Embedded EEG Feature Selection for Multi-Dimension Emotion Recognition via Local and Global Label Relevance

Xueyuan Xu, Fulin Wei, Tianyuan Jia, Li Zhuo, Hui Zhang, Xiaoguang Li, and Xia Wu, *Senior Member, IEEE*

***Abstract*— Due to the problem of a small amount of EEG samples and relatively high dimensionality of electroencephalogram (EEG) features, feature selection plays an essential role in EEG-based emotion recognition. However, current EEG-based emotion recognition studies utilize a problem transformation approach to transform multi-dimension emotional labels into single-dimension labels, and then implement commonly used single-label feature selection methods to search feature subsets, which ignores the relations between different emotional dimensions. To tackle the problem, we propose an efficient EEG feature selection method for multi-dimension emotion recognition (EFSMDER) via local and global label relevance. First, to capture the local label correlations, EFSMDER implements orthogonal regression to map the original EEG feature space into a low-dimension space. Then, it employs the global label correlations in the original multi-dimension emotional label space to effectively construct the label information in the low-dimension space. With the aid of local and global relevance information, EFSMDER can conduct representational EEG feature subset selection. Three EEG emotional databases with multi-dimension emotional labels were used for performance comparison between EFSMDER and fourteen state-of-the-art methods, and the EFSMDER method achieves the best multi-dimension classification accuracies of 86.43, 84.80, and 97.86 percent on the DREAMER, DEAP, and HDED datasets, respectively.**

***Index Terms*— Electroencephalogram, feature selection, multi-dimension emotional labels, global relevance.**

I. Introduction

ELECTROENCEPHALOGRAM (EEG) is a non-trauma and portable measurement of brain activity and could rapidly respond to different emotional states [1], [2]. Recently, due to the high temporal resolution and cost-effectiveness of EEG, EEG-based emotion recognition attracts extensive attention in the field of brain-computer interface. To accurately portray different emotional states, various types of feature extraction approaches have been used to analyze the non-stationary and nonlinear EEG signals, including higher-order crossing (HOC) [3], differential entropy (DE) [4], rational asymmetry (RASM) [5], non-stationary index (NSI) [6], etc.

With the rapid development of EEG signal acquisition equipment, the number of electrodes available for emotion recognition is growing and a mass of EEG features can be extracted from the plentiful electrodes [7], [8]. Nevertheless, the associated EEG features are often high-dimensional and inevitably contain irrelevant, redundant, and noise information, which can easily deteriorate the emotion recognition performance due to the relatively small amount of EEG samples [9], [10]. To select informative features and remove irrelevant features from the high-dimensional EEG data, multiple feature selection approaches have been implemented in the EEG-based emotion recognition task [11], [12], [13].

According to feature subset evaluation and search mechanism, the EEG feature selection methods could be roughly categorized into three groups: filter, wrapper, and embedded approaches [14]. The filter methods evaluate the importance of EEG features in emotion recognition according to the statistical properties of the EEG data. But these methods are independent of the learning algorithm and the feature selection performance is always unsatisfactory [15]. To solve the problem, multiple studies are dedicated to implementing wrapper approaches. The wrapper methods adopt the learning results of a specific classifier as the evaluation index for the EEG feature subset and can frequently obtain higher prediction performance than the filter methods [14]. However, the

Manuscript received 22 August 2023; revised 11 December 2023; accepted 12 January 2024. Date of publication 18 January 2024; date of current version 26 January 2024. This work was supported in part by the Special Fund for Research on National Major Research Instruments of the Nature Science Foundation of China under Grant 62227801, in part by the Beijing Natural Science Foundation under Grant 4212037 and Grant 4244087, in part by the China Postdoctoral Science Foundation under Grant 2022M720332, and in part by the Beijing Postdoctoral Foundation under Grant 2023-zz-85. *(Corresponding author: Xia Wu.)*

Xueyuan Xu, Li Zhuo, Hui Zhang, and Xiaoguang Li are with the Faculty of Information Technology and the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China (e-mail: xxy@bjut.edu.cn; zhuoli@bjut.edu.cn; huizhang@bjut.edu.cn; lxg@bjut.edu.cn).

Fulin Wei, Tianyuan Jia, and Xia Wu are with the School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China, and also with the Guangdong Artificial Intelligence and Digital Economy Laboratory, Guangzhou 511442, China (e-mail: weifulin@mail.bnu.edu.cn; 201921210018@mail.bnu.edu.cn; wuxia@bnu.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3355488

wrapper methods often require multiple trails and have a large computational cost [15]. Nowadays, embedded methods have attracted much attention from researchers as an alternative strategy for addressing the filter issue. Embedded methods, incorporating the search for EEG feature subsets into the optimization problem, can take advantages of both filter and wrapper approaches [16], such as less computational cost and better EEG feature selection capability.

To analyze the multi-dimension emotional labels, the existing EEG-based emotion recognition studies often adopt the strategy of problem transformation. In other words, they first transform multi-dimension emotional labels into single-dimension labels and then implement the existing EEG feature selection models to deal with the transformed single-dimension labels, i.e., valence-arousal labels (0,0),(0,1), (1,0), and (1,1) could be changed into 1, 2, 3, and 4. Nevertheless, one major shortcoming of the strategy is that it ignores the label correlations among the EEG emotional data, including feature-label relevance (local) and label-label relevance (global). The local correlations are essential to accurately construct the statistical model between the EEG features and multi-dimension emotional labels, and the global correlations can be utilized to search discriminative information in the multi-dimension emotional labels.

To address the above issue, we propose a novel embedded EEG feature selection method for multi-dimension emotion recognition (EFSMDER). To capture the local label relevance, EFSMDER implements orthogonal regression to map the original EEG feature space into a low-dimension space. Then, EFSMDER employs the global label relevance in the original multi-dimension emotional label space to effectively construct the label information in the low-dimension space.

Moreover, the main contributions of our work are as follows:

- EEG-based multi-dimension emotion recognition task is implemented through a multi-label learning strategy, which is distinguished from the approach of transforming the multi-dimension emotion recognition task into single-label learning sub-problems adopted by most of the existing studies. Our strategy is conducive to exploiting effectively the label relevance among the multiple dimensions.
- A novel embedded EEG feature selection method is proposed for multi-dimension emotion recognition. Multi-dimensional affective labels are reconstructed with global label correlation maximization restriction, which is instrumental in exploring shared and specific information. With the aid of the emotional label correlations (local and global) and the global feature redundancy, EFSMDER can select an informative and non-redundant EEG feature subset for multi-dimension emotion recognition.
- On the basis of the average feature weights acquired by EFSMDER, the effects of various EEG feature types in multi-dimension emotion recognition were compared. The experimental results show that the EEG features extracted from the time-frequency domain are more effective representations of multi-dimensional emotion than those extracted from the time domain and frequency domain. The discovery may be useful guidance and largely facilitate the relevant research on EEG emotional feature extraction.
- Three EEG emotional databases DREAMER, DEAP, and HDED with various numbers of electrodes (14, 32, and 90) are utilized to verify the validity of EFSMDER in the multi-dimension emotion recognition task. The experimental results demonstrate that the EEG feature subsets selected by EFSMDER can obtain the best performance on six evaluation metrics in contrast with those of fourteen state-of-the-art feature selection methods.

The remainder of the article is arranged as follows. In Section II, the details of the proposed EFSMDER framework are presented. Then, Section III provides an effective but simple solving algorithm of EFSMDER. The experimental details are explained in Section IV. The experimental results are shown and discussed in Section V. The conclusion and future works are provided in Section VI.

II. The Proposed Framework

A. Notations

First of all, notations and definitions are briefly introduced in this section. Throughout this article, vectors and matrices are represented by lower-case boldface letters (e.g., $\boldsymbol{a}$, $\boldsymbol{b}$, ...) and capital letters (e.g., A, B, ...), respectively. For an arbitrary matrix S, the transpose, trace, and vectorization are represented as S^T, $tr(S)$, and $vec(S)$, respectively. $\boldsymbol{s}_i$ and $\boldsymbol{s}_{.j}$, and $\boldsymbol{s}_{ij}$ respectively represent the ith row, the jth column, and the (i, j) entry of the matrix S. The Frobenius norm of the matrix S is denoted as $\|S\|_F = \sqrt{tr\left(S^T S\right)}$. The operator $\circ$ denotes the Hadamard product/entry-wise product. $\mathbf{1}_n = (1, \ldots, 1)^T \in \mathbb{R}^{n\times 1}$ is a column vector. I_n denotes an $n \times n$ identity matrix.

Let (X,Y) be the multi-dimension emotional data set. $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d]^T \in \mathbb{R}^{\mathrm{d}\times n}$ is the EEG feature matrix where $\boldsymbol{x}_d \in \mathbb{R}^{1\times n}$. $Y = \left[\boldsymbol{y}_{.1}, \boldsymbol{y}_{.2}, \ldots, \boldsymbol{y}_{.k}\right] \in \{-1, 1\}^{n\times k}$ is the multi-dimension emotional label matrix where $\boldsymbol{y}_{.i} = \{y_{1i}, \ldots, y_{ni}\}^T \in \{-1, 1\}^{n\times 1}$. Each row and column of X denote one feature dimension and a sample, respectively. Each column of Y represents one emotional dimension.

B. Problem Formulation

To obtain informative and non-redundant feature subsets for multi-dimension emotion recognition, a novel embedded EEG feature selection method is proposed in this section. The proposed EFSMDER framework is defined as follow:

$$\min_{W,\Theta,V} F(X, W, \Theta, V) + \lambda C(Y, V) + \gamma\Omega(\Theta) \qquad (1)$$

where W, Θ, and V respectively represent the projection matrix, the feature weighting matrix, and the latent semantics of the multiple label information. λ and γ are both tradeoff

parameters. The terms F, C, and Ω denote the EEG feature mapping function, the multi-dimension emotional label learning function, and the EEG feature redundancy function, respectively. Firstly, the EEG feature mapping function is used to capture the local label correlations between EEG features and emotional labels. Additionally, the multi-dimension emotional label learning function is employed to explore the global emotional label correlations. Finally, the EEG feature redundancy function is adopted to mine the redundancy between EEG features from a global view. The definitions of the terms F, C, and Ω will be introduced in the following sections.

C. Exploiting Local Label Correlations

Inspired by our previous works [17], the orthogonal regression model is employed to preserve more local information in the projection subspace W. The term F can be formulated as follow:

$$F(X, W, \Theta, V) = \left\| X^T \Theta W + \mathbf{1}_n \boldsymbol{b}^T - V \right\|_F^2 + \eta tr\left(V^T L V\right) \quad \text{s.t. } W^T W = I_c, \boldsymbol{\theta}^T \mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0 \tag{2}$$

where $\boldsymbol{b} \in \mathbb{R}^{c\times 1}$ is the bias term and η ($\eta > 0$) is a tradeoff parameter. Different from sparse based feature selection methods, a feature weighting vector $\boldsymbol{\theta} \in \mathbb{R}^{d\times 1}$ ($\boldsymbol{\theta} \geq 0$, $\boldsymbol{\theta}^T \mathbf{1}_d = 1$) is utilized to assess the importance of EEG features in the multi-dimension emotional recognition task. $\Theta \in \mathbb{R}^{d\times d}$ is a diagonal matrix with $\Theta_{ii} = \boldsymbol{\theta}_i$. Next, the term $\mathrm{Tr}\left(V^T L V\right)$ is employed to retain that the local geometry structures are consistent between the EEG original feature space X and the low-dimensional space V [18]. $L = G - S$ denotes the graph Laplacian matrix ($L \in \mathbb{R}^{n\times n}$). S denotes the affinity graph of X, and G denotes a diagonal matrix with $G_{ii} = \sum_{j=1}^{n} S_{ij}$. The affinity graph S is computed by a heat kernel. The element S_{ij} in S is the similarity value of two instances $\boldsymbol{x}_{.i}$ and $\boldsymbol{x}_{.j}$. S_{ij} is defined as follow:

$$S_{ij} = \begin{cases} \exp\left(-\dfrac{\left\|\boldsymbol{x}_{.i} - \boldsymbol{x}_{.j}\right\|^2}{\sigma^2}\right) & \boldsymbol{x}_{.i} \in \mathcal{N}_p\left(\boldsymbol{x}_{.j}\right) \text{ or } \boldsymbol{x}_{.j} \in \mathcal{N}_p\left(\boldsymbol{x}_{.i}\right) \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where σ and $\mathcal{N}_p\left(\boldsymbol{x}_j\right)$ represent the graph construction parameter and the set of Top-P nearest neighbors of the instance $\boldsymbol{x}_{.j}$.

D. Exploring Global Label Correlations

$\boldsymbol{v}_{.i}$ and $\boldsymbol{v}_{.j}$ in the projection matrix V denote the coefficients of pairwise labels $\boldsymbol{y}_{.i}$ and $\boldsymbol{y}_{.j}$. If the pairwise emotional labels $\boldsymbol{y}_{.i}$ and $\boldsymbol{y}_{.j}$ are strongly similar, their predictions (i.e, $\boldsymbol{v}_{.i}$ and $\boldsymbol{v}_{.j}$) should be similar, and vise versa. To achieve the goal, the global label information is taken into consideration during the multi-dimension emotional label learning process. A regularizer for the coefficient matrix V is defined as:

$$\sum_{i=1}^{k}\sum_{j=1}^{k} R_{ij} \boldsymbol{v}_{.i}^T \boldsymbol{v}_{.j} \tag{4}$$

where $R_{ij} = 1 - Z_{ij}$, and the element Z_{ij} in Z denotes the label relevance between pairwise labels $\boldsymbol{y}_{.i}$ and $\boldsymbol{y}_{.j}$. The element Z_{ij} is computed by the cosine similarity to mine second-order correlation among multi-dimension emotional labels. It is obvious that R is a positive semidefinite matrix.

Plugging Eq. (4) into $\|Y - V\|_F^2$, and then the C can be formulated as the following:

$$C(Y, V) = \|Y - V\|_F^2 + \beta \,\mathrm{tr}\left(\mathrm{RV^T V}\right) \tag{5}$$

where β ($\beta > 0$) denotes a tradeoff parameter. Eq. (5) is employed to construct the label information in the low-dimension space V from a global view of emotional label correlations.

E. Evaluating Global Feature Redundancy

Furthermore, we define a global feature redundancy matrix A to assess the correlations among the EEG features. A can be calculated as follow:

$$A_{i,j} = \left(O_{i,j}\right)^2 = \left(\frac{\boldsymbol{f}_i^T \boldsymbol{f}_j}{\|\boldsymbol{f}_i\| \, \|\boldsymbol{f}_j\|}\right)^2 \tag{6}$$

where $\boldsymbol{f}_i \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{f}_j \in \mathbb{R}^{n\times 1}$ represent i-th and j-th centralized features of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ ($i, j = 1, 2, \ldots, d$). $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ are defined as follows:

$$\begin{cases} \boldsymbol{f}_i = H \boldsymbol{x}_i^T \\ \boldsymbol{f}_j = H \boldsymbol{x}_j^T \end{cases} \tag{7}$$

where $H = I_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n{}^T$. Eq. (6) can be reformulated as

$$O = DF^T FD = (FD)^T FD \tag{8}$$

where $F = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_d]$. D is a diagonal matrix with $D_{i,i} = \frac{1}{\|\boldsymbol{f}_i\|}$ ($i = 1, 2, \ldots, d$). Because the matrix O is positive semi-definite and $A = O \circ O$, the global feature redundancy matrix A is non-negative and positive semi-definite [19]. Hence, the term Ω can be defined as:

$$\Omega(\Theta) = \boldsymbol{\theta}^T A \boldsymbol{\theta} \quad \text{s.t. } \boldsymbol{\theta}^T \mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0 \tag{9}$$

F. The Final Objective Function of EFSMDER

Based on the above discussion, the final objective function of EFSMDER can be formulated as follows:

$$\min_{W, \boldsymbol{b}, \Theta, V} \left\| X^T \Theta W + \mathbf{1}_n \boldsymbol{b}^T - V \right\|_F^2 + \alpha \|Y - V\|_F^2 + \eta tr\left(V^T L V\right) + \lambda \boldsymbol{\theta}^T A \boldsymbol{\theta} + \beta \,\mathrm{tr}\left(\mathrm{RV^T V}\right) \quad \text{s.t. } W^T W = I_c, \boldsymbol{\theta}^T \mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0 \tag{10}$$

where α, η, λ, and β are tradeoff parameters. The proposed EFSMDER framework is shown in Fig. 1.

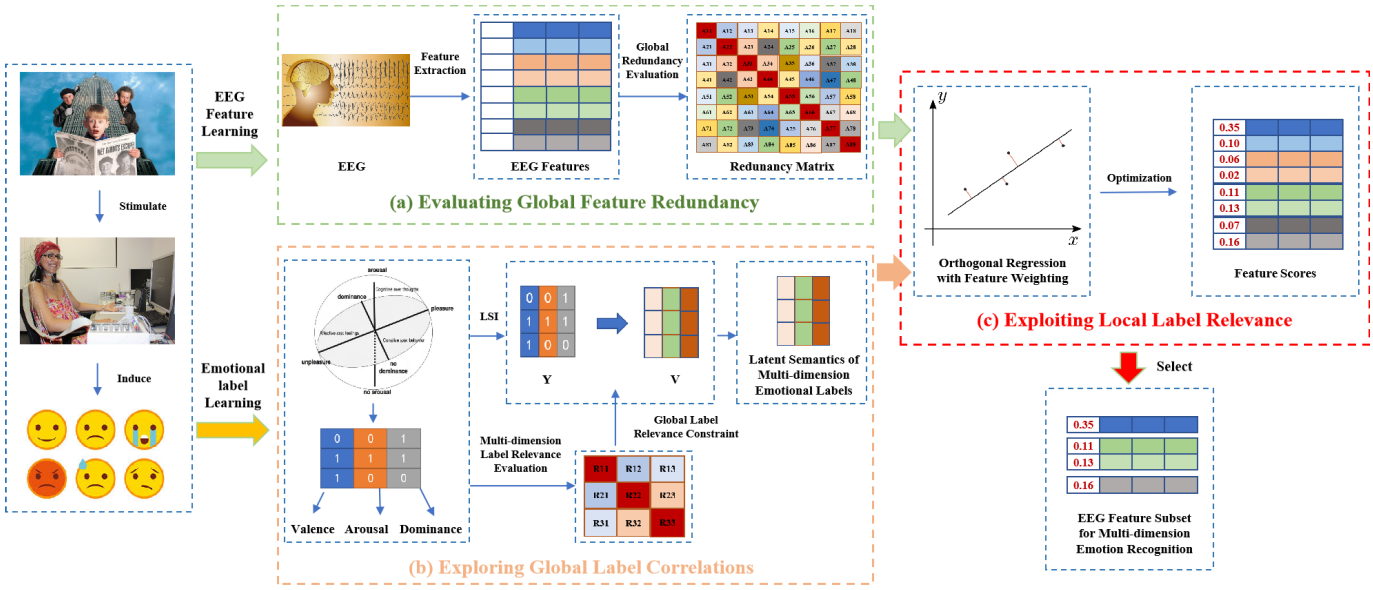

Fig. 1. The proposed EFSMDER framework consists of the following three sections: (a) evaluating global feature redundancy; (b) exploring global label correlations; (3)exploiting local label relevance.

III. Optimization Strategy

Setting the partial derivative of Eq. (10) w.r.t $\boldsymbol{b}$ to zero, we can obtain $\boldsymbol{b} = \frac{1}{n}\left(V^T\mathbf{1}_n - W^T\Theta X\mathbf{1}_n\right)$. Substituting the above expression into Eq. (10), the optimization problem will become

$$\begin{aligned}\min_{W,\Theta,V} &\left\|HX^T\Theta W - HV\right\|_F^2 + \alpha\|Y-V\|_F^2 \\ &+ \eta tr\left(V^T LV\right) + \lambda\boldsymbol{\theta}^T A\boldsymbol{\theta} + \beta\,\mathrm{tr}\left(\mathrm{RV^TV}\right) \\ \text{s.t. } & W^TW = I_c, \boldsymbol{\theta}^T\mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0\end{aligned} \tag{11}$$

In Eq. (11), there are three variables (W, Θ, and V) that need to be solved. The alternative optimization approach is adapted to obtain solutions for the above variables. A specific optimization strategy is introduced as follows.

A. Update W

When Θ and V are fixed, Eq. (11) can be briefly written as

$$\min_{W^TW=I_c} \mathrm{tr}\left(W^TJW - 2W^TM\right) \tag{12}$$

in which $J = \Theta XHX^T\Theta^T$ and $M = \Theta XHV$.

Eq. (12) is the quadratic problem on the Stiefel manifold (QPSM). To solve the problem, the generalized power iteration (GPI) method [20] is adopted. More details can be referred to [20].

B. Update Θ

When W and V are fixed, the subproblem that only relates to Θ in Eq. (11) can be reformulated in the vector form as

$$\begin{aligned}\min_{\Theta} &\left[\mathrm{tr}\left(\Theta XHX^T\Theta WW^T\right) + \lambda\boldsymbol{\theta}^TA\boldsymbol{\theta} - \mathrm{tr}\left(2\Theta XHVW^T\right)\right] \\ \text{s.t. } & \boldsymbol{\theta}^T\mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0\end{aligned} \tag{13}$$

Eq. (13) can be converted to

$$\begin{aligned}\min_{\boldsymbol{\theta}} &\left[\boldsymbol{\theta}^T\left[\left(XHX^T\right)^T \circ \left(WW^T\right) + \lambda A\right]\boldsymbol{\theta} - \boldsymbol{\theta}^Ts\right] \\ \text{s.t. } & \boldsymbol{\theta}^T\mathbf{1}_d = 1, \boldsymbol{\theta} \geq 0\end{aligned} \tag{14}$$

Eq. (14) can be rewritten as

$$\min_{\boldsymbol{\theta}^T\mathbf{1}_d=1,\boldsymbol{\theta}\geq 0} \boldsymbol{\theta}^TQ\boldsymbol{\theta} - \boldsymbol{\theta}^Ts \tag{15}$$

in which

$$\begin{cases} Q = \left(XH^TX^T\right) \circ \left(WW^T\right) + \lambda A \\ s = diag\left(2XHVW^T\right) \end{cases} \tag{16}$$

The objective function in Eq. (15) is an equality and inequality constrained optimization problem, which can be solved by the general augmented Lagrangian multiplier (ALM) method. Subsequently, Eq. (15) can be changed into

$$\min_{\boldsymbol{\theta}^T\mathbf{1}_d=1,\boldsymbol{v}\geq 0,\boldsymbol{v}=\boldsymbol{\theta}} \boldsymbol{\theta}^TQ\boldsymbol{\theta} - \boldsymbol{\theta}^Ts \tag{17}$$

We now write the augmented lagrangian function of Eq. (15) as

$$\begin{aligned}&L\left(\boldsymbol{\theta}, \boldsymbol{v}, \mu, \boldsymbol{\delta}_1, \delta_2\right) \\ &= \boldsymbol{\theta}^TQ\boldsymbol{\theta} - \boldsymbol{\theta}^Ts + \frac{\mu}{2}\left\|\boldsymbol{\theta} - \boldsymbol{v} + \frac{1}{\mu}\boldsymbol{\delta}_1\right\|_F^2 \\ &+ \frac{\mu}{2}\left(\boldsymbol{\theta}^T\mathbf{1}_d - 1 + \frac{1}{\mu}\delta_2\right)^2 \quad \text{s.t. } \boldsymbol{v} \geq 0\end{aligned} \tag{18}$$

where $\boldsymbol{v}$ and $\boldsymbol{\delta}_1$ are both column vectors, and μ denotes the Lagrangian multiplier. When $\boldsymbol{v}$ is fixed, Eq. (18) becomes

$$\min_{\boldsymbol{\theta}} \frac{1}{2}\boldsymbol{\theta}^TE\boldsymbol{\theta} - \boldsymbol{\theta}^Tf \tag{19}$$

in which

$$\begin{cases} E = 2Q + \mu I_d + \mu \mathbf{1}_d \mathbf{1}_d^T \\ \boldsymbol{f} = \mu \boldsymbol{v} + \mu \mathbf{1}_d - \delta_2 \mathbf{1}_d - \boldsymbol{\delta}_1 + \boldsymbol{s} \end{cases} \tag{20}$$

We can derive that $\boldsymbol{\theta} = E^{-1}\boldsymbol{f}$.

With the fixed $\boldsymbol{\theta}$, Eq. (18) can be converted to

$$\min_{\boldsymbol{v} \geq 0} \left\| \boldsymbol{v} - \left(\boldsymbol{\theta} + \frac{1}{\mu} \boldsymbol{\delta}_1 \right) \right\|^2 \tag{21}$$

Solving Eq. (21), we could get

$$\boldsymbol{v} = pos\left(\hat{\boldsymbol{\theta}} + \frac{1}{\mu} \boldsymbol{\delta}_1 \right) \tag{22}$$

where $pos\,(t)$ is a function that assigns 0 to each negative element of t.

C. Update V

When Θ and W are fixed, we take the partial derivative w.r.t V and set the result to zero, and we have

$$2\left[H^T (V - X^T \Theta W) + \alpha (V - Y) + \eta L V + \beta V R \right] = 0 \tag{23}$$

Eq. (23) can be reformulated as

$$(H^T + \eta L)V + V(\alpha I_k + \beta R) = H^T X^T \Theta W + \alpha Y \tag{24}$$

Eq. (24) can be formulated as follow

$$MV + VN = P \tag{25}$$

where

$$\begin{cases} M = H^T + \eta L \\ N = \alpha I_k + \beta R \\ P = H^T X^T \Theta W + \alpha Y \end{cases} \tag{26}$$

It can be seen that Eq. (25) is the Sylvester equation. We employ the lyap function in Matlab to obtain the optimal solution for the variable V in the Sylvester equation.

Based on the above analysis, the detailed optimization procedures of Eq. (10) are given in Algorithm 1. The EFSMDER chooses an alternating optimization strategy to iteratively update the three variables W, Θ, and V until convergence. The output $\boldsymbol{\theta}$ can be utilized to evaluate the importance of each EEG feature in the multi-dimension emotion recognition task. Finally, the m discriminative and non-redundant EEG features with the top rankings in the feature score vector $\boldsymbol{\theta}$ are chosen.

Algorithm 1 EEG Feature Selection for Multi-Dimension Emotion Recognition (EFSMDER)

Input: The EEG feature matrix $X \in \mathbb{R}^{d \times n}$, the multi-dimension emotional label matrix $Y \in \mathbb{R}^{n \times k}$. $p > 1$, $\theta_i = \frac{1}{d}\,(1 \leq i \leq d)$, $\boldsymbol{v} = \boldsymbol{\theta}$, $\delta_2 = 0$, $u > 0$, $\boldsymbol{\delta}_1 = (0, 0, \ldots, 0)^T \in \mathbb{R}^{d \times 1}$.

Output: EEG feature score vector $\boldsymbol{\theta}$.

Initial $\Theta \in \mathbb{R}^{d \times d}$ satisfying $\boldsymbol{\theta}^T \mathbf{1}_d = 1$, and $\boldsymbol{\theta} \geq 0$. $H = I_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n{}^T$. Initial W and V randomly.
repeat
  Calculate $J = \Theta X H X^T \Theta^T$ and $M = \Theta X H V^T$
  Update W via GPI.
  repeat
    Update Q and s via Eq. (16)
    Update E by $E = 2Q + \mu I_d + \mu \mathbf{1}_d \mathbf{1}_d^T$;
    Update $\boldsymbol{f}$ by $\boldsymbol{f} = \mu \boldsymbol{v} + \mu \mathbf{1}_d - \delta_2 \mathbf{1}_d - \boldsymbol{\delta}_1 + \boldsymbol{s}$;
    Update $\boldsymbol{\theta}$ by $\boldsymbol{\theta} = E^{-1}\boldsymbol{f}$;
    Update $\boldsymbol{v}$ by $\boldsymbol{v} = \text{pos}\left(\boldsymbol{\theta} + \frac{1}{\mu}\boldsymbol{\delta}_1 \right)$;
    Update $\boldsymbol{\delta}_1$ by $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_1 + \mu\,(\boldsymbol{\theta} - \boldsymbol{v})$;
    Update δ_2 by $\delta_2 = \delta_2 + \mu\left(\boldsymbol{\theta}^T \mathbf{1}_d - 1 \right)$;
    Update μ by $\mu = p\mu$;
  until Convergence;
  Update Θ via $\Theta = diag(\boldsymbol{\theta})$;
  Update V by solving Eq. (24);
until Convergence;
return $\boldsymbol{\theta}$ for EEG feature selection.

IV. Experimental Details

The experimental details are introduced in this section, such as EEG databases, feature extraction, experimental setup, and evaluation metrics.

A. Database Description

We conduct extensive experiments to evaluate the validity and reliability of the proposed EFSMDER method on three EEG emotional databases, including DREAMER [21], DEAP [22], and HDED [13]. All the databases adopt the multi-dimension emotion model to represent human emotions, i.e. valence-arousal-dominance model.

The DREAMER and DEAP databases respectively adopted 14 and 32 EEG electrodes to simultaneously record EEG signals during the video stimulation. Hence, both DEAP and DREAMER are low-density EEG databases. To further evaluate the ability of the high-density EEG feature selection, high-density EEG recordings (HDED) were recorded via a 128-channel HydroCel Geodesic Sensor Net, and HDED adopted an experimental scheme similar to the SEED database. To the EEG recordings in the HDED database, since the 90 channels near the center of the brain (Cz electrode) on the topographical map can basically cover the entire brain area, we only utilize the corresponding 90 channels to extract the EEG features. The experimental protocols for constructing the three databases are described in detail in [13], [21], and [22]. The basic information of the three databases is shown in Table I.

B. Feature Extraction From EEG Recordings

In this work, the EEG signals were filtered out the noise by a 1-50 Hz band-pass filter. Independent component analysis was then used to suppress muscular and eye movement artifacts. It should be noted that an entire trial was treated as a sample

TABLE I
COMPARISONS AMONG THE THREE EEG DATABASES

Data set	DREAMER	DEAP	HDED
Channel no.	14	32	90
Subject no.	23	32	16
Video no.	18	40	12
Sample no.	414	1280	192
Stimulus materials	film clips	music videos	film clips

for EEG feature extraction. In other words, trials were not segmented into several segments to increase the sample size in our experiment.

According to the feature extraction and analysis researches in the previous literature [12], [23], [24], the following thirteen kinds of features were extracted for EEG-based multi-dimension emotion recognition: NSI [6], HOC [3], spectral entropy [25], shannon entropy [26], C0 complexity [27], DE [4], absolute power (AP), the absolute power ratio of the beta band to the theta band(AP_β/AP_θ) [28], the amplitude of the Hilbert transform of intrinsic mode functions (AHTIMF), the instantaneous phase of the Hilbert transform of intrinsic mode functions (IPHTIMF) [23], differential asymmetry (DASM) [4], RASM [5], and function connectivity (FC). The detailed definitions of above features can refer to [3], [4], [23], [25], [26], [27], and [28].

To extract the three kinds of EEG features in the frequency domain (DE, AP, AP_β/AP_θ), the original EEG signals were divided into five bands: delta band (1)-4 Hz), theta band (4)-8 Hz), alpha band (8)-13 Hz), beta band (13)-30 Hz), and gamma band (30-50 Hz). Due to the advantage of the autoregressive model in the spectral estimation of nonstationary signals [29], the autoregressive model was employed to extract the EEG feature AP. The EEG features AHTIMF and IPHTIMF, employed to describe dynamical changes in the non-stationary and nonlinear EEG recordings, are two commonly used time-frequency features. The intrinsic mode functions (IMF) were decomposed by the empirical mode decomposition method [30], and then the AHTIMF and IPHTIMF are extracted from the IMFs. DASM and RASM were calculated from the symmetric channels on the left and right hemispheres. Pearson correlation coefficient was utilized to calculate the functional connectivity between pairwise electrodes. The number of the above thirteen EEG features extracted from the three databases is shown in Table II.

TABLE II
THE DIMENSION OF THE THIRTEEN EEG FEATURES EXTRACTED FROM THE THREE DATABASES

Feature type	DREAMER	DEAP	HDED
NSI	14	32	90
HOC	140	320	900
spectral entropy	14	32	90
shannon entropy	14	32	90
C0 complexity	14	32	90
AP	70	160	450
AP_β/AP_θ	14	32	90
DE	70	160	450
AHTIMF	70	160	450
IPHTIMF	70	160	450
DASM	35	70	205
RASM	35	70	205
FC	91	496	4005
Total	651	1756	7565

C. Experimental Setup

To fully evaluate the performance of EFSMDER in the emotion recognition task, fourteen state-of-the-art feature selection methods are compared, including three single-label filter methods (i.e., ReliefF [31], CMIM [32], and mRMR [33]), six single-label embedded methods (i.e., FSOR [12], RFS [34], embedded supervised feature selection (ESFS) [35], robust and pragmatic multi-class feature selection (RPMFS) [36], subspace sparsity discriminant feature selection (SDFS) [37], GRMOR [13]), and five multi-label embedded feature selection methods (i.e., feature selection method via manifold regularization (MDFS) [38], multi-label feature selection using multi-criteria decision making (MFS-MCDM) [39], global relevance and redundancy optimization (GRRO) [40], multi-label graph-based feature selection (MGFS) [41], and shared common mode between features and labels (SCMFS) [42]).

According to the self-evaluated value of each emotional dimension, the threshold, classified the EEG instances into high and low levels, was set to 5. Hence, the 3-dimension emotional labels were divided into 8 categories. Multi-label k-nearest neighbor (ML-KNN) [43] is one representative multi-label learning method and was used as a base classifier in the multi-dimension emotion recognition task. The number of nearest neighbors and smooth are respectively set to 10 and 1. KNN and support vector machines (SVM) were adopted as single-dimension emotion recognition classifiers of each emotion dimension (i.e. valence, arousal, and dominance). 70% of the subjects were randomly selected as training sets and then the remaining 30% were treated as test sets. Note that a cross-subject experiment setting was conducted. To avoid possible bias, 50 independent realizations were conducted, and then the average was regarded as the final emotion recognition results.

D. Performance Metrics

The performance metric accuracy was implemented to analyze the single-dimension affective computing ability. Additionally, six performance metrics were employed to compare the multi-dimension emotion recognition performance and redundant information removal performance from various aspects, including one feature redundancy evaluation metric redundancy, two label-based evaluation metrics macro-F1 and micro-F1, and three evaluation example-based metrics average precision, coverage, and hamming loss. The definitions of the six metrics are described in [13] and [38].

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Performance Comparison in the Single-Dimension Emotion Recognition

Almost 10% of the original number of all features were chosen by the above feature selection methods, and then the results in the single-dimension emotion recognition task

TABLE III
THE COMPARISONS OF AVERAGE CLASSIFICATION ACCURACY (%) OF KNN ON 3 BENCHMARK DATASETS

Methods	Discrete Labels								
	DREAMER			DEAP			HDED		
	Valence	Arousal	Dominance	Valence	Arousal	Dominance	Valence	Arousal	Dominance
mRMR	60.16	73.87	77.74	61.48	64.77	64.64	86.87	63.97	68.45
ReliefF	63.71	72.34	77.42	58.26	63.02	64.24	58.10	68.97	68.79
CMIM	58.63	74.03	76.85	59.43	62.71	63.80	87.93	65.00	68.28
RFS	62.58	73.95	77.26	59.90	64.97	64.71	86.72	67.41	69.48
SDFS	61.13	73.71	78.06	56.72	62.42	62.42	82.93	66.72	68.45
ESFS	64.68	74.03	77.66	59.77	63.28	64.24	85.00	64.48	64.48
RPMFS	62.74	71.94	77.82	63.36	63.05	65.05	55.86	62.59	63.97
FSOR	62.18	71.77	77.50	60.08	64.11	65.18	86.72	64.31	66.72
SCMFS	60.48	70.97	76.77	62.24	63.52	65.49	58.28	57.93	65.17
GRRO	63.79	75.48	79.27	63.93	65.76	66.98	93.62	77.76	74.14
MDFS	65.81	75.81	79.35	63.91	65.47	66.54	93.79	79.31	73.97
MFS_MCDM	60.48	71.85	76.85	58.67	65.00	65.60	72.76	62.76	69.31
MGFS	61.94	73.06	77.66	61.85	64.32	64.87	92.24	76.21	70.86
GRMOR	61.77	75.16	79.44	63.83	65.13	65.55	90.34	69.83	72.07
EFSMDER	**69.68**	**78.37**	**83.89**	**67.08**	**66.35**	**69.40**	**95.69**	**81.03**	**76.17**

TABLE IV
THE COMPARISONS OF AVERAGE CLASSIFICATION ACCURACY (%) OF SVM ON 3 BENCHMARK DATASETS

Methods	Discrete Labels								
	DREAMER			DEAP			HDED		
	Valence	Arousal	Dominance	Valence	Arousal	Dominance	Valence	Arousal	Dominance
mRMR	61.13	74.52	76.37	56.38	62.34	61.85	71.72	56.72	62.59
ReliefF	60.16	71.37	77.18	54.87	60.52	60.10	69.66	60.86	59.66
CMIM	59.44	71.85	75.81	55.89	62.71	61.82	67.41	58.28	58.10
RFS	61.45	72.90	74.44	57.66	64.82	65.00	79.31	64.48	67.59
SDFS	60.89	72.82	74.92	54.79	62.24	62.01	79.48	62.41	59.31
ESFS	59.76	70.97	76.29	57.89	63.10	63.83	71.03	57.41	59.48
RPMFS	58.55	73.71	76.77	57.99	62.66	62.68	65.17	61.21	64.31
FSOR	60.48	71.94	76.21	60.21	62.50	62.55	73.97	62.41	60.17
SCMFS	60.56	72.58	76.61	56.80	63.44	64.66	61.55	53.10	56.72
GRRO	63.15	74.19	78.79	60.78	63.82	65.76	83.45	68.62	69.48
MDFS	63.31	73.79	79.84	62.42	64.49	66.07	83.79	71.38	70.34
MFS_MCDM	57.34	70.08	77.42	59.38	62.86	64.14	70.69	68.28	64.14
MGFS	58.95	71.69	75.65	59.79	63.72	63.54	81.38	62.24	62.59
GRMOR	62.18	73.31	78.39	60.94	64.38	65.47	78.62	67.76	68.62
EFSMDER	**66.35**	**76.48**	**82.16**	**65.35**	**66.03**	**68.89**	**86.72**	**73.62**	**73.44**

have been summarized in Table III and Table IV. As shown in Table III and Table IV, the results of the accumulative two classifiers (KNN and SVM) exhibited the better emotion recognition accuracies of the proposed method, in comparison to the other 14 feature selection methods, which demonstrated the emotion recognition performance depends on our method rather than on the combination of feature selection method and certain classifier model.

Additionally, we compare the single-dimension emotion recognition results with several recent studies [44], [45], [46], [47], [48], [49], [50], [51], [52]. Most of the existing studies used subject-dependent settings and segmented an entire trial into several windows to increase the number of samples. Since we conduct a cross-subject experiment setting and regard an entire trial as a sample, we only compare with the studies adopting similar experimental settings. The above experimental settings are more challenging. To the best of our knowledge, there are few cross-subject studies on emotion recognition for DREAMER datasets, and HDED is not a public dataset. Table V shows the accuracy comparison of EFSMDER against existing studies on the DEAP dataset. As shown in Table V, EFSMDER achieved a higher accuracy of 67.08% for valence and 66.35% for arousal on the DEAP dataset than existing studies.

TABLE V
ACCURACY (%) COMPARISON WITH EXISTING WORKS ON THE DEAP DATASET

Studies	Valence	Arousal
Lew et al. [44]	56.78	56.60
Pandey et al. [45]	62.50	61.25
Rayatdoost et al. [46]	59.22	55.70
Pandey et al. [47]	61.50	58.50
Li et al. [48]	64.20	58.40
Miguel et al. [49]	64.00	59.00
Li et al. [50]	62.66	-
She et al. [51]	65.59	-
He et al. [52]	64.33	63.25
Our work	**67.08**	**66.35**

B. Performance Comparison in the Multi-Dimension Emotion Recognition

For the multi-dimension feature selection model evaluation on the above three EEG databases, each dimension of the 3-dimension emotional model is divided into two categories

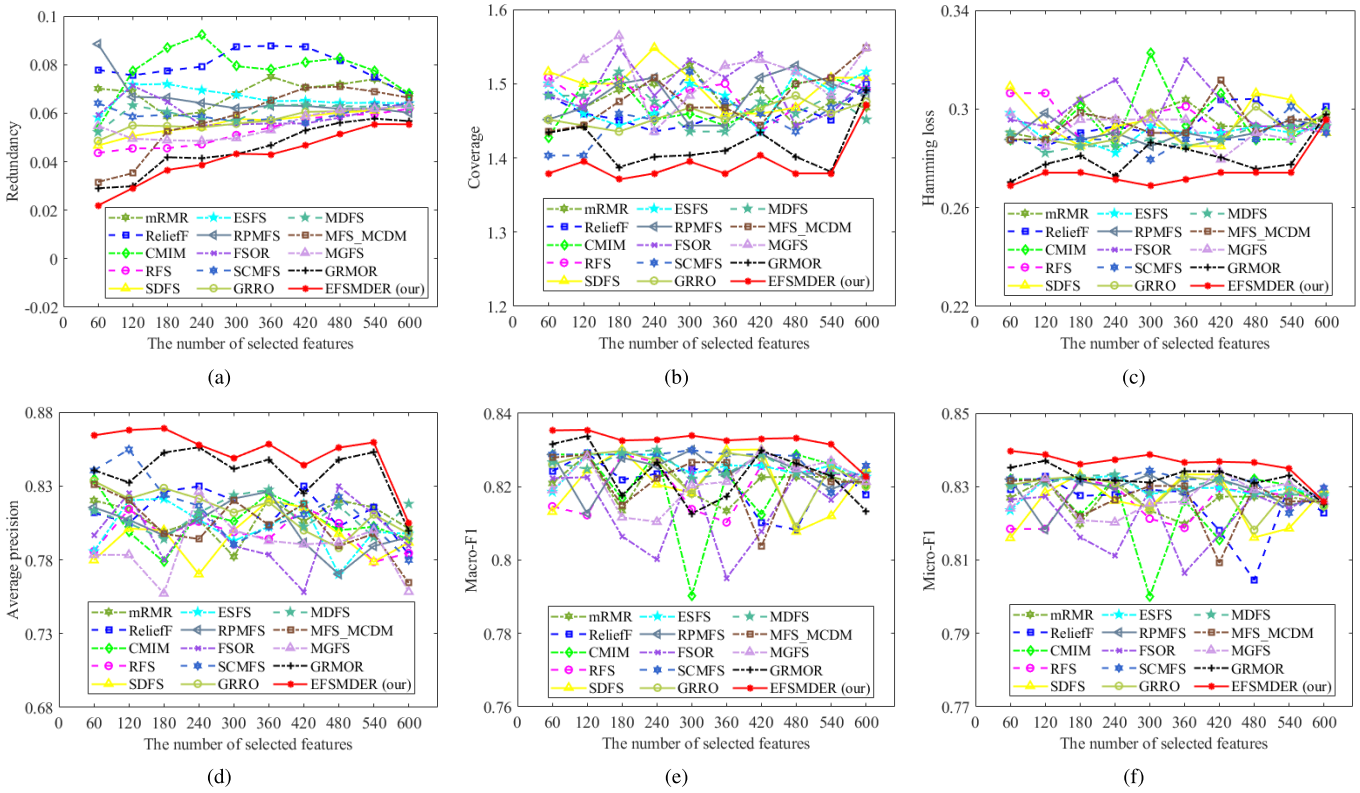

Fig. 2. Multi-dimension emotion classification performance with different number of selected features on the DREAMER data set: (a) Redundancy; (b) Coverage; (c) Hamming loss; (d) Average precision; (e) Macro-F1; (f) Micro-F1.

TABLE VI
THE COMPARISONS OF AVERAGE CLASSIFICATION ACCURACY (%) OF ML-KNN ON 3 BENCHMARK DATASETS

Methods	Multi-dimension Labels		
	DREAMER	DEAP	HDED
mRMR	82.02	82.05	90.17
ReliefF	81.19	79.64	74.36
CMIM	83.45	80.28	92.95
RFS	78.57	82.39	87.18
SDFS	77.98	79.11	91.03
ESFS	78.57	79.90	93.16
RPMFS	81.55	82.54	71.15
FSOR	79.64	79.98	93.59
SCMFS	84.05	82.16	91.45
GRRO	83.33	80.92	90.81
MDFS	81.31	83.26	95.73
MFS_MCDM	83.10	79.75	80.77
MGFS	78.33	82.88	88.03
GRMOR	81.07	81.86	92.31
EFSMDER	**86.43**	**84.80**	**97.86**

TABLE VII
FRIEDMAN TEST RESULTS (15 COMPARING METHODS, 3 DATABASES, SIGNIFICANCE LEVEL $\alpha = 0.05$)

Evaluation metric	F_F	Critical value
Redundancy	4.5969	
Coverage	3.7931	
Hamming loss	6.6598	
Average precision	3.2610	≈ 2.064
Macro-F1	2.7637	
Micro-F1	2.2070	

(high/low) and then an 8-class emotion recognition task is performed. The comparison results of the fifteen algorithms in terms of the six evaluation metrics are shown in Fig. 2, Fig. 3, and Fig. 4. For each figure, the horizontal axis denotes the feature number in the selected EEG feature subsets and the vertical axis denotes the results of each evaluation metric. The red line in the above figures represent EFSMDER.

As shown in Fig. 2(a-c), Fig. 3(a-c), and Fig. 4(a-c), the smaller the value, the better the multi-dimension emotion recognition performance. Additionally, as shown in Fig. 2(d-f), Fig. 3(d-f), and Fig. 4(d-f), the larger the value, the better the multi-dimension emotion recognition performance. Compared with the values of the other fourteen methods, our proposed method almost keeps the maximum values with the different number of selected EEG features, which indicates the optimal multi-dimension emotion recognition results of the EFSMDER method. Table VI shows quantitative comparison results of average classification accuracy. Almost 10% of the original number of all features were selected by 15 methods, and then the average multi-dimension classification accuracies were obtained by ML-KNN. Overall, the results in Fig. 2, Fig. 3, Fig. 4, and Table VI demonstrate that the selected EEG feature subsets of the EFSMDER method can obtain the best performance.

To further analyze the difference in the multi-dimension emotion recognition results of the fifteen methods, the Friedman test is employed. The significance level in the Friedman test is set to 0.05. Table VII shows the Friedman statistical significance test results of six evaluation metrics. The results

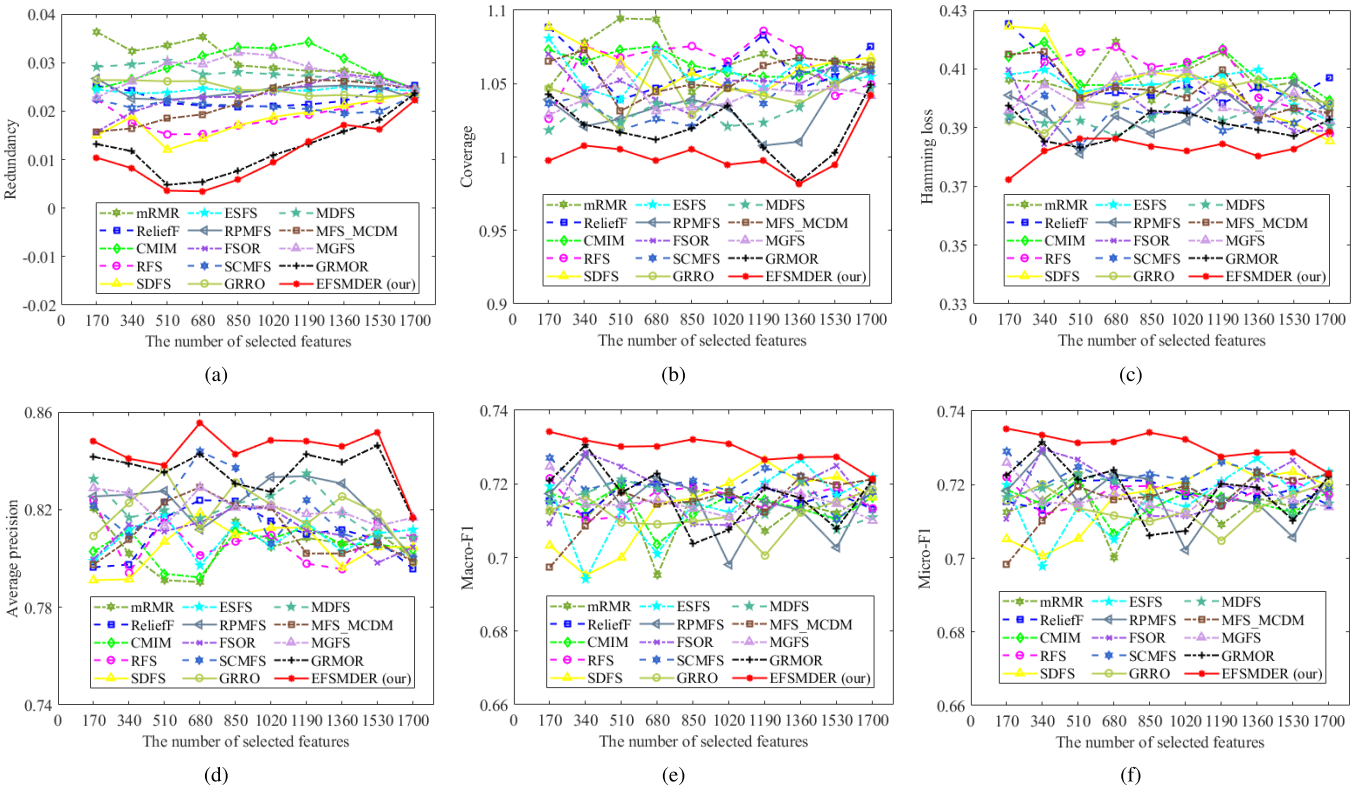

Fig. 3. Multi-dimension emotion classification performance with different number of selected features on the DEAP data set: (a) Redundancy; (b) Coverage; (c) Hamming loss; (d) Average precision; (e) Macro-F1; (f) Micro-F1.

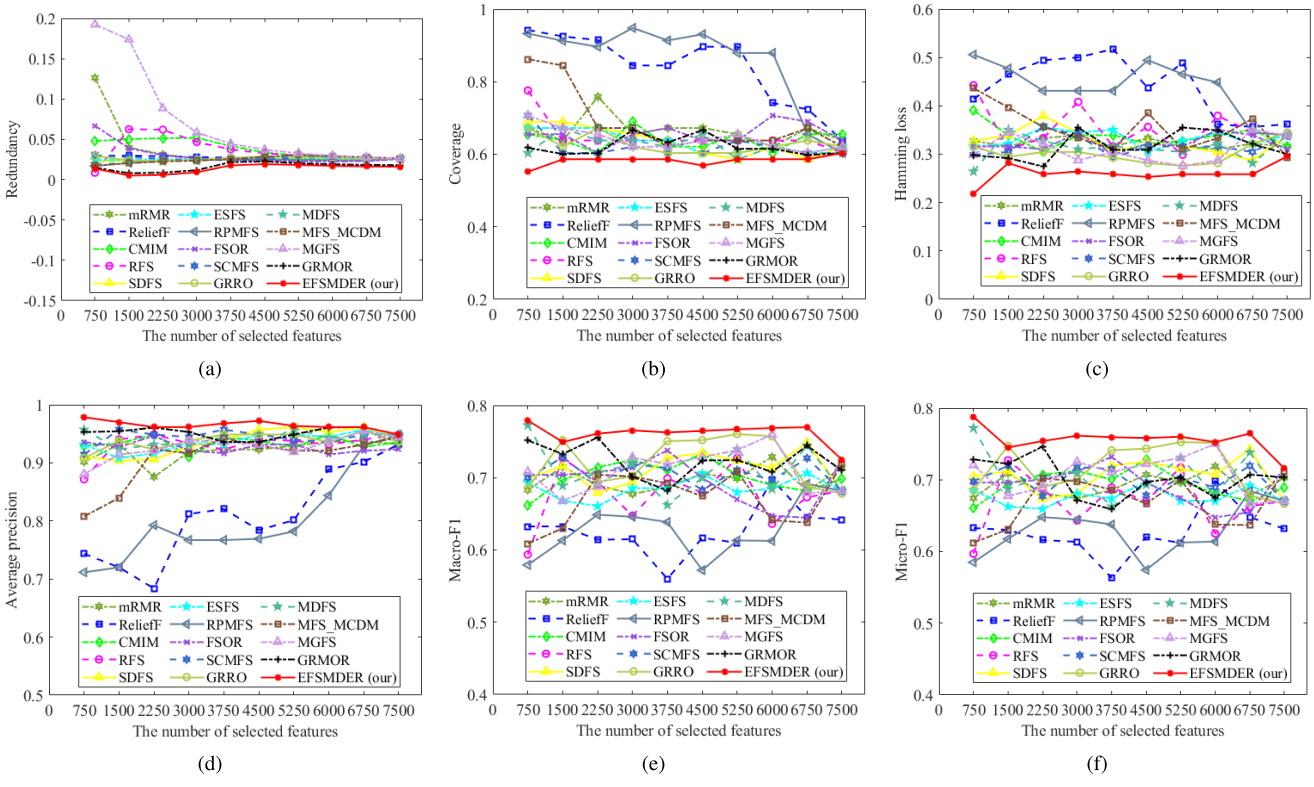

Fig. 4. Multi-dimension emotion classification performance with different number of selected features on the HDED data set: (a) Redundancy; (b) Coverage; (c) Hamming loss; (d) Average precision; (e) Macro-F1; (f) Micro-F1.

in Table VII indicate that the null hypothesis is rejected and the multi-dimension emotion recognition performance of the fifteen methods has a significant difference. To complete the multi-dimension emotion recognition performance

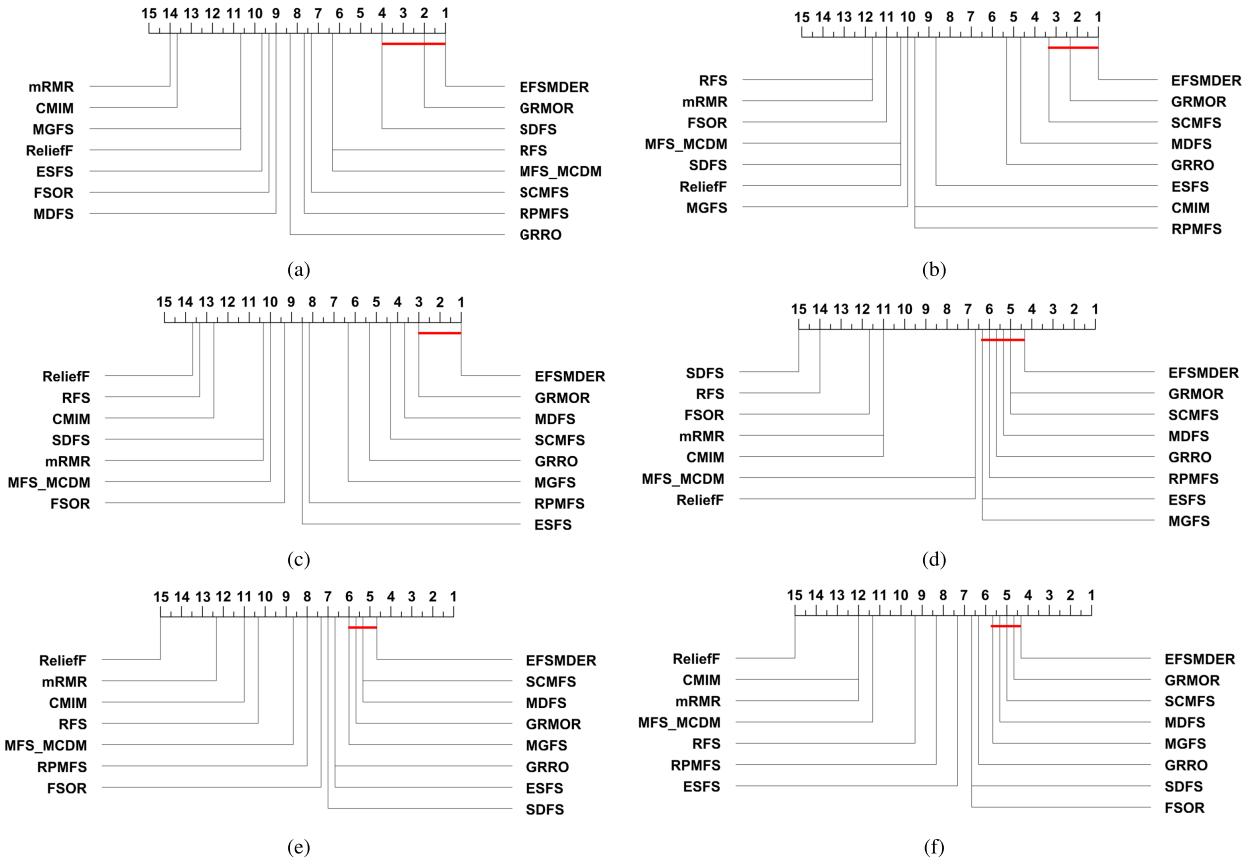

Fig. 5. The Nemenyi test results ($CD = 12.3830$, significance level $\alpha = 0.05$): (a) Redundancy; (b) Coverage; (c) Hamming loss; (d) Average precision; (e) Macro-F1; (f) Micro-F1.

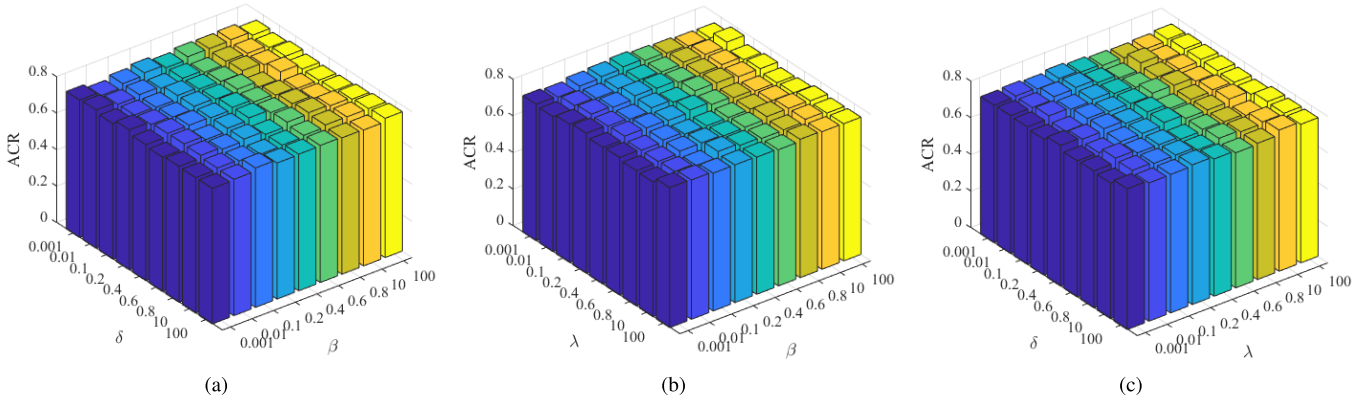

Fig. 6. The parameter sensitivity (under the varying λ, δ, and β) and convergence of EFSMDER on the DEAP data set.

comparison, the Nemenyi test is adopted as a certain post-hoc test and the EFSMDER method is chosen as the control method. For the Nemenyi test, the critical difference (CD), controlled the family-wise error rate, is computed as follows:

$$CD = q_\alpha \sqrt{\frac{nc(nc+1)}{6nd}} \tag{27}$$

where nc and nd represent the number of feature selection methods and databases. The q_α is 3.164 at significance level $\alpha = 0.05$. The value of CD is 4.2841 ($nc = 15$, $nd = 3$).

Fig. 5 shows the Nemenyi statistical significance test results of six evaluation metrics. As shown in Fig. 5, the EFSMDER method ranks 1st among the fifteen methods on all the evaluation metrics, which demonstrates that the EFSMDER method can obtain highly competitive multi-dimension emotion recognition performance against other feature selection methods.

Specifically, we made the following observations.

(1) As shown in Fig. 5, EFSMDER, GRMOR, MDFS, GRRO, and SCMFS have good performance in terms of the majority performance metrics. Only the GRMOR method is a single-label feature selection method. Hence, compared with converting the multi-label directly into the one-dimension label, the joint analysis of multi-label conduce to mining the

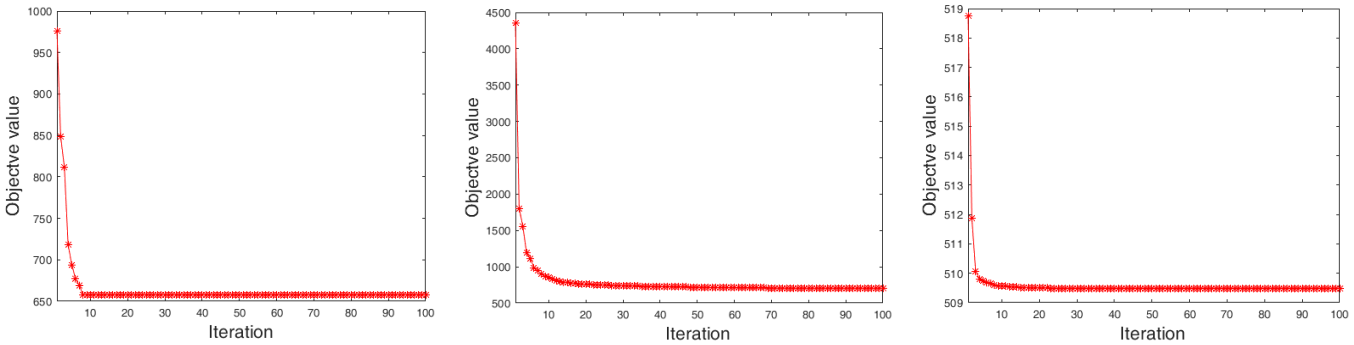

Fig. 7. Convergence of the EFSMDER algorithm on the three data sets: (a) DREAMER; (b) DEAP; (c) HDED.

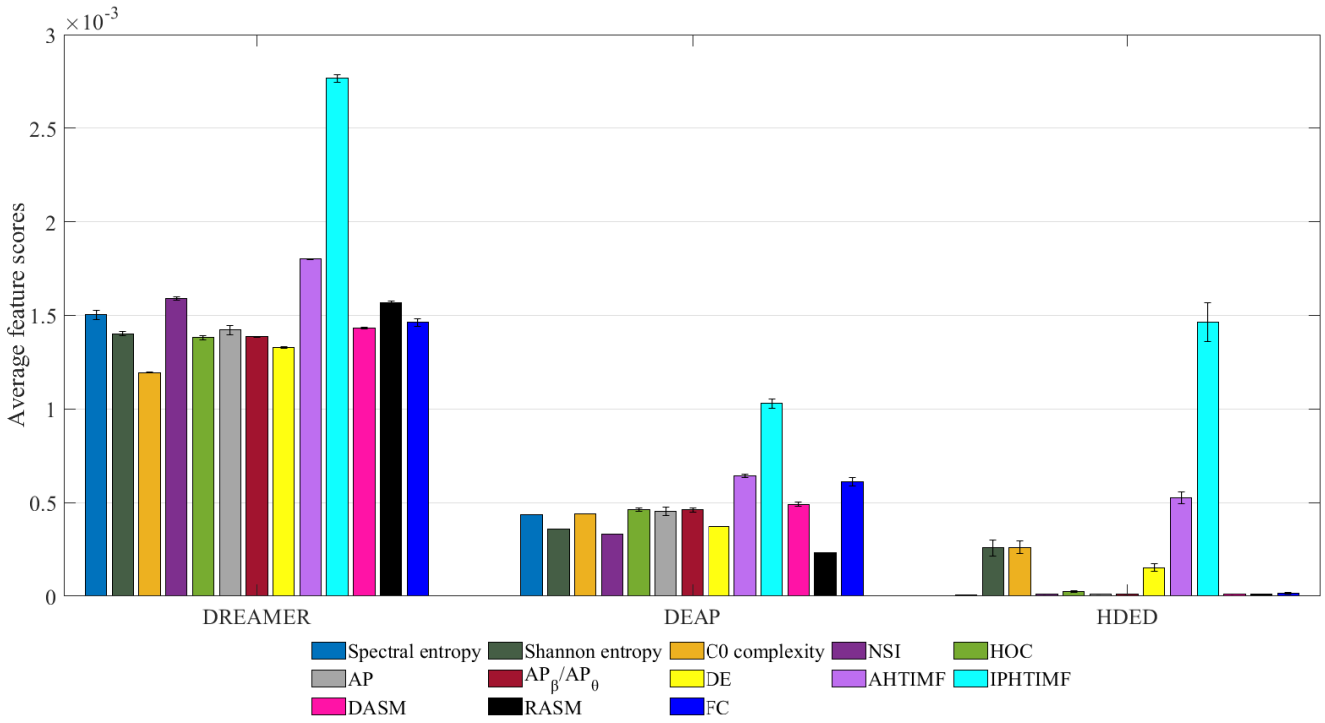

Fig. 8. The importance of different EEG feature types on the multi-dimension emotion recognition task is evaluated based on the feature scores ranked by the proposed EFSMDER method.

latent semantic information in the multi-dimension emotional labels, and the latent semantic information is beneficial for multi-dimension emotion recognition.

(2) From Fig. 2, Fig. 3, and Fig. 4, it can be observed that the performance of FSOR is worse than that of most feature selection methods and the GRMOR method have a better classification performance than the FSOR method, which indicates the effect of the global redundancy term on redundant information removal from the original EEG features.

(3) The EEG feature subsets, selected by the proposed EFSMDER method, consistently yielded a better performance than the GRMOR in terms of six performance metrics. The results demonstrate the multi-dimension emotional label learning function can contribute to improving feature selection performance and the global relevance in multi-dimension emotional labels is worth considering.

C. Parameter Sensitivity Analysis and Convergence Demonstration

To avoid the influence of the value of the tradeoff parameter α on the item $tr\left(\mathrm{RV^TV}\right)$, the value of α is set to 1. Three parameters λ, β, and η are used to regulate the impacts of each term. In this section, the parameter sensitivity analysis for EFSMDER over the three parameters is conducted. The values of the parameters are all adjusted in the range of $\{0.001, 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, 10, 100\}$. With one parameter fixed at 0.1, other parameters are searched in the above set. Due to space limitations, we only show the parameter sensitivity results for the DEAP database. The 3-D histograms in terms of ACR values on the ML-KNN classifier are shown in Fig. 6. As shown in Fig. 6, the values of ACR keep nearly unchanged when two parameters are varying, which demonstrates that the multi-dimension emotion recognition

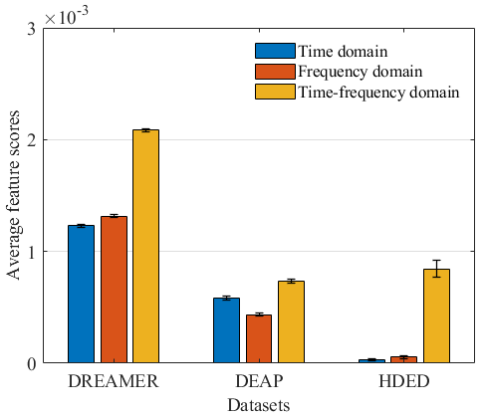

Fig. 9. Average feature scores of EEG features extracted from the time domain, frequency domain, and time-frequency domain.

performance on the DEAP database is not very sensitive to the balance parameters.

Furthermore, we conduct the convergence speed analysis of our iterative optimization algorithm. The convergence curves of the objective value on the three databases are shown in Fig. 7. The balance parameters λ, β, and η are all fixed at 10. As shown in Fig. 7, the proposed EFSMDER algorithm converges quickly within a few iterations, which indicates the effectiveness of our iterative optimization algorithm.

D. Feature Ranking Analysis

Lastly, to analyze the impacts of different kinds of EEG features on multi-dimension emotion recognition performance, the average feature weights acquired by EFSMDER are shown in Fig. 8. It should be noted that a histogram in Fig. 8 denotes the mean of all the feature weights of a type of EEG feature. As shown in Fig. 8, the average feature weight values of the AHTIMF and IPHTIMF are higher than those of other EEG feature types, which indicates the two EEG feature types have a greater influence on multi-dimension emotion recognition performance. It is obvious that the two EEG feature types are both extracted from the time-frequency domain.

To further compare various EEG feature extracted approaches, the average feature scores of EEG features extracted from the time domain, frequency domain, and time-frequency domain are computed and shown in Fig. 9. The values of EEG features in time-frequency domain are the largest, which indicates the advantages of time-frequency domain analysis for multi-dimension EEG affective computing. According to the surveys [24], [53], most of the EEG-based emotion researches often adopt the EEG features in the time domain or frequency domain. Hence, the discovery may provide a suggestion for future researches on EEG emotional feature extraction.

VI. Conclusion and Future Work

Via merging simultaneously local and global label relevance in orthogonal regression, an embedded EEG feature selection framework is proposed to select discriminative and non-redundant EEG features for multi-dimension emotion recognition. To evaluate the performance of EFSMDER in the multi-dimension emotion recognition task, three EEG emotional databases are employed and fourteen state-of-the-art feature selection methods are adopted as comparison methods. The experimental results of six performance metrics demonstrate the effective performance of EFSMDER. By analyzing the feature weights ranked by the EFSMDER method, we found that the EEG features extracted from the time-frequency domain are effective representations of multi-dimension emotion.

In future work, we will focus on the effects of different time-frequency domain analysis methods on the multi-dimension emotion recognition task. Next, missing multi-dimension emotion learning is an interesting research direction.

References

[1] Y. Peng, H. Liu, J. Li, J. Huang, B.-L. Lu, and W. Kong, "Cross-session emotion recognition by joint label-common and label-specific EEG features exploration," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 759–768, 2023.

[2] M. Wu, W. Teng, C. Fan, S. Pei, P. Li, and Z. Lv, "An investigation of olfactory-enhanced video on EEG-based emotion recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1602–1613, 2023.

[3] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[4] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 81–84.

[5] Y.-P. Lin et al., "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[6] E. Kroupi, A. Yazdani, and T. Ebrahimi, "EEG correlates of different emotional states elicited during watching music videos," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2011, pp. 457–466.

[7] H. Becker, J. Fleureau, P. Guillotel, F. Wendling, I. Merlet, and L. Albera, "Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 244–257, Apr. 2020.

[8] H. Wang, X. Wu, and L. Yao, "Identifying cortical brain directed connectivity networks from high-density EEG for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1489–1500, Jul. 2022.

[9] F. Wang et al., "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, Art. no. 107506.

[10] C. Zhan, D. She, S. Zhao, M.-M. Cheng, and J. Yang, "Zero-shot emotion recognition via affective structural embedding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1151–1160.

[11] 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.

[12] X. Xu, F. Wei, Z. Zhu, J. Liu, and X. Wu, "EEG feature selection using orthogonal regression: Application to emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1239–1243.

[13] X. Xu, T. Jia, Q. Li, F. Wei, L. Ye, and X. Wu, "EEG feature selection via global redundancy minimization for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 421–435, Jan. 2023.

[14] C. Hou, F. Nie, D. Yi, and Y. Wu, "Feature selection via joint embedding learning and sparse regression," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1324–1329.

[15] R. Zhang, F. Nie, X. Li, and X. Wei, "Feature selection with multi-view data: A survey," *Inf. Fusion*, vol. 50, pp. 158–167, Oct. 2019.

[16] J. Ma and T. W. S. Chow, "Topic-based instance and feature selection in multilabel classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 315–329, Jan. 2022.

[17] X. Wu, X. Xu, J. Liu, H. Wang, B. Hu, and F. Nie, "Supervised feature selection with orthogonal regression and feature weighting," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1831–1838, May 2021.

[18] L. Jian, J. Li, K. Shu, and H. Liu, "Multi-label informed feature selection," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 1627–1633.

[19] D. Wang, F. Nie, and H. Huang, "Feature selection via global redundancy minimization," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 10, pp. 2743–2755, Oct. 2015.

[20] F. Nie, R. Zhang, and X. Li, "A generalized power iteration method for solving quadratic problem on the Stiefel manifold," *Sci. China Inf. Sci.*, vol. 60, no. 11, Nov. 2017, Art. no. 112101.

[21] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[22] S. Koelstra et al., "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[23] 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.

[24] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul. 2019.

[25] A. Zhang, B. Yang, and L. Huang, "Feature extraction of EEG signals using power spectral entropy," in *Proc. Int. Conf. Biomed. Eng. Informat.*, May 2008, pp. 435–439.

[26] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[27] Y. Zhou, L. Xie, G. Yu, F. Liu, Y. Zhao, and Y. Huang, "The study of C0 complexity on epileptic absence seizure," in *Proc. 7th Asian–Pacific Conf. Med. Biol. Eng.* Cham, Switzerland: Springer, 2008, pp. 420–425.

[28] S. Koehler et al., "Increased EEG power density in alpha and theta bands in adult ADHD patients," *J. Neural Transmiss.*, vol. 116, no. 1, pp. 97–104, Jan. 2009.

[29] Z. Li et al., "The recognition of multiple anxiety levels based on electroencephalograph," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 519–529, Jan. 2022.

[30] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 112–114, Feb. 2004.

[31] I. Kononenko, "Estimating attributes: Analysis and extensions of relief," in *Proc. Eur. Conf. Mach. Learn.* Cham, Switzerland: Springer, 1994, pp. 171–182.

[32] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, no. 11, pp. 1531–1555, 2004.

[33] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, Apr. 2005.

[34] F. Nie, H. Huang, X. Cai, and C. H. 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] L. Chen, J. Tang, and B. Li, "Embedded supervised feature selection for multi-class data," in *Proc. SIAM Int. Conf. Data Mining*, 2017, pp. 516–524.

[36] X. Cai, F. Nie, and H. Huang, "Exact top-k feature selection via $\ell_{2,0}$-norm constraint," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1240–1246.

[37] Z. Wang, F. Nie, L. Tian, R. Wang, and X. Li, "Discriminative feature selection via a structured sparse subspace learning module," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, 2020, pp. 3009–3015.

[38] J. Zhang, Z. Luo, C. Li, C. Zhou, and S. Li, "Manifold regularized discriminative feature selection for multi-label learning," *Pattern Recognit.*, vol. 95, pp. 136–150, Nov. 2019.

[39] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "MFS-MCDM: Multi-label feature selection using multi-criteria decision making," *Knowl.-Based Syst.*, vol. 206, Oct. 2020, Art. no. 106365.

[40] J. Zhang, Y. Lin, M. Jiang, S. Li, Y. Tang, and K. C. Tan, "Multi-label feature selection via global relevance and redundancy optimization," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 2512–2518.

[41] A. Hashemi, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality," *Expert Syst. Appl.*, vol. 142, Mar. 2020, Art. no. 113024.

[42] L. Hu, Y. Li, W. Gao, P. Zhang, and J. Hu, "Multi-label feature selection with shared common mode," *Pattern Recognit.*, vol. 104, Aug. 2020, Art. no. 107344.

[43] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[44] W. L. Lew et al., "EEG-based emotion recognition using spatial–temporal representation via Bi-GRU," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 116–119.

[45] P. Pandey and K. R. Seeja, "Subject independent emotion recognition from EEG using VMD and deep learning," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1730–1738, May 2022.

[46] S. Rayatdoost and M. Soleymani, "Cross-corpus EEG-based emotion recognition," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2018, pp. 1–6.

[47] P. Pandey and K. R. Seeja, "Subject independent emotion recognition system for people with facial deformity: An EEG based approach," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 2, pp. 2311–2320, Feb. 2021.

[48] X. Li, P. Zhang, D. Song, G. Yu, Y. Hou, and B. Hu, "EEG based emotion identification using unsupervised deep feature learning," in *Proc. Workshop Neuro-Physiol. Methods IR Res. (SIGIR)*, Aug. 2015, pp. 1–2. [Online]. Available: https://oro.open.ac.uk/44132/

[49] M. Arevalillo-Herráez, M. Cobos, S. Roger, and M. García-Pineda, "Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from EEG signals," *Sensors*, vol. 19, no. 13, p. 2999, Jul. 2019.

[50] 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.

[51] Q. She, C. Zhang, F. Fang, Y. Ma, and Y. Zhang, "Multisource associate domain adaptation for cross-subject and cross-session EEG emotion recognition," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[52] Z. He, Y. Zhong, and J. Pan, "An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105048.

[53] H. A. Gonzalez et al., "Hardware acceleration of EEG-based emotion classification systems: A comprehensive survey," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 3, pp. 412–442, Jun. 2021.