



Noise Robustness Low-Rank Learning Algorithm for Electroencephalogram Signal Classification

Ming Gao¹, Runmin Liu² and Jie Mao^{2*}

¹ College of Sports Science and Technology, Wuhan Sports University, Wuhan, China, ² College of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, China

OPEN ACCESS

Edited by:

Kaijian Xia,
Changshu No.1 People's Hospital,
China

Reviewed by:

Yufeng Yao,
Changshu Institute of Technology,
China

Tongguang Ni,
Changzhou University, China

*Correspondence:

Jie Mao
maojie@whsu.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 18 October 2021

Accepted: 05 November 2021

Published: 24 November 2021

Citation:

Gao M, Liu R and Mao J (2021)
Noise Robustness Low-Rank
Learning Algorithm for
Electroencephalogram Signal
Classification.
Front. Neurosci. 15:797378.
doi: 10.3389/fnins.2021.797378

Electroencephalogram (EEG) is often used in clinical epilepsy treatment to monitor electrical signal changes in the brain of patients with epilepsy. With the development of signal processing and artificial intelligence technology, artificial intelligence classification method plays an important role in the automatic recognition of epilepsy EEG signals. However, traditional classifiers are easily affected by impurities and noise in epileptic EEG signals. To solve this problem, this paper develops a noise robustness low-rank learning (NRLRL) algorithm for EEG signal classification. NRLRL establishes a low-rank subspace to connect the original data space and label space. Making full use of supervision information, it considers the local information preservation of samples to ensure the low-rank representation of within-class compactness and between-classes dispersion. The asymmetric least squares support vector machine (aLS-SVM) is embedded into the objective function of NRLRL. The aLS-SVM finds the maximum quantile distance between the two classes of samples based on the pinball loss function, which further improves the noise robustness of the model. Several classification experiments with different noise intensity are designed on the Bonn data set, and the experiment results verify the effectiveness of the NRLRL algorithm.

Keywords: electroencephalogram, epilepsy, noise robustness, low-rank learning, pinball loss function

INTRODUCTION

Brain computer interface (BCI) is a system that collects the signals from the brain to communicate with computers or other devices (Gummadavelli et al., 2018; Jiang et al., 2020). As an efficient way for the human brain to directly communicate with peripheral devices, the BCI does not need to rely on the peripheral nervous system and muscles. Electroencephalogram (EEG) signals, as a biomarker, play an important role in BCI. EEG is often used in clinical diagnosis to determine the presence and type of epilepsy (Fahimi et al., 2019; Jiang et al., 2019). The epileptic seizure process has several different periods: interictal, pre-seizure, and seizure. The waveform, frequency, and signal characteristics of different stages are different in EEG. Based on the analysis of the characteristics of epilepsy EEG, many studies can generally be divided into two directions: epilepsy detection and epilepsy prediction (Jiang et al., 2017; Gu et al., 2021). The epilepsy detection algorithm uses signal processing, machine learning, and deep learning to extract signal features, and distinguishes the EEG signals between the interictal period and the seizure period. The epilepsy prediction algorithm distinguishes the EEG signals in the pre-seizure period and the seizure period.

The prediction task is more difficult than the detection task. First of all, there is no uniform definition of epilepsy prediction and internal standards in the industry. Secondly, compared with the EEG signal in the seizure period, the signal pattern of the EEG signals in the pre-seizure period and the EEG signals in the intermittent period are more similar, so the algorithm is required to be more robust.

Both epilepsy detection and epilepsy prediction are essentially classification tasks in machine learning (Ni et al., 2020). Several studies focus on the classification of EEG information, which involve both epilepsy detection tasks and epilepsy prediction tasks. Zhou et al. (2018) represented the epilepsy EEG signal into a two-dimensional image, and then they constructed a convolutional neural network to automatically learn the transformed image. This method borrows the idea of image processing to analyze the EEG signals and broaden the method of signal processing. Birjandtalab et al. (2017) performed non-linear dimensionality reduction on EEG signals after extracting time-frequency features. This feature processing method can reflect the non-linear relationship of the data in the process of low-dimensional mapping. Ramakrishnan and Muthanantha (2018) computed the approximate entropy value, the maximum Lyapunov component, and the correlation coefficient dimension on different sub-bands of epilepsy EEG signals, and introduced fuzzy rules to fuzzify the features. The authors believe that fuzzy rules are the natural choice of using human professional knowledge to build machine learning systems, which is closely related to people's way of thinking. Wang et al. (2017) explored multiple bands of the EEG signal by considering the maximum and standard deviation characteristics of each band. Then they constructed the feature vector of the EEG and used a one-to-one self-organization strategy to create a high high-precision epilepsy detection system. Sun et al. (2019) intercepted and analyzed the pre-seizure data, and they used recurrent autoencoders on multivariate signals to extracted EEG features. Liu et al. (2019) transformed the EEG signal into spectral data through the combination of dimensionality reduction and short-time Fourier transform. Then, the authors constructed a shallow convolutional neural network (CNN) network to automatically learn data features. Yu et al. (2020) used the local mean decomposition (LMD) method to obtain the feature matrix of the EEG signals, and they used a CNN model to implement feature extraction and combined Bayesian linear discriminant analysis to obtain the prediction result.

In supervised learning, the support vector machine (SVM) represented by least squares regression (LSR) is a simple and effective method. The core idea of LSR is to learn the non-linear projection from the original data to the feature space, and the obtained projection vector of the original data is also used as the data representation in the label space. For example, discriminative LSR method include multiclass classification (Xiang et al., 2012), groupwise retargeted LSR method (Ling and Geng, 2019), regularized label relaxation linear regression (Fang et al., 2018), double relaxed regression for classification (Han et al., 2019), and so on. For epilepsy data, the scalp EEG data will have more impurity signals and noise signals. Moreover, dimensional explosion and information

redundancy problems are common in EEG signals. Learning a discriminatively compact data representation is a very critical problem in pattern recognition. At present, there are many methods based on subspace learning and least squares classifier to learn good classifiers. For example, to combine projection learning with the task of exploring label information, Meng et al. (2020) proposed a constrained discriminative projection learning for joint optimization of subspace learning and classification problems, which used low-rank constraints to learn robust subspaces to connect the original visual features and target output. Subspace learning essentially tries to find a suitable low-dimensional space in which the discriminative representation of the original features is preserved as much as possible. In recent years, low-rank learning has achieved relatively good results in matrix analysis, data recovery, and data denoising. At the same time, low-rank representation is an effective means to describe the structure of high-dimensional data, and it is a generalized form of sparsity in matrix space. That is, low-rank representation can describe the low-dimensional subspace structure of high-dimensional data, thus its component in the subspace becomes the most important factor in characterizing the data. In addition, low-rank representation effectively introduces low-rank constraints into the data matrix, which can help to construct discriminative feature subspaces and eliminate outliers. Inspired by this idea, the noise robustness low-rank learning (NRLRL) algorithm is proposed for EEG signal classification. NRLRL learns a low-rank subspace that connects the original data space and the label space. It fully considers the correlation information and local structure of samples, and it guarantees the minimum rank of the coefficient matrix constructed of data under its self-expression. By integrating the multi-class asymmetric least squares SVM classifier with low-rank representation, NRLRL is insensitive to noise and outliers. The experiments performed on noisy EEG signals are shown that our algorithm is noise robust. NRLRL has several advantages as follows: (1) since the low-rank representation follows the minimum rank criterion, NRLRL is robust when reconstructing the original data with noise and outliers. (2) By full use of supervised information and pinball loss function, an asymmetric least square SVM is jointly learned into our objection function, so that NRLRL explores a robust classifier in the framework of low-rank learning. (3) Local constraints based on low-rank representations are used based on supervision information. The criteria of low-rank representations for minimum within-class and maximum between-classes are adopted to capture the discriminative structure of the data.

BACKGROUND

Low-Rank Representation

Give a set of data samples $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$, each sample $\mathbf{x}_i \in \mathbf{R}^d$ in \mathbf{X} can be represented as a linear combination of atoms from a dictionary \mathbf{A} :

$$\mathbf{X} = \mathbf{A}\mathbf{C}, \quad (1)$$

where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_n] \in \mathbf{R}^{n \times m}$ is the coefficient matrix rank representation.

As a common practice in low-rank learning, the dictionary \mathbf{A} is set to \mathbf{X} , i.e.,

$$\mathbf{X} = \mathbf{XC}, \tag{2}$$

Eq. (2) uses the data set itself to represent the data, which is called the self-expression of the data. Each data sample in data set \mathbf{X} can be represented by:

$$\mathbf{x}_i = \sum_{j \neq i} C_{ij} \mathbf{x}_j, \tag{3}$$

By minimizing the rank of the coefficient matrix \mathbf{C} , Eq. (1) can be written by:

$$\begin{aligned} \min_{\mathbf{C}} \text{rank}(\mathbf{C}), \\ \text{s.t. } \mathbf{X} = \mathbf{XC}. \end{aligned} \tag{4}$$

where $\text{rank}(\mathbf{C})$ is the rank function of \mathbf{C} .

Considering the existence of noise and outliers in the data sample, the structure of the original data \mathbf{X} is taken as two parts: one is the linear combination of the dictionary \mathbf{X} and a low-rank coefficient matrix \mathbf{C} , and the second part is noise (error) matrix, i.e.,

$$\mathbf{X} = \mathbf{XC} + \mathbf{E}, \tag{5}$$

Then the low-rating representation can be defined as follows:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}} \text{rank}(\mathbf{C}) + \mu \|\mathbf{E}\|_{2,0}, \\ \text{s.t. } \mathbf{X} = \mathbf{XC} + \mathbf{E}, \end{aligned} \tag{6}$$

where $\|\cdot\|_{2,0}$ means the $\ell_{2,0}$ -norm operator. μ is the trade-off parameter. In Eq. (6), the term $\|\mathbf{E}\|_{2,0}$ encourages the sparseness of the error components.

The low-rank optimization problem of Eq. (6) is a non-convex NP-hard problem. To find its unique optimal solution, it is necessary to perform convex relaxation of Eq. (6). The kernel norm is the best convex approximation of the rank function on the unit sphere in the matrix spectral norm (Candès and Recht, 2009). Therefore, the convex kernel norm can be used to approximate the non-convex rank function, and the $\ell_{2,0}$ norm can be relaxed to its $\ell_{2,1}$ norm (Raghuandan et al., 2010). Then Eq. (6) can be written as the following convex optimization problem

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}} \|\mathbf{C}\|_* + \mu \|\mathbf{E}\|_{2,1}, \\ \text{s.t. } \mathbf{X} = \mathbf{XC} + \mathbf{E}, \end{aligned} \tag{7}$$

where $\|\cdot\|_*$ means the nuclear norm operator, and $\|\cdot\|_{2,1}$ means the $\ell_{2,1}$ -norm operator.

Asymmetric Least Squares Support Vector Machine

The loss function in the least squares SVM (LS-SVM) pays attention to both the correctly classified and incorrectly classified samples. It minimizes the squared error of the classifier as follows,

$$l(\mathbf{x}) = (1 - \mathbf{w}^T \mathbf{x} - b)^2. \tag{8}$$

In fact, the above loss function is noise sensitive, especially the noises around the separation hyperplane. Many extensions of least squares loss function have been proposed to solve this problem, such as iteratively reweighted least square (Leski, 2015) and asymmetric square function (Leski, 2015), asymmetric squared loss (Huang et al., 2014). Using the statistical property to lower quantile value, the asymmetric squared loss is defined as:

$$L_p^{aLS}(\mathbf{x}) = \begin{cases} p(1 - \mathbf{w}^T \mathbf{x} + b)^2, & \text{if } 1 - \mathbf{w}^T \mathbf{x} + b \geq 0 \\ (1 - p)(1 - \mathbf{w}^T \mathbf{x} + b)^2, & \text{if } 1 - \mathbf{w}^T \mathbf{x} + b < 0 \end{cases} \tag{9}$$

where \mathbf{w} and b are the hyperplane parameter and bias parameter of SVM classifier, respectively. p is the lower quantile value parameter.

The aLS-SVM uses the expectile distance and maximizes the expectile distance between different classes. The aLS-SVM has the following optimization problem:

$$\min_{\mathbf{w}, b} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\alpha}{2} \sum_{i=1}^n L_p^{aLS}(1 - y_i(\mathbf{w}^T \mathbf{x}_i + b))^2. \tag{10}$$

where α is the regularization parameter. This optimization problem can be solved by quadratic programming method.

NOISE ROBUSTNESS LOW-RANK LEARNING ALGORITHM

The Noise Robustness Low-Rank Learning Model

Given a set of data points $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ and their labels $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$ are distributed in K classes. $\mathbf{y}_k = [y_{1,k}, y_{2,k}, \dots, y_{n,k}]$ is the class label vector of n training samples associated with the k -th class. Considering the influence of noise or outliers, the main goal of our algorithm is to find the lowest rank representation \mathbf{C} and the best classifier based on \mathbf{C} .

First, to increase discrimination capability, the local preservation with label embedding is incorporated into the learning process. Different from the traditional local preservation term in low-rank learning, the label information is embedded into the k -nearest neighborhood relationships. For sample \mathbf{x}_i , its low-rank is expressed as \mathbf{c}_i . Without considering the label information of the sample, if \mathbf{x}_j is in the k -nearest neighbor of \mathbf{x}_i , their corresponding low-rank representations \mathbf{c}_j and \mathbf{c}_i should be closer to each other. Obviously, this strategy is not suitable for classification tasks. Based on the basic classification principles of within-class compactness and between-classes separation, the label information is introduced into the k -nearest neighborhood relationships. The within-class matrix \mathbf{B}_{within} and between-classes matrix $\mathbf{B}_{between}$ are accordingly defined, and their elements can be defined as,

$$B_{within,ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t), & \text{if } \mathbf{x}_i \in N(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N(\mathbf{x}_i), \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$B_{between,ij} = \begin{cases} -\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t), & \text{if } \mathbf{x}_i \in N(\mathbf{x}_j) \text{ or} \\ & \mathbf{x}_j \in N(\mathbf{x}_i), y_i \neq y_j, \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where $N(\mathbf{x}_j)$ returns the k -nearest neighbors of \mathbf{x}_j .

In the NRLRL, the original data is projected into a low-dimensional subspace by low-rank representation. NRLRL shows the similarity within the class and the difference between classes of the data. To achieve this goal, the label embedded local preservation term is defined as,

$$\begin{aligned} \min_{\mathbf{C}} &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 (B_{within,ij} - B_{between,ij}) \\ &= Tr(\mathbf{C}\mathbf{B}_{within}\mathbf{C}^T) - Tr(\mathbf{C}\mathbf{B}_{between}\mathbf{C}^T) \\ &= Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T) \end{aligned} \quad (13)$$

where $\mathbf{L} = \mathbf{L}_{between} - \mathbf{L}_{within}$, $\mathbf{L}_{between} = \mathbf{B}_{between} - \mathbf{B}_{between}$, $\mathbf{L}_{within} = \mathbf{B}_{within} - \mathbf{B}_{within}$. $\mathbf{B}_{between}$ and \mathbf{B}_{within} are diagonal matrices, and their elements are $\mathbf{B}_{between,ii} = \sum_{j=1}^n \mathbf{B}_{between,ij}$, $\mathbf{B}_{within,ii} = \sum_{j=1}^n \mathbf{B}_{within,ij}$, separately. $Tr(\cdot)$ is the trace operator.

The label embedded local preservation term is to ensure that if the nearest neighbors \mathbf{x}_i and \mathbf{x}_j are from the same class, their low-rank codes \mathbf{c}_i and \mathbf{c}_j are also close to each other. At the same time, if the nearest neighbors \mathbf{x}_i and \mathbf{x}_j are from different classes, their low-rank codes \mathbf{c}_i and \mathbf{c}_j are separated as much as possible. In such a low-rank learning stage, the ideal non-linear local structure of EEG data is preserved.

To promote the discriminative ability of low-rank representation vectors, a multi-class aLS-SVM classification term is embedded into the NRLRL algorithm. The multi-class aLS-SVM classification term $L(\mathbf{C})$ includes two parts

$$\begin{aligned} L(\mathbf{C}, \mathbf{W}, \mathbf{b}) &= \sum_{k=1}^K (f(\mathbf{C}, \mathbf{y}_k, \mathbf{w}_k, b_k) + \alpha \|\mathbf{w}_k\|_2^2), \\ f(\mathbf{C}, \mathbf{y}_k, \mathbf{w}_k, b_k) &= \sum_{i=1}^n l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k), \end{aligned} \quad (14)$$

where $l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k)$ is the loss function associated with the k -th aLS-SVM.

The squared pinball loss in NRLRL can be written as

$$l(\mu) = \begin{cases} p\mu^2, & \mu \geq 0, \\ (1-p)\mu^2, & \mu < 0, \end{cases} \quad (15)$$

where $\mu_{k,i} = y_{k,i}(\mathbf{w}_k^T \mathbf{c}_i + b_k) - 1$.

Bedding the local preservation with label embedding term and multi-class aLS-SVM classification term into the Eq.(7), the objective function of NRLRL can be written as

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}, \mathbf{W}, \mathbf{b}} & \|\mathbf{C}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \gamma L(\mathbf{C}, \mathbf{W}, \mathbf{b}) + \eta Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T), \\ \text{s.t. } & \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ & \mathbf{1}_n^T \mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (16)$$

where λ, γ , and η are regularization parameters.

The loss function term is decomposed into the sum of the loss term of each sample, and it can be seen that the contribution of each data sample to the objective function is linearly cumulative. $L(\mathbf{C}, \mathbf{W}, \mathbf{b})$ includes a class-by-class loss function

term $l(\mathbf{c}_i, y_{k,i}, \mathbf{w}_k, b_k)$ on the low-rank representation, so that the obtained lowest-rank representations are highly correlated within the class. EEG signals belonging to the same class usually contain the common discriminative features. The lowest-rank representation obtained by Eq. (16) has the characteristics of strong within-class correlation and between-class difference for classification tasks.

To reduce the time costs, the Frobenius norm is used to replace the nuclear norm, the objective function of NRLRL can be rewritten as:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}, \mathbf{W}, \mathbf{b}} & \|\mathbf{C}\|_F^2 + \lambda \|\mathbf{E}\|_{2,1} + \gamma L(\mathbf{C}, \mathbf{W}, \mathbf{b}) + \eta Tr(\mathbf{C}\mathbf{L}\mathbf{C}^T), \\ \text{s.t. } & \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ & \mathbf{1}_n^T \mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (17)$$

For simplicity of expression, combining the two terms $\|\mathbf{C}\|_F^2$ and $Tr(\mathbf{C}^T \mathbf{L} \mathbf{C})$, Eq. (17) can be written as:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}, \mathbf{W}, \mathbf{b}} & \lambda \|\mathbf{E}\|_{2,1} + \gamma L(\mathbf{C}, \mathbf{W}, \mathbf{b}) + Tr(\mathbf{C}(\eta \mathbf{L} + \mathbf{I})\mathbf{C}^T), \\ \text{s.t. } & \mathbf{X} = \mathbf{X}\mathbf{C} + \mathbf{E}, \\ & \mathbf{1}_n^T \mathbf{C} = \mathbf{1}_n^T, \end{aligned} \quad (18)$$

From Eq. (18), we can see that the NRLRL algorithm consists of three sub-problems, namely the parameters of \mathbf{C} , \mathbf{E} , and aLS-SVM classifier. These three sub-problems can be solved alternately until the NRLRL algorithm converges. We use the alternating direction multipliers method (ADMM) (Luo et al., 2017) to solve Eq. (18). The augmented Lagrangian function corresponding to Eq. (18) can be written as:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{E}, \mathbf{W}, \mathbf{b}} & \lambda \|\mathbf{E}\|_{2,1} + \gamma L(\mathbf{C}, \mathbf{W}, \mathbf{b}) + Tr(\mathbf{C}(\eta \mathbf{L} + \mathbf{I})\mathbf{C}^T) \\ & + Tr[\theta^T (\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E})] + Tr[\delta^T (\mathbf{1}_n^T \mathbf{C} - \mathbf{1}_n^T)] \\ & + \frac{\mu}{2} (\|\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}\|_F^2 + \|\mathbf{1}_n^T \mathbf{C} - \mathbf{1}_n^T\|_F^2) \end{aligned} \quad (19)$$

where θ and δ are Lagrange multipliers, and μ is the penalty parameter.

Optimization of the Objective Function

According to the ADMM algorithm, the parameters in Eq. (19) can be updated alternately, that is, when one parameter is updated, other parameters are fixed until the NRLRL algorithm converges.

(1) Update \mathbf{C} by fixing \mathbf{E} , \mathbf{w}_k , and b_k . Eq. (19) is converted to the following problem:

$$\begin{aligned} L(\mathbf{c}_i) &= \gamma \left\| 1 - y_{k,i}(\mathbf{w}_k^T \mathbf{c}_i + b_k) \right\|_2^2 + Tr(\mathbf{c}_i(\eta \mathbf{L} + \mathbf{I})\mathbf{c}_i^T) \\ &+ Tr[\theta^T (\mathbf{x}_i - \mathbf{x}_i \mathbf{c}_i) + \delta^T \mathbf{1}_n^T \mathbf{c}_i] \\ &+ \frac{\mu}{2} (\|\mathbf{x}_i - \mathbf{x}_i \mathbf{c}_i - \mathbf{e}_i\|_F^2 + \|\mathbf{1}_n^T \mathbf{c}_i - \mathbf{1}_n^T\|_F^2). \end{aligned} \quad (20)$$

Let the derivative of Eq. (19) with respect to c_i be zero, the solution of c_i is:

$$c_i = \left(2(\eta\mathbf{L} + \mathbf{I}) + \mu(\mathbf{x}_i^T \mathbf{x}_i + \mathbf{1}_n \mathbf{1}_n^T) + 2\gamma \sum \mathbf{w}_k \mathbf{w}_k^T \right)^{-1} (\mu(\mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{e}_i + \mathbf{1}_n \mathbf{1}_n^T) + \mathbf{x}_i^T \theta - \mathbf{1}_n \delta + 2\gamma \sum \mathbf{w}_k (y_{k,i} - b_k)) \quad (21)$$

(2) Update \mathbf{E} by fixing \mathbf{C} , \mathbf{w}_k , and b_k . Using the same calculation and reduction strategy by Liu et al. (2013). Eq. (19) is converted to the following problem:

$$\min_{\mathbf{E}} \frac{\lambda}{\mu} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \left\| \mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{C} + \frac{1}{\mu}\theta) \right\|_F^2, \quad (22)$$

Then the solution of \mathbf{E} can be obtained by:

$$\mathbf{E}(:, i) = \begin{cases} \frac{\|\theta_i\| - \lambda}{\|\theta_i\|} \theta_i, & \text{if } \frac{\lambda}{\mu} < \|\theta_i\|, \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

where θ_i is the i th column vector of the matrix θ .

(3) Update \mathbf{w}_k and b_k by fixing \mathbf{E} and \mathbf{C} . Eq. (19) is converted to the following problem:

$$\min_{\mathbf{w}_k, b_k} \sum_{k=1}^K \left\{ \alpha \|\mathbf{w}_k\|_2^2 + \sum_{i=1}^n l(c_i, y_{i,k}, \mathbf{w}_k, b_k) \right\}, \quad (24)$$

Eq. (24) is a multi-class aLS-SVM problem, and the optimal parameters \mathbf{w}_k and b_k can be obtained by the aLS-SVM algorithm.

The training procedure of NRLRL is summarized in **Algorithm 1**.

ALGORITHM 1 | The training procedure of NRLRL is summarized.

Algorithm NRLRL: noise robustness low-rank learning algorithm

Input: Training samples \mathbf{X} , label matrix \mathbf{Y} .

Output: lowest-rank representation matrix \mathbf{C} , multi-class aLS-SVM classifier.

$t = 1$

While not convergence or $t <$ maximum number of iterations

Update $\mathbf{C}(t)$ by solving Eq. (21);

Update $\mathbf{E}(t)$ by solving Eq. (23);

Update $\mathbf{w}(t)$ and $b(t)$ by solving aLS-SVM algorithm;

$t = t + 1$

end while

EXPERIMENT

Datasets and Experimental Settings

This study used EEG signals are from Bonn University. The data set consists of five subsets (groups A to E), each of which consists of 100 EEG segments with a single channel duration of 23.6 s and 4,097 samples. The fragments in groups A-B were taken from

five healthy subjects, and the fragments in groups {C, D, E} were taken from patients with epilepsy. The groups C and D recorded the signal during the intermittent period of epileptic seizures. The group E recorded the signal during the seizure. The signals of the five groups of EEG data are shown in **Figure 1**. In the experiment, the 4,097 data points were divided into three data blocks to obtain the research samples, that is, a data block is a sample, representing the EEG information in about 8 s. Therefore, the sample size of this paper is $3 \times 100 = 300$ in each group, and each sample has 1,365 features of sampling points. We design two types of classification tasks on the Bonn dataset. One is the binary classification task: non-epileptic condition (sets {A, B, C, D, E}) and epileptic condition (set E). The other is the three classes of classification task: normal (sets {A, B}), interictal (sets {C, D}), and ictal (set E).

We use the following algorithms as the comparison algorithms: DLSR (Xiang et al., 2012), LC-KSVD (Jiang et al., 2013), SRRS (Li and Fu, 2016), LRS (Kong et al., 2017), aLS-SVM (Huang et al., 2014), and LRDLR (Chen et al., 2020). The parameters of these comparison methods are set to their default settings. In NRLRL, the dictionary size is set from {40, ..., 320}, three regularization parameters are set from $\{2^{-5}, \dots, 2^3\}$, the k -nearest neighbor parameter is set from {3, ..., 11}, and the pinball loss parameter p is set from {40, ..., 360}. We adopt the one-by-one strategy to select the optimal parameters.

Following the method of references (Huang et al., 2014; Gu et al., 2019, 2020), 20 and 50% samples are randomly selected and common Gaussian white noise is added. To test the sensitivity of the classifier to noise intensity, the intensity of Gaussian white noise is divided into three types: the mean value is 0, and the variance is set to 5, 10, and 15% of the sample features, respectively. For example, the *noise* (20%, 10%) indicates that 20% of the samples in the Bonn dataset contain Gaussian white noise, and the variance of the noise is 10% of the sample features.

Classification Result Comparison

First, we perform experiments on the binary classification task. We compare all algorithms in indexes of specificity, sensitivity, and accuracy on the noisy Bonn dataset. The experimental results of the binary classification task are shown in **Tables 1–3**.

From the experimental results, it can be seen that (1) with the increase of noise intensity, the specificity, sensitivity, and accuracy of all algorithms show a decline in varying degrees. It can be seen that the characteristic noise of samples will seriously affect the classification effect of the classifier. Especially the DLSR and LC-KSVD algorithms do not consider the impact of the noise sample interference on the classification surface. As the noise intensity increases, the classification result decreases rapidly. (2) SRRS, LRS, aLS-SVM, LRDLR, and NRLRL algorithms are all noise-insensitive classification algorithms, therefore the classification results are significantly better than conventional classification algorithms. The proposed NRLRL algorithm achieves the best results in classification performance. The NRLRL algorithm removes the influence of noise on the sample in the lowest rank representation, and it uses the pinball loss function to obtain a noise-insensitive classification classifier

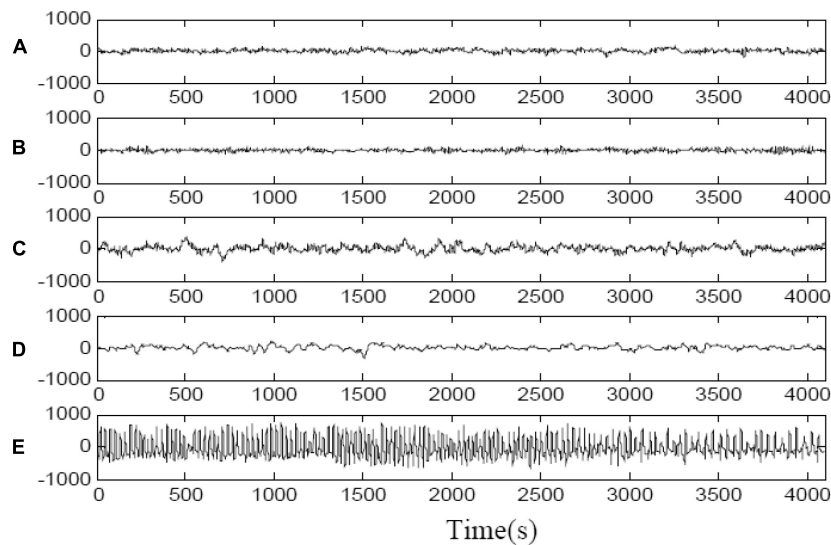


FIGURE 1 | Sample electroencephalogram (EEG) signals in each group in Bonn dataset. **(A)** Epileptic EEG signals measured from healthy people with eyes open. **(B)** Epileptic EEG signals measured from healthy people with eyes closed. **(C)** Epileptic EEG signals obtained in hippocampal formation of the opposite hemisphere of brain during seizure free intervals. **(D)** Epileptic EEG signals obtained from within epileptogenic zone during seizure free intervals. **(E)** Epileptic EEG signals measured during seizure activity.

by maximizing the distance between the two classes of quantile distances. In addition, the NRLRL algorithm can mine the geometric structure of samples in low-dimensional space by low-rank learning, and fully considers the correlation information and subspace structure between samples. Therefore, the within-class similarity and between-class differences of the data are more prominent, which makes the NRLRL algorithm obtain good classification performance in the presence of noise.

Then, we perform experiments on three class classification task, i.e., classification of EEG data in normal, interictal and epileptic periods. Similarly to the above experiment procedure, we compare all algorithms in indexes of specificity, sensitivity, and accuracy on the noisy Bonn dataset. The experimental results of three classification task are shown in **Tables 4–6**. From these results, it can be observed that since the complexity of the three class classification is higher than that of the binary class classification, the results of three class classification are lower than those of binary class classification. The proposed NRLRL algorithm achieves the best results of specificity, sensitivity, and accuracy.

TABLE 1 | The specificity results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
<i>noise (20%, 5%)</i>	92.93	93.06	95.08	95.06	95.57	95.50	96.55
<i>noise (20%, 10%)</i>	92.61	92.70	95.05	95.21	95.32	95.61	96.53
<i>noise (20%, 15%)</i>	92.04	92.87	95.06	95.01	94.80	95.46	96.51
<i>noise (50%, 5%)</i>	90.97	91.11	94.61	94.96	95.06	95.30	96.40
<i>noise (50%, 10%)</i>	90.40	90.54	94.36	94.29	94.62	95.23	96.36
<i>noise (50%, 15%)</i>	89.84	90.32	93.99	94.31	94.49	95.06	96.33

The bold values mean the best values in comparison experiments.

Parameter Analysis

Here we discuss the key parameters in the NRLRL algorithm on binary classification task *noise (20%, 10%)* and three class classification *noise (50%, 10%)*.

The *k*-nearest parameter *k*-nearest is an important parameter in NRLRL. It determines the neighbor relationship between samples. The *k*-nearest parameter is set from {3, ..., 11}. The

TABLE 2 | The sensitivity results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
<i>noise (20%, 5%)</i>	92.94	93.49	95.31	95.50	95.74	95.96	96.65
<i>noise (20%, 10%)</i>	92.63	93.18	95.24	95.39	95.52	95.91	96.63
<i>noise (20%, 15%)</i>	92.08	92.91	95.27	95.25	95.40	95.84	96.58
<i>noise (50%, 5%)</i>	90.04	91.63	94.71	95.11	95.28	95.54	96.52
<i>noise (50%, 10%)</i>	90.42	91.02	94.52	94.31	94.72	95.31	96.48
<i>noise (50%, 15%)</i>	89.95	90.47	94.16	94.39	94.77	95.15	96.46

The bold values mean the best values in comparison experiments.

TABLE 3 | The accuracy results of binary classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
<i>noise (20%, 5%)</i>	92.93	93.22	95.24	95.20	95.57	95.69	96.58
<i>noise (20%, 10%)</i>	92.61	92.72	95.16	95.33	95.34	95.76	96.57
<i>noise (20%, 15%)</i>	92.07	92.88	95.19	95.19	94.99	95.48	96.54
<i>noise (50%, 5%)</i>	90.03	91.29	94.62	95.02	95.18	95.33	96.42
<i>noise (50%, 10%)</i>	90.41	90.69	94.47	94.28	94.63	95.26	96.41
<i>noise (50%, 15%)</i>	89.90	90.46	94.06	94.38	94.57	95.09	96.39

The bold values mean the best values in comparison experiments.

TABLE 4 | The specificity results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.41	89.43	92.45	92.78	92.66	93.31	94.18
noise (20%, 10%)	89.20	89.15	91.06	92.26	91.65	93.27	94.11
noise (20%, 15%)	88.34	88.45	91.72	92.04	91.90	93.08	93.96
noise (50%, 5%)	87.59	87.28	92.92	91.77	91.27	92.41	93.69
noise (50%, 10%)	86.85	86.87	91.67	91.46	92.19	92.44	93.62
noise (50%, 15%)	85.30	85.35	91.41	91.13	91.15	92.83	93.50

The bold values mean the best values in comparison experiments.

TABLE 5 | The sensitivity results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.51	89.51	92.60	92.85	92.69	93.39	94.20
noise (20%, 10%)	89.22	89.26	92.1	92.31	92.74	93.32	94.14
noise (20%, 15%)	88.66	88.49	91.99	92.03	92.08	93.27	93.99
noise (50%, 5%)	87.69	87.33	92.53	91.78	91.42	92.55	93.72
noise (50%, 10%)	86.88	86.81	91.79	91.51	92.21	92.54	93.65
noise (50%, 15%)	86.49	86.42	91.49	91.19	91.28	92.85	93.53

The bold values mean the best values in comparison experiments.

TABLE 6 | The accuracy results of three class classification task on the noisy Bonn dataset.

	DLSR	LC-KSVD	SRRS	LRSD	aLS-SVM	LRDLSR	NRLRL
noise (20%, 5%)	89.46	89.48	92.48	92.80	92.67	93.32	94.18
noise (20%, 10%)	89.20	89.17	92.04	92.28	92.61	93.29	94.01
noise (20%, 15%)	88.54	88.48	91.77	92.03	91.92	93.19	93.95
noise (50%, 5%)	87.57	87.28	92.47	91.64	91.37	92.49	93.70
noise (50%, 10%)	86.86	86.82	91.63	91.43	92.18	92.47	93.63
noise (50%, 15%)	86.40	86.39	91.43	91.14	91.22	92.82	93.52

The bold values mean the best values in comparison experiments.

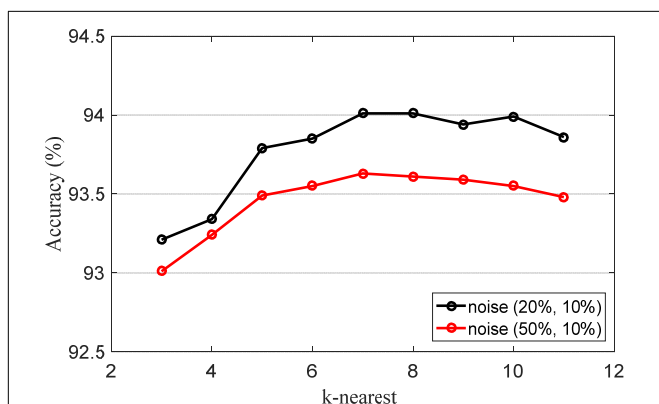


FIGURE 2 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different k -nearest.

classification accuracies of NRLRL with different k -nearest are shown in **Figure 2**. When $k = 7$, the classification accuracy is the

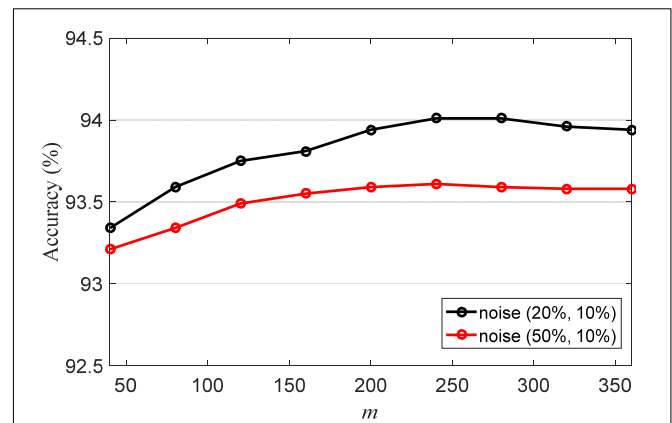


FIGURE 3 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different m .

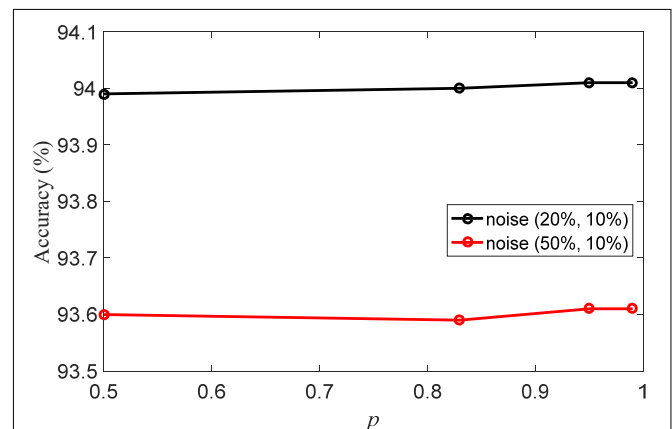


FIGURE 4 | Classification accuracies of noise robustness low-rank learning (NRLRL) with different p .

highest. The appropriate value of k can reflect the local structural information of the sample to the greatest extent. From the results in **Figure 2**, we can set $k = 7$ in the NRLRL algorithm for noisy Bonn dataset.

Another important parameter is m , which is the size of the matrix C . The classification accuracies of NRLRL with different m are shown in **Figure 3**. The parameter m controls the data structure of the low-rank space. When m is too small, the low-rank representation related to the data is not enough to model its structure in the low-rank space. When m is too large, the redundant information will produce errors of low-rank representation. From the results in **Figure 3**, we can set $m = 240$.

The pinball loss parameter p is the important parameter in aLS-SVM classifier in NRLRL. The value range of the pinball loss parameter is $\{0.5, 0.83, 0.95, 0.99\}$. The classification accuracies of NRLRL with different p are shown in **Figure 4**. With different values of p , the NRLRL algorithm has achieved high classification accuracy. It shows that the NRLRL algorithm is not sensitive to the p parameter, so the value of p can be fixed to 0.95 in the experiment.

CONCLUSION

In this study, the NRLRL algorithm is proposed for EEG signal classification. Different from noise-insensitive SVM learns the classification hyperplane in the original space or kernel space, NRLRL learns a low-rank subspace as the transformation from the original data space to the label space, to improve the overall classification effect. By introducing the criteria of low-rank representations for minimum within-class and maximum between-classes, the discriminative ability of the model has been greatly improved. The pinball loss function is also helpful to improve the noise insensitivity of the model. The effectiveness of the proposed algorithm is verified on the noisy Bonn EEG dataset. Since our algorithm directly uses the EEG sample point as the input features, we will consider applying various feature extraction methods to the NRLRL algorithm in the next stage. The seizure data is often insufficient in epilepsy detection or prediction tasks. To obtain an effective algorithm model, the down-sampling strategy is often performed to make the class balanced. This strategy will cause the loss of signal data. Therefore, research on appropriate imbalanced data classification methods is the focus of the next stage.

REFERENCES

- Birjandtalab, J., Pouyan, M. B., Cogan, D., Nourani, M., and Harvey, J. (2017). Automated seizure detection using limited-channel EEG and non-linear dimension reduction. *Comput. Biol. Med.* 82, 49–58. doi: 10.1016/j.combiomed.2017.01.011
- Candès, E. J., and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.* 9:717. doi: 10.1007/s10208-009-9045-5
- Chen, Z., Wu, X. J., and Kittler, J. (2020). Low-rank discriminative least squares regression for image classification. *Signal Process.* 173:107485. doi: 10.1016/j.sigpro.2020.107485
- Fahimi, F., Zhang, Z., Goh, W. B., Lee, T. S., Ang, K. K., and Guan, C. (2019). Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. *J. Neural Eng.* 16:026007. doi: 10.1088/1741-2552/aaf3f6
- Fang, X., Xu, Y., Li, X., Lai, Z., Wong, W. K., and Fang, B. (2018). Regularized label relaxation linear regression. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 1006–1018. doi: 10.1109/TNNLS.2017.2648880
- Gu, X. Q., Ni, T. G., and Fan, Y. Q. (2019). A fast and robust support vector machine with anti-noise convex hull and its application in large-scale ncRNA data classification. *IEEE Access* 7, 134730–134741. doi: 10.1109/ACCESS.2019.2941986
- Gu, X. Q., Ni, T. G., Fan, Y. Q., and Wang, W. B. (2020). Scalable kernel convex hull online support vector machine for intelligent network traffic classification. *Ann. Telecommun.* 75, 471–486. doi: 10.1007/s12243-020-00767-2
- Gu, X. Q., Zhang, C., and Ni, T. G. (2021). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699
- Gummadavelli, A., Zaveri, H. P., Spencer, D. D., and Gerrard, J. L. (2018). Expanding brain-computer interfaces for controlling epilepsy networks: novel thalamic responsive neurostimulation in refractory epilepsy. *Front. Neurosci.* 12:474. doi: 10.3389/fnins.2018.00474
- Han, N., Wu, J., Fang, X. Z., Wong, W. K., Xu, Y., Yang, J., et al. (2019). Double relaxed regression for image classification. *IEEE Trans. Circ. Syst. Video Technol.* 30, 307–319. doi: 10.1109/TCSVT.2018.2890511
- Huang, X., Lei, S., and Suykens, J. (2014). Asymmetric least squares support vector machine classifiers. *Computat. Statist. Data Anal.* 70, 395–405. doi: 10.1016/j.csa.2013.09.015

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/benfulcher/hctsaTutorial_BonnEEG.

AUTHOR CONTRIBUTIONS

MG conceived and designed the proposed model and wrote the manuscript. RL and JM performed the experiment. All authors read and approved the manuscript.

FUNDING

This work was supported in part by the Scientific Research Program of Education Department of Hubei Province, China (D20184101), the Higher Education Reform Project of Hubei Province, China (201707), the “East Lake Scholar” of Wuhan Sport University Fund, China, and the Hubei Provincial University Specialty subject group construction Special fund, China.

- Jiang, Y. Z., Chung, F., and Wang, S. (2019). Recognition of multiclass epileptic EEG signals based on knowledge and label space inductive transfer. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 630–642. doi: 10.1109/TNSRE.2019.2904708
- Jiang, Y. Z., Deng, Z. H., Chung, F., Wang, G. J., Qian, P. J., Choi, K., et al. (2017). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405
- Jiang, Y. Z., Gu, X., Ji, D., Qian, P., Xue, J., Zhang, Y., et al. (2020). Smart diagnosis: a multiple-source transfer TSK fuzzy system for EEG seizure identification. *ACM Trans. Multimed. Comput. Commun. Appl.* 16, 1–21. doi: 10.1145/3340240
- Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88
- Kong, W., Liu, Y., Jiang, B., Dai, G., and Xu, L. (2017). “A new EEG signal processing method based on low-rank and sparse decomposition,” in *Proceedings of the Cognitive Systems and Signal Processing. ICCSIP 2016. Communications in Computer and Information Science*, eds F. Sun, H. Liu, and D. Hu (Singapore: Springer), doi: 10.1007/978-981-10-5230-9_54
- Leski, J. M. (2015). Fuzzy ($c+p$)-means clustering and its application to a fuzzy rule-based classifier: toward good generalization and good interpretability. *IEEE Trans. Fuzzy Syst.* 23, 802–812. doi: 10.1109/TFUZZ.2014.2327995
- Li, S., and Fu, Y. (2016). Learning robust and discriminative subspace with low-rank constraints. *IEEE Trans. Neural Netw. Learn. Syst.* 27, 2160–2173. doi: 10.1109/TNNLS.2015.2464090
- Ling, M. G., and Geng, X. (2019). Soft video parsing by label distribution learning. *Front. Comput. Sci.* 13, 302–317. doi: 10.1007/s11704-018-8015-y
- Liu, G., Lin, Z., Yan, S., Yu, Y., and Ma, Y. (2013). Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 171–184. doi: 10.1109/TPAMI.2012.88
- Liu, Q., Cai, J., Fan, S. Z., Abbod, M. F., Shieh, J. S., Kung, Y., et al. (2019). Spectrum analysis of EEG signals using CNN to model patient’s consciousness level based on anesthesiologists’ experience. *IEEE Access* 7, 53731–53742. doi: 10.1109/ACCESS.2019.2912273
- Luo, M., Zhang, L., Liu, J., Guo, J., and Zheng, Q. (2017). Distributed extreme learning machine with alternating direction method of multiplier. *Neurocomputing* 261, 164–170. doi: 10.1016/j.neucom.2016.03.112

- Meng, M., Lan, M., Yu, J., Wu, J., and Tao, D. (2020). Constrained discriminative projection learning for image classification. *IEEE Trans. Image Process.* 29, 186–198. doi: 10.1109/TIP.2019.2926774
- Ni, T. G., Gu, X. Q., and Zhang, C. (2020). An intelligence EEG signal recognition method via noise insensitive TSK fuzzy system based on interclass competitive learning. *Front. Neurosci.* 14:837. doi: 10.3389/fnins.2020.00837
- Raghunandan, H. K., Andrea, M., and Sewoong, O. (2010). Matrix completion from noisy entries. *J. Mach. Learn. Res.* 11, 2057–2078. doi: 10.1007/s10883-010-9100-1
- Ramakrishnan, S., and Muthanatha, M. A. S. (2018). Epileptic seizure detection using fuzzy-rules-based sub-band specific features and layered multi-class SVM. *Pattern Anal. Appl.* 22, 1161–1176. doi: 10.1007/s10044-018-0691-6
- Sun, L., Jin, B., Yang, H., Tong, J., Liu, C., and Xiong, H. (2019). Unsupervised EEG feature extraction based on echo state network. *Inform. Sci.* 475, 1–17.
- Wang, Y., Li, Z., Feng, L., Zheng, C., and Zhang, W. (2017). Automatic detection of epilepsy and seizure using multiclass sparse extreme learning machine classification. *Comput. Math. Methods Med.* 2017:6849360. doi: 10.1155/2017/6849360
- Xiang, S. M., Nie, F. P., Meng, G. F., Pan, C., and Zhang, C. (2012). Discriminative least squares regression for multiclass classification and feature selection. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 1738–1754. doi: 10.1109/TNNLS.2012.2212721
- Yu, Z., Nie, W., Zhou, W., Xu, F., Yuan, S., Leng, Y., et al. (2020). Epileptic seizure prediction based on local mean decomposition and deep convolutional neural network. *J. Supercomput.* 76, 3462–3476. doi: 10.1007/s11227-018-2600-6
- Zhou, M., Tian, C., Cao, R., Wang, B., Yan, N., Ting, H., et al. (2018). Epileptic seizure detection based on EEG signals and CNN. *Front. Neuroinform.* 12:95. doi: 10.3389/fninf.2018.00095

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

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

Copyright © 2021 Gao, Liu and Mao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.