

S³LRR: A Unified Model for Joint Discriminative Subspace Identification and Semisupervised EEG Emotion Recognition

Yong Peng¹, Member, IEEE, Yikai Zhang², Wanzeng Kong³, Member, IEEE, Feiping Nie⁴, Senior Member, IEEE, Bao-Liang Lu⁵, Fellow, IEEE, and Andrzej Cichocki⁶, Life Fellow, IEEE

Abstract—Emotion recognition from electroencephalogram (EEG) data has been a research spotlight in both academic and industrial communities, which lays a solid foundation to achieve harmonic human-machine interaction. However, most of the existing studies either directly performed classification on primary EEG features or employed a two-stage paradigm of “feature transformation plus classification” for emotion recognition. The former usually cannot obtain promising performance, while the latter inevitably breaks the connection between feature transformation and recognition. In this article, we propose a simple yet effective model named semisupervised sparse low-rank regression (S³LRR) to unify the discriminative subspace identification and semisupervised emotion recognition together. Specifically, S³LRR is formulated by decomposing the projection matrix in least square regression (LSR) into two factor matrices, which complete the discriminative subspace identification and connect the subspace EEG data representation with emotional states. Experimental studies on the benchmark SEED_V dataset show that the emotion recognition performance is greatly improved by the joint learning mechanism of S³LRR.

Furthermore, S³LRR exhibits additional abilities in affective activation patterns exploration and EEG feature selection.

Index Terms—Discriminative subspace identification, electroencephalogram (EEG), emotion recognition, low-rank regression, semisupervised classification.

I. INTRODUCTION

EMOTIONAL intelligence along with the logical intelligence are considered as the two complementary aspects to achieve artificial intelligence, which primarily aims to enable machine the ability of recognizing the emotional states of human beings. Compared with the widely used data modalities such as image, video, speech, and text [1]–[3], electroencephalogram (EEG) has its unique advantages of high time resolution and difficult to camouflage in emotion recognition since it is directly generated from the neural activities of central nervous system [4]. Therefore, EEG provides a new path for objective emotion recognition and some other brain-computer interface applications [5], which have been drawing a lot of attention from academic and industrial communities in past decades.

Currently, the general pipeline for the stimulus-evoked EEG emotion recognition consists of three stages of *preprocessing*, *feature extraction and transformation*, and *classification*. Since EEG is weak and easily contaminated by various electrophysiological artifacts during the data collection process, in the preprocessing stage, we mainly aim to remove these artifacts such as electromyogram, electrocardiogram, and electrooculogram to obtain purified EEG data for subsequent analysis [6]. Usually, different types of EEG features can be extracted from time domain, frequency domain, time-frequency domain, and spatial domain [7], [8], among which the power spectral density (PSD) and differential entropy (DE) [9] are especially widely used in EEG-based emotion recognition. Most of the time, machine learning-based methods are used to transform the primary EEG features in order to further enhance their discriminative ability [10]. Finally, classifiers, such as support vector machine (SVM) and sparse representation, are employed for recognizing the emotional states [11], [12].

However, only a few deep learning models have the ability to perform emotion recognition in an end-to-end manner,

Manuscript received December 19, 2021; revised March 23, 2022; accepted March 29, 2022. Date of publication April 8, 2022; date of current version April 21, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0116800, in part by the National Natural Science Foundation of China under Grant 61971173 and Grant U20B2074, in part by the Natural Science Foundation of Zhejiang Province under Grant LY21F030005, in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant GK209907299001-008, in part by the CAAC Key Laboratory of Flight Techniques and Flight Safety under Grant FZ2021KF16, and in part by the Guangxi Key Laboratory of Optoelectronic Information Processing (Guilin University of Electronic Technology) under Grant GD21202. The Associate Editor coordinating the review process was Bardia Yousefi. (*Corresponding author: Wanzeng Kong.*)

Yong Peng is with the School of Computer Science and Technology and the Zhejiang Key Laboratory of Brain-Machine Collaborative Intelligence, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the Key Laboratory of Flight Techniques and Flight Safety, Civil Aviation Administration of China (CAAC), Guanghan 618307, China.

Yikai Zhang and Wanzeng Kong are with the School of Computer Science and Technology and the Zhejiang Key Laboratory of Brain-Machine Collaborative Intelligence, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: kongwanzeng@hdu.edu.cn).

Feiping Nie is with the School of Artificial Intelligence, OPTics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi’an 710072, China.

Bao-Liang Lu is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

Andrzej Cichocki is with the Center for Computational and Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, 143026 Moscow, Russia.

Digital Object Identifier 10.1109/TIM.2022.3165741

1557-9662 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

which directly takes raw EEG data as the input and outputs the recognition results [13]. Most of the existing studies performed classification on either primary EEG features or transformed EEG features to determine the emotional states [14]–[16]. Obviously, the former paradigm usually cannot obtain promising recognition performance, while the latter two-stage paradigm breaks the connection between feature transformation and classification. In this article, we propose a simple yet effective model termed semisupervised sparse low-rank regression (S^3LRR) for joint discriminative subspace identification and semisupervised emotion recognition from EEG. Mathematically, it is formulated by replacing the projection matrix in least square regression (LSR) with the multiplication of two factor matrices. Functionally, one factor matrix acts as exploring a discriminative subspace to make the data more separable and the other bridges the EEG data of subspace representation with emotional states. Besides, S^3LRR is implemented in the semisupervised paradigm in which the soft label matrix of unlabeled samples is jointly estimated to facilitate the discriminative subspace identification. We conduct extensive experiments on a benchmark emotional EEG dataset, and the results of recognition accuracy, feature selection performance, and affective activation patterns exploration show the effectiveness of our proposed S^3LRR model.

Compared with the existing studies, we summarize the main contributions of this article as follows.

- 1) We propose a new machine learning model termed S^3LRR to unify the two tasks of discriminative subspace identification and emotional state recognition together. In particular, S^3LRR uses the multiplication of two factor matrices to replace the single projection matrix in LSR, which works for the discriminative subspace identification and emotion recognition.
- 2) We implement S^3LRR in the semisupervised paradigm that is more appropriate for cross-session emotion recognition. The immediate benefit is that the discriminative subspace identification can be effectively guided by the estimation of soft labels of unlabeled samples. These two objectives can be jointly optimized toward the optimum.
- 3) We enforce the multiplication of two factor matrices in S^3LRR to be row sparse, which not only assigns it the ability of efficient EEG feature selection but also provides us an efficient tool for affective activation patterns mining based on the quantitative feature importance measure.

The remainder of this article is structured as follows. In Section II, we provide some background knowledge on EEG-based emotion recognition. Section III introduces the S^3LRR model formulation and its optimization in detail. Experimental studies are conducted in Section IV. Section V concludes the whole article and points out the future work.

Notations: In this article, we use *Delta*, *Theta*, *Alpha*, *Beta*, and *Gamma* to denote the EEG frequency bands. Greek letters, such as θ and λ , represent the model variables or parameters. Matrices and vectors are, respectively, denoted by boldface uppercase and lowercase letters. The $\ell_{2,1}$ -norm of matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ is defined as $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^m (\sum_{j=1}^n m_{ij}^2)^{1/2} = \sum_{i=1}^m$

$\|\mathbf{m}^i\|_2$, where \mathbf{m}^i is the i th row of \mathbf{M} . In particular, $\mathbf{1}_n$ denotes an all-one column vector and the subscript n indicates its length.

II. BACKGROUND

In this section, we provide a brief introduction to recent advances in EEG-based emotion.

At present, emotion recognition from EEG is mainly based on machine learning methods; therefore, we make a review along the path of EEG feature extraction and emotional state classification. In [17], the popular EEG features for emotion recognition were extensively reviewed. For example, the time-domain EEG features, such as the statistics, event-related potential, energy, and high-order zero crossings, are the most intuitive since EEG data are directly collected in time domain. After transforming it from time domain to frequency domain by Fourier transform (FT), we can extract the features such as PSD, DE, and higher order spectrum, which are usually more stable than time-domain features. Since EEG data are nonstationary, wavelet transform and short-time FT are usually used to extract the time–frequency domain features to capture the local frequency information. To make better use of the multichannel property of EEG data, features, such as differential asymmetry, rational asymmetry, and connectivity, can be built to explore the spatial information [18]–[20].

The machine learning models in emotional EEG data processing can be roughly categorized into linear and nonlinear ones. To select the most beneficial samples to label, Wu and Huang [21] proposed two multitask active learning models for affect estimation in the 3-D space of valence, arousal, and dominance. In [22], considering the complementary effect of activation features (i.e., PSD and DE) and network patterns (i.e., C-Coefficient, SP-Length, G-Efficient, and L-Efficient), a feature fusion approach was adopted to combine them for emotion recognition. Based on the hypergraph theory, Liang *et al.* [23] proposed to divide the EEG-based hypergraph into a specific number of clusters, with each cluster corresponding to one emotional state. Though some linear models were extended to nonlinear ones, such as the SVM with radial basis function (RBF) kernel [24], kernel Fisher’s discriminant analysis [25], and transfer component analysis [26], by kernel trick to enhance their nonlinear modeling ability, existing nonlinear models mainly utilized neural networks for feature learning. In [27], two different types of random networks, random functional vector link and extreme learning machine, were used for cross-session emotion recognition from EEG. Compared with the shallow ones, deep neural networks show more powerful nonlinear learning abilities. Deep belief network (DBN) was used for cross-session EEG emotion recognition and the mean absolute weight distribution of the trained DBNs provides clue for critical EEG frequency bands identification [28]. Song *et al.* [29] proposed a dynamic graph convolutional neural network (DGCNN) to learn the intrinsic relationship among EEG channels. Based on the observation that different brain regions sampled by EEG electrodes may be related to different brain functions, a sparse DGCNN model was proposed by taking the localized

and sparse functional relations among electrodes into consideration [30]. In [31], a deep learning model was proposed to suppress the cross-subject differences by simultaneously minimizing the classification error on the source subject and aligning the EEG data discrepancies between source and target subjects. Though deep learning models achieved promising results in diverse EEG-based applications, there also have some limitations such as the black-box training mode, complicated to implement, time-consuming to train and requiring a lot of training samples [13].

III. METHOD

In this section, we first formulate the objective function of S³LRR and then introduce its optimization method. Moreover, some discussions on S³LRR and one extended model are provided.

A. Model Formulation

In semisupervised learning, we are usually given an EEG data collection matrix $\mathbf{X} = [\mathbf{X}_l, \mathbf{X}_u] \in \mathbb{R}^{d \times n}$ consisting of l labeled and u unlabeled samples. $\mathbf{Y}_l \in \mathbb{R}^{l \times c}$ is the label indicator matrix of labeled samples, which uses the one-hot encoding to represent the emotional state membership of samples. In particular, if EEG sample $\mathbf{x}_i|_{i=1}^l$ is from the j th emotional state and $\mathbf{y}^i \in \mathbb{R}^{1 \times c}$ is the i th row of \mathbf{Y}_l , then the j th element of \mathbf{y}^i is one and all the others of \mathbf{y}^i are zeros. \mathbf{Y}_u is an unknown label matrix corresponding to the unlabeled samples, and $\mathbf{Y} = [\mathbf{Y}_l; \mathbf{Y}_u] \in \mathbb{R}^{n \times c}$ is the combined label matrix corresponding to \mathbf{X} . Here, d is the dimensionality of samples, c is the number of emotional states, and $n = l + u$ is the total number of EEG samples. Our task is to estimate $\mathbf{Y}_u \in \mathbb{R}^{u \times c}$ as accurate as possible given \mathbf{X} and \mathbf{Y}_l .

Usually, connections between EEG data matrix and the emotional label matrix are directly built. For example, if the ℓ_2 -norm regularized LSR is used in the supervised manner, we have the following objective:

$$\min_{\mathbf{W}} \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}_l\|_2^2 + \lambda \|\mathbf{W}\|_2^2 \quad (1)$$

based on which we can fit the projection matrix $\mathbf{W} \in \mathbb{R}^{d \times c}$ by $(\mathbf{X}_l, \mathbf{Y}_l)$. Then, the prediction \mathbf{Y}_u can be obtained by $\mathbf{X}_u^T \mathbf{W}$. By extending (1) into semisupervised learning, we have

$$\min_{\mathbf{W}, \mathbf{Y}_u} \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_2^2 + \lambda \|\mathbf{W}\|_2^2, \quad \text{s.t. } \mathbf{Y}_u \geq \mathbf{0}, \quad \mathbf{Y}_u \mathbf{1}_c = \mathbf{1}_u. \quad (2)$$

The second constraint means that the summation of elements in each of \mathbf{Y}_u should be one. Together with the nonnegativity constraint, the elements in each row of \mathbf{Y}_u can be considered as the probabilities of a sample belonging to different emotional states. Therefore, we can directly determine the emotional state of each sample by checking the location of the largest value in each row of \mathbf{Y}_u . For example, if the third row of \mathbf{Y}_u is [0.04, 0.81, 0.01, 0.11, 0.03], then the third unlabeled sample should be categorized into the second state. Obviously, the improvements from supervised version to semisupervised version are two folds. One is the incorporation of unlabeled samples into the learning process, and the other is that \mathbf{Y}_u

is treated as a variable and jointly optimized with the other model variable \mathbf{W} .

However, establishing direct connection between EEG data matrix and the label matrix is too rigorous for the projection matrix to well capture the properties of EEG data since the complexity of EEG data makes it usually not so easy to handle. An ideal way might be first projecting EEG data into a discriminative subspace to enhance its separability and then mapping such subspace data representation to an emotional label matrix. To this end, as shown in Fig. 1, we propose a new model termed S³LRR to seamlessly unify the discriminative subspace identification and semisupervised emotion recognition together, which can effectively avoid the suboptimality limitation caused by the two-stage manner of “feature extraction/transformation plus classification.”

Supposing that $\mathbf{A} \in \mathbb{R}^{d \times s}$ is the projection matrix to induce a discriminative subspace and $\mathbf{B} \in \mathbb{R}^{s \times c}$ is the matrix to bridge the subspace data representation with the label information, where s is the subspace dimensionality. The objective function of our S³LRR model can be obtained by mathematically replacing \mathbf{W} in (2) with \mathbf{AB} , namely,

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{Y}_u} \|\mathbf{X}^T \mathbf{AB} - \mathbf{Y}\|_2^2 + \frac{\lambda}{2} \|\mathbf{AB}\|_{2,1} \\ \text{s.t. } \mathbf{Y}_u \geq \mathbf{0}, \quad \mathbf{Y}_u \mathbf{1}_c = \mathbf{1}_u. \quad (3)$$

Here, we use the $\ell_{2,1}$ -norm instead of the ℓ_2 -norm in order to enforce the row sparsity of \mathbf{AB} , which potentially achieves the adaptive feature weighting. Based on the definition of $\ell_{2,1}$ -norm, (3) is equivalent to

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{Y}_u} \|\mathbf{X}^T \mathbf{AB} - \mathbf{Y}\|_2^2 + \lambda \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{DAB}) \\ \text{s.t. } \mathbf{Y}_u \geq \mathbf{0}, \quad \mathbf{Y}_u \mathbf{1}_c = \mathbf{1}_u \quad (4)$$

where $\mathbf{D} \in \mathbb{R}^{d \times d}$ is a diagonal matrix and its i th diagonal element is defined as

$$d_{ii} = \frac{1}{2\|\mathbf{g}^i\|_2}, \quad i = 1, 2, \dots, d. \quad (5)$$

Here, \mathbf{g}^i is the i th row of matrix $\mathbf{G} = \mathbf{AB}$. $\|\mathbf{g}^i\|_2$ is the ℓ_2 -norm of the i th row of \mathbf{G} , which is defined by $(g_{i1}^2 + g_{i2}^2 + \dots + g_{ic}^2)^{1/2}$.

B. Model Optimization

There are three variables, i.e., \mathbf{A} , \mathbf{B} , and \mathbf{Y}_u , in the S³LRR model objective function (4). In the following, we propose to update them in an alternate manner.

- 1) Update \mathbf{Y}_u with \mathbf{A} and \mathbf{B} fixed. By denoting $\mathbf{M} \triangleq \mathbf{X}_u^T \mathbf{AB}$, we have the subobjective function of \mathbf{Y}_u as

$$\min_{\mathbf{Y}_u} \|\mathbf{M} - \mathbf{Y}_u\|_2^2, \quad \text{s.t. } \mathbf{Y}_u \geq \mathbf{0}, \quad \mathbf{Y}_u \mathbf{1}_c = \mathbf{1}_u. \quad (6)$$

By row-wisely decoupling the above objective function and denoting $\mathbf{y}^i|_{i=1}^u$ as the i th row of \mathbf{Y}_u , we have

$$\min_{\mathbf{y}^i} \|\mathbf{m}^i - \mathbf{y}^i\|_2^2, \quad \text{s.t. } \mathbf{y}^i \geq \mathbf{0}, \quad \mathbf{y}^i \mathbf{1}_c = \mathbf{1} \quad (7)$$

which specifies an Euclidean distance defined on a simplex constraint [32]. The detailed optimization method to (7) is provided in Appendix A.

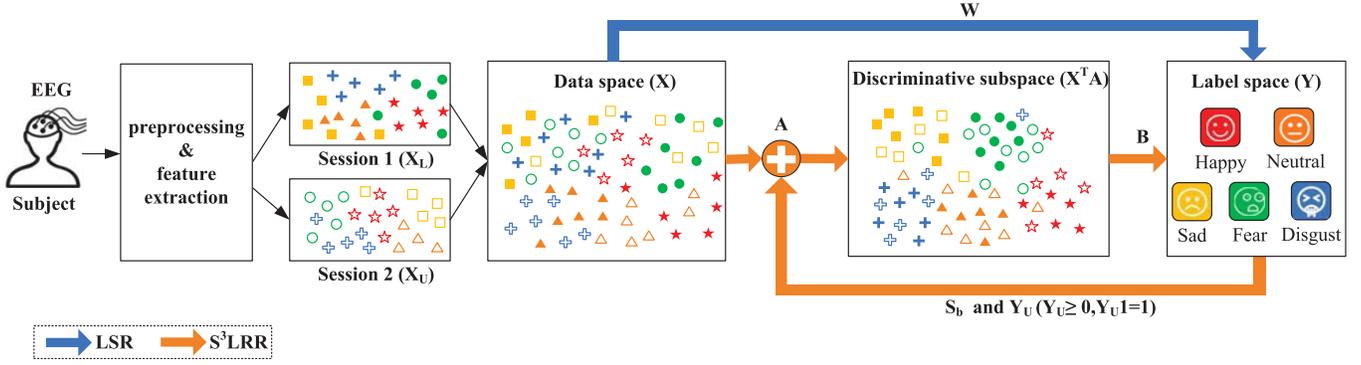


Fig. 1. General framework of our proposed S^3LRR model.

- 2) Update \mathbf{B} with \mathbf{Y}_u and \mathbf{A} fixed. Taking the derivative of (4) with respect to \mathbf{B} and setting it to zero, we have

$$\mathbf{B} = (\mathbf{A}^T (\mathbf{X}\mathbf{X}^T + \lambda\mathbf{D})\mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}\mathbf{Y}. \quad (8)$$

- 3) Update \mathbf{A} with \mathbf{Y}_u and \mathbf{B} fixed. Substituting (8) back into (4), we achieve the subobjective function in terms of variable \mathbf{A} as

$$\max_{\mathbf{A}} \text{Tr} \left((\mathbf{A}^T (\mathbf{X}\mathbf{X}^T + \lambda\mathbf{D})\mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}\mathbf{Y}\mathbf{Y}^T \mathbf{X}^T \mathbf{A} \right). \quad (9)$$

Note that

$$\mathbf{S}_t = \mathbf{X}\mathbf{X}^T, \quad \mathbf{S}_b = \mathbf{X}\mathbf{Y}\mathbf{Y}^T \mathbf{X}^T \quad (10)$$

where \mathbf{S}_t and \mathbf{S}_b are the total-class scatter matrix and the between-class scatter matrix defined in a linear discriminant analysis (LDA), respectively. Therefore, the solution of problem (9) is

$$\mathbf{A}^* = \arg \max_{\mathbf{A}} \left\{ \text{Tr} \left((\mathbf{A}^T (\mathbf{S}_t + \lambda\mathbf{D})\mathbf{A})^{-1} \mathbf{A}^T \mathbf{S}_b \mathbf{A} \right) \right\}. \quad (11)$$

Its global optimal solution is the top s eigenvectors of $(\mathbf{S}_t + \lambda\mathbf{D})^{-1} \mathbf{S}_b$ corresponding to the nonzero eigenvalues.

According to (5), the diagonal matrix \mathbf{D} should be updated when \mathbf{A} and \mathbf{B} are updated. Since the between-class scatter matrix \mathbf{S}_b relies on the estimation of \mathbf{Y} , we should also update it when \mathbf{Y}_u is obtained. As a whole, we summarize the complete optimization procedure to objective function (4) in Algorithm 1.

C. Discussions on S^3LRR

In the following, we first summarize the main characteristics of S^3LRR and then explain the differences between it and one related model, sparse low-rank regression (SLRR) [33].

On the characteristics of S^3LRR , they are listed as follows.

- 1) Functionally, matrix \mathbf{A} aims at exploring a discriminative subspace where the EEG samples are easier to separate and matrix \mathbf{B} performs the mapping from subspace EEG data representation to the corresponding label matrix.
- 2) Mathematically, the model objective of S^3LRR is formulated by replacing \mathbf{W} with the multiplication of two factor matrices, i.e., \mathbf{A} and \mathbf{B} . This makes that our

Algorithm 1 Optimization to S^3LRR Objective Function

Input: data matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$, label matrix $\mathbf{Y}_l \in \mathbb{R}^{l \times c}$, low-rank parameter s and regularization parameter λ ;

Output: projection matrices $\mathbf{A} \in \mathbb{R}^{d \times s}$ and $\mathbf{B} \in \mathbb{R}^{s \times c}$, and label matrix $\mathbf{Y}_u \in \mathbb{R}^{u \times c}$.

- 1: Initialize $t = 0$, $\mathbf{Y}_u^{(t)} = \frac{\mathbf{1}_u \mathbf{1}^T}{u} \in \mathbb{R}^{u \times c}$ and $\mathbf{D}^{(t)} \in \mathbb{R}^{d \times d}$ as an identity matrix;
- 2: **while** not converged **do**
- 3: Calculate $\mathbf{A}^{(t+1)}$ by equation (11);
- 4: Calculate $\mathbf{B}^{(t+1)}$ by equation (8);
- 5: Update the diagonal matrix $\mathbf{D}^{(t+1)}$ where its i th diagonal element is $\frac{1}{2\|(\mathbf{A}^{(t+1)}\mathbf{B}^{(t+1)})^T\|_2}$;
- 6: Calculate $\mathbf{Y}_u^{(t+1)}$ by solving (7) with Algorithm 3;
- 7: Update \mathbf{S}_b based on $\mathbf{Y}^{(t+1)} = [\mathbf{Y}_l; \mathbf{Y}_u^{(t+1)}]$;
- 8: $t = t + 1$;
- 9: **end while**

S^3LRR model has a succinct objective function, which is also easy to optimize.

- 3) Inspired by the model optimization, we realize that the projection matrix \mathbf{A} has an explicit meaning of essentially performing the LDA operation.
- 4) Under the semisupervised learning paradigm, the soft label matrix \mathbf{Y}_u of unlabeled samples is jointly optimized with the other variables. In particular, based on \mathbf{Y}_u , the between-class scatter matrix \mathbf{S}_b could be estimated for better updating matrix \mathbf{A} . This desirable property of S^3LRR is explicitly highlighted in Fig. 1.

The connections as well as differences between S^3LRR and SLRR are summarized in the following.

- 1) From the model formulation perspective, our proposed S^3LRR model is inspired by the existing SLRR model. Both of them aim to perform joint discriminative subspace exploration and recognition.
- 2) S^3LRR is a semisupervised extension of SLRR, which involves the unlabeled EEG samples into the learning process and therefore is more appropriate for the cross-session EEG-based emotion recognition, that is, S^3LRR can jointly estimate the label information of unlabeled samples and the other model variables.

- 3) Based on the learned combined projection matrix \mathbf{AB} , S³LRR has the ability to perform out-of-sample prediction on unseen EEG samples. Therefore, it is a pure semisupervised model.
- 4) We assign the combined projection matrix \mathbf{AB} a unique sense of meaning in EEG-based emotion recognition, based on which we can explore the affective activation patterns on critical EEG frequency bands and brain regions. To simplify the notations, we still use $\mathbf{G} \in \mathbb{R}^{d \times c}$ to denote the multiplication of obtained optimal matrices \mathbf{A} and \mathbf{B} . Suppose that $\boldsymbol{\theta} \in \mathbb{R}^d$ is a vector to characterize the importance of different EEG feature dimensions in recognizing different emotional states. Inspired by [34], the importance of each feature dimension can be measured by its normalized ℓ_2 -norm, i.e.,

$$\theta_i = \frac{\|\mathbf{g}^i\|_2}{\sum_{j=1}^d \|\mathbf{g}^j\|_2}, \quad i = 1, 2, \dots, d. \quad (12)$$

Besides, since there exists the coupling relationship between each feature dimension and each EEG frequency band (channel) [35], we can automatically perform the critical EEG frequency bands and channels identification according to the quantitative feature importance vector $\boldsymbol{\theta}$. Considering that we have an emotional EEG dataset consisting of P frequency bands and Q channels, then, for the $p|_{p=1}^P$ th EEG frequency band, its importance can be calculated by

$$\omega(p) = \theta_{(p-1)*Q+1} + \theta_{(p-1)*Q+2} + \dots + \theta_{p*Q}. \quad (13)$$

Similarly, the importance of the $q|_{q=1}^Q$ th EEG channel is

$$\psi(q) = \theta_q + \theta_{q+Q} + \dots + \theta_{q+(P-1)*Q}. \quad (14)$$

As stated by some existing studies, the affective EEG activation patterns exploration provides more insights into the understanding of neural mechanism in emotion expression [35], [36]. Besides, this might provide underlying theoretical support for customizing the emotion-related EEG data acquisition devices.

D. Extension From S³LRR to S²LRR

If we do not explicitly impose the $\ell_{2,1}$ -norm-based feature weighting on the combined projection matrix, the general ℓ_2 -norm can be used to shrink the elements in \mathbf{AB} . Then, we get an extended model, named semisupervised low-rank regression (S²LRR), whose objective function is

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}, \mathbf{Y}_u} \quad & \|\mathbf{X}^T \mathbf{AB} - \mathbf{Y}\|_2^2 + \lambda \|\mathbf{AB}\|_2^2 \\ \text{s.t.} \quad & \mathbf{Y}_u \geq \mathbf{0}, \quad \mathbf{Y}_u \mathbf{1}_c = \mathbf{1}_u. \end{aligned} \quad (15)$$

The only difference between objective functions (15) and (3) is whether the intermediate variable \mathbf{D} is involved. In other words, we can treat \mathbf{D} as an identity matrix in S²LRR. Then, the updating rules to \mathbf{A} and \mathbf{B} are

$$\mathbf{A}^* = \arg \max_{\mathbf{A}} \left\{ \text{Tr} \left((\mathbf{A}^T (\mathbf{S}_t + \lambda \mathbf{I}) \mathbf{A})^{-1} \mathbf{A}^T \mathbf{S}_b \mathbf{A} \right) \right\} \quad (16)$$

and

$$\mathbf{B} = (\mathbf{A}^T (\mathbf{X}\mathbf{X}^T + \lambda \mathbf{I}) \mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}\mathbf{Y}. \quad (17)$$

Algorithm 2 Optimization Procedure to S²LRR Objective

Input: data matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$, label matrix $\mathbf{Y}_l \in \mathbb{R}^{l \times c}$, low-rank parameter s and regularization parameter λ ;

Output: projection matrices $\mathbf{A} \in \mathbb{R}^{d \times s}$ and $\mathbf{B} \in \mathbb{R}^{s \times c}$, and label matrix $\mathbf{Y}_u \in \mathbb{R}^{u \times c}$.

- 1: Initialize $t = 0$, $\mathbf{Y}_u^{(t)} = \frac{\mathbf{1}_u \mathbf{1}_c^T}{u} \in \mathbb{R}^{u \times c}$;
 - 2: **while** not converged **do**
 - 3: Calculate $\mathbf{A}^{(t+1)}$ by equation (16);
 - 4: Calculate $\mathbf{B}^{(t+1)}$ by equation (17);
 - 5: Calculate $\mathbf{Y}_u^{(t+1)}$ by solving (7) with Algorithm 3;
 - 6: Update \mathbf{S}_b based on $\mathbf{Y}^{(t+1)} = [\mathbf{Y}_l; \mathbf{Y}_u^{(t+1)}]$;
 - 7: $t = t + 1$;
 - 8: **end while**
-

Here, we directly provide its optimization procedure in Algorithm 2 instead of repeating the detailed derivations step by step.

IV. EXPERIMENTS

In this section, we try to answer the following questions by experiments: 1) whether the joint learning mechanism employed by S³LRR is better than directly bridging EEG data with emotional label matrix by a single projection? 2) how the learned combined projection matrix \mathbf{AB} explores the activation EEG patterns related to the occurrence of affective effect? and 3) whether S³LRR is competent for selecting discriminative EEG features?

A. Dataset and Experimental Setup

In the following experiments, we used the publicly available emotional dataset SEED_V <https://bcmi.sjtu.edu.cn/seed/seed-v.html> [37]. In SEED_V, five different emotional states of *happy*, *sad*, *disgust*, *neutral*, and *fear* were evoked by the corresponding movie clips; 20 subjects were recruited to participate in the EEG data collection experiments and EEG data of 16 subjects were made public. Each subject was asked to participate in the experiments three times. In each experiment, the subjects watched 15 video clips in which three clips correspond to one emotional state. During watching the video clips, EEG data of subjects were recorded by a 62-channel ESI NeuroScan system. After downsampling the raw EEG data to 200 Hz, the DE features were extracted from the five frequency bands, *delta* (1–4 Hz), *theta* (4–8 Hz), *Alpha* (8–14 Hz), *Beta* (14–31 Hz), and *Gamma* (31–50 Hz) bands. The definition of DE is

$$h(X) = - \int_x f(x) \ln f(x) dx \quad (18)$$

where X is a random variable with probability density function $f(x)$ [38]. By assuming that the EEG data follow the Gaussian distribution, i.e., $f(x) = \mathcal{N}(x; \mu, \sigma^2)$, we calculate its DE by:

$$\begin{aligned} h(X) &= - \int_{-\infty}^{\infty} f(x) \ln \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2} dx \\ &= \frac{1}{2} \ln(2\pi\sigma^2) + \frac{\text{Var}(X)}{2\sigma^2} = \frac{1}{2} \ln(2\pi e\sigma^2). \end{aligned} \quad (19)$$

TABLE I
CROSS-SESSION EMOTION RECOGNITION ACCURACIES (%) OF COMPARED MODELS

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s16	Avg.
session1→session2																	
sSVM	81.70	39.74	45.66	83.18	61.18	69.13	66.73	69.13	69.32	61.37	53.23	77.08	86.32	80.78	58.60	58.41	66.35
SDA	50.65	64.14	63.77	76.89	64.51	62.85	79.11	65.62	79.85	59.70	62.11	61.55	88.72	62.11	80.78	78.00	68.77
RLSR2	74.49	59.33	57.49	78.00	72.27	61.55	65.99	65.62	73.57	63.40	65.25	74.86	89.28	88.17	59.70	77.26	70.39
RLSR	83.36	59.52	68.95	80.04	76.71	68.39	68.76	70.43	72.46	56.19	66.54	75.97	83.18	73.94	61.18	79.11	71.55
S ² LRR	87.80	62.66	70.24	81.70	76.16	72.09	81.70	72.27	75.05	66.36	59.52	77.82	85.21	84.47	77.45	79.85	75.65
S ³ LRR	87.80	75.60	68.95	85.77	76.16	72.64	81.89	78.19	81.15	67.10	66.91	82.99	90.76	78.56	79.30	82.07	78.49
session1→session3																	
sSVM	52.75	39.10	53.74	74.88	68.89	54.24	79.03	56.24	84.86	67.55	80.87	58.90	69.72	52.08	47.75	35.27	60.99
SDA	72.71	33.28	64.24	80.87	71.88	30.28	77.37	68.55	86.36	59.90	62.36	77.37	66.72	58.40	56.74	46.09	63.95
RLSR2	65.89	66.06	53.91	87.52	67.89	48.75	88.85	73.88	82.53	46.59	79.87	88.02	87.69	59.73	42.10	59.07	68.65
RLSR	67.89	63.89	65.06	89.52	74.54	46.76	91.35	68.55	91.01	48.92	86.19	78.70	67.89	62.56	53.24	62.90	69.94
S ² LRR	66.72	65.72	71.55	87.85	74.54	51.58	83.03	75.54	80.20	49.08	83.19	85.86	82.20	58.07	56.91	67.05	71.19
S ³ LRR	73.54	67.22	75.71	90.02	76.04	57.40	91.35	76.54	92.18	50.58	93.84	89.02	92.85	64.56	60.23	61.23	75.77
session2→session3																	
sSVM	89.68	79.20	61.56	74.88	75.21	54.58	69.55	62.06	84.86	66.72	80.87	83.36	83.86	40.93	47.25	57.74	69.52
SDA	83.69	83.03	73.21	79.03	47.09	70.72	87.02	93.18	91.01	53.58	68.89	81.70	70.22	69.05	57.40	63.89	73.80
RLSR2	96.51	86.86	69.55	69.88	57.57	68.05	85.36	87.02	87.02	34.78	72.88	82.53	77.20	65.72	61.90	72.38	73.45
RLSR	89.85	89.35	69.88	76.21	62.73	66.39	88.35	81.20	81.36	47.75	69.55	83.36	81.36	57.40	65.39	70.55	73.79
S ² LRR	92.68	91.18	73.88	79.03	55.24	71.05	82.53	90.02	88.52	41.60	81.03	91.51	79.53	50.08	61.06	76.04	75.31
S ³ LRR	96.51	94.68	75.04	80.87	64.73	72.55	93.51	92.85	90.52	47.75	77.54	91.51	86.36	58.07	65.72	76.04	79.02

Note: s1, s2, ..., s16 are the indices of the 16 subjects in SEED_V.

By concatenating the 62 points of each of the five frequency bands together, the dimensionality of each sample vector is 310. Since the video clips in each session are slightly different in length, we have 681, 541, and 601 samples in the three sessions.

Since S³LRR is a semisupervised model, we compare it with semisupervised SVM (sSVM) with linear kernel and some related models including: 1) a two-stage strategy of performing semisupervised discriminant analysis (SDA) first and then sSVM [39]; 2) the rescaled LSR (RLSR) model [40]; and 3) the RLSR with no explicit feature weighting (RLSR2), which actually imposes the ℓ_2 -norm on the projection matrix. Also, S²LRR is also included in the comparison. The regularization parameters involved in respective models were searched from $\{2^{-10}, 2^{-9}, \dots, 2^{10}\}$. The rank parameter s in both S³LRR and S²LRR is always fixed as $c - 1$, which is 4, since $c = 5$ for the SEED_V dataset. We performed cross-session emotion recognition experiments in chronological order, and therefore, we have three recognition tasks for each subject, i.e., “session 1 → session 2,” “session 1 → session 3,” and “session 2 → session 3.” In the case of “session 1 → session 2,” samples from session 1 are fully labeled, while samples from session 2 are unlabeled. Our task is to estimate the labels of these unlabeled samples as accurate as possible.

B. Recognition Results and Analysis

The cross-session emotion recognition results are provided in Table I, where the best accuracy in each case is highlighted in bold. From these results, besides the obvious conclusion that our proposed S³LRR model obtained the best average performance in comparison with the other models, we have the following observations.

- 1) By pairwise comparing the results obtained by RLSR and S³LRR, we find that S³LRR made considerable improvements of 6.94%, 5.83%, and 5.22% in the three cross-session emotion recognition tasks. Therefore, we conclude that our joint learning mechanism is better than directly bridging EEG data with label indicator matrix. The EEG data representation in subspace representation is of higher separability than its original representation.
- 2) SDA in our experiments performed first the SDA and then the classification by sSVM. Such a two-stage paradigm breaks the inner connections of these two operations and prevents them from well matching each other. In both S²LRR and S³LRR, the label estimation of unlabeled samples is jointly completed with the optimization of the other model variables, i.e., the two factor matrices. In particular, the underlying connection between \mathbf{Y}_u and subspace projection matrix \mathbf{A} is explicitly considered. Therefore, both S²LRR and S³LRR obtained the superior performance to SDA.
- 3) Based on our experimental results, RLSR is better than RLSR2 and S³LRR makes improvements in comparison with S²LRR in terms of the average performance. This shows that the explicit feature weighting (selection) is beneficial for improving the emotion recognition performance. Since EEG data are typically multirhythm and multichannel and each frequency-domain feature dimension can be backtracked to a certain EEG frequency band and channel, these numerical accuracies depict that there might be only partial EEG frequency bands and channels contribute significantly to emotion expression at the macro level. In Section IV-C, we will provide



Fig. 2. Emotion recognition results (%) of compared models represented by confusion matrices. (a) sSVM. (b) SDA. (c) RLSR2. (d) RLSR. (e) S²LRR. (f) S³LRR.

the detailed analysis on the affective activation patterns explored by S³LRR.

In Fig. 2, we organize the recognition results in the form of confusion matrices, from which we can gain more insights into the EEG-based emotion recognition. From each model, we know the average recognition accuracy on each emotional state and the misclassification rates of each emotional state into the others. Besides, by comparing the confusion matrix of S³LRR with those of the other models, we can clearly see the performance improvement brought by S³LRR. For example, S³LRR obtained the highest recognition accuracy (83.96%) on the *fear* state and the lowest accuracy (63.95%) on the *sad* state. For the *fear* state, S³LRR improves the accuracy by 5.63% in comparison with S²LRR. On average, 83.96% EEG samples belonging to the *fear* state were correctly recognized by S³LRR, while 3.86%, 4.84%, 2.38%, and 4.96% of them were misclassified as *sad*, *neutral*, *happy*, and *disgust*, respectively.

C. Affective Activation Patterns Exploration

As stated in Section III-C, once the S³LRR model is fitted by given EEG data, we can obtain the quantitative measure of feature importance values by (12). As shown in Fig. 3(a), we plot the ranked feature importance values by averaging all the 48 cross-session emotion recognition cases, from which we observe that different feature dimensions contribute differently in emotion recognition.

Considering that different feature dimensions are extracted from different EEG frequency bands and channels, we perform further investigation on which EEG frequency bands

and channels are more important from the perspective of contributing to accurate emotion recognition. Based on (13), we divide these features into five groups corresponding to the five frequency bands. From Fig. 3(b) and (c), we find that the *Gamma* band contributes the most in emotion recognition. This result is consistent with the one obtained by existing studies, which used the trial-and-error manner [18], [41], that is, they tried each of the five EEG frequency bands and then found that the features from the *Gamma* band lead to the highest recognition accuracy.

Similarly, the general consensus is that different brain regions correlate differently to the emotional expression. According to (14), each EEG channel is quantitatively assigned a value to characterize its importance. In Fig. 4(a), we list the top ten channels that are considered as the most important ones in differentiating the emotional states. By projecting the importance values of these 62 channels onto the brain topology, the critical brain regions correlated more to emotion recognition are adaptively obtained, as shown in Fig. 4(b). We generally conclude that the frontal and left/right temporal lobes might be correlated more to the emotion expression. It is worth mentioning that the results in Fig. 4 correspond to the average effect in terms of all the five EEG frequency bands. In Fig. 5, we provide the topographical view of critical brain regions in emotion recognition corresponding to different EEG frequency bands. Since the *Gamma* and the *Delta* bands, respectively, take the primary and secondary places in emotion recognition, the brain topologies of these two frequency bands are closer to the average result.

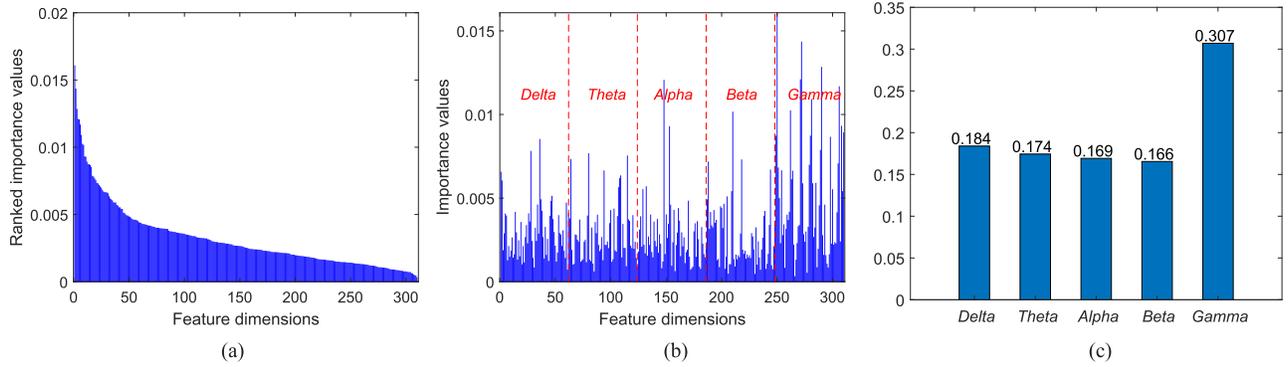


Fig. 3. Identification of critical EEG frequency bands in emotion recognition. (a) Ranked feature importance values. (b) Average feature weights. (c) Importance of frequency bands.

TABLE II
EMOTION RECOGNITION RESULTS (%) BY DIFFERENT FEATURE SELECTION MODELS

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s16	Avg.
session1→session2																	
mRMR	75.23	57.67	53.79	88.72	72.64	51.13	65.06	74.86	54.53	57.12	60.26	80.41	86.32	70.43	56.75	41.22	65.38
L21	63.59	64.33	56.01	78.74	61.37	55.82	66.91	72.83	64.33	64.88	53.79	79.85	87.62	77.82	56.38	38.26	65.16
RRPC	75.97	66.95	61.92	74.68	65.80	62.48	82.81	65.99	61.55	52.50	50.65	75.05	81.14	70.24	68.21	48.24	66.51
RLSR	65.62	69.69	64.51	80.22	52.13	81.33	78.93	66.26	75.42	53.42	57.67	78.00	87.62	74.31	58.96	47.69	68.24
S ³ LRR	85.40	65.80	65.43	88.72	80.22	71.90	83.73	68.58	71.16	54.90	68.02	77.08	87.62	79.30	66.54	64.51	73.68
session1→session3																	
mRMR	53.74	58.24	64.39	76.37	59.57	35.77	69.72	64.73	82.70	52.58	76.87	67.72	54.08	39.43	51.58	45.92	59.59
L21	60.73	56.57	55.74	69.55	59.57	30.12	71.21	58.07	83.36	58.24	83.53	76.04	59.23	37.77	42.93	38.10	58.80
RRPC	66.89	49.75	53.08	82.20	65.56	53.41	78.87	70.55	59.57	58.07	64.39	64.39	68.55	61.40	44.93	37.94	61.22
RLSR	61.06	30.12	54.91	93.18	48.75	69.55	72.55	73.71	84.19	55.57	68.72	68.89	68.55	66.39	48.59	51.08	63.49
S ³ LRR	77.04	69.22	67.22	80.70	68.05	47.09	90.68	70.55	85.86	59.90	77.54	84.53	67.89	66.39	51.25	54.74	69.92
session2→session3																	
mRMR	74.21	75.71	57.24	64.23	61.56	63.23	97.84	76.54	72.21	64.23	62.90	73.71	80.20	35.94	63.73	47.25	66.92
L21	84.69	79.20	63.73	76.71	34.61	47.59	80.03	70.05	74.88	44.09	64.23	79.70	80.03	51.08	47.42	53.41	64.47
RRPC	75.37	74.88	69.38	71.05	56.57	66.39	76.54	76.04	77.54	51.08	61.40	81.86	66.89	50.58	50.42	70.38	67.27
RLSR	77.70	73.21	71.55	82.53	45.76	65.39	84.53	72.55	76.71	61.40	64.39	70.72	72.38	49.92	51.75	71.71	68.26
S ³ LRR	91.02	77.54	75.04	84.86	65.22	70.88	93.01	81.53	86.86	60.23	64.89	83.36	84.69	54.91	70.72	81.86	76.66

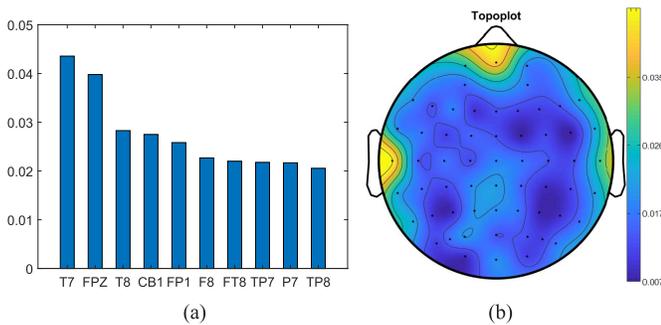


Fig. 4. Identification of critical EEG channels and brain regions in emotion recognition. (a) Top ten channels. (b) Critical brain regions.

D. Feature Selection by S³LRR

As discussed in Section III-C, the importance of the i th feature dimension can be quantitatively measured by $\theta_i|_{i=1}^d$, which is deservedly appropriate for determining discriminative features. In the following, we evaluate the effectiveness of S³LRR in EEG feature selection by comparing it with some

widely used methods, including the minimal-redundancy-maximal-relevance criterion (mRMR) [42], the $\ell_{2,1}$ -norm [34], the max-relevance and min-redundancy criterion based on Pearson's correlation (RRPC) coefficient [43], and RLSR [40], [44]. The former two methods are supervised feature selection methods and the latter two methods are semisupervised methods. The involved parameters in respective models were set as suggested by their original papers. SVM with linear kernel was used to classify the newly formed EEG data by selected features, whose regularization parameter was searched from $\{2^{-10}, 2^{-9}, \dots, 2^{10}\}$. For each model, we set the number of selected features as 10, 20, 50, 100, and 200, respectively; then, the best result as well as the corresponding numbers of selected features are reported.

In Table II, we show the emotion recognition results obtained by the compared feature selection models, where the best accuracy in each recognition case is highlighted in bold. Accordingly, we use triples to represent the numbers of selected features when these models achieved the best performance in Table III. For example, the first

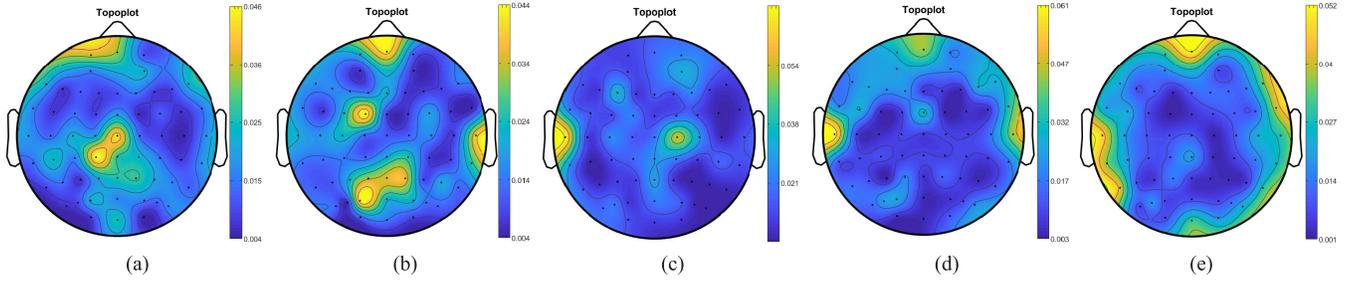


Fig. 5. Topographical view of critical brain regions corresponding to different EEG frequency bands in emotion recognition. (a) *Delta*. (b) *Theta*. (c) *Alpha*. (d) *Beta*. (e) *Gamma*.

triple (10, 20, and 200) means that the numbers of selected features are, respectively, 10, 20, and 200 when mRMR achieved the best accuracies in the three cross-session emotion recognition tasks. On the whole, S³LRR obtained the best performance among these five feature selection methods. Besides, we find that the average performance of semisupervised methods is better than that of the supervised ones because involving unlabeled samples into the learning process can make them better capture the data properties. Furthermore, the best results in some cases were obtained when the number of selected features is much less than 200. For example, in the case of “subject 1: session 1→session 3,” the best accuracy of S³LRR, 77.04%, is obtained when the number of selected features is 20. This further explains that different EEG feature dimensions contribute differently to emotion recognition. Accordingly, different EEG frequency bands and channels correlate differently to the occurrence and change of the affective effect. From the perspective of pattern recognition, features can be divided into three groups, i.e., discriminative, redundant, and noisy features, according to their different recognition abilities. Discriminative features are beneficial to correctly recognizing the emotional states, while the noisy features are harmful to improving the recognition performance. Redundant features are in between, which meaninglessly increases the length of sample dimensionality. For EEG-based emotion recognition, feature selection models are expected to preserve discriminative features, suppress redundant features, and remove noisy features.

E. Algorithm Properties

In the following, we analyze the properties of S³LRR from the two perspectives of parameter sensitivity and convergence.

In S³LRR, the regularization parameter λ controls the row sparsity of the combined matrix \mathbf{AB} . The larger λ , the sparser the rows of \mathbf{AB} . In Fig. 6, we show how the emotion recognition accuracies of S³LRR change in terms of different λ s in the three cases on subject 1. From Fig. 6, we generally conclude that S³LRR is not very sensitive to λ and it achieves satisfactory accuracies with many candidate λ s. Similar results can be found on the remaining subjects.

On the convergence of S³LRR, its model objective function is no longer convex since we have the multiplication form of two factor matrices, \mathbf{A} and \mathbf{B} . Therefore, we introduce an auxiliary matrix \mathbf{D} to facilitate the optimization. Since

TABLE III
NUMBERS OF SELECTED FEATURES CORRESPONDING TO THE BEST ACCURACIES OBTAINED BY THE FEATURE SELECTION MODELS

	mRMR	L21	RRPC	RLSR	S ³ LRR
s1	(10,20,200)	(200,10,100)	(200,200,200)	(10,50,200)	(200,20,10)
s2	(20,50,200)	(100,100,200)	(100,100,200)	(50,200,200)	(50,100,100)
s3	(100,20,20)	(10,10,10)	(200,200,10)	(10,10,50)	(10,10,100)
s4	(200,200,200)	(100,100,20)	(200,200,100)	(200,200,10)	(100,50,100)
s5	(200,20,20)	(100,50,100)	(50,100,20)	(200,50,200)	(100,100,100)
s6	(200,20,10)	(10,100,20)	(200,200,50)	(20,100,20)	(200,200,10)
s7	(100,100,100)	(100,50,200)	(200,200,200)	(200,20,20)	(100,100,10)
s8	(100,200,100)	(20,200,50)	(200,200,200)	(100,100,10)	(200,10,20)
s9	(100,200,200)	(200,50,200)	(200,200,200)	(100,10,200)	(200,200,100)
s10	(200,200,50)	(100,10,20)	(200,200,100)	(20,10,20)	(100,50,10)
s11	(100,50,200)	(10,100,200)	(20,200,200)	(200,10,200)	(10,50,200)
s12	(200,200,200)	(50,50,100)	(200,100,200)	(20,20,20)	(100,200,100)
s13	(200,200,200)	(50,50,50)	(200,200,200)	(50,200,200)	(200,200,100)
s14	(200,200,20)	(200,100,20)	(100,200,200)	(200,10,10)	(200,10,50)
s15	(200,200,200)	(100,10,20)	(200,20,50)	(200,50,100)	(200,10,20)
s16	(20,10,50)	(200,200,20)	(50,100,200)	(200,50,100)	(50,200,200)

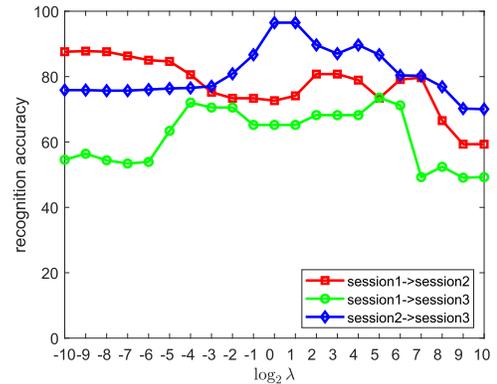


Fig. 6. Recognition performance of S³LRR in terms of different λ s on subject 1.

the auxiliary matrix \mathbf{D} also involves the variables \mathbf{A} and \mathbf{B} , we have to iteratively update the variables \mathbf{A} , \mathbf{B} , and \mathbf{D} . However, we declare that the optimization procedure described in Algorithm 1 has good convergence property. In Appendix B, we theoretically prove that the S³LRR objective function values monotonically decrease according to our proposed optimization method in Algorithm 1. Apart from the theoretical proof, in Fig. 7, we experimentally show the convergence curves of S³LRR on the three emotion recognition cases

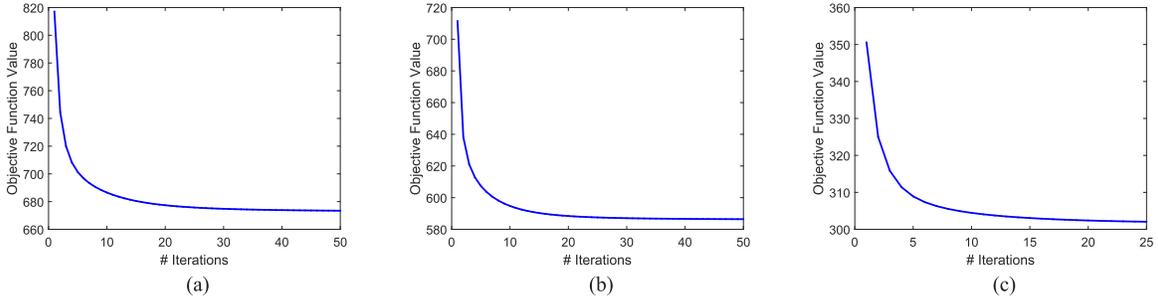


Fig. 7. Convergence property of S^3LRR in the three cases of subject 1: (a) session1 \rightarrow session2, (b) session1 \rightarrow session3, and (c) session2 \rightarrow session3.

on subject 1. From Fig. 7, we find that S^3LRR has a fast converge rate, usually within 30 iterations.

V. CONCLUSION

In this article, we proposed a unified model termed S^3LRR to implement joint discriminative subspace identification and semisupervised EEG emotion recognition. The merits of S^3LRR are: 1) effectively avoiding the limitation of breaking feature extraction/transformation and emotion recognition into two isolated stages; 2) jointly optimizing the soft label matrix of unlabeled samples and the subspace projection matrix; and 3) providing a quantitative way to explore the affective activation patterns on critical EEG frequency bands and brain regions in emotion expression. Experimental results demonstrated that S^3LRR exhibits excellent performance in improving the emotion recognition accuracy and selecting discriminative EEG features. On average, we conclude that the *Gamma* frequency band and the brain regions of frontal and left/right temporal lobes are more correlated with the occurrence of affective effect. In the future, we will consider further enhancing the emotion recognition performance from two aspects, i.e., incorporating multiple EEG features to better capture the EEG data properties and improving the S^3LRR to make it have the nonlinear learning ability.

APPENDIX A OPTIMIZATION TO OBJECTIVE (7)

To simplify the following derivations, we use \mathbf{m}_i and \mathbf{y}_i to, respectively, denote the transpose of \mathbf{m}^i and \mathbf{y}^i . Then, the Lagrangian function of problem (7) is

$$\mathcal{L}(\mathbf{y}_i, \eta, \boldsymbol{\beta}) = \|\mathbf{y}_i - \mathbf{m}_i\|_2^2 - \eta(\mathbf{y}_i^T \mathbf{1}_c - 1) - \boldsymbol{\beta}^T \mathbf{y}_i \quad (20)$$

where η and $\boldsymbol{\beta} \in \mathbb{R}^c$ are Lagrangian multipliers in scalar and vector forms, respectively. In the following, we provide an analysis that both the Lagrangian multipliers can be determined. Suppose that the optimal solution to the proximal problem (7) is \mathbf{y}_i^* , and the associated Lagrangian multipliers are η^* and $\boldsymbol{\beta}^*$. Then, according to the Karush-Kuhn-Tucker (KKT) condition, we have the following equations and inequalities:

$$\begin{cases} \forall j, & y_{ij}^* - m_{ij} - \eta^* - \beta_j^* = 0 & (21) \\ \forall j, & y_{ij}^* \geq 0 & (22) \\ \forall j, & \beta_j^* \geq 0 & (23) \\ \forall j, & y_{ij}^* \beta_j^* = 0 & (24) \end{cases}$$

where y_{ij}^* is the j th scalar element of vector \mathbf{y}_i^* . Equation (21) can be rewritten in the vector form as

$$\mathbf{y}_i^* - \mathbf{m}_i - \eta^* \mathbf{1}_c - \boldsymbol{\beta}^* = \mathbf{0}. \quad (25)$$

Considering the constraint $\mathbf{y}_i^{*T} \mathbf{1} = 1$, the above equation can be reformulated into

$$\eta^* = \frac{1 - \mathbf{1}_c^T \mathbf{m}_i - \mathbf{1}^T \boldsymbol{\beta}^*}{c}. \quad (26)$$

By substituting (26) into (25), we have

$$\mathbf{y}_i^* = \mathbf{m}_i - \frac{\mathbf{1}\mathbf{1}^T}{c} \mathbf{m}_i + \frac{1}{c} \mathbf{1} - \frac{\mathbf{1}^T \boldsymbol{\beta}^*}{c} \mathbf{1} + \boldsymbol{\beta}^*. \quad (27)$$

Denote $\bar{\boldsymbol{\beta}}^* = (\mathbf{1}^T \boldsymbol{\beta}^* / c)$ and $\mathbf{q} = \mathbf{m}_i - (\mathbf{1}\mathbf{1}^T / c) \mathbf{m}_i + (1/c) \mathbf{1}$, and the above equation can be rewritten as

$$\mathbf{y}_i^* = \mathbf{q} + \boldsymbol{\beta}^* - \bar{\boldsymbol{\beta}}^* \mathbf{1}. \quad (28)$$

Therefore, for each $j = 1, \dots, c$, we have

$$y_{ij}^* = q_j + \beta_j^* - \bar{\beta}^*. \quad (29)$$

According to (22)–(24) and (29), we know $q_j + \beta_j^* - \bar{\beta}^* = (q_j - \bar{\beta}^*)_+$, where $(f(\cdot))_+ = \max(f(\cdot), 0)$. Therefore, we have

$$y_{ij}^* = (q_j - \bar{\beta}^*)_+. \quad (30)$$

Now, if the optimal $\bar{\boldsymbol{\beta}}^*$ can be determined, the optimal solution \mathbf{y}_i^* can be obtained from (30). Equation (29) can be rewritten as $\beta_j^* = y_{ij}^* + \bar{\beta}^* - q_j$ such that $\beta_j^* = (\bar{\beta}^* - q_j)_+$. Therefore, $\bar{\boldsymbol{\beta}}^*$ can be calculated as

$$\bar{\boldsymbol{\beta}}^* = \frac{1}{c} \sum_{j=1}^c (\bar{\beta}^* - q_j)_+. \quad (31)$$

According to the constraint $\mathbf{y}_i^{*T} \mathbf{1} = 1$ and (30), we define the following function:

$$f(\bar{\boldsymbol{\beta}}) = \sum_{j=1}^c (q_j - \bar{\beta})_+ - 1 \quad (32)$$

and the optimal $\bar{\boldsymbol{\beta}}^*$ should satisfy $f(\bar{\boldsymbol{\beta}}^*) = 0$. When (32) equals zero, the optimal $\bar{\boldsymbol{\beta}}^*$ can be obtained via Newton method, namely,

$$\bar{\boldsymbol{\beta}}^{(k+1)} = \bar{\boldsymbol{\beta}}^{(k)} - \frac{f(\bar{\boldsymbol{\beta}}^{(k)})}{f'(\bar{\boldsymbol{\beta}}^{(k)})}. \quad (33)$$

We know that $f(\bar{\boldsymbol{\beta}})$ is a piecewise linear and monotonically increasing function. When $q_j \geq \bar{\beta}$, $f(\bar{\boldsymbol{\beta}}) = \sum_{j=1}^c q_j - \bar{\beta} - 1$,

Algorithm 3 Algorithm to Solve Objective Function (7)

Input: vector $\mathbf{m}_i \in \mathbb{R}^c$;

Output: vector $\mathbf{y}_i \in \mathbb{R}^c$.

- 1: Compute $\mathbf{q} = \mathbf{m}_i - \frac{\mathbf{1}\mathbf{1}^T}{c}\mathbf{m}_i + \frac{1}{c}\mathbf{1}$;
 - 2: Use Newton's method to obtain the root $\bar{\beta}^*$ of (32);
 - 3: Obtain the optimal solution $y_{ij}^* = (q_j - \bar{\beta}^*)_+$ for $j = 1, \dots, c$;
-

and we have $f'(\bar{\beta}) = -1$. When $q_j \leq \bar{\beta}$, $f(\bar{\beta}) = -1$ and its derivative $f'(\bar{\beta}) = 0$. Therefore, we can obtain $f'(\bar{\beta})$ by counting the number of positive values in $(q_j - \bar{\beta})|_{j=1}^c$. Consequently, the optimization procedure to problem (7) is provided in Algorithm 3.

APPENDIX B

PROOF TO THE CONVERGENCE OF ALGORITHM 1

Proof: Since the calculation of \mathbf{A} and \mathbf{B} is coupled, we first prove that the updating of these two variables can guarantee the convergence. In the t th iteration, we have

$$\begin{aligned} & \{\mathbf{A}^{(t+1)}, \mathbf{B}^{(t+1)}, \mathbf{Y}^{(t+1)}\} \\ &= \arg \min_{\mathbf{A}, \mathbf{B}, \mathbf{Y}_u} \|\mathbf{Y} - \mathbf{X}^T \mathbf{A} \mathbf{B}\|_2^2 + \frac{\lambda}{2} \text{Tr}(\mathbf{B}^T \mathbf{A}^T \mathbf{D}^{(t)} \mathbf{A} \mathbf{B}). \end{aligned} \quad (34)$$

That is,

$$\begin{aligned} & \|\mathbf{Y}^{(t+1)} - \mathbf{X}^T \mathbf{A}^{(t+1)} \mathbf{B}^{(t+1)}\|_2^2 \\ &+ \frac{\lambda}{2} \text{Tr}(\mathbf{B}^{(t+1)T} \mathbf{A}^{(t+1)T} \mathbf{D}^{(t)} \mathbf{A}^{(t+1)} \mathbf{B}^{(t+1)}) \\ &\leq \|\mathbf{Y}^{(t)} - \mathbf{X}^T \mathbf{A}^{(t)} \mathbf{B}^{(t)}\|_2^2 + \frac{\lambda}{2} \text{Tr}(\mathbf{B}^{(t)T} \mathbf{A}^{(t)T} \mathbf{D}^{(t)} \mathbf{A}^{(t)} \mathbf{B}^{(t)}). \end{aligned}$$

Denote $\mathbf{G}^{(t)} = \mathbf{A}^{(t)} \mathbf{B}^{(t)}$ and $\mathbf{G}^{(t+1)} = \mathbf{A}^{(t+1)} \mathbf{B}^{(t+1)}$. According to the definition of matrix \mathbf{D} , the above equation can be rewritten as

$$\begin{aligned} & \|\mathbf{Y}^{(t+1)} - \mathbf{X}^T \mathbf{G}^{(t+1)}\|_2^2 + \lambda \sum_{i=1}^d \frac{\|\mathbf{g}^{i(t+1)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2} \\ &\leq \|\mathbf{Y}^{(t)} - \mathbf{X}^T \mathbf{G}^{(t)}\|_2^2 + \lambda \sum_{i=1}^d \frac{\|\mathbf{g}^{i(t)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2} \end{aligned} \quad (35)$$

where $\mathbf{g}^{i(t)}$ and $\mathbf{g}^{i(t+1)}$ are the i th row of matrix $\mathbf{G}^{(t)}$ and $\mathbf{G}^{(t+1)}$, respectively.

For any two nonnegative values a and b , there is

$$a - \frac{a^2}{2b} \leq b - \frac{b^2}{2b}. \quad (36)$$

By denoting $a = \|\mathbf{g}^{i(t+1)}\|_2$ and $b = \|\mathbf{g}^{i(t)}\|_2$, we have

$$\|\mathbf{g}^{i(t+1)}\|_2 - \frac{\|\mathbf{g}^{i(t+1)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2} \leq \|\mathbf{g}^{i(t)}\|_2 - \frac{\|\mathbf{g}^{i(t)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2}.$$

Therefore, summing up the above d inequalities and multiplying the summation with the regularization parameter λ ,

we obtain

$$\begin{aligned} & \lambda \sum_{i=1}^d \left(\|\mathbf{g}^{i(t+1)}\|_2 - \frac{\|\mathbf{g}^{i(t+1)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2} \right) \\ &\leq \lambda \sum_{i=1}^d \left(\|\mathbf{g}^{i(t)}\|_2 - \frac{\|\mathbf{g}^{i(t)}\|_2^2}{2\|\mathbf{g}^{i(t)}\|_2} \right). \end{aligned} \quad (37)$$

Combining (35) and (37), we get

$$\begin{aligned} & \|\mathbf{Y}^{(t+1)} - \mathbf{X}^T \mathbf{G}^{(t+1)}\|_2^2 + \lambda \sum_{i=1}^d \|\mathbf{g}^{i(t+1)}\|_2 \\ &\leq \|\mathbf{Y}^{(t+1)} - \mathbf{X}^T \mathbf{G}^{(t)}\|_2^2 + \lambda \sum_{i=1}^d \|\mathbf{g}^{i(t)}\|_2. \end{aligned} \quad (38)$$

Therefore, we have

$$\begin{aligned} & \|\mathbf{Y}^{(t+1)} - \mathbf{X}^T \mathbf{G}^{(t+1)}\|_2^2 + \lambda \|\mathbf{G}^{(t+1)}\|_{2,1} \\ &\leq \|\mathbf{Y}^{(t)} - \mathbf{X}^T \mathbf{G}^{(t)}\|_2^2 + \lambda \|\mathbf{G}^{(t)}\|_{2,1}. \end{aligned} \quad (39)$$

Specifically, variables \mathbf{A} and \mathbf{B} are updated according to gradient. Variable \mathbf{Y}_u is updated according to the Lagrangian multiplier method in which the multipliers can be uniquely determined. We conclude that Algorithm 1 monotonically decreases the objective function (4) in each iteration. \square

REFERENCES

- [1] B. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018.
- [2] M. B. Akçay and K. Oğuz, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers," *Speech Commun.*, vol. 116, pp. 56–76, Jan. 2020.
- [3] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowl. Inf. Syst.*, vol. 62, no. 8, pp. 2937–2987, Aug. 2020.
- [4] S. K. Khare, V. Bajaj, and G. R. Sinha, "Adaptive tunable Q wavelet transform-based emotion identification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9609–9617, Dec. 2020.
- [5] H. Wang *et al.*, "Performance enhancement of P300 detection by multiscale-CNN," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 2507012.
- [6] X. Chen *et al.*, "ReMAE: User-friendly toolbox for removing muscle artifacts from EEG," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2105–2119, May 2020.
- [7] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.
- [8] Z. Pei, H. Wang, A. Bezerianos, and J. Li, "EEG-based multiclass workload identification using feature fusion and selection," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 4001108.
- [9] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 6627–6630.
- [10] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cogn. Developmental Syst.*, vol. 14, no. 1, pp. 4–19, Mar. 2020.
- [11] S. Liu *et al.*, "Study on an effective cross-stimulus emotion recognition model using EEGs based on feature selection and support vector machine," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 5, pp. 721–726, May 2018.
- [12] W. Kong, X. Song, and J. Sun, "Emotion recognition based on sparse representation of phase synchronization features," *Multimedia Tools Appl.*, vol. 80, no. 14, pp. 21203–21217, Jun. 2021.
- [13] S. Gong, K. Xing, A. Cichocki, and J. Li, "Deep learning in EEG: Advance of the last ten-year critical period," *IEEE Trans. Cognit. Develop. Syst.*, early access, May 13, 2021, doi: 10.1109/TCDS.2021.3079712.

- [14] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Develop. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.
- [15] A. Mert and A. Akan, "Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform," *Digit. Signal Process.*, vol. 81, pp. 106–115, Oct. 2018.
- [16] D.-W. Chen *et al.*, "A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition," *Sensors*, vol. 19, no. 7, p. 1631, Apr. 2019.
- [17] M. Yu *et al.*, "A review of EEG features for emotion recognition," *Scientia Sinica Informationis*, vol. 49, no. 9, pp. 1097–1118, Sep. 2019.
- [18] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul/Sep. 2017.
- [19] X. Sun, B. Hu, X. Zheng, Y. Yin, and C. Ji, "Emotion classification based on brain functional connectivity network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 2082–2089.
- [20] C. Luo *et al.*, "A survey of brain network analysis by electroencephalographic signals," *Cognit. Neurodyn.*, vol. 16, no. 1, pp. 17–41, Feb. 2022.
- [21] D. Wu and J. Huang, "Affect estimation in 3D space using multi-task active learning for regression," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 16–27, Jan. 2022.
- [22] P. Li *et al.*, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.
- [23] Z. Liang, S. Oba, and S. Ishii, "An unsupervised EEG decoding system for human emotion recognition," *Neural Netw.*, vol. 116, pp. 257–268, Aug. 2019.
- [24] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Syst. Appl.*, vol. 47, pp. 35–41, Apr. 2016.
- [25] Y.-H. Liu, W.-T. Cheng, Y.-T. Hsiao, C.-T. Wu, and M.-D. Jeng, "EEG-based emotion recognition based on kernel Fisher's discriminant analysis and spectral powers," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2014, pp. 2221–2225.
- [26] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu, "Transfer components between subjects for EEG-based emotion recognition," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 917–922.
- [27] Y. Peng, Q. Li, W. Kong, F. Qin, J. Zhang, and A. Cichocki, "A joint optimization framework to semi-supervised RVFL and ELM networks for efficient data classification," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106756.
- [28] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
- [29] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul/Sep. 2020.
- [30] G. Zhang, M. Yu, Y.-J. Liu, G. Zhao, D. Zhang, and W. Zheng, "SparseDGCNN: Recognizing emotion from multichannel EEG signals," *IEEE Trans. Affect. Comput.*, early access, Jan. 13, 2021, doi: [10.1109/TAFFC.2021.3051332](https://doi.org/10.1109/TAFFC.2021.3051332).
- [31] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.
- [32] Y. Peng, X. Zhu, F. Nie, W. Kong, and Y. Ge, "Fuzzy graph clustering," *Inf. Sci.*, vol. 571, pp. 38–49, Sep. 2021.
- [33] X. Cai, C. Ding, F. Nie, and H. Huang, "On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 1124–1132.
- [34] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $\ell_{2,1}$ -norms minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1813–1821.
- [35] Y. Peng, F. Qin, W. Kong, Y. Ge, F. Nie, and A. Cichocki, "GFIL: A unified framework for the importance analysis of features, frequency bands and channels in EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, early access, May 21, 2021, doi: [10.1109/TCDS.2021.3082803](https://doi.org/10.1109/TCDS.2021.3082803).
- [36] Y. Peng, Q. Li, W. Kong, J. Zhang, B.-L. Lu, and A. Cichocki, "Joint semi-supervised feature auto-weighting and classification model for EEG-based cross-subject sleep quality evaluation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 946–950.
- [37] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition," *IEEE Trans. Cogn. Developmental Syst.*, early access, Apr. 5, 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9395500>
- [38] J. Wu, *Essentials of Pattern Recognition: An Accessible Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2020.
- [39] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–7.
- [40] X. Chen, G. Yuan, F. Nie, and J. Z. Huang, "Semi-supervised feature selection via rescaled linear regression," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 1525–1531.
- [41] Y. Peng and B.-L. Lu, "Discriminative manifold extreme learning machine and applications to image and EEG signal classification," *Neurocomputing*, vol. 174, pp. 265–277, Jan. 2016.
- [42] S. Ramírez-Gallego *et al.*, "Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data," *Int. J. Intell. Syst.*, vol. 32, no. 2, pp. 134–152, 2017.
- [43] J. Xu, B. Tang, H. He, and H. Man, "Semisupervised feature selection based on relevance and redundancy criteria," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 9, pp. 1974–1984, May 2016.
- [44] X. Chen, G. Yuan, F. Nie, and Z. Ming, "Semi-supervised feature selection via sparse rescaled linear square regression," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 165–176, Jan. 2020.



Yong Peng (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2015.

He is currently a Full Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China. He has authored (coauthored) more than 30 SCI-indexed journal articles, such as *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING (TNSRE)*, *TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS (TCDS)*, *TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT (TIM)*, *Information Sciences*, *Neural Networks*, and *Knowledge-Based Systems*. His main research interests include machine learning, pattern recognition, and electroencephalogram (EEG)-based brain–computer interfaces.

Dr. Peng received the President Prize from the Chinese Academy of Sciences in 2009 and the Third Prize from the Chinese Institute of Electronics in 2018.



Yikai Zhang received the B.S. degree from the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China, in 2021, where he is currently pursuing the M.S. degree.

His research interests include machine learning, pattern recognition, and electroencephalogram (EEG)-based brain–computer interfaces.



Wanzeng Kong (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2008.

He was a Visiting Research Associate with the Department of Biomedical Engineering, University of Minnesota, Twin Cities, Minneapolis, MN, USA, from 2012 to 2013. He is currently a Full Professor with the School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou. His current research interests include biomedical signal processing, brain–computer interface, cognitive computing, and pattern recognition.



Feiping Nie (Senior Member, IEEE) received the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in 2009.

He is currently a Full Professor with Northwestern Polytechnical University, Xi'an, China. He has authored more than 100 articles in top journals and conferences, such as TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), *International Journal of Computer Vision (IJCV)*, TRANSACTIONS ON IMAGE PROCESSING (TIP), TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS (TNNLS)/(TNN), TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), *Bioinformatics*, International Conference on Machine Learning (ICML), Conference on Neural Information Processing Systems (NIPS), ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), International Joint Conference on Artificial Intelligence (IJCAI), Association for the Advancement of Artificial Intelligence (AAAI), International Conference on Computer Vision (ICCV), Conference on Computer Vision and Pattern Recognition (CVPR), and ACM Multimedia (MM). His articles have been cited more than 10000 times. His research interests include machine learning and its applications, such as pattern recognition, data mining, computer vision, image processing, and information retrieval.

Dr. Nie is currently an associate editor or a PC member for several prestigious journals and conferences in the related fields.



Bao-Liang Lu (Fellow, IEEE) received the Dr.Eng. degree in electrical engineering from Kyoto University, Kyoto, Japan, in 1994.

From 1994 to 1999, he was a Frontier Researcher with the Bio-Mimetic Control Research Center, Institute of Physical and Chemical Research (RIKEN), Nagoya, Japan, and a Research Scientist with the RIKEN Brain Science Institute, Wako, Japan, from 1999 to 2002. Since 2002, he has been a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include brainlike computing, neural networks, machine learning, brain-computer interaction, and affective computing.

Dr. Lu was a Steering Committee Member of the IEEE TRANSACTIONS ON AFFECTIVE COMPUTING. He received the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT Outstanding Paper Award in 2018. He is currently an Associate Editor of IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS and *Journal of Neural Engineering*.



Andrzej Cichocki (Life Fellow, IEEE) received the M.Sc. (Hons.), Ph.D., and Dr.Sc. (Habilitation) degrees in electrical engineering from the Warsaw University of Technology, Warszawa, Poland, in 1972, 1976, and 1982, respectively.

He spent several years at the University of Erlangen-Nuremberg, Erlangen, Germany, as an Alexander-von-Humboldt Research Fellow and a Guest Professor. He was a Senior Team Leader and the Head of the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako, Japan. He is currently a Professor with the Skolkovo Institute of Science and Technology, Moscow, Russia, and an Adjunct Professor with Hangzhou Dianzi University, Hangzhou, China. He has authored more than 500 technical journal articles and six monographs in English (two of them translated to Chinese). His current research interests include multiway blind source separation, tensor decompositions, tensor networks, deep learning, human-robot interactions, and brain-computer interfaces.

Dr. Cichocki served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and *Journal of Neuroscience Methods*. He is also the Founding Editor-in-Chief of the *Journal Computational Intelligence and Neuroscience*.