

Optimization

In the following, the alternating optimization approach is used to update the parameters $\{\tilde{\mathbf{Q}}, \mathbf{D}, \mu\}$ in Equation (9).

- (1) Update step for $\tilde{\mathbf{Q}}$. With \mathbf{D} and μ fixed, and with the known $\tilde{\mathbf{W}}_w$ and $\tilde{\mathbf{W}}_b$, the optimization of $\tilde{\mathbf{Q}}$ can be solved by

$$(\mu^2 \tilde{\mathbf{W}}_w - \mu \tilde{\mathbf{W}}_b) \tilde{\mathbf{Q}} = \gamma \tilde{\mathbf{Q}}, \quad (10)$$

The projection matrix $\tilde{\mathbf{Q}}$ is constituted by the feature vector corresponding to the first d minimum eigenvalues of Equation (10).

- (2) Update step for \mathbf{D} . With the definition of

$$\begin{aligned} \tilde{\mathbf{X}} &= \begin{bmatrix} \mathbf{X}^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathbf{X}^R \end{bmatrix}, \quad \delta_{l_j}^r = [\delta_{l_j}(s_1^r), \delta_{l_j}(s_2^r), \dots, \delta_{l_j}(s_{n_r}^r)], \\ \zeta_{l_j}^r &= [\zeta_{l_j}(s_1^r), \zeta_{l_j}(s_2^r), \dots, \zeta_{l_j}(s_{n_r}^r)], \quad \Omega = [\delta_{l_j}^1, \delta_{l_j}^2, \dots, \delta_{l_j}^R], \\ \text{and } \Theta &= [\zeta_{l_j}^1, \zeta_{l_j}^2, \dots, \zeta_{l_j}^R], \text{ and with the known } \tilde{\mathbf{Q}} \text{ and } \mu, \end{aligned}$$

Equation (9) can be written as

$$\begin{aligned} \max_{\mathbf{D}} & \mu^2 \text{tr}[(\tilde{\mathbf{Q}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T \mathbf{D} \Omega)(\tilde{\mathbf{Q}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T \mathbf{D} \Omega)^T] \\ & - \mu \text{tr}[(\tilde{\mathbf{Q}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T \mathbf{D} \Theta)(\tilde{\mathbf{Q}}^T \tilde{\mathbf{X}} - \tilde{\mathbf{Q}}^T \mathbf{D} \Theta)^T], \end{aligned} \quad (11)$$

For each column of $\tilde{\mathbf{X}}$, i.e., $\tilde{\mathbf{X}}_k$, the optimization of \mathbf{D} can be solved by the following problem:

$$\frac{\partial L(\tilde{\mathbf{X}}_k)}{\partial \mathbf{D}} = 2\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T \mathbf{D}(\Omega\Omega^T + \Theta\Theta^T) - 2\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T \tilde{\mathbf{X}}_k(\Omega^T + \Theta^T), \quad (12)$$

Then, \mathbf{D} can be updated by

$$\mathbf{D} = \mathbf{D} - \frac{\lambda_D}{n} \sum_k \frac{\partial L(\tilde{\mathbf{X}}_k)}{\partial \mathbf{D}}, \quad (13)$$

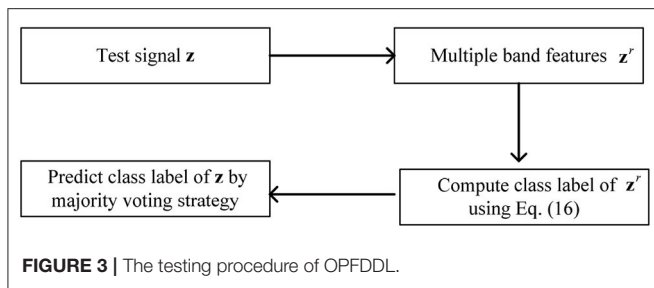
where λ_D is the step size.

- (3) Update step for \mathbf{S} . When \mathbf{D} , $\tilde{\mathbf{Q}}$, and μ are learned, the sparse code \mathbf{s} for each signal \mathbf{x} in the subspace can be obtained by

$$\min_{\mathbf{s}} \|\mathbf{Q}^T \mathbf{x} - \mathbf{D}\mathbf{s}\|_F^2 + \lambda \|\mathbf{s}\|, \quad (14)$$

- (4) Update step for μ . By recalling that the matrixes, $\tilde{\mathbf{W}}_b$ and $\tilde{\mathbf{W}}_w$ can be built according to the obtained $\tilde{\mathbf{Q}}$ and \mathbf{S} . When $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$ are learned, the solution of μ is $\frac{\partial L}{\partial \mu} = 0$, and it can be obtained by a closed-form solution

$$\mu = \frac{\text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_b \tilde{\mathbf{Q}})}{2\text{tr}(\tilde{\mathbf{Q}}^T \tilde{\mathbf{W}}_w \tilde{\mathbf{Q}})}, \quad (15)$$



Based on the above analysis, the implementation process of OPFDDL is described in Algorithm 1. We initialize the sub-dictionary for each class by the K-SVD algorithm, and then, we integrate them to form the initialization dictionary \mathbf{D} .

When the projection matrix \mathbf{Q} and dictionary \mathbf{D} are learned, we perform the following procedure to run testing work. The testing procedure of OPFDDL is illustrated in **Figure 3**. For each testing EEG signal \mathbf{z} , its r th frequency band feature is denoted as \mathbf{z}^r . We map \mathbf{z}^r into projection subspace using \mathbf{Q}^r and classify its class label according to the smallest reconstruction error on each class as follows:

$$y^r = \arg \min \left\| (\mathbf{Q}^r)^T \mathbf{z}^r - \mathbf{D}_j \tilde{\mathbf{D}}_j \mathbf{z}^r \right\|_2, \quad (16)$$

where $\tilde{\mathbf{D}}_j = (\mathbf{D}_j^T \mathbf{D}_j)^{-1} \mathbf{D}_j^T$ is the pseudo-inverse of dictionary \mathbf{D}_j of class j .

Finally, we use the majority voting to identify the class label of signal \mathbf{z} , i.e.,

$$y = \arg \max_j \Delta_j, \quad (17)$$

where Δ_j is the number of votes for class j .

Algorithm 1 The OPFDDL algorithm

Input: Multiple frequency band EEG signals $\mathbf{X}^r (r = 1, 2, \dots, R)$ with their class labels.

Output: Projection matrix \mathbf{Q} and dictionary \mathbf{D} .

Initialization: Initialize \mathbf{D} and \mathbf{S} using the K-SVD algorithm and initialize $\tilde{\mathbf{Q}}$ such that $\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T = \mathbf{1}$.

Repeat

Sparse codes update: Compute sparse code \mathbf{s} for each training sample using Equation (14);

Projection matrix update: Compute $\tilde{\mathbf{Q}}$ using Equation (10);

Dictionary update: Compute \mathbf{D} using Equations (11–13);

Adaptive weight update: Compute μ using Equation (15);

Until convergence

EXPERIMENT

Experimental Settings

Following the study of Li Y. et al. (2019), we used extracted methods of three features on the SEED dataset, including

differential entropy (DE), power spectral density (PSD), and fractal dimension (FD). We investigated the EEG features over all frequency bands per second with no overlap in each channel. We used the random 10 trials in each subject for model training and the rest 5 trials for testing. The classification performance corresponding to each period is recorded for each subject. For DREAMER dataset, to balance the number and length of the segments, we divided the 60-s EEG signals into 59 blocks with an overlap rate of 50%. The DE feature extraction method was carried out and 14-dimensional features for each frequency band were obtained. For each subject, we trained our model using the random 12 trials and the rest 6 trials for testing.

We compared our proposed model with five machine learning methods, including SVM (Cortes and Vapnik, 1995), K-SVD (Aharon et al., 2006), PCB-ICL-TSK (Ni et al., 2020b), DDL (Zhou et al., 2012), and dictionary pair learning (DPL) (Ameri et al., 2016). The Gaussian kernel and Gaussian fuzzy membership were used in SVM and PCB-ICL-TSK, respectively. The parameters in comparison methods were set according to the default settings in corresponding methods. In OPFDDL, the dimension of the projection subspace was set as 90% of the dimension of the EEG signal features. The number of atoms in each class was selected in {10, 15, 20, 25, 30, 35}. The λ parameter in Equation (2) was set as 0.01. We used the 5-fold cross-validation method to select the optimal parameters, and we performed five independent runs to evaluate the classification accuracy of all methods.

Experiment Results on the SEED Dataset

In this subsection, we performed the comparison experiments on the SEED dataset using various combinations of frequency bands and various features. The average accuracy performances of all methods with three feature methods are summarized in **Table 2**. From these results, we have the following observations: (1) Under different frequency band combinations, the results of total frequency bands of all methods are the best. For example, the classification accuracy of OPFDDL using all frequency bands is 6.70, 3.83, and 2.17% higher than that using frequency bands $\beta + \gamma$, $\alpha + \beta + \gamma$, and $\theta + \alpha + \beta + \gamma$. The classification accuracy of SVM using all frequency bands is 3.57, 2.90, and 1.76% higher than that using frequency bands $\beta + \gamma$, $\alpha + \beta + \gamma$, and $\theta + \alpha + \beta + \gamma$. In addition, in most cases, the SDs of all methods are small in all five bands. It demonstrates that multiple bands are helpful for EEG-based ER, due to that the features of each band have discrimination ability and five bands are complementary for distinguishing EEG emotions. (2) The classification performance of the three features is comparable. The performance of the DE feature is slightly better and shows an advantage in most of the cases. The classification accuracy of OPFDDL using the DE feature is 88.87%. It indicates that the DE feature is suitable to deal with EEG emotion signals. (3) OPFDDL outperforms all comparison methods, especially in the case of all five bands. It is because that OPFDDL can effectively integrate band-independent information and inter-band correlation information. The encouraging results indicate that direct concatenation of five frequency bands of EEG data

TABLE 2 | The average accuracies (SDs) of all methods under four combinations of frequency bands and three feature methods.

Features	Methods	$\beta + \gamma$	$\alpha + \beta + \gamma$	$\theta + \alpha + \beta + \gamma$	$\delta + \theta + \alpha + \beta + \gamma$
PSD	SVM	77.41	78.08	79.22	80.98
		(12.65)	(11.21)	(11.01)	(10.38)
	KSVD	77.03	78.00	79.46	81.28
		(12.28)	(11.76)	(11.28)	(10.47)
	PCB-ICL-TSK	77.53	80.10	80.29	81.26
		(11.76)	(11.29)	(10.63)	(10.23)
	DDL	79.79	80.85	81.78	83.16
		(10.49)	(10.56)	(10.01)	(9.45)
DPL	80.46	81.12	82.28	83.42	
	(10.58)	(10.02)	(9.21)	(8.64)	
DE	SVM	81.00	83.27	84.93	87.10
		(10.26)	(9.75)	(9.32)	(8.32)
	KSVD	79.58	80.46	80.89	81.89
		(12.76)	(12.88)	(11.43)	(10.93)
	PCB-ICL-TSK	78.32	78.83	80.07	82.06
		(12.77)	(12.05)	(12.11)	(10.85)
	DDL	79.66	80.55	81.36	82.49
		(11.63)	(11.09)	(10.35)	(10.04)
DPL	80.26	81.52	83.75	84.19	
	(10.56)	(10.09)	(9.43)	(9.11)	
FD	SVM	81.76	82.12	83.37	84.34
		(10.74)	(10.12)	(9.21)	(8.65)
	KSVD	82.09	85.34	86.92	88.87
		(10.30)	(9.65)	(9.21)	(8.29)
	PCB-ICL-TSK	78.02	78.94	79.42	81.25
		(12.43)	(12.32)	(11.86)	(11.01)
	DDL	77.56	78.18	79.79	81.61
		(12.48)	(12.02)	(11.63)	(11.12)
DPL	78.34	80.43	80.98	81.65	
	(11.57)	(11.21)	(10.68)	(9.95)	
OPFDDL	80.21	81.23	82.46	83.64	
	(10.75)	(10.12)	(9.24)	(9.15)	
OPFDDL	80.74	81.81	82.79	84.09	
	(10.35)	(10.07)	(9.31)	(9.17)	
OPFDDL	81.94	83.96	85.61	87.98	
	(10.14)	(9.64)	(9.32)	(9.01)	

The highest average accuracy is shown in bold type.

cannot well-exploit the inherent distinguishing characteristics of data. Considering the common information of multiband and band-specific information shared in each band, it is important to jointly learn multiple band representations. (4) Compared with K-SVD and ODFDL, OPFDDL generates the shared common dictionary on all frequency bands in the projected subspace, which can maintain the data structure of multiple frequency bands. In addition, based on the Fisher discrimination criterion of maximizing within-class compactness and minimizing between-class separation, OPFDDL can well

learn the discriminative dictionary from the cooperation of multiple frequency bands.

To further validate the discrimination ability of OPFDDL, in **Figure 4**, we reported the confusion matrix of the OPFDDL model with the DE feature. As shown in **Figure 4**, OPFDDL achieves a better classification performance on positive and neutral emotions than negative emotions in all cases. The performance result of OPFDDL is similar to that in references (Zheng and Lu, 2015; Li Y. et al., 2019). This suggests that subjects may have different EEG signals when experiencing negative

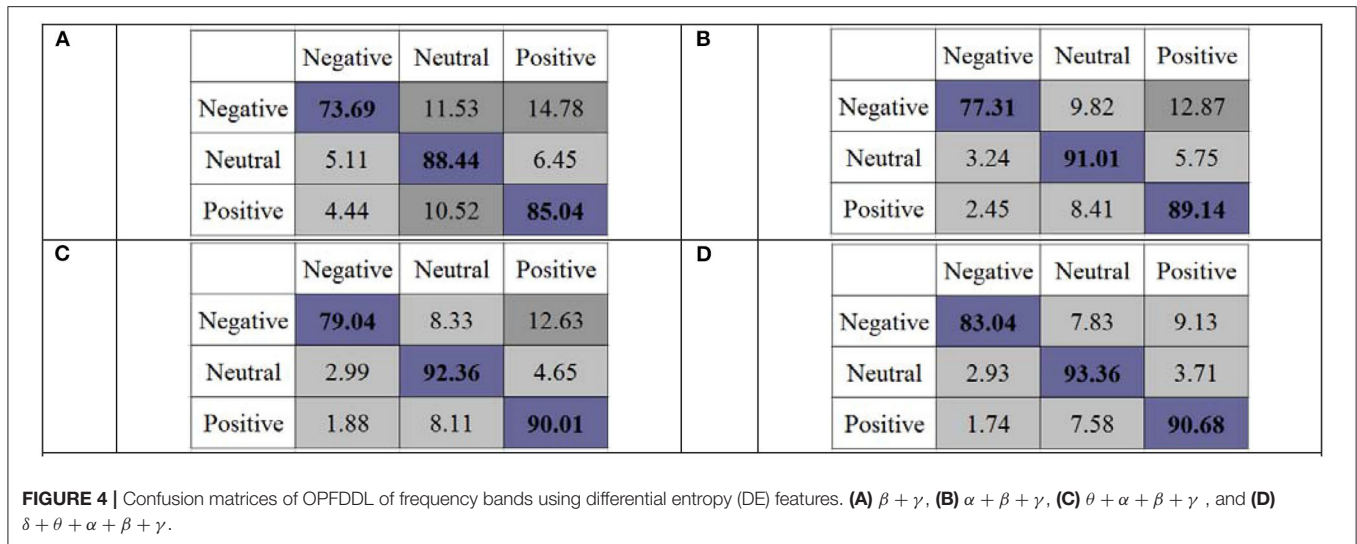


TABLE 3 | Average accuracies (SDs) of all methods using DE feature on DREAMER dataset.

Methods	Valence	Arousal	Dominance
SVM	86.35 (5.21)	86.19 (5.04)	86.43 (4.89)
KSVD	86.77 (5.28)	86.94 (5.08)	87.38 (4.89)
PCB-ICL-TSK	86.79 (5.26)	87.66 (5.02)	86.90 (4.87)
DDL	87.54 (5.26)	87.85 (5.00)	87.99 (4.93)
DPL	87.61 (5.21)	88.33 (4.94)	87.98 (4.90)
OPFDDL	89.84 (5.19)	90.11 (4.90)	89.96 (4.74)

The highest average accuracy is shown in bold type.

emotions and have similar EEG signals when experiencing positive and natural emotions. In addition, it can be seen that OPFDDL achieves the best classification performance (see **Figure 4D**) which uses all five frequency bands.

Experiment Results on the DREAMER Dataset

In this subsection, we performed the comparison experiments on the DREAMER dataset. We performed the comparison experiment using the DE feature. Similar to the SEED dataset, our model is compared with the abovementioned five methods. In the experiment, we verified the performance of OPFDDL according to valence, arousal, and dominance. The accuracy performance of all methods under the frequency band $\theta + \alpha + \beta$ is shown in **Table 3**. From the results, we can see that OPFDDL obtained the highest accuracies of 89.84, 90.11, and 89.96% in terms of arousal, valence, and dominance. Similar to the performance

TABLE 4 | Average accuracies of each subject for OPFDDL model using differential entropy (DE) feature on DREAMER dataset.

	Valence	Arousal	Dominance
Subject 1	93.52	93.13	95.65
Subject 2	90.45	90.08	86.04
Subject 3	86.26	85.53	81.46
Subject 4	96.87	96.32	96.84
Subject 5	93.71	94.56	95.22
Subject 6	81.05	81.04	78.04
Subject 7	84.46	84.02	82.64
Subject 8	85.42	84.97	86.73
Subject 9	81.27	82.43	85.05
Subject 10	88.95	88.04	90.43
Subject 11	82.43	83.10	87.65
Subject 12	92.48	92.15	93.04
Subject 13	89.26	92.07	90.65
Subject 14	95.19	94.66	94.86
Subject 15	98.53	98.75	98.05
Subject 16	87.16	86.78	86.43
Subject 17	87.84	89.01	88.05
Subject 18	89.45	89.02	87.21
Subject 19	94.77	95.43	94.05
Subject 20	91.01	88.03	89.23
Subject 21	93.16	94.20	94.07
Subject 22	94.83	96.04	95.11
Subject 23	88.35	93.21	92.75

results on the SEED dataset, OPFDDL performed best among all comparison methods. Based on the Fisher discrimination criterion, OPFDDL can well-learn the intrinsic relationships of EEG bands and can obtain the discriminative dictionary from multiple frequency bands cooperation in the projection subspace. In addition, the joint optimization strategy, which

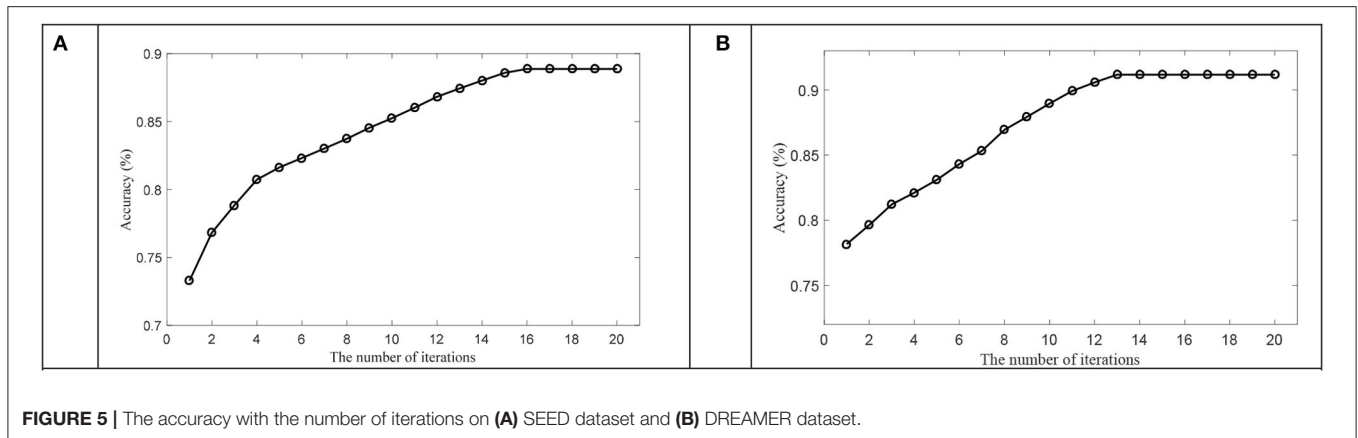


FIGURE 5 | The accuracy with the number of iterations on (A) SEED dataset and (B) DREAMER dataset.

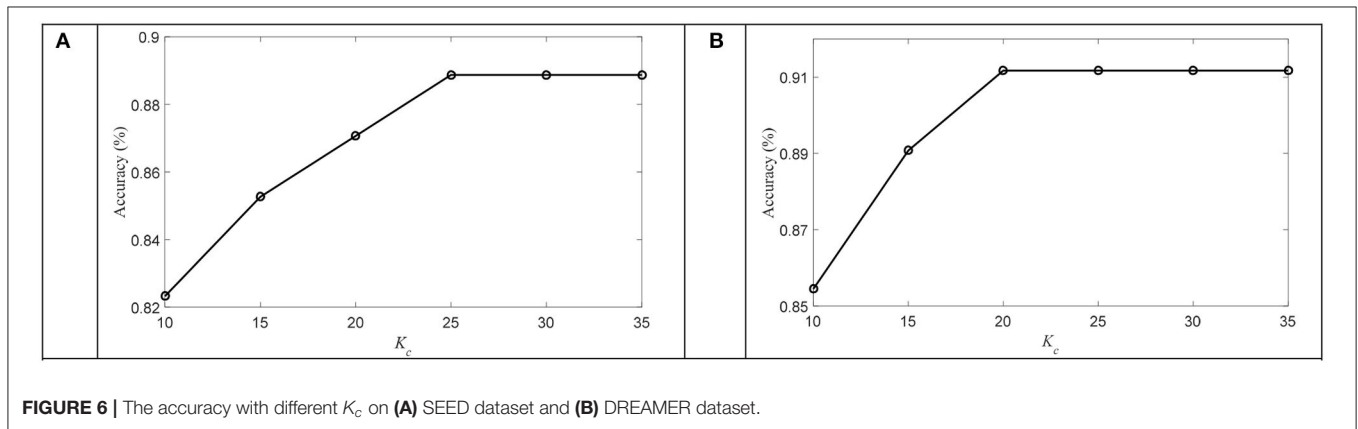


FIGURE 6 | The accuracy with different K_c on (A) SEED dataset and (B) DREAMER dataset.

addresses the shared projection subspace and dictionary learning, also can incrementally enhance the recognition performance of our proposed model. Thus, our proposed model can utilize more distinctive representations of multiple frequency bands of EEG signals.

Then, we recorded the average accuracies of each subject for the OPFDDL model using the DE feature in terms of arousal, valence, and dominance. The experimental results are shown in **Table 4**. The proposed OPFDDL model had achieved satisfactory recognition performance for all three dimensions of arousal, valence, and dominance. Based on the structure of dictionary learning and the principles of projection and Fisher discrimination criterion, OPFDDL can make better use of discriminative information of different frequency band data and has stronger generalization ability, so it can be effectively used in EEG emotion classification task.

Parameter Variations

In this subsection, we first discussed the convergence of OPFDDL using the DE feature of total frequency bands on SEED and DREAMER datasets. The threshold for iteration stop was set as 10^{-3} . **Figure 5** plots the accuracy that varies with the number of iterations on one subject in two datasets. The results verify the convergence of OPFDDL. It can be seen that the OPFDDL model can achieve convergence within 20 iterations.

Then, we discussed the number of atoms used in OPFDDL. The number of atoms in each class K_c was increased from 10 to 35 in increments of 5. **Figure 6** plots the accuracy that varies with the parameter K_c . The results show that after an initial dramatic increase, the classification accuracy of OPFDDL becomes stable after $K_c = 20$. In addition, the variation trend of accuracy is consistent on two datasets. Thus, the classification performance of OPFDDL is acceptable for small dictionary sizes.

CONCLUSIONS

Most previous machine learning methods focus on extracting feature representations for total frequency bands together without considering specific discriminative information of different frequency bands. In this study, we propose collaborative learning of multiple frequency bands for EEG-based ER. In particular, our model is an integration of projection and dictionary learning based on the Fisher discrimination criterion. For subspace projection optimization, a shared subspace is employed for each frequency band such that the band-specific representations and shared band-invariant information can be simultaneously utilized. For dictionary learning optimization, a shared dictionary is learned from the projected subspace where the Fisher discrimination criterion is used to minimize

within-class sparse reconstruction error and maximize between-class sparse reconstruction error. The joint learning strategy allows the model to extend easily. Consequently, we obtain a discriminative dictionary with a small size. We have performed the experiments and proved the performance of OPFDDL on two real-world EEG emotion datasets, i.e., SEED and DREAMER. For further studies, we will try to utilize and test more discriminative sparse representation criteria in our model. In addition, we only consider subject-dependent classification in EEG emotion identification. Applying this model to subject-independent classification is a challenging work.

DATA AVAILABILITY STATEMENT

The SEED dataset analyzed for this study can be found in this link (<http://bcmi.sjtu.edu.cn/~seed/seed.html>). The DREAMER

dataset analyzed for this study can be found in this link (<https://zenodo.org/record/546113>).

AUTHOR CONTRIBUTIONS

XG, JZho, and JZhu conceived and developed the theoretical framework of the study. All authors carried out experiment and data process, and drafted the study.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61806026 and by the Natural Science Foundation of Jiangsu Province under Grant No. BK20180956.

REFERENCES

- Abolghasemi, V., and Ferdowsi, S. (2015). EEG-fMRI: dictionary learning for removal of ballistocardiogram artifact from EEG. *Biomed. Signal Process. Control* 18, 186–194. doi: 10.1016/j.bspc.2015.01.001
- Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199
- Ameri, R., Pouyan, A., and Abolghasemi, V. (2016). Projective dictionary pair learning for EEG signal classification in brain computer interface applications. *Neurocomputing* 218, 382–389. doi: 10.1016/j.neucom.2016.08.082
- Barthélemy, Q., Gouy-Pailler, C., Isaac, Y., Souloumiac, A., Larue, A., and Mars, J. (2013). Multivariate temporal dictionary learning for EEG. *J. Neurosci. Methods* 215, 19–28. doi: 10.1016/j.jneumeth.2013.02.001
- Bhatti, A., Majid, M., Anwar, S., and Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Comput. Hum. Behav.* 65, 267–275. doi: 10.1016/j.chb.2016.08.029
- Brouwer, A., Zander, T., van Erp, J., Korteling, J., and Bronkhorst, A. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front. Neurosci.* 9:136. doi: 10.3389/978-2-88919-613-5
- Cortes, C., and Vapnik, V. (1995). Support vector machine. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Edgar, P., Torres, E. A., Hernández-Lvarez, M., and Sang, G. Y. (2020). EEG-based BCI emotion recognition: a survey. *Sensors* 20:5083. doi: 10.3390/s20185083
- Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H., and Dickhaus, H. (2012). Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Computer Methods Prog. Biomed.* 108, 10–19. doi: 10.1016/j.cmpb.2011.11.005
- Geng, S., Zhou, W., Yuan, Q., Cai, D., and Zeng, Y. (2011). EEG non-linear feature extraction using correlation dimension and hurst exponent. *Neurol. Res.* 33, 908–912. doi: 10.1179/1743132811Y.0000000041
- Gong, C., Zhou, P., Han, J., and Xu, D. (2019). Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection. *IEEE Trans. Image Process.* 28, 265–278. doi: 10.1109/TIP.2018.2867198
- Gu, X., Zhang, C., and Ni, T. (2020). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2020.3006699. [Epub ahead of print].
- Huang, J., Li, Y., Chen, B., Lin, C., and Yao, B. (2020). An intelligent EEG classification methodology based on sparse representation enhanced deep learning networks. *Front. Neurosci.* 14:808. doi: 10.3389/fnins.2020.00808
- Hwang, S., Hong, K., Son, G., and Byun, H. (2020). Learning CNN features from DE features for EEG-based emotion recognition. *Pattern Anal. Applicat.* 23, 1323–1335. doi: 10.1007/s10044-019-00860-w
- Jiang, Z., Lin, Z., and Davis, L. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88
- Kanoga, S., Kanemura, A., and Asoh, H. (2019). Multi-scale dictionary learning for ocular artifact reduction from single-channel electroencephalograms. *Neurocomputing* 347, 240–250. doi: 10.1016/j.neucom.2019.02.060
- Kashefpoor, M., Rabbani, H., and Barekatin, M. (2019). Supervised dictionary learning of EEG signals for mild cognitive impairment diagnosis. *Biomed. Signal Process. Control* 53:101559. doi: 10.1016/j.bspc.2019.101559
- Katsigiannis, S., and Ramzan, N. (2018). Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low cost off-the-shelf devices. *IEEE J. Biomed. Health Informat.* 22, 98–107. doi: 10.1109/JBHI.2017.2688239
- Ko, K., Yang, H., and Sim, K. (2009). Emotion recognition using EEG signals with relative power values and bayesian network. *Int. J. Control Automat. Syst.* 7:865. doi: 10.1007/s12555-009-0521-0
- Kutepov, I., Dobriyan, V., Zhigalov, M., Stepanov, M., Krysko, A., Yakovleva, T., et al. (2020). EEG analysis in patients with schizophrenia based on Lyapunov exponents. *Informat. Med. Unlock.* 18:100289. doi: 10.1016/j.imu.2020.100289
- Li, L., Li, S., and Fu, Y. (2013). “Discriminative dictionary learning with low-rank regularization for face recognition,” in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (Shanghai: IEEE).
- Li, X., Fan, H., Wang, H., and Wang, L. (2019). Common spatial patterns combined with phase synchronization information for classification of EEG signals. *Biomed. Signal Process. Control* 52, 248–256. doi: 10.1016/j.bspc.2019.04.034
- Li, Y., Zheng, W., Cui, Z., Zong, Y., and Ge, S. (2019). EEG emotion recognition based on graph regularized sparse linear regression. *Neural Process. Lett.* 49, 555–571. doi: 10.1007/s11063-018-9829-1
- Nakisa, B., Rastgoo, M., Tjondronegoro, D., and Chandrana, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Syst. Applicat.* 93, 143–155. doi: 10.1016/j.eswa.2017.09.062
- Ni, T., Gu, X., and Jiang, Y. (2020a). Transfer discriminative dictionary learning with label consistency for classification of EEG signals of epilepsy. *J. Ambient Intell. Hum. Comput.* doi: 10.1007/s12652-020-02620-9. [Epub ahead of print].
- Ni, T., Gu, X., and Zhang, C. (2020b). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Front. Neurosci.* 14:837. doi: 10.3389/fnins.2020.00837
- 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
- Petrantonakis, P., and Hadjileontiadis, L. (2010). Emotion recognition from EEG using higher order crossings. *IEEE Trans. Informat. Technol. Biomed.* 14, 186–197. doi: 10.1109/TITB.2009.2034649

- Song, T., Zheng, W., Song, P., and Cui, Z. (2020). EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans. Affect. Comput.* 11, 532–541. doi: 10.1109/TAFFC.2018.2817622
- Sreeja, S., and Himanshu, D. (2020). Distance-based weighted sparse representation to classify motor imagery EEG signals for BCI applications. *Multimed. Tools Applicat.* 79, 13775–13793. doi: 10.1007/s11042-019-08602-0
- Wei, C., Chen, L., Song, Z., Lou, X., and Li, D. (2020). EEG-based emotion recognition using simple recurrent units network and ensemble learning. *Biomed. Signal Process. Control* 58:101756. doi: 10.1016/j.bspc.2019.101756
- Wu, T., Yan, G., Yang, B., and Sun, H. (2008). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement* 41, 618–625. doi: 10.1016/j.measurement.2007.07.007
- Zhang, G., Yang, J., Zheng, Y., Luo, Z., and Zhang, J. (2021). Optimal discriminative feature and dictionary learning for image set classification. *Informat. Sci.* 547, 498–513. doi: 10.1016/j.ins.2020.08.066
- Zheng, W., Liu, W., Lu, Y., Lu, B., and Cichocki, A. (2019). Emotion meter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybernet.* 49, 1110–1122. doi: 10.1109/TCYB.2018.2797176
- 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. Autom. Ment. Dev.* 7, 162–175. doi: 10.1109/TAMD.2015.2431497
- Zheng, Z., and Sun, H. (2019). Jointly discriminative projection and dictionary learning for domain adaptive collaborative representation-based classification. *Pattern Recognit.* 90, 325–336. doi: 10.1016/j.patcog.2019.01.004
- Zhou, W., Yang, Y., and Yu, Z. (2012). “Discriminative dictionary learning for EEG signal classification in brain-computer interface,” in *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)* (Guangzhou), 1582–1585. doi: 10.1109/ICARCV.2012.6485382
- 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.
- Copyright © 2021 Gu, Fan, Zhou and Zhu. 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.