Su, Jie, Qing, Linbo, He, Xiaohai, Zhang, Hang, Zhou, Jing and Peng, Yonghong (2018) A New Regularized Matrix Discriminant Analysis (R-MDA) Enabled Human-Centered EEG Monitoring Systems. IEEE Access (99). p. 1. ISSN 2169-3536

Downloaded from: http://sure.sunderland.ac.uk/id/eprint/8954/

Usage guidelines

Please refer to the usage guidelines at http://sure.sunderland.ac.uk/policies.html or alternatively contact sure@sunderland.ac.uk.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

A New Regularized Matrix Discriminant Analysis (R-MDA) Enabled Human-Centered EEG Monitoring Systems

JIE SU[1], LINBO QING[1],(Member, IEEE), XIAOHAI HE[1],(Member, IEEE), HANG ZHANG[2], JING ZHOU[3],AND YONGHONG PENG[4],(Member, IEEE).

[1]College of Electronics and Information Engineering, Sichuan University, Chengdu, 610065 China
[2]School of Electrical and Computer Engineering, Cornell University, Ithaca, 14850 USA
[3]Department of Biomedical Engineering, School of Materials Science and Engineering, South China University of Technology, Guangzhou, 510640 China
[4]Faculty of Computer Science, University of Sunderland, St Peters Campus, Sunderland, SR6 0DD UK

Corresponding author: Linbo Qing (e-mail:qing_lb@scu.edu.cn).

ABSTRACT The wider use of wearable devices for EEG data capturing provides a very useful way for the monitoring and self-management of human health. However, the large volumes of data with high dimensions cause computational complexity in EEG data processing and poses a great challenge to the use of wearable EEG devices in healthcare. This study proposes a new approach to extract the structural information of EEG data and tackle the curse of dimensionality of EEG data. A set of methods for dimensionality reduction (DR) like linear discriminant analysis (LDA) and their improved methods have been developed for EEG processing in the literature. However, existing LDA-related methods suffer from the singularity problem or expensive computational cost, and none of existing methods take into consideration the structure of the projection matrix, which is crucial for the extraction of structural information of EEG data. In this paper, a new method called regularized matrix discriminant analysis (R-MDA) is proposed for EEG feature representation and dimensionality reduction. In R-MDA, the EEG data is represented as data matrix, and projection vectors are reshaped to be a set of projection matrices stacking together. By reformulating LDA as a least-square formulation and imposing specified constraint on each projection matrix, the new R-MDA has been constructed to effectively reduce EEG dimensions and capturing the structural information of EEG data. Experimental results demonstrate that this new R-MDA outperforms the existing LDA-related methods, including achieving improved accuracy with significant dimensionality reduction of EEG data. This offers an effective way to enable wearable EEG devices be applicable in human-centered health monitoring.

INDEX TERMS Human-Centered Health Monitoring, Wearable EEG, Dimensionality Reduction, Regularized Matrix Discriminant Analysis, Projection Matrices Stack, Machine Learning Algorithms, Pattern Recognition, Signal Processing

I. INTRODUCTION

THE electroencephalogram (EEG) becomes one of the most commonly used noninvasive method for the Brain Computer Interface (BCI) due to its strong reliability performance and low cost [1]–[3]. Through multi-channel electrodes placed on the scalp, the weak bioelectricity signals generated by the human brain can be collected. After amplification and preprocessing, the signals are recorded digitally, which reflects neuronal activities of the brain. EEG signals play a vital role in a wide range of application areas, such as healthcare (epilepsy and sleepy disorder diagnosis) [2], [4], human emotions recognition [5], [6], Motor imagery [7], controlling functional neuroprostheses [3] and monitoring drivers' vigilance [8]. However, the conventional EEG causes inconveniences to people who are undergoing EEG monitoring. It can only be used for disease diagnosis after the onset of the patient, rather than performing real-time disease prevention. These issues may restrict the use of EEG to some

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

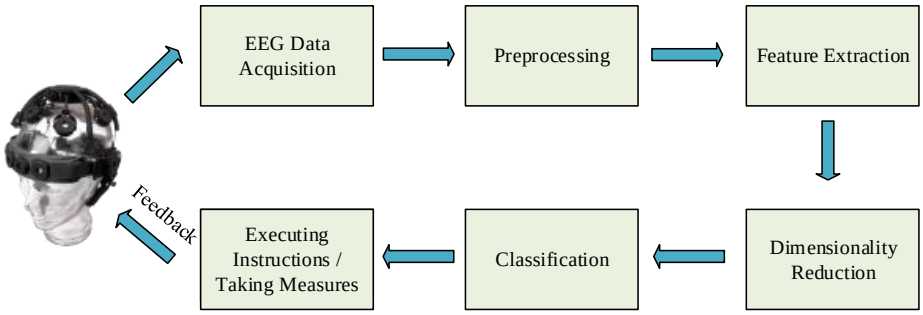


FIGURE 1: Human-centered wearable EEG in closed-loop system.

extent [2].

In recent years, the EEG systems is transforming from machine-centered conventional EEG to the human-centered wearable EEG, and the effective use of EEG data captured from wearable devices provides an effective way for the monitoring and self-management of human health. Due to its small size and portability, wearable EEG can perform real-time operation without restricting subjects' movements, thus facilitating health monitoring and self-management of health. The wearable EEG for health monitoring can be presented in a closed-loop system [1], [2], [9], [10], as shown in Fig. 1. In this system, human is considered to be a part of the whole closed-loop system. Through operations of EEG data acquisition, preprocessing and pattern recognition, the wearable EEG system automatically give appropriate feedback according to the predictive analytics result. For instance, when a seizure is likely to occur, wearable EEG epilepsy treatment closed-loop system can take preventative action via electrical neural stimulation or anti-epileptic drug release [2], [4].

In order to enhance the performance of wearable EEG monitoring systems, we need to develop the abilities in EEG data acquisition and EEG data processing. To achieve better performance, it is often to require increasing recording channels and adopting higher sampling frequencies for EEG data acquisition [2]. This of course produces larger volumes of EEG data with higher dimensions. It thus increases computational complexity and poses a great challenge to power consumption of wearable EEG devices. In the meanwhile, with voltage fluctuations at multiple channels over a period of time, it is commonly expected that the voltage values of the adjacent channels and time points are highly correlated for EEG [11], [12]. Therefore, it is imperative to reduce EEG data dimensions while preserving the latent information and discovering structural correlations of EEG data. One intuitive way of resolving this issue is to perform dimensionality reduction (DR) with structural correlations preserved.

In this paper, we propose a new regularized matrix discriminant analysis (R-MDA) method to tackle the issue of feature representation and DR in wearable EEG system. Extensive experiments have been performed in this study which demonstrate the performance of the new R-MDA on both synthetic data and EEG alcoholism data. The experiments show that the new proposed R-MDA approach has achieved competitive performance on feature representation and DR for EEG data. The key contributions of our work are summarized as follows.

- A new R-MDA approach has been proposed, based on wearable EEG data analytics, for human-centered health monitoring system.
- The latent structures and correlations of EEG data samples in matrix form is captured in this new method by considering the projection vector as matrix.
- By imposing specified constraint on each projection matrix, the correlations among the rows and columns of the projection matrix can be acquired effectively.
- The constraint and the least-square formulation is integrated in a unified model to solve the problem of singularity and expensive computational cost.

II. RELATED WORK

DR methods include principal components analysis (PCA), linear discriminant analysis (LDA), canonical correlation analysis (CCA) [13] andd so on. LDA is a commonly used classical DR method for dimensionality reduction and classification of EEG data [6], [9], [14], having achieved good results in EEG data processing. Classical LDA method, designed for data in vector form, formulates the optimization problem as an eigen-decomposition problem, which may lead to the undersampled or singularity problem in scatter matrices when there are limited available data samples [15], [16]. Even when the data satisfies the conditions for eigen-decomposition, it may suffer from expensive computational cost during the decomposition procedure [15], [17]. Additionally, as the voltage values of the adjacent channels and time points are highly correlated, there exist strong correlations among the rows and columns of the EEG matrix instead of feature covariance in the vector. The illustration of a typical EEG recordings is shown in Fig 2. Using traditional vector-based LDA methods to process EEG data, we have to reshape them into vectors, which may destroy the topological correlations and structures between different channels in

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

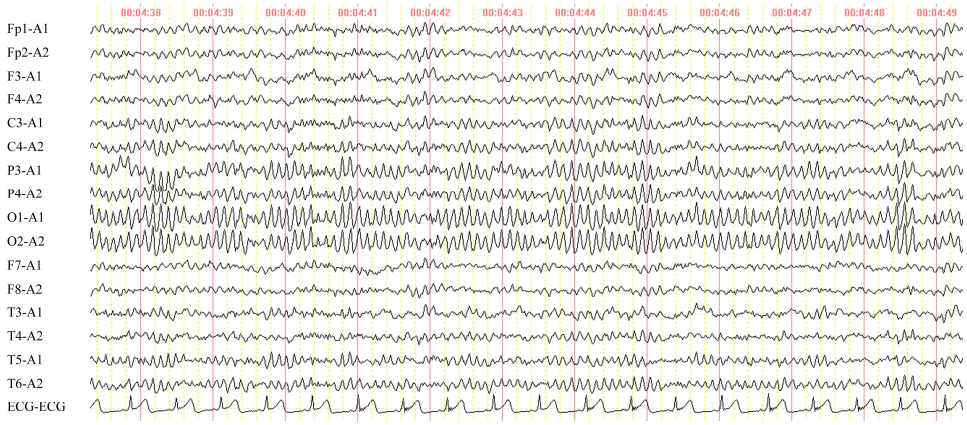


FIGURE 2: Illustration of a typical EEG recordings.

EEG data [15], [16]. To tackle this issue, several methods, especially matrix-based LDA and its improved models, have been proposed to perform dimensionality reduction for matrices directly.

Two-dimensional linear discriminant analysis (2DLDA) [15], [16] was proposed to work with data in matrix representation and can conquer the singularity problem. A Bayes optimal matrix-variate formulation of LDA based on a matrix-variate model (MVLDA) [18] was proposed for the spatio-spectral EEG patterns. Regularized MVLDA (R-MVLDA) [19] integrates vector-variate and matrix-variate approaches, and allows the estimated scatter matrices to adapt to the EEG data characteristics. Matrix-based LDA and row-column duality (D-MLDA) and penalized D-MLDA (D-MPDA) [20] were proposed for constructing and selecting of discriminant space-time-scale features of EEG. L2-norm 2DLDA (L2-2DLDA) and L1-norm 2DLDA (L1-2DLDA) [21] were proposed, which are more robust to outliers and noises than 2DLDA.

These LDA extension methods have been successfully used for feature representation and dimensionality reduction of EEG data, as they can consider the latent correctional information behind EEG data. However, there still exist several problems for these LDA extension methods. Firstly, most of the current methods do take into consideration that data samples themselves can be represented as matrices, but they ignore the possibility that the projection vector could be in matrix form, which means there may exist correlations in the columns or rows if we reshape projection vector into projection matrix. In certain circumstances, there do exist structural correlations among the columns and rows of the projection matrix, which should be taken into consideration in constructing the LDA models. Secondly, these matrix-based LDA methods consider the optimization problem as a two-dimensional eigen-decomposition problem. Although it can solve the singularity problem, it still suffers from expensive computational cost problem caused by eigen-decomposition. Hence, it is imperative to derive matrix discriminant analysis method to solve the above issues in a unified model.

In this paper, we propose a new regularized matrix discriminant analysis (R-MDA) method for extracting the EEG matrix data patterns, which not only can make use of the structural information of projection matrices, but also remove the singularity problem and the computational cost in eigen-decomposition. We firstly represent EEG data samples as data matrices, and in the meanwhile we represent the conventional project vectors as project matrices correspondingly. Taking into consideration the structural information of each projection matrix, we imposing specified constraint on each of them to capture the latent correctional information. We then formulate LDA as a least-square problem with a certain indicator matrix [17]. By considering the least-square formulation and imposing specified constraints together, we could form a unified model to solve the above issues.

III. PRELIMINARIES AND NOTATIONS

In this Section, the mathematical notations are given for introducing our method and corresponding numerical solver. Lower case letters (e.g., x), bold lower case letters (e.g., $\mathbf{x}$) and bold upper case letters (e.g., $\mathbf{X}$) are used to represent the scalar values, vectors and matrices, respectively. For a matrix $\mathbf{X} \in \mathbb{R}^{p\times q}$ of rank r where $r \leq \min(p, q)$, $\mathbf{X}_{ij}$ represents its (i, j)-entity. $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $||\mathbf{X}||_*$ is the nuclear norm of a matrix $\mathbf{X}$, where $||\mathbf{X}||_* = \sum_{i=1}^{n} \sigma_i$ (σ_i is the i^{th} singular value in matrix $\mathbf{X}$). For model training, the training dataset is presented with n samples $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the input vector and $y_i \in \{1, 2, \cdots, k\}$, $(k \geq 2)$ is the class label of the i-th sample. In multivariate linear regression (MLR), there is a k-tuple of separating functions $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_k(\mathbf{x}))$, for any $\mathbf{x} \in \mathbb{R}^d$. $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \cdots, \tilde{\mathbf{x}}_n] \in \mathbb{R}^{d\times n}$ is denoted as the centered data matrix $\mathbf{X}$, where $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ and $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_{ij}) \in \mathbb{R}^{n\times k}$ is denoted as the centered indicator matrix $\mathbf{Y}$, where $\tilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \bar{\mathbf{Y}}_j$ and $\bar{\mathbf{Y}}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_{ij}$. The weight vectors $\{\mathbf{w}_j\}_{j=1}^{k} \in \mathbb{R}^d$ is denoted to construct the k linear models, $f_j(\mathbf{x}) = \mathbf{x}^T\mathbf{w}_j$, for $j = 1, \cdots, k$.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

IV. REGULARIZED MATRIX DISCRIMINANT ANALYSIS

In this section, we analyze the issues mentioned in the introduction with more details and then derive the proposed R-MDA. An Alternating Direction Method of Multipliers (ADMM) algorithm is introduced for model training.

A. R-MDA MODEL

A EEG data with n samples, where each sample can be represented as $\mathbf{X}_i \in \mathbb{R}^{p\times q}$, for $i = 1, ..., n$. However, classical vector-based LDA methods [13] have to reshape each data matrix $\mathbf{X}_i$ to vector $\mathbf{x}_i$ by operation of $\text{vec}(\mathbf{X}_i) = \mathbf{x}_i \in \mathbb{R}^d$, where $d = p \times q$. These data vectors together can form a data matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$. With projection operation $\mathbf{X}^\top\mathbf{W}$, the original data vectors $\mathbf{x}_i$ are projected onto a low-dimensional subspace, where $\mathbf{W} \in \mathbb{R}^{d\times k}$, $(k < d)$ is projection matrix and the projected data vectors are represented as $\mathbf{x}_i^{'} \in \mathbb{R}^k$, for $i = 1, \cdots, n$. The way for classical methods to get projection vector $\mathbf{w}_j$ is to perform eigen-decomposition on certain matrices. For example, solving LDA can be equivalently formulated as follows

$$\arg\min_{\mathbf{w}} \frac{\mathbf{w}^\top\mathbf{S}_B\mathbf{w}}{\mathbf{w}^\top\mathbf{S}_W\mathbf{w}}, \tag{1}$$

where $\mathbf{w}$ is the projection vector, $\mathbf{S}_B$ is the scatter matrix for intra-class distance, and $\mathbf{S}_W$ is the scatter matrix for inter-class distance. The problem in Eq. (1) can be typically solved by eigen-decomposition on scatter matrices $\mathbf{S}_B$ and $\mathbf{S}_W$.

However, when data dimension is high, such as EEG alcoholism data with size of 256×64 , this approach would suffer from expensive computational cost on decomposing the matrix. Furthermore, when the number of the data samples is limited, the scatter matrix may be singular, which causes problem to eigen-decomposition. The over-fitting problem may also be inevitable in these cases.

In this study, we propose a unified model called regularized matrix discriminant analysis (R-MDA) to tackle the above issues. We represent data samples as their original matrix form $\mathbf{X}_i$, $(i = 1, \cdots, n)$ to retain the structural correlations. Correspondingly, each column of the conventional projection matrix $\mathbf{w}_j$ is reshaped to a matrix $\mathbf{W}_j$. Thus, the classical projection matrix is reshaped to a set of projection matrices stacking together, shown in Fig. 3, where $\mathbf{W}_j \in \mathbb{R}^{p\times q}$ and $j = 1, ..., k$. As the structural correlations within matrix can be captured by taking low-rank assumption on the matrix data [11], [12]. We impose low-rank regularization on each $\mathbf{W}_j$ to capture the correctional information behind the reshaped projection matrices. The illustration of our proposed R-MDA method is shown in Fig. 3 (b).

With the model configuration, we figure out that equivalently formulating the eigen-decomposition as a least-square problem [17] would be the best way. By specifically constructing an indicator matrix under a mild condition [17], we could derive the following objective function to replace the eigen-decomposition.

$$L(\mathbf{W}) = \frac{1}{2}\|\tilde{\mathbf{X}}^T\mathbf{W} - \tilde{\mathbf{Y}}\|_F^2 = \frac{1}{2}\sum_{j=1}^{k}\sum_{i=1}^{n}\|f_j(\tilde{\mathbf{x}}_i) - \tilde{\mathbf{Y}}_{ij}\|^2. \tag{2}$$

where $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_{ij}) \in \mathbb{R}^{n\times k}$ is a centered specific indicator matrix, which is defined as follows

$$\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_{ij}) = \begin{cases} \sqrt{\frac{n}{n_j}} - \sqrt{\frac{n_j}{n}} & \text{if } y_i = j, \\ -\sqrt{\frac{n_j}{n}} & \text{otherwise,} \end{cases} \tag{3}$$

where n is the number of total samples, n_j is the number of the j^{th} class's samples. With the above equations, the projection matrix $\mathbf{W} \in \mathbb{R}^{d\times k}$ can be computed by the following optimization problem

$$\mathbf{W} = \arg\min_{\mathbf{W}} \frac{1}{2}\sum_{j=1}^{k}\left(\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\mathbf{x}}_i^T\mathbf{w}_j - \tilde{\mathbf{Y}}_{ij}\|_2^2\right), \tag{4}$$

where $\tilde{\mathbf{x}}_i \in \mathbb{R}^d$ is the feature vector of the i^{th} sample, $\mathbf{w}_j$ is the j^{th} $(j = 1, 2, \cdots, k)$ column of the projection matrix $\mathbf{W}$ (the j^{th} projection vector), and $\tilde{\mathbf{Y}}_{ij}$ is a pre-specified value determined by Eq. (3). Equipped with Eq. (4), we could avoid computationally expensive eigen-decomposition and singularity of scatter matrix. More importantly, it provides an effective way to impose low-rank constraint on the projection matrix, which can help further capture the structures and correlations among the rows and columns. Specifically, these structures and correlations mean that there exist certain linear combinations among the rows and columns of the projection matrix. We reshape the feature vectors $\tilde{\mathbf{x}}_i \in \mathbb{R}^d$ into feature matrices $\tilde{\mathbf{X}}_i \in \mathbb{R}^{p\times q}$, where $p \times q = d$, $\tilde{\mathbf{x}}_i = \text{vec}(\tilde{\mathbf{X}}_i^T)$. In the meanwhile, each column of the conventional projection matrix ($\mathbf{w}_j \in \mathbb{R}^d$) is reshaped into a matrix $\mathbf{W}_j \in \mathbb{R}^{p\times q}$, where $p \times q = d$, $\mathbf{w}_j = \text{vec}(\mathbf{W}_j^T)$. The item $\tilde{\mathbf{x}}_i^T\mathbf{w}_j$ in Eq. (4) can be represented as $\tilde{\mathbf{x}}_i^T\mathbf{w}_j = \text{vec}(\tilde{\mathbf{X}}_i^T)^T\text{vec}(\mathbf{W}_j^T) = \text{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j)$. The nuclear norm [12], [23] has been shown to be the best convex approximation of the matrix rank over the unit ball of matrices, thus it is always employed to penalize the projection matrix, which is the known best approximation of low-rank constraint. We then impose a nuclear norm constraint on each $\mathbf{W}_j$ to capture the correlation of linear combinations. Therefore, the final objective function for our R-MDA can be formulated as follows

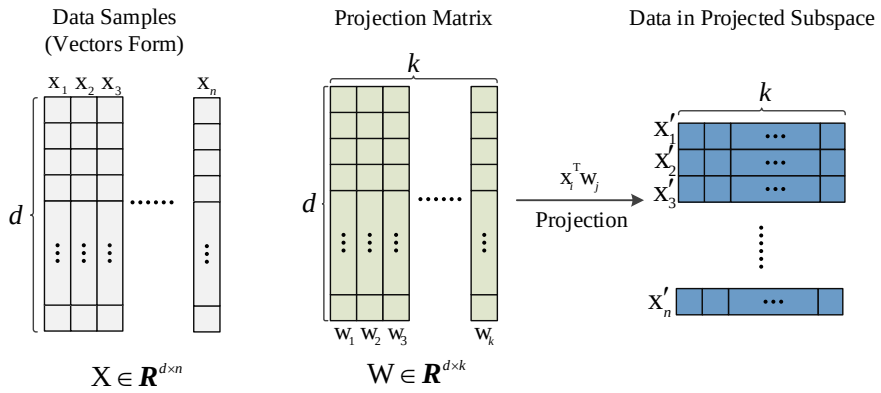
$$L(\mathbf{W}_{(1,2,\ldots,k)}) = \frac{1}{2n}\sum_{j=1}^{k}\sum_{i=1}^{n}\|\text{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j) - \tilde{\mathbf{Y}}_{ij}\|_2^2 + \tau\sum_{j=1}^{k}\|\mathbf{W}_j\|_*, \tag{5}$$

where τ is a hyper-parameter which is determined by cross-validation. In our R-MDA framework, we assume that each $\mathbf{W}_j$ $(j = 1, 2, \cdots, k)$ is independent with each other. Thus, we can solve the j^{th} sub-objective functions separately to obtain optimized parameters. For each $\mathbf{W}_j$, the sub-objective function can be formulated as follows

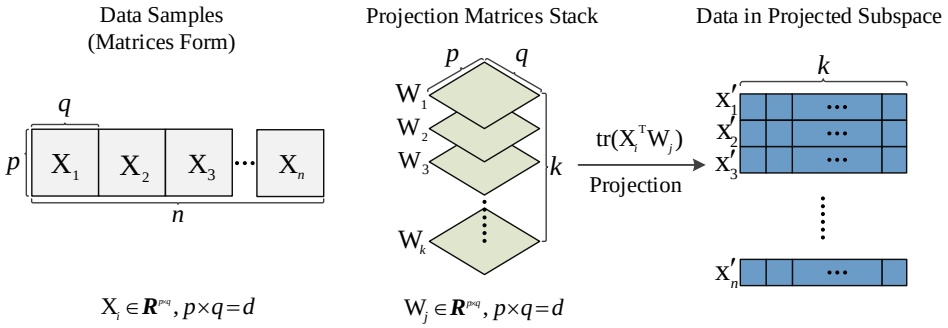
$$L(\mathbf{W}_j) = \frac{1}{2n}\sum_{i=1}^{n}\|\text{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j) - \tilde{\mathbf{Y}}_{ij}\|_2^2 + \tau\|\mathbf{W}_j\|_*. \tag{6}$$

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access



(a) The classical LDA approach



(b) The proposed R-MDA approach

FIGURE 3: The classical LDA approach and the proposed R-MDA approach.

B. SOLVER

The Alternating Direction Method of Multipliers (ADMM) [24] is adopted to solve the optimization problem. Firstly, we can equivalently write the Eq. (6) as follows

$$\begin{aligned} \arg\min_{\mathbf{W}_j,\mathbf{S}_j} \quad & H(\mathbf{W}_j)+G(\mathbf{S}_j), \\ \text{s.t.} \quad & \mathbf{W}_j-\mathbf{S}_j=\mathbf{0}, \end{aligned} \tag{7}$$

where $H(\mathbf{W}_j) = \frac{1}{2n}\sum_{i=1}^{n}\|\mathrm{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j)-\tilde{\mathbf{Y}}_{ij}\|_2^2$, and $G(\mathbf{S}_j)=\tau\|\mathbf{S}_j\|_*$. We then apply augmented Lagrangian multiplier on Eq. (7), and the objective function is derived as follows

$$\begin{aligned} L(\mathbf{W}_j,\mathbf{S}_j,\mathbf{\Lambda}_j) =& \frac{1}{2n}\sum_{i=1}^{n}\|\mathrm{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j)-\tilde{\mathbf{Y}}_{ij}\|_2^2+\tau\|\mathbf{S}_j\|_* \\ &+\mathrm{tr}[\mathbf{\Lambda}_j^T(\mathbf{S}_j-\mathbf{W}_j)]+\frac{\rho}{2}\|\mathbf{S}_j-\mathbf{W}_j\|_F^2, \end{aligned} \tag{8}$$

where $\mathbf{\Lambda}_j$ is a Lagrangian multiplier matrix and $\rho > 0$ is a hyper-parameter. The optimization problem in Eq. (8) can be divided into two sub-problems in terms of $\mathbf{W}_j$ and the auxiliary variable $\mathbf{S}_j$.

1) OPTIMIZATION FOR AUXILIARY VARIABLE $\mathbf{S}_j$

The first sub-problem for solving $\mathbf{S}_j$ can be written in the following equivalent way

$$\mathbf{S}_j=\arg\min_{\mathbf{S}_j}\tau\|\mathbf{S}_j\|_*+\mathrm{tr}(\mathbf{\Lambda}_j^T\mathbf{S}_j)+\frac{\rho}{2}\|\mathbf{S}_j-\mathbf{W}_j\|_F^2, \tag{9}$$

with the help of the shrinkage thresholding (SVT) [25] operator, we can get analytical solution of $\mathbf{S}_j^{(t)}$ in the t-th iteration as follows

$$\mathbf{S}_j^{(t)}=\frac{1}{\rho}\mathcal{D}_\tau(\rho\mathbf{W}_j^{(t)}-\mathbf{\Lambda}_j^{(t)}), \tag{10}$$

where $\rho\mathbf{W}_j-\mathbf{\Lambda}_j=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is a Singular Value Decomposition (SVD) form. For any $\tau>0$, $\mathcal{D}_\tau(\cdot)=\mathbf{U}S_\tau(\mathbf{\Sigma})\mathbf{V}^T$ is called the singular value thresholding (SVT) operator, where $S_\tau(\mathbf{\Sigma})_{ii}=\max(\mathbf{\Sigma}_{ii}-\tau,0)$.

2) OPTIMIZATION FOR THE PROJECTION MATRIX $\mathbf{W}_j$

The second sub-problem for solving $\mathbf{W}_j$, the expression can be equivalently written as follows

$$\begin{aligned} \mathbf{W}_j =& \arg\min_{\mathbf{W}_j}\frac{1}{2n}\sum_{i=1}^{n}\|\mathrm{tr}(\tilde{\mathbf{X}}_i^T\mathbf{W}_j)-\tilde{\mathbf{Y}}_{ij}\|_2^2 \\ &-\mathrm{tr}(\mathbf{\Lambda}_j^T\mathbf{W}_j)+\frac{\rho}{2}\|\mathbf{S}_j-\mathbf{W}_j\|_F^2. \end{aligned} \tag{11}$$

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

The problem in Eq. (11) is a well-defined convex optimization problem, and here we resort to gradient descent to solve it. In each iteration, we can perform the following operation to update $\mathbf{W}_j$

$$\mathbf{W}_j := \mathbf{W}_j - \alpha \left(\frac{\partial L_{\mathbf{W}_j}}{\partial \mathbf{W}_j} \right), \tag{12}$$

where $\alpha > 0$ is a hyper-parameter, called learning rate. The partial derivation of $\mathbf{W}_j$ is formulated as

$$\frac{\partial L_{\mathbf{W}_j}}{\partial \mathbf{W}_j} = \frac{1}{n} \sum_{i=1}^{n} [\mathrm{tr}(\tilde{\mathbf{X}}_i^T \mathbf{W}_j) - \tilde{\mathbf{Y}}_{ij}] \tilde{\mathbf{X}}_i - \mathbf{\Lambda}_j + \rho(\mathbf{W}_j - \mathbf{S}_j). \tag{13}$$

The main steps for solving the optimization problem in Eq. (8) have been presented, which solves $\mathbf{W}_j$ and $\mathbf{S}_j$ iteratively. Additionally, the Lagrangian parameter $\mathbf{\Lambda}_j$ is updated in a single gradient step as follows

$$\mathbf{\Lambda}_j = \widehat{\mathbf{\Lambda}}_j^{(t)} - \rho(\mathbf{W}_j^{(t)} - \mathbf{S}_j^{(t)}). \tag{14}$$

The whole flow is summarized in Algorithm 1, where ADMM is applied here.

Algorithm 1 ADMM for Problem in Eq. (8)

1: **for** $j = 1, 2, 3..., k$ **do**
2: Initialize $\mathbf{S}_j^{(-1)} = \widehat{\mathbf{S}}_j^{(0)} \in \mathbb{R}^{p \times q}$, $\mathbf{\Lambda}_j^{(-1)} = \widehat{\mathbf{\Lambda}}_j \in \mathbb{R}^{p \times q}$, $\rho > 0$.
3: **for** $t = 1, 2, 3...$ **do**
4: $\mathbf{W}_j^{(t)} = \arg\min_{\mathbf{W}_j} \frac{1}{2n} \sum_{i=1}^{n} \|\mathrm{tr}(\tilde{\mathbf{X}}_i^T \mathbf{W}_j) - \tilde{\mathbf{Y}}_{ij}\|_2^2 - \mathrm{tr}(\mathbf{\Lambda}_j^T \mathbf{W}_j) + \frac{\rho}{2} \|\mathbf{S}_j - \mathbf{W}_j\|_F^2$
5: $\mathbf{S}_j^{(t)} = \arg\min_{\mathbf{S}_j} \tau \|\mathbf{S}_j\|_* + \mathrm{tr}(\mathbf{\Lambda}_j^T \mathbf{S}_j) + \frac{\rho}{2} \|\mathbf{S}_j - \mathbf{W}_j\|_F^2$
6: $\mathbf{\Lambda}_j^{(t)} = \widehat{\mathbf{\Lambda}}_j^{(t)} - \rho(\mathbf{W}_j^{(t)} - \mathbf{S}_j^{(t)})$
7: **end for**
8: **end for**

V. EXPERIMENTS

In order to verify the performance of our proposed method, we conduct comparative studies. The proposed R-MDA method is implemented by Matlab R2015b in a machine with Inter(R)Core(TM)i5-7500 3.40GHz CPU and 8GB RAM. We study the performance of our R-MDA with the comparison of three wide-used methods: 1) Least Square Linear Discriminant Analysis (LS-LDA) [17]; 2) Two-Dimensional Linear Discriminant Analysis (2DLDA) [15]. 3)Robust L1-norm two-dimensional linear discriminant analysis (L12DLDA) [21].

To evaluate and compare the performance, we apply these methods to both synthetic dataset and EEG alcoholism dataset. We firstly conduct experiments on synthetic low-rank data to verify our model is effective on low-rank data. We consider the illustrative examples by examining different kinds of geometric and natural shapes on the regression matrix. Secondly, we apply our R-MDA on the EEG alcoholism dataset, where each sample can be represented as a matrix with size 256×64.

A. SYNTHETIC DATA ANALYSIS

In order to verify that R-MDA is effective on low-rank EEG data, we firstly conduct experiments on synthetic low-rank data. We start with elaborating the illustrative examples by examining three different signal shapes (Square, Triangle and Butterfly), where each of them is represented as a 64×64 signal matrix. The illustration of three signal shapes (Square,Triangle and Butterfly) are shown in Fig. 4.

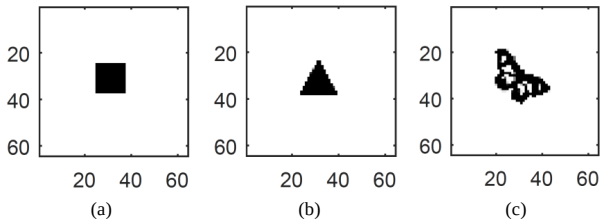


FIGURE 4: Three signal shapes, respective of Square, Triangle and Butterfly.

Based on the signal shapes shown in Fig. 4, we randomly generate 1000 samples for 10 rounds. We firstly randomly generate 10 matrices with the size of 64×64, where each matrix corresponds to one type of class. After duplicating 100 matrices for each type of class, we add random Gaussian noise (The probability distribution function is $P(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{-(x-\mu)^2}{2\sigma^2})$) on all 1000 matrices, after which we can get $\mathbf{P}_i$ matrices, $(i = 1, 2, \cdots, 1000)$. Hence, the i^{th} sample can be generated by the following equation

$$\mathbf{X}_i = \mathbf{S}^\top \mathbf{P}_i, \tag{15}$$

where $\mathbf{S}$ is the regression matrix illustrated by a signal shape, shown in Fig. 4. $(\mathbf{X}_i, y_i)$ is a randomly generated illustrative sample, where $y_i \in \{1, 2, \cdots, 10\}$. In each round, half of the samples are used for training and the rest are for testing. We apply support vector machine (SVM), which is considered as one of the powerful linear classifiers, to classify the projected samples. We then compute the mean and standard deviation of accuracies in 10 rounds on regression matrix $\mathbf{S}$ for each approach. In the experiments, all the hyper parameters are chosen via cross validation. The detailed comparison of our R-MDA with other methods are shown in Table. 1.

As shown in the Table. 1, it can be observed that the proposed R-MDA method outperforms LS-LDA, 2DLDA and L12DLDA remarkably. Though LS-LDA can tackle the undersampled problem, it treats each sample as a vector, resulting in the loss of structural information. 2DLDA and L12DLDA claim that they treat each sample as a matrix directly; however, they do not consider each projection vector itself could be a matrix, which leads to information loss during training. Although L12DLDA usually deemed as an effective way to deal with outliers and noises, it fails to capture the latent correctional information behind low-rank matrix. Particularly, R-MDA can keep the accuracy around 90% with only 10 dimensional features, while 2DLDA and

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

TABLE 1: The accuracies on synthetic data.

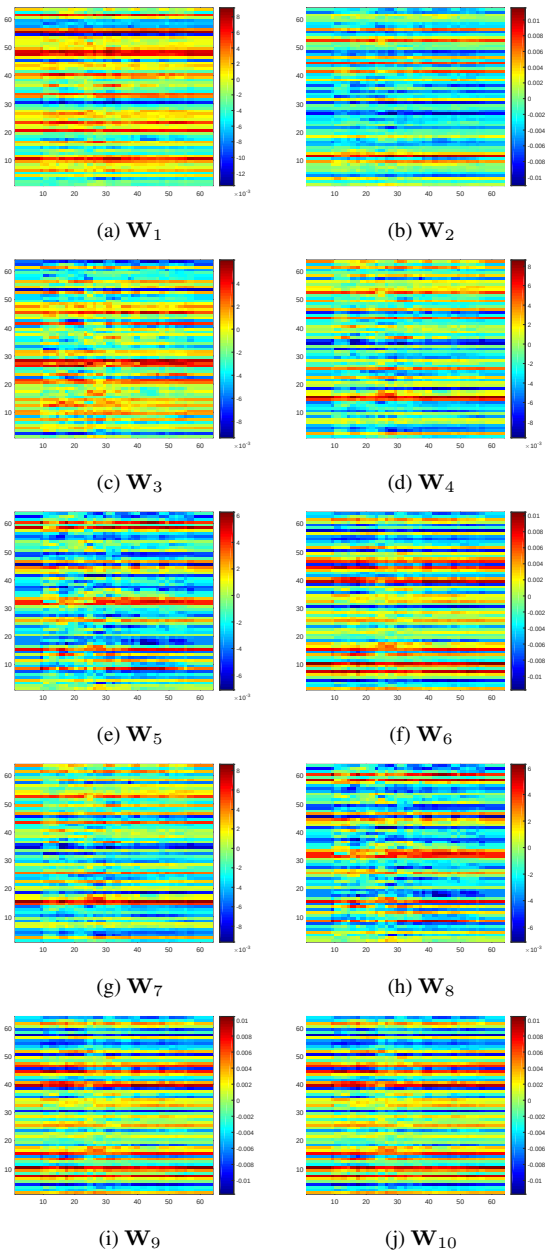
Regression Matrix	SVM		LS-LDA+SVM [17]		2DLDA+SVM [15]		L12DLDA+SVM(L12DLDA) [21]		Ours(RMDA+SVM)	
S	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions
Square	0.4124 ± 0.0977	4096	0.0808 ± 0.0465	**10**	0.6200 ± 0.2253	640	0.5962 ± 0.2096	640	**0.9140**± 0.0886	**10**
Triangle	0.5524 ± 0.1773	4096	0.1108 ± 0.0106	**10**	0.7652 ± 0.1924	640	0.6674 ± 0.2452	640	**0.8844**± 0.0828	**10**
Butterfly	0.5146 ± 0.2361	4096	0.0836 ± 0.0361	**10**	0.8162 ± 0.2405	640	0.6808 ± 0.2541	640	**0.9222**± 0.0645	**10**

L12DLDA need $p \times k = 640$ (p is the height of illustrative samples, where $p = 64$ and $k = 10$) dimensional features to achieve only around 80% and 70% accuracies, respectively. Compared with 2DLDA and L12DLDA, the proposed R-MDA can achieve higher accuracy with lower feature dimensions. Additionally, in R-MDA, each projection vector $\mathbf{w}_j \in \mathbb{R}^{4096}$ is reshaped into a projection matrix $\mathbf{W}_j \in \mathbb{R}^{64\times 64}$. Thus the conventional projection matrix is represented as a set of projection matrices stacking together. The number of projection matrices is determined by the number of categories, namely $k = 10$. The visualized projection matrices stack $\mathbf{W}_j$, $(j = 1, 2, \cdots, 10)$ in the first round experiment based on the Butterfly signal shape (an visualization example of projection matrices stack) are given in Fig. 5. We can observe from Fig. 5 that the colors among adjacent rows and columns are very close in each $\mathbf{W}_j$, which means there do exist strong structural correlations among rows and columns of projection matrices. By imposing low-rank regularization on each projection matrix, R-MDA can obtain the correctional information.

B. EEG DATA ANALYSIS

We further evaluate R-MDA on EEG alcoholism data, and conduct comparison study. We directly adopt EEG alcoholism dataset [1], which arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It consists of two groups of subjects: alcoholic and control. For each subject, 64 channels of electrodes are placed to record voltage values at 256 time points (sampled at 256 Hz, 3.9-msec epoch) per second in three different matching conditions. All the data can be presented as matrices with the size of 256×64. The example plots of an alcoholic and control subject are visualized in Fig. 6. The plots indicate voltage, time, and channel and are averaged over 10 runs for the single stimulus condition. The version of the EEG alcoholism dataset we choosing is the Large Data Set. The training dataset contains 10 alcoholic and 10 control subjects, getting 10 runs per subject per condition. The testing dataset uses the same subjects, but with 10-out-of-samples running per subject per condition. So for both training and testing dataset, there exist 600 samples (300 alcoholic samples and 300 control samples), respectively.

We conduct experiments for 10 rounds to verify the stability and effectiveness of the proposed method. In order to show the advantages of our method in facing undersam-



(a) $\mathbf{W}_1$ (b) $\mathbf{W}_2$ (c) $\mathbf{W}_3$ (d) $\mathbf{W}_4$ (e) $\mathbf{W}_5$ (f) $\mathbf{W}_6$ (g) $\mathbf{W}_7$ (h) $\mathbf{W}_8$ (i) $\mathbf{W}_9$ (j) $\mathbf{W}_{10}$

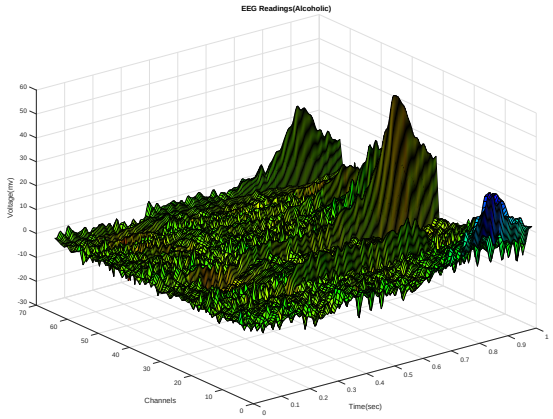
FIGURE 5: Visualizations of the projection matrices stack $\mathbf{W}_j$ in the first round experiment based on the Butterfly signal shape.

[1] http://kdd.ics.uci.edu/databases/eeg/eeg.html

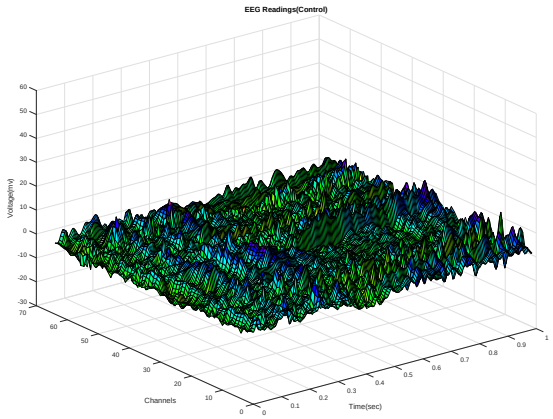
2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

pled problem, we randomly choose 300, 200, 150, 100, 50 samples from the training dataset for training and all data from the testing dataset for testing in each round. We then compute the mean and standard deviation of accuracies in 10 rounds on different training settings for each approach. In the experiments, all the hyper parameters are selected via cross validation. The experimental results are shown in Table. 2. And the line chart of accuracies on EEG alcoholism data for different approaches is shown in Fig. 7.



(a) EEG readings (alcoholic)



(b) EEG readings (control)

FIGURE 6: The example plots of an alcoholic and control subject.

As shown in Table. 2, the R-MDA not only reduces the feature dimensions to an extreme low level, only 2 dimensional features, but also achieves the best classification accuracies on EEG alcoholism dataset for all different training settings. LS-LDA reduces data dimensions to 2 as well, but it reshapes EEG matrices into vectors. It destroys the correlations behind data, resulting in the loss of structural information. Otherwise, both 2DLDA and L12DLDA project EEG samples from 16384 (256×64) dimensional space to 512 (256×2) dimensional subspace. Compared with 2DLDA and L12DLDA, our R-MDA can achieve the highest accuracy with lowest dimensional features in different training settings. Also, with efficient capture of the correlations among the rows and columns of the projection matrix, our method can restrict the parameter searching space, which allows R-MDA to get superior performance among these methods even when there are only 50 data samples for training. From the Fig. 7, it can be observed that with a decrease in the number of training samples, the superiority of our proposed R-MDA method becomes more remarkable.

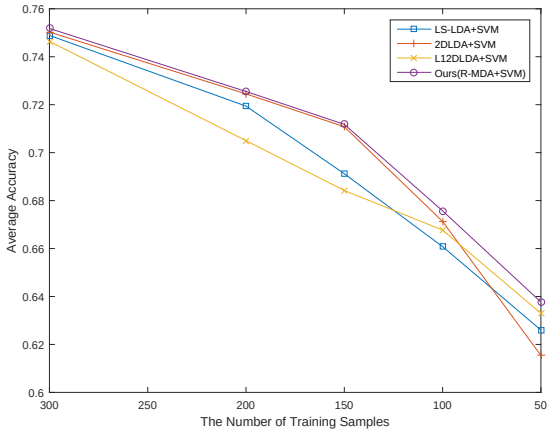


FIGURE 7: The line chart of accuracies on EEG alcoholism data for different approaches.

VI. CONCLUSION

To deal with the contradiction between increasing volumes of EEG data and power consumption in human-centered wearable EEG systems, we propose a new regularized matrix discriminant analysis (R-MDA) method to tackle EEG feature representation and DR problem. It is showed that the proposed R-MDA provides an efficient and effective way to process datasets with EEG data. It can avoid the computational expensive eigen-decomposition and the undersampled problem. In addition, the R-MDA method is capable of capturing the correlations among the rows and columns of the projection matrices efficiently , which is crucial for extraction of structural information of EEG data. The experimental results of applying R-MDA , respectively, on synthetic dataset and EEG alcoholism dataset show that the new R-MDA has achieved competitive performance for processing EEG data with improved classification accuracy and significantly improved dimensionality reduction. In the future study, we will investigate the possible correlations among the projection matrices, and construct regularized tensor discriminant analysis model based on tensor minimization for EEG signal patterns.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

TABLE 2: The accuracies on EEG alcoholism data.

Training Samples	LS-LDA+SVM [17]		2DLDA+SVM [15]		L12DLDA+SVM [21]		Ours(RMDA+SVM)	
	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions	Accuracy	Projected Samples Dimensions
300	0.7489 ± 0.0103	**2**	0.7503 ± 0.0087	512	0.7464 ± 0.0230	512	**0.7518**± 0.0093	**2**
200	0.7194 ± 0.0183	**2**	0.7245 ± 0.0118	512	0.7050 ± 0.0137	512	**0.7255**± 0.0158	**2**
150	0.6912 ± 0.0170	**2**	0.7108 ± 0.0149	512	0.6842 ± 0.0186	512	**0.7118**± 0.0131	**2**
100	0.6608 ± 0.0256	**2**	0.6713 ± 0.1924	512	0.6677 ± 0.0258	512	**0.6757**± 0.0234	**2**
50	0.6260 ± 0.0294	**2**	0.6154 ± 0.0334	512	0.6330 ± 0.0312	512	**0.6377**± 0.0247	**2**

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61471248), the Chengdu Science and Technology Project (2016-XT00-00015-GX), the Scientific Research Project of Sichuan Education Department (18ZB0355) and the Technology Program of Public Wellbeing of Chengdu (2015-HM01-00293-SF).

REFERENCES

[1] A. J. Casson, S. Smith, J. S. Duncan, et al, "Wearable EEG: what is it, why is it needed and what does it entail?," in Engineering in Medicine and Biology Society (EMBS), 30th Annual International Conference of the IEEE. IEEE, 2008

[2] A. J. Casson, D. C. Yates, S. J. M. Smith, et al, "Wearable electroencephalography," IEEE Eng. Med. Biol. Mag., vol. 29, no. 3, pp. 44–56, May, 2010, DOI: 10.1109/MEMB.2010.936545.

[3] L.-D. Liao, C.-Y. Chen, I.-J. Wang, et al, "Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors," J. Neuroeng. Rehabil., vol. 9, no. 1, Jan., 2012, DOI: 10.1186/1743-0003-9-5.

[4] B. K. Karumuri, A. Gupta, R. Liu, et al, "Classification of pre-ictal and interictal periods based on EEG frequency features in epilepsy," in Biomedical Engineering Conference (SBEC), 2016 32nd Southern. IEEE, 2016

[5] M. Murugappan, R. Nagarajan, S. Yaacob, "Comparison of different wavelet features from EEG signals for classifying human emotions," in Industrial Electronics and Applications (ISIEA), 2009 IEEE Symposium on. IEEE, 2009

[6] A. Bhardwaj, I. Vlachos, P. Jain, et al, "Classification of human emotions from EEG signals using SVM and LDA classifiers," in Signal Processing and Integrated Networks (SPIN), 2015 2nd International Conference on. IEEE, 2015

[7] S.-L. Wu, Y.-T. Liu, K.-P. Chou, et al, "A motor imagery based brain-computer interface system via swarm-optimized fuzzy integral and its application," in Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on. IEEE, 2016

[8] C.-T. Lin, C.-H. Chuang, C.-S. Huang, et al, "Wireless and wearable EEG system for evaluating driver vigilance," IEEE Trans. Biomed. Circuits Syst., vol. 8, no. 2, pp. 165–176, May, 2014, DOI: 10.1109/TBCAS.2014.2316224.

[9] R. Xu, N. Jiang, C. Lin, et al, "Enhanced low-latency detection of motor intention from EEG for closed-Loop brain-computer interface applications," IEEE Trans. Biomed. Eng., vol. 61, no. 2, pp. 288–296, Feb., 2014, DOI: 10.1109/TBME.2013.2294203.

[10] W. He, Y. Zhao, H.-Y Tang , et al, "A wireless BCI and BMI system for wearable robots," IEEE Trans. Syst. Man Cybern., Syst., vol. 46, no. 7, pp. 936–946, Jul., 2016, DOI: 10.1109/TSMC.2015.2506618 .

[11] H. Zhou and L.-X. Li, "Regularized matrix regression," J. R. Stat. Soc. Series B Stat. Methodol., vol. 76, no. 2, pp. 463–483, Mar., 2014, DOI: 10.1111/rssb.12031.

[12] L. Luo, Y.-B. Xie, Z.-H. Zhang, et al, "Support matrix machines," in International Conference on Machine Learning (ICML), ICML'15 Proceedings of the 32nd International Conference on. PMLR, 2015

[13] R. O. Duda, P. E. Hart, D. G. Stork, "Fisher linear discriminant," in Pattern classification, 2nd ed. New York, NY, USA: John Wiley & Sons, 2001.

[14] A. Subasi, M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," Expert Syst. Appl., vol. 37, no. 12, pp. 8659–8666, Dec., 2010, DOI: 10.1016/j.eswa.2010.06.065.

[15] J.-P. Ye, R. Janardan, Q. Li, "Two-dimensional linear discriminant analysis," in Advances in Neural Information Processing Systems 17 (NIPS), Neural Information Processing Systems. DBLP, 2004

[16] W.-S. Zheng, J. H. Lai, S. Z. Lin, "1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis better than matrix-based?," Pattern Recogn., vol. 41, no. 7, pp. 2156–2172, Jul., 2008, DOI: 10.1016/j.patcog.2007.11.025.

[17] J.-P. Ye, "Least squares linear discriminant analysis," in International conference on Machine learning (ICML), ICML '07 Proceedings of the 24th International Conference on. PMLR, 2007

[18] M. S. Mahanta, A. S. Aghaei, K. N. Plataniotis, "A Bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals," in Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference on. IEEE, 2012

[19] M. S. Mahanta, A. S. Aghaei, K. N. Plataniotis, "Regularized LDA based on separable scatter matrices for classification of spatio-spectral EEG patterns," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013

[20] J. Spinnato, M.-C. Roubaud, B. Burle, et al, "Finding EEG space-time-scale localized features using Matrix-based penalized discriminant analysis," in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014

[21] C.-N. Li, Y.-H. Shao, N.-Y. Deng, "Robust L1-norm two-dimensional linear discriminant analysis," Neural Networks, vol. 65, no. C, pp. 92–104, May, 2015, DOI: 10.1016/j.neunet.2015.01.003.

[22] A. Majumdar, A. Gogna, R. Ward, "A low-rank matrix recovery approach for energy efficient EEG acquisition for a wireless body area network," Sensors, vol. 14, no. 9, pp. 15729–15748, Sep., 2014, DOI: 10.3390/s1409157295.

[23] F. R. Bach, "Consistency of trace norm minimization," J. Mach. Learn Res., vol. 9, pp. 1019–1048, Jun., 2008.

[24] B. He, X-.M,. Yuan, "On non-ergodic convergence rate of couglas–rachford alternating direction method of multipliers," Numer. Math., vol. 130, no. 3, pp. 567–577, Jul., 2015.

[25] J.-F. Cai, E. J. candès, Z.-W. Shen, "A singular value thresholding algorithm for matrix completion," SIAM J. Optimiz., vol. 20, no. 4, pp. 1956–1982, Jan., 2010, DOI: 10.1137/080738970.

JIE SU received her B.S. degree in Electronic Science and Technology from Chengdu University of Information Technology in 2012. She received her M.S. degree in Signal and Information Processing from Sichuan University in 2015. She is currently pursuing her Ph.D. degree in Communication and Information System with Sichuan University, Chengdu, China. Her research interests include image processing, machine learning, pattern recognition, computer vision and deep learning.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2803806, IEEE Access

LINBO QING (M'16) is an associate professor with the College of Electronics and Information Engineering at Sichuan University. He received his B.S. degree in Electronic Information Science and Technology from Sichuan University, China in 2003. In 2008, he received his Ph.D. degree in Communication and Information System from the same university. His main research interests include image processing, video coding and transmission, and information theory. He is the corresponding author of this paper.

XIAOHAI HE (M'16) received his B.S. and M.S. degrees in electrical engineering from Sichuan University, Chengdu, China, in 1985 and 1991 respectively. In 2002, he received his Ph.D. degree in biomedical engineering from the same university. He is currently a professor with the College of Electronics and Information Engineering, at Sichuan University. His research interests include image processing, pattern recognition, computer vision, image communication, and software engineering. He is a Senior Member of the Chinese Institute of Electronics. He is an editor of the Journal of Information and Electronic Engineering and an editor of the Journal of Data Acquisition & Processing.

HANG ZHANG received his B.Eng degree from School of Electronics & Information Engineering at Sichuan University in 2015. He received his M.Phil. degree from Department of Computer Science & Engineering at The Chinese University of Hong Kong in 2017. He is currently pursuing his Ph.D. degree in School of Electrical and Computer Engineering at Cornell University. His research interests include theories and algorithms of deep learning, machine learning and their applications in robotics, autonomous systems and autonomous driving. He is a recipient of Best Paper Award at the International Symposium on Physical Design (ISPD) 2017. He is a recipient of First Place Award in CAD contest at the International Conference on Computer-aided Design (ICCAD) 2015 and Second Place Award in CAD contest at the ISPD 2016. He is also a recipient of Cornell Graduate School Fellowship.

JING ZHOU is a lecturer with the Biomedical Engieering at South China University of Technlogy. In 2015, she received her Ph.D. degree in Biomedical Engineering from the same university. Her main research interests include detecting and processing of physiological signals and study on sleep medicine.

YONGHONG PENG (M'02) is a Professor of Data Science and the leader for Data Science Research at the University of Sunderland, United Kingdom. His research areas include Data Science, Machine Learning, Data Mining and Artificial Intelligence. He is the Chair for the Big Data Task Force (BDTF), and a member of Data Mining and Big Data Analytics Technical Committee of IEEE computational intelligence society (CIS). He is also a founding member of the Technical Committee on Big Data (TCBD) of IEEE Communications and an advisory board member for IEEE Special Interest Group (SIG) on Big Data for Cyber Security and Privacy. Prof Peng is an Associate Editor for IEEE Transaction on Big Data, and an Academic Editor of PeerJ and PeerJ Computer Science.

2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.