



EEG Feature Selection via Stacked Deep Embedded Regression With Joint Sparsity

Kui Jiang¹, Jiayi Tang¹, Yulong Wang¹, Chengyu Qiu¹, Yuanpeng Zhang^{1*} and Chuang Lin^{2*}

¹ Department of Medical Informatics of Medical (Nursing) School, Nantong University, Nantong, China, ² Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

OPEN ACCESS

Edited by:

Mohammad Khosravi,
Persian Gulf University, Iran

Reviewed by:

Shan Zhong,
Changshu Institute of Technology,
China
Hongru Zhao,
Soochow University, China

*Correspondence:

Yuanpeng Zhang
maxbirdzhang@ntu.edu.cn
Chuang Lin
chuang.lin@siat.ac.cn

Specialty section:

This article was submitted to
Neuroprosthetics,
a section of the journal
Frontiers in Neuroscience

Received: 12 June 2020

Accepted: 16 July 2020

Published: 06 August 2020

Citation:

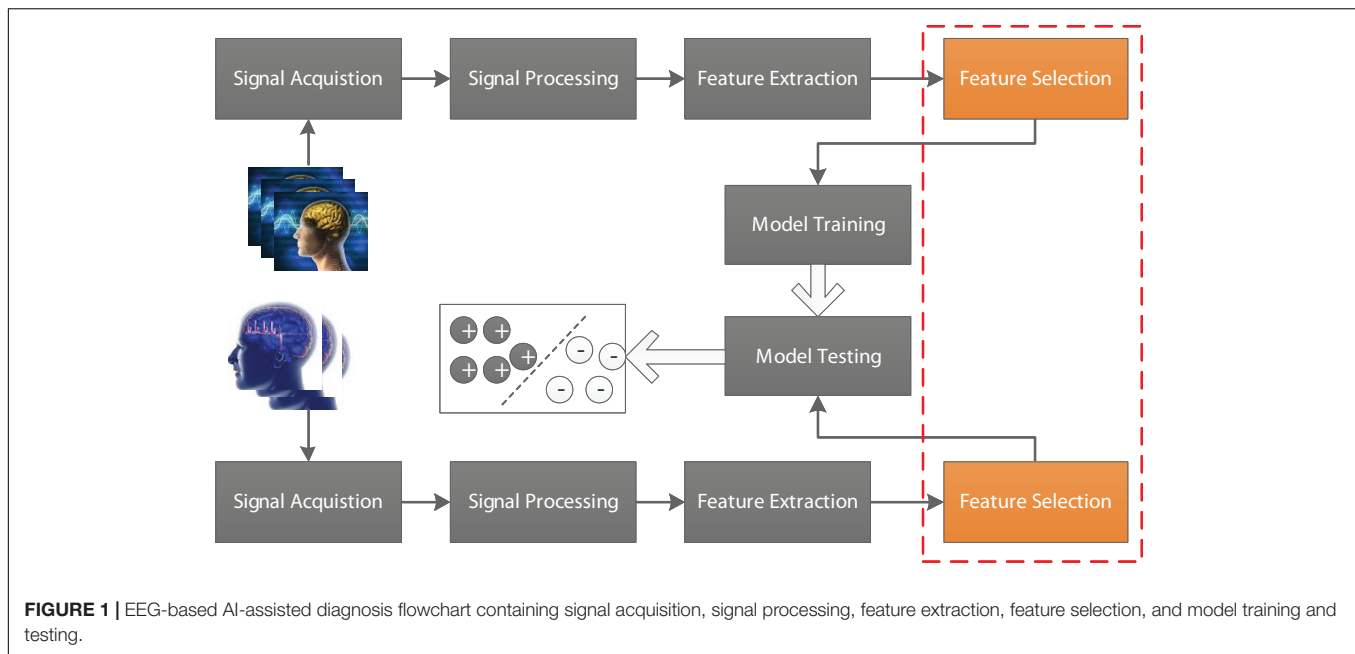
Jiang K, Tang J, Wang Y, Qiu C,
Zhang Y and Lin C (2020) EEG
Feature Selection via Stacked Deep
Embedded Regression With Joint
Sparsity. *Front. Neurosci.* 14:829.
doi: 10.3389/fnins.2020.00829

In the field of brain-computer interface (BCI), selecting efficient and robust features is very seductive for artificial intelligence (AI)-assisted clinical diagnosis. In this study, based on an embedded feature selection model, we construct a stacked deep structure for feature selection in a layer-by-layer manner. Its promising performance is guaranteed by the stacked generalized principle that random projections added into the original features can help us to continuously open the manifold structure existing in the original feature space in a stacked way. With such benefits, the original input feature space becomes more linearly separable. We use the epilepsy EEG data provided by the University of Bonn to evaluate our model. Based on the EEG data, we construct three classification tasks. On each task, we use different feature selection models to select features and then use two classifiers to perform classification based on the selected features. Our experimental results show that features selected by our new structure are more meaningful and helpful to the classifier hence generates better performance than benchmarking models.

Keywords: brain-computer interface, feature selection, stacked deep structure, stacked generalized principle, EEG

INTRODUCTION

Electroencephalogram (EEG) as a biomarker plays an important role in the brain-computer interface (BCI) (Wang et al., 2013; Zheng, 2017; Mammone et al., 2019; Nakamura et al., 2020). For example, EEG signals are often used to determine the presence and type of epilepsy in clinical diagnosis (Rieke et al., 2003; Yetik et al., 2005; Adeli et al., 2007; Lopes da Silva, 2008; Coito et al., 2016; Parvez and Paul, 2016; Peker et al., 2016; Panwar et al., 2019). In recent years, with the rapid development of artificial intelligence technology, AI-assisted diagnosis has attracted more and more attention and achieved unprecedented success in many scenarios including BCI (Agarwal et al., 2018; Wu et al., 2018). In general, a standard EEG-based AI-assisted diagnosis flowchart is illustrated in **Figure 1**, which contains signal acquisition, signal processing, feature extraction, feature selection and model training and testing. As we know that original features extracted from EEG signals cannot be directly used for model training because they are often represented in very high-dimensional feature space. Therefore, feature selection is usually performed before model training. In this study, we focus on how to selection effective features to guarantee high-efficiency AI-assisted clinical diagnosis.



To the best of our knowledge, most of the existing feature selection models belong to one of three main catalog, i.e., filter, embedded, and wrapper (Visalakshi and Radha, 2014; Ang et al., 2016; Shah and Patel, 2016; Saputra Rangkuti et al., 2018). In filter models, feature selection depends on the intrinsic properties and the relevancies existing among features. That is to say, filter models are independent of classifiers. Some of the most commonly-used filter models include mRMR (Peng et al., 2005), F-statistic (Habbema and Hermans, 1977), Chi-square and information gain (Raileanu and Stoffel, 2004), t -test (Raileanu and Stoffel, 2004) and Relief (Kira and Rendell, 1992), etc. All of them perform feature selection by making use of global statistical information such as the relevance/sensitivity/correlation of a feature w.r.t the class label distribution of the data. In wrapper models, feature selection is around classifiers providing them subsets of features and receiving their feedback. Different from filter models, wrapper models are tightly coupled with a specific classifier. Some representative models include CFS (Hall and Smith, 1999) and RFE-SVM (Guyon et al., 2002), etc. In embedded models, feature selection is considered as an optimization problem and integrating into a specific classifier so that the selected features have a seductive effect on the corresponding classification task. For example, Nie et al. (2010) integrated l_2 , l_1 -norm into a robust loss function and proposed an efficient and robust model (renamed as E-JS-Regression) to perform feature selection. Their experimental results on several biomedical data indicated that E-JS-Regression won better performance than both filter models and wrapper models.

In ensemble learning (Webb and Zheng, 2004; Minku et al., 2010; Chen et al., 2017; Liu et al., 2019; Zhu et al., 2020), stacking is a popular classifier combination strategy which takes the outputs of other classifier as input to train a generalizer. In Wolpert (1992) proposed the stacked generalization principle which indicated that the outputs can

help to open the manifold of data distribution. In our previous work (Zhang et al., 2018), we made use of this principle and proposed a deep TSK fuzzy system. Therefore, in this study, based on this principle and by taking E-JS-Regression as the basic component, we will construct a layer-by-layer stacked deep structure for feature extraction. The new model is termed as SDE-JS-Regression. In SDE-JS-Regression, each component is connected in a layer-by-layer manner, the output of the previous layer is transformed by random projection as a random shift and then added into the input space. The new input space is considered as the input to the next component. In such a way, the manifold in the training space is continuously opened. The contribution of this study is summarized as follows:

- (i) Based on E-JS-Regression proposed by Nie et al. we construct a stacked deep structure for feature selection in a layer-by-layer manner so as to add random projections into the original features so that the manifold structure existing in the original feature space is continuously opened in a stacked way. Therefore, according to the stacked generalized principle, the original input feature space becomes more linearly separable.
- (ii) We build three classification tasks from epilepsy EEG data provided by the University of Bonn and introduce different kinds of feature selection methods to demonstrate the promising performance of our proposed method.

DATA AND METHODS

Data

The epilepsy EEG data downloaded from the University of Bonn will be used to evaluate our proposed feature selection

model. This dataset consists of 5 groups of subsets (from group A to group E), where each group is composed of 100 single channel EEG segments during 23.6 s duration. Segments in group A and group B are collected from 5 healthy subjects, while segments in the rest groups are collected from epileptics. **Table 1** lists the data structure and collection conditions. Additionally, **Figure 2** (Zhang et al., 2020) illustrates the amplitudes during the collection procedure of one subject in each group.

Methods

In this section, we will give technical details of our proposed method including its framework, optimization, and algorithm steps. Before we do that, we first summarize the following used notations and some commonly-used definitions.

Notations and Definitions

We use $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{d \times n}$ and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T \in R^{n \times c}$ to represent a training set, where $\mathbf{x}_i = [x_1, x_2, \dots, x_d]^T \in R^d$ represents a training sample and $y_i \in R^c$ is the corresponding label vector of \mathbf{x}_i , $1 \leq i \leq n$. For matrix \mathbf{B} , we use b_{ij} to represent its element in the i -th row and j -th column, b^i and b_j to represent its i -th row and j -th column, respectively. The $l_{2,1}$ -norm of matrix \mathbf{B} is defined as:

$$\|\mathbf{B}_{2,1}\| = \sum_{i=1}^n \left(\sum_{j=1}^m b_{ij}^2 \right)^{1/2} = \sum_{i=1}^n \|b^i\| \quad (1)$$

Structure of SDE-JS-Regression

In Nie et al. (2010) proposed an efficient and robust embedded regression model for feature selection via joint $l_{2,1}$ -norm sparsity (simplified as E-JS-Regression). Since l_2 -norm based loss function is sensitive to outliers, they used a $l_{2,1}$ -norm based loss function to remove outliers. Additionally, they also used a $l_{2,1}$ -norm to regularize the transformation matrix to select features with joint sparsity. That is to say, each feature either has small

scores for all samples or has large scores for all samples. The objective function is defined as:

$$\min_{\mathbf{W}} J(\mathbf{W}) = \|\mathbf{X}^T \mathbf{W} - \mathbf{Y}\|_{2,1} + \theta \|\mathbf{W}\|_{2,1} \quad (2)$$

where θ is the regularized parameter, $\mathbf{W} \in R^{d \times c}$. The stacked generalized principle as an ensemble learning strategy can provide an efficient way for model combination. Although the stacked generalized principle is not as widely used as boosting and bagging, its great innovation has been successful in many application scenarios. In this study, we take E-JS-Regression as a basic component to construct a stacked deep embedded regression model for EEG feature selection. **Figure 3** illustrates the stacked deep structure of our proposed model.

The stacked deep structure is composed of m basic components linked in a layer-by-layer manner. To be specific, when the first component is fixed, the input to the subsequent components consists of two parts: the original input features and the output of the previous layer/component. How to fuse these two parts is very important in this study. Referring to the stacked generalized principle, we randomly generate a project of the output of the previous layer as a random shift and then integrate the random shift into the original input features. Therefore, the input of component s ($1 < s \leq m$) \mathbf{X}_s can be obtained by the following equation,

$$\mathbf{X}_s^T = \mathbf{X}^T + \sigma \mathbf{Y}_{s-1} \mathbf{Z} \quad (3)$$

where $\mathbf{Z} \in R^{c \times d}$ is a random projection matrix in which each element is in the range of $[0, 1]$, σ is a positive regularized parameter. By virtue of this structure, all components (E-JS-Regression) are stacked and bridged by adding the original features to a continuous random shift to form the proposed feature selection model SDE-JS-Regression.

The benefits we inherit from the stacked deep structure lie in that the random projections added into the original features can help us to continuously open the manifold structure existing in the original feature space in a stacked way. With such benefits, the input feature space becomes more linearly separable.

Optimization of SDE-JS-Regression

By substituting (3) into (2), the optimization of SDE-JS-Regression can be considered as solving m subproblems. The s -th subproblem can be formulated as follows,

$$\min_{\mathbf{W}} J(\mathbf{W}) = \frac{1}{\theta} \|\mathbf{X}^T + \sigma \mathbf{Y}_{s-1} \mathbf{Z}\mathbf{W} - \mathbf{Y}\|_{2,1} + \|\mathbf{W}\|_{2,1} \quad (4)$$

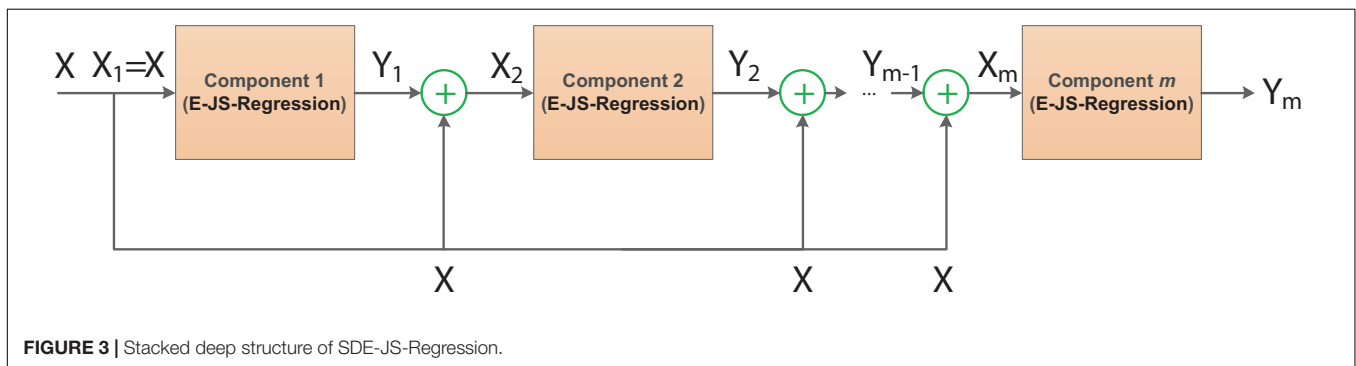
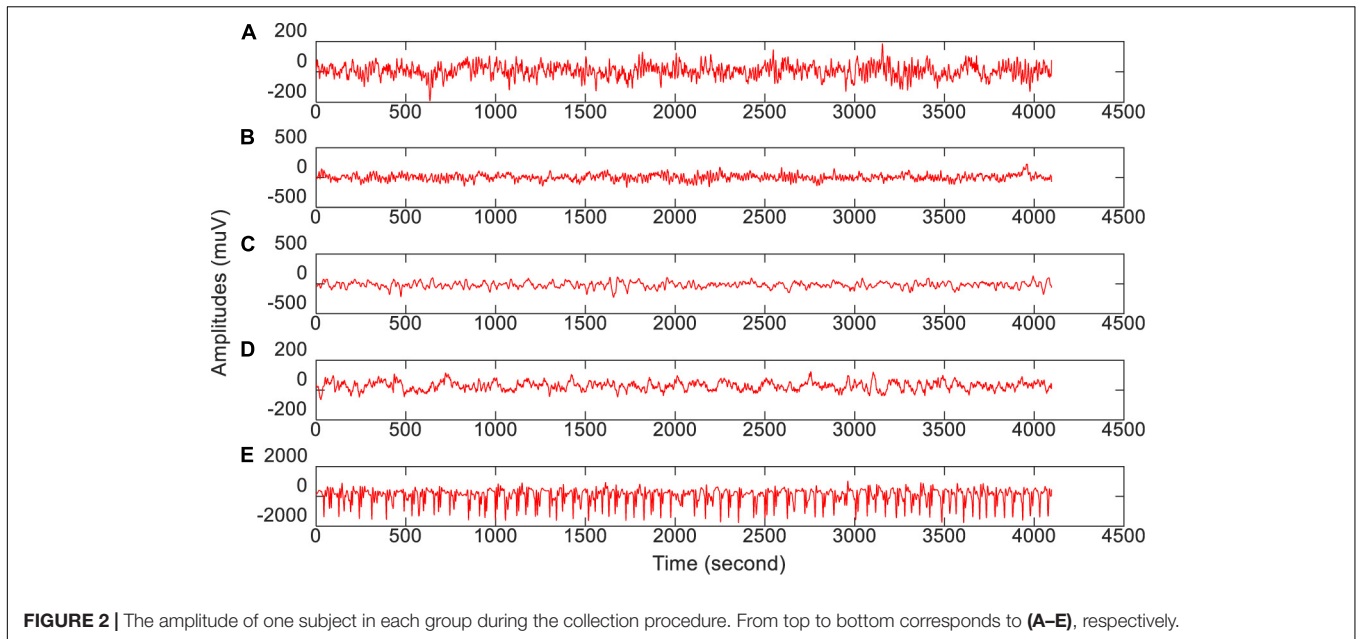
which is equivalent to the following problem,

$$\min_{\mathbf{W}, \mathbf{Q}} J(\mathbf{W}) = \|\mathbf{Q}\|_{2,1} + \|\mathbf{W}\|_{2,1} \quad (5)$$

$$\text{s.t. } (\mathbf{X}^T + \sigma \mathbf{Y}_{s-1} \mathbf{Z})\mathbf{W} + \theta \mathbf{Q} = \mathbf{Y} \quad (6)$$

TABLE 1 | Data structure and collection conditions of epilepsy EEG segments.

Volunteers	Groups	#Channels	#Features	Collection conditions
Health	A	100	4097	Signals were recorded when volunteer subjects were relaxed in awaken state with eyes open.
	B	100	4097	Signals were recorded when volunteer subjects were relaxed in awaken state with eyes closed.
Epileptic	C	100	4097	Signals were recorded from the hippocampal formation of the opposite hemisphere of brain.
	D	100	4097	Signals were recorded within epileptogenic zone during seizure free intervals.
	E	100	4097	Signals were recorded during seizure activity.



By equivalent transformation, we have:

$$\min_{\mathbf{W}, \mathbf{Q}} J(\mathbf{W}) = \left\| \begin{bmatrix} \mathbf{W} \\ \mathbf{Q} \end{bmatrix} \right\|_{2,1} \quad (7)$$

$$\text{s.t. } [\mathbf{X}^T + \sigma \mathbf{Y}_{s-1} \mathbf{Z} \theta \mathbf{I}] \begin{bmatrix} \mathbf{W} \\ \mathbf{Q} \end{bmatrix} = \mathbf{Y} \quad (8)$$

where $\mathbf{I} \in R^{n \times n}$ is a identity matrix. Let $h = n + d$, $\mathbf{K} = [\mathbf{X}^T + \sigma \mathbf{Y}_{s-1} \mathbf{Z} \theta \mathbf{I}] \in R^{n \times h}$ and $\mathbf{V} = \begin{bmatrix} \mathbf{W} \\ \mathbf{Q} \end{bmatrix} \in R^{h \times c}$, then the optimization problem in (7) can be updated as follows,

$$\min_{\mathbf{V}} J(\mathbf{V}) = \|\mathbf{V}\|_{2,1} \quad (9)$$

$$\text{s.t. } \mathbf{KV} = \mathbf{Y} \quad (10)$$

By introducing Lagrangian multiplies Δ , the corresponding Lagrangian function of (9) is formulated as follows,

$$L(\mathbf{V}) = \|\mathbf{V}\|_{2,1} - \text{Tr}(\Delta^T (\mathbf{KV} - \mathbf{Y})) \quad (11)$$

By setting the partial derivative of $L(\mathbf{V})$ w.r.t \mathbf{V} to 0, i.e.,

$$\frac{\partial L(\mathbf{V})}{\partial \mathbf{V}} = 2\mathbf{G}\mathbf{V} - \mathbf{K}^T \Delta = 0 \quad (12)$$

where $\mathbf{G} \in R^{h \times h}$ is a diagonal matrix in which the i -th diagonal element is:

$$g_{ii} = \frac{1}{2 \|\mathbf{v}^i\|_2} \quad (13)$$

Thus, by multiplying the two sides of (12) by \mathbf{KG}^{-1} , and making use of the constraint $\mathbf{KV} = \mathbf{Y}$, we have:

$$\begin{aligned} 2\mathbf{KV} - \mathbf{KG}^{-1} \mathbf{K}^T \Delta &= 0 \\ \Rightarrow 2\mathbf{Y} - \mathbf{KG}^{-1} \mathbf{K}^T \Delta &= 0 \\ \Rightarrow \Delta &= 2(\mathbf{KG}^{-1} \mathbf{K}^T)^{-1} \mathbf{Y} \end{aligned} \quad (14)$$

By substituting (14) into (12), we obtain \mathbf{V} as:

$$\mathbf{V} = \mathbf{G}^{-1} \mathbf{K}^T (\mathbf{KG}^{-1} \mathbf{K}^T)^{-1} \mathbf{Y} \quad (15)$$

Algorithm of SDE-JS-Regression

The detailed algorithm steps of SDE-JS-Regression are listed in Algorithm 1. When the transformation matrix $\mathbf{W} \in R^{d \times c}$ is

obtained by SDE-JS-Regression, we compute the sum of each column vector \mathbf{w}_j , then sort the elements in the final column vector from largest to smallest. In such a way, we obtain the feature ranking list, which can guide feature selection.

RESULTS

In this section, we will report our experimental settings and results.

Setups

To fairly evaluate the feature selection performance of SDE-JS-Regression, we introduce several types of feature selection models, i.e., E-JS-Regression (Nie et al., 2010), mRMR (Peng et al., 2005), RFE-SVM (Guyon et al., 2002), and Relief (Kira and Rendell, 1992) for benchmarking testing. A brief introduction of each benchmarking model is summarized as follows.

- E-JS-Regression: It is an embedded feature selection model and also the basic component of our proposed method. Its involved regularized parameter γ will be determined by 5-CV in our experiments.
- mRMR: It is a filtering feature selection model based on minimum redundancy and maximum relevancy. The redundancy is measured by mutual information.
- RFE-SVM: It is a wrapper feature selection model combining with the SVM classifier to achieve recursive feature elimination. Parameters in SVM are all determined by 5-CV.
- Relief: It is also a filtering feature selection model, which assigns a weight to each feature depending on the relevance between features and classes. The number of nearest neighbors is set to 10 in our experiments.

Algorithm 1: SDE-JS-Regression

Input:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{d \times n} \text{ and}$$

$$\mathbf{Y} = [y_1, y_2, \dots, y_n]^T \in R^{n \times c}$$

θ, σ and m

Output:

\mathbf{W}

Procedure:

Set $t \leftarrow 0$

Initialize $\mathbf{G}^{(t)} \in R^{h \times h}$ as an identity matrix

Set $s \leftarrow 1$

Set $\mathbf{Y}_0 = \mathbf{0}$

$$\text{Compute } \mathbf{K} = [\mathbf{X}^T + \sigma \mathbf{Y}_0 \mathbf{Z} \theta \mathbf{I}] \in R^{n \times h}$$

Repeat

$$\text{Compute } \mathbf{V}^{(t+1)} = (\mathbf{G}^{(t)})^{-1} \mathbf{K}^T (\mathbf{K} (\mathbf{G}^{(t)})^{-1} \mathbf{K}^T)^{-1} \mathbf{Y}$$

Compute $\mathbf{G}^{(t+1)}$, where the i -th diagonal element is

$$g_{ii} = \frac{1}{2 \|\mathbf{v}^{(t+1)}\|_2}$$

Set $t \leftarrow t + 1$

Until $|J^{(t+1)}(\mathbf{V}) - J^{(t)}(\mathbf{V})| <$

Extract \mathbf{W} from \mathbf{V}

$$\text{Compute } \mathbf{Y}_1 = \mathbf{X}^T \mathbf{W}$$

For $s = 1$ to m

Set $t \leftarrow 0$

Initialize $\mathbf{G}^{(t)} \in R^{h \times h}$ as an identity matrix

Randomly generate $\mathbf{Z} \in R^{c \times d}$, where each element is in the range of $[0, 1]$

$$\text{Compute } \mathbf{K} = [\mathbf{X}^T + \sigma \mathbf{Y}_s \mathbf{Z} \theta \mathbf{I}] \in R^{n \times h}$$

Repeat

$$\text{Compute } \mathbf{V}^{(t+1)} = (\mathbf{G}^{(t)})^{-1} \mathbf{K}^T (\mathbf{K} (\mathbf{G}^{(t)})^{-1} \mathbf{K}^T)^{-1} \mathbf{Y}$$

Compute $\mathbf{G}^{(t+1)}$, where the i -th diagonal element

$$\text{is } g_{ii} = \frac{1}{2v_2^{i(t+1)}}$$

Set $t \leftarrow t + 1$

Until $|J^{(t+1)}(\mathbf{V}) - J^{(t)}(\mathbf{V})| <$

Extract \mathbf{W} from \mathbf{V}

$$\text{Compute } \mathbf{Y}_s = \mathbf{X}^T \mathbf{W}$$

End

When the feature ranking list generated by each model is obtained, the Gaussian kernel based SVM (Chang and Lin, 2011) and Ridge regression (Ridge) (Yang and Wen, 2018) are employed to perform classification tasks. Based on the epilepsy EEG data shown in **Table 1**, we construct 3 classification tasks (see **Table 2**).

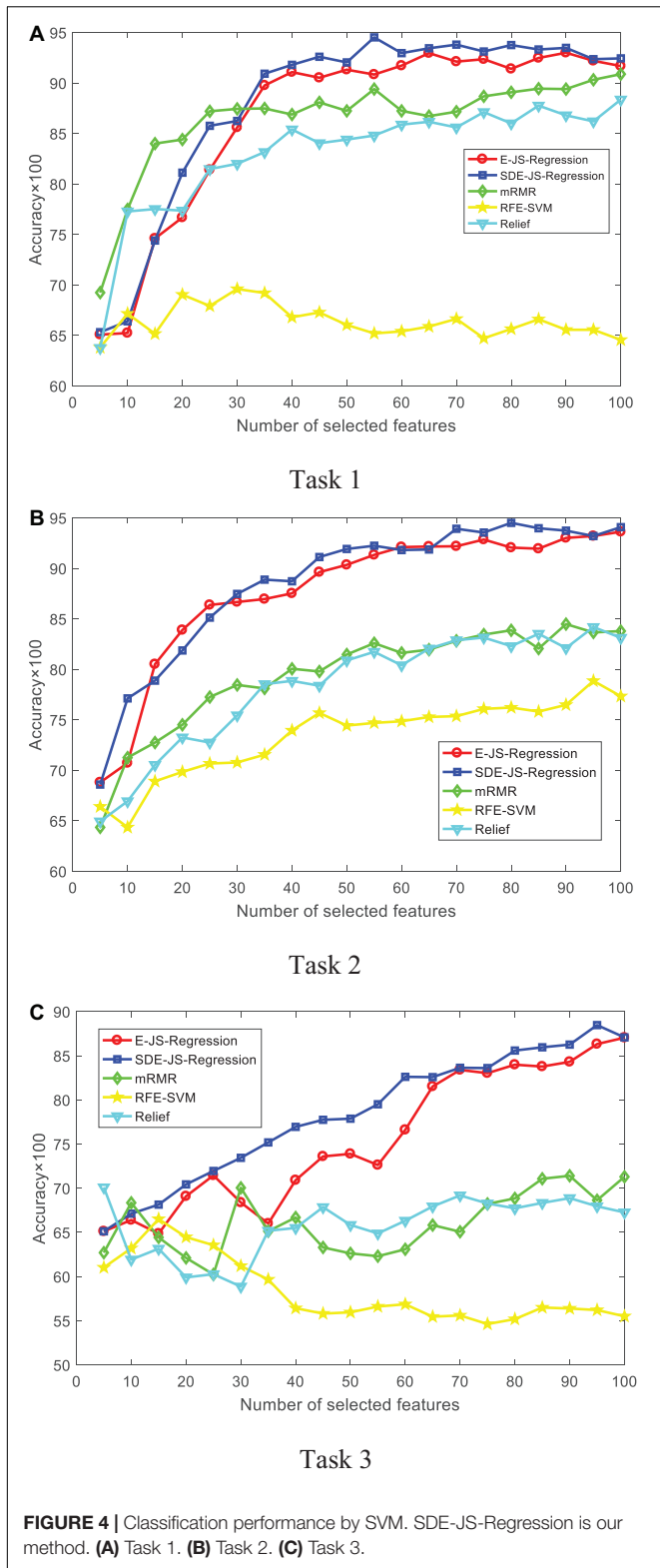
For each task, 75% samples are used for training and 25% samples are used for testing. Parameters (kernel width and slack variable) in the Gaussian kernel based SVM and the regularized parameter in Ridge are determined by 5-CV on the training set. Testing procedure is repeated 100 times and the average results in terms of *Accuracy* are recorded, where *Accuracy* is defined as the ratio of the number of correctly classified samples to the number of all samples.

Experimental Results

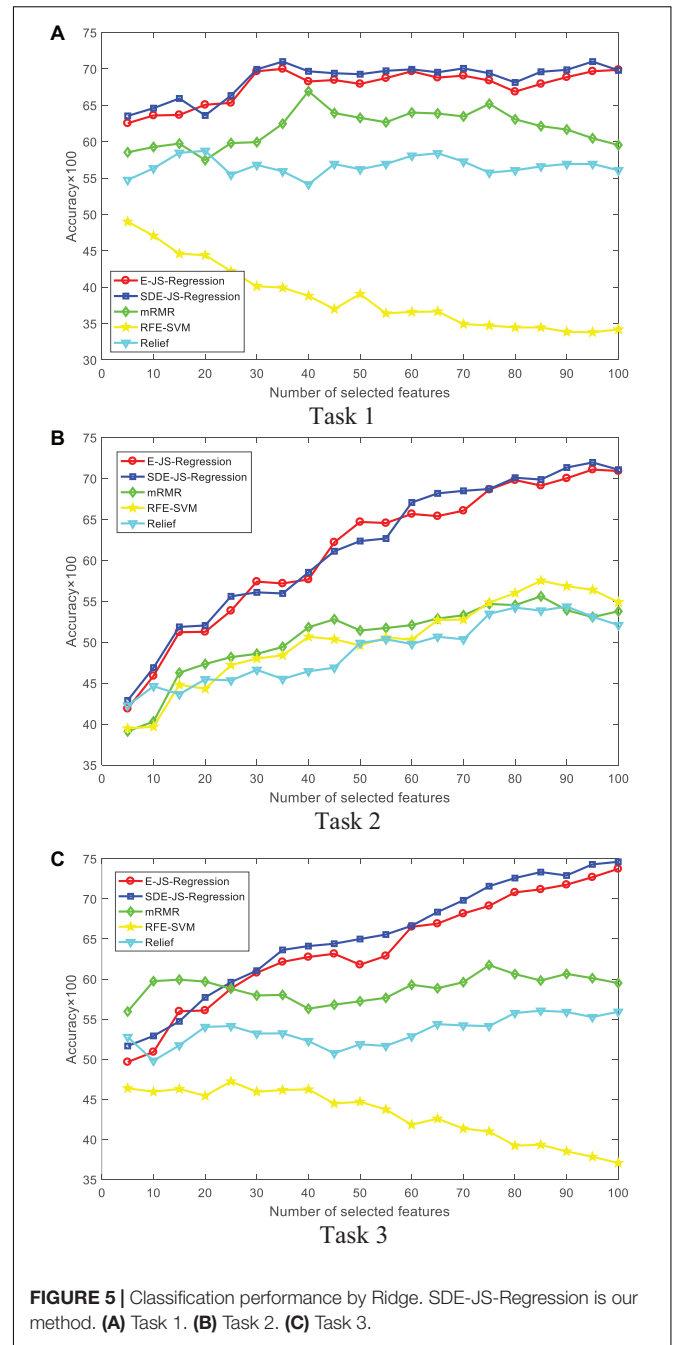
In this section, we report our experimental results from two main aspects, i.e., classification performance of selected features and the parameter analysis. **Figures 4, 5** show the classification performance of five models with different numbers of features (from 5 to 100, step size is 5) selected from the corresponding

TABLE 2 | Three classification tasks for selected features.

No. of tasks	Volunteers	Groups	#Sizes	#Classes	Task description
Task 1	Health	A, B	200	2	Classify healthy subjects to the eye-opening group and the eye-closing group.
Task 2	Epileptic	C, D and E	300	3	Classify epileptic subjects to non-seizure period, interseizure period and seizure period.
Task 3	Health/Epileptic	A, B, C, D and E	500	2	Classify all subjects into healthy group and epileptic group.



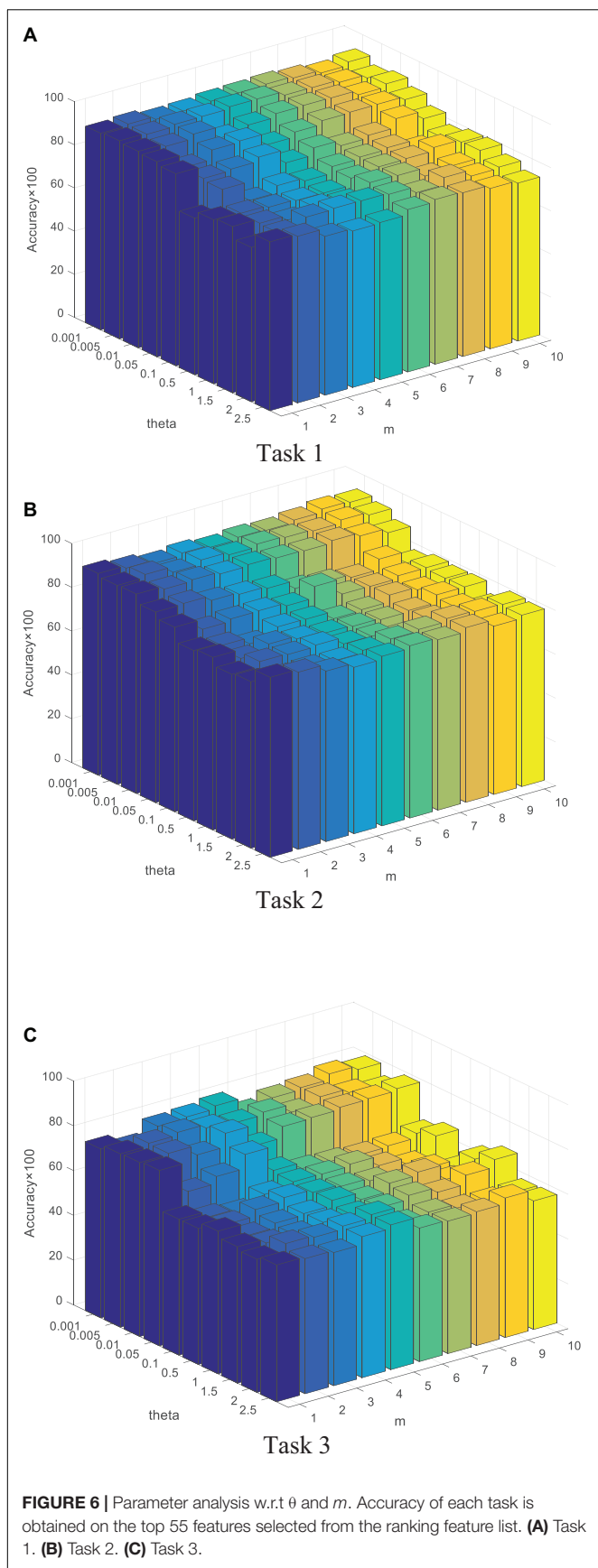
ranking list. **Figure 6** shows parameter analysis results w.r.t the regularized parameter θ and the number of components m , where θ is searched from the range [0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1,



1.5, 2, 2.5] and m is searched from the range [1–10]. Accuracy of each task is obtained on the top 55 features selected from the ranking feature list.

DISCUSSION

From the comparative results of three classification tasks shown in **Figures 4, 5**, we observe that SDE-JS-Regression performs better than the benchmarking models, especially mRMR, RFE-SVM, and Relief. On task 3, regardless of SVM or Ridge,



SDE-JS-Regression always perform better than E-JS-Regression when the number of selected top features is bigger than 15. More characteristics are exhibited from the following aspects.

- From our experimental results, we find that features obtained from embedded feature selection models (SDE-JS-Regression and E-JS-Regression) are more inductive to the classifier than filter models (mRMR and Relief) and wrapper models (RFE-SVM). This is because embedded feature selection models minimize the classification training errors during the procedure of feature selection. Therefore, for our epilepsy classification tasks via EEG signals, embedded feature selection models are more suitable.
- On the three classification tasks, especially task 3, SDE-JS-Regression achieves better performance than E-JS-Regression, which indicates that our stacked deep structure can indeed help to select more classification additive features and hence improve the classification performance. As we stated before, the benefits we inherit from the stacked deep structure lie in that the random projections added into the original features can help us to continuously open the manifold structure existing in the original feature space in a stacked way. With such benefits, the input feature space becomes more linearly separable.
- From **Figure 5**, with respect to θ , we observe that SDE-JS-Regression performs well in its range of $[0.001, 0.05]$. With the further increase of θ from 0.05 to 2.5, the classification performance begins to decrease. However, although the performance begins to decline when θ is in the range of $[0.05, 2.5]$, the performance of SDE-JS-Regression does not show a significant change. Therefore, our proposed SDE-JS-Regression seems to be robust to θ . For our three EEG classification tasks, θ can be set from 0.001 to 0.05.
- The number of layers (components) in the structure of SDE-JS-Regression determines the number of random shifts added into the input feature space. As we can see from **Figure 5** that “the more layers the better performance” is not holds. On the three tasks, 4–6 layers can guarantee a relatively good performance. Too many random shifts can lead to distribution distortion of the training set.

CONCLUSION

In this study, we propose a feature selection model SDE-JS-Regression for AI-assisted clinical diagnosis through EEG signals. SDE-JS-Regression is quite different from the existing embedded models due to its stacked deep structure that is constructed in a layer-by-layer manner based on the stacked generalized principle. SDE-JS-Regression is derived from E-JS-Regression but performs better than E-JS-Regression since that random projections added into the original features can help us to continuously open the manifold structure existing in the original feature space in a stacked way so that the original input feature space becomes more linearly separable. We construct three classification tasks based on the selected features to evaluate the

effectiveness of SDE-JS-Regression. Experimental results show that features selected by SDE-JS-Regression are more meaningful and helpful to the classifier hence generates better performance than benchmarking models. This study is not without limitations. For example, how to effectively determine the number of layers is very important. Therefore, in addition to CV, a new finding strategy will be desired in our coming studies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.meb.unibonn.de/epileptologie/science/physik/eedata.html>.

AUTHOR CONTRIBUTIONS

YZ, KJ, and CL designed the whole algorithm and experiments. JT, YW, and CQ contributed on code implementation.

REFERENCES

- Adeli, H., Ghosh-Dastidar, S., and Dadmehr, N. (2007). A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy. *IEEE Trans. Biomed. Eng.* 54, 205–211. doi: 10.1109/tbme.2006.886855
- Agarwal, A., Dowsley, R., McKinney, N. D., and Wu, D. (2018). Privacy-preserving linear regression for brain-computer interface applications, in *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, 5277–5278.
- Ang, J. C., Mirzal, A., Haron, H., and Hamed, H. N. A. (2016). Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 13, 971–989. doi: 10.1109/tcb.2015.2478454
- Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* 2, 1–27. doi: 10.1145/1961189.1961199
- Chen, C., Dong, D., Qi, B., Petersen, I. R., and Rabbitt, H. (2017). Quantum ensemble classification: a sampling-based learning control approach. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 1345–1359. doi: 10.1109/tnls.2016.2540719
- Coito, A., Michel, C. M., van Mierlo, P., Vulliemoz, S., and Plomp, G. (2016). Directed functional brain connectivity based on EEG source imaging: methodology and application to temporal lobe epilepsy. *IEEE Trans. Biomed. Eng.* 63, 2619–2628. doi: 10.1109/tbme.2016.2619665
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422.
- Habbema, J., and Hermans, J. (1977). Selection of variables in discriminant analysis by F-statistic and error rate. *Technometrics* 19, 487–493. doi: 10.1080/00401706.1977.10489590
- Hall, M. A., and Smith, L. A. (1999). “Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper,” in *Proceedings of the FLAIRS Conference*, Hollywood, FL, 235–239.
- Kira, K., and Rendell, L. A. (1992). “A Practical Approach to Feature Selection,” in *Proceedings of the Ninth International Workshop on Machine Learning*, San Francisco, CA, 249–256. doi: 10.1016/b978-1-55860-247-2.50037-1
- Liu, W., Wu, W., Wang, Y., Fu, Y., and Lin, Y. (2019). Selective ensemble learning method for belief-rule-base classification system based on PAES. *Big Data Mining Anal.* 2, 306–318. doi: 10.26599/bdma.2019.9020008
- Lopes da Silva, F. H. (2008). The Impact of EEG/MEG Signal Processing and Modeling in the Diagnostic and Management of Epilepsy. *IEEE Rev. Biomed. Eng.* 1, 143–156. doi: 10.1109/rbme.2008.2008246
- All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant No. 817017938, by Jiangsu Post-doctoral Research Funding Program under Grant No. 2020Z020, and by the Shenzhen Basic Research under Grant Nos. JCYJ20170413152804728 and JCYJ2018050718250885.

ACKNOWLEDGMENTS

We thank the reviewers whose comments and suggestions helped improve this manuscript. We also thank Feiping Nie for sharing the source code of E-JS-Regression.

- Mammone, N., De Salvo, S., Bonanno, L., and Ieracitano, C. (2019). Brain Network Analysis of Compressive Sensed High-Density EEG Signals in AD and MCI subjects. *IEEE Trans. Indus. Inform.* 15, 527–536. doi: 10.1109/tii.2018.2868431
- Minku, L. L., White, A. P., and Yao, X. (2010). The impact of diversity on online ensemble learning in the presence of concept drift. *IEEE Trans. Knowledge Data Eng.* 22, 730–742. doi: 10.1109/tkde.2009.156
- Nakamura, T., Alqurashi, Y. D., Morrell, M. J., and Mandic, D. P. (2020). Hearables: automatic overnight sleep monitoring with standardized in-ear EEG sensor. *IEEE Trans. Biomed. Eng.* 67, 203–212. doi: 10.1109/tbme.2019.2911423
- Nie, F., Huang, H., Cai, X., and Ding, C. H. (2010). Efficient and robust feature selection via joint l2, l1-norms minimization, in *Advances in Neural Information Processing Systems*, eds M. I. Jordan, Y. LeCun, and S. A. Solla (Cambridge, MA: MIT Press), 1813–1821.
- Panwar, S., Joshi, S. D., Gupta, A., and Agarwal, P. (2019). Automated epilepsy diagnosis Using EEG with test set evaluation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1106–1116. doi: 10.1109/tnsre.2019.2914603
- Parvez, M. Z., and Paul, M. (2016). Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation. *IEEE Trans. Neural Syst. Rehabil. Eng.* 24, 158–168. doi: 10.1109/tnsre.2015.2458982
- Peker, M., Sen, B., and Delen, D. (2016). A novel method for automated diagnosis of epilepsy using complex-valued classifiers. *IEEE J. Biomed. Health Inform.* 20, 108–118. doi: 10.1109/jbhi.2014.2387795
- Peng, H., Long, F., and Ding, C. H. Q. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/tpami.2005.159
- Raileanu, L. E., and Stoffel, K. (2004). Theoretical comparison between the gini index and information gain criteria. *Ann. Math. Artif. Intell.* 41, 77–93. doi: 10.1023/b:amai.0000018580.96245.c6
- Rieke, C., Mormann, F., Andrzejak, R. G., Kreuz, T., David, P., Elger, C. E., et al. (2003). Discerning nonstationarity from nonlinearity in seizure-free and pre-seizure EEG recordings from epilepsy patients. *IEEE Trans. Biomed. Eng.* 50, 634–639. doi: 10.1109/tbme.2003.810684
- Saputra Rangkuti, F. R., Fauzi, M. A., Sari, Y. A., and Sari, E. D. L. (2018). “Sentiment analysis on movie reviews using ensemble features and pearson correlation based feature selection,” in *Proceedings of the 2018 International Conference on Sustainable Information Engineering and Technology (SIET)*, Malang, 88–91.
- Shah, F. P., and Patel, V. (2016). “A review on feature selection and feature extraction for text classification,” in *Proceedings of the 2016 International*

- Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 2264–2268.
- Visalakshi, S., and Radha, V. (2014). “A literature review of feature selection techniques and applications: review of feature selection in data mining,” in *Proceedings of the 2014 IEEE International Conference on Computational Intelligence and Computing Research*, Coimbatore, 1–6.
- Wang, D., Miao, D., and Blohm, G. (2013). A New Method for EEG-Based Concealed Information Test. *IEEE Trans. Inform. Forensics Security* 8, 520–527. doi: 10.1109/tifs.2013.2244884
- Webb, G. I., and Zheng, Z. (2004). Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. *IEEE Trans. Knowledge Data Eng.* 16, 980–991. doi: 10.1109/tkde.2004.29
- Wolpert, D. H. (1992). Stacked generalization. *Neural Netw.* 5, 241–259.
- Wu, D., King, J., Chuang, C., Lin, C., and Jung, T. (2018). Spatial Filtering for EEG-Based Regression Problems in Brain–Computer Interface (BCI). *IEEE Trans. Fuzzy Syst.* 26, 771–781. doi: 10.1109/tfuzz.2017.2688423
- Yang, X., and Wen, W. (2018). Ridge and Lasso Regression Models for Cross-Version Defect Prediction. *IEEE Trans. Reliabil.* 67, 885–896. doi: 10.1109/tr.2018.2847353
- Yetik, I. S., Nehorai, A., Lewine, J. D., and Muravchik, C. H. (2005). Distinguishing between moving and stationary sources using EEG/MEG measurements with an application to epilepsy. *IEEE Trans. Biomed. Eng.* 52, 471–479. doi: 10.1109/tbme.2004.843289
- Zhang, Y., Ishibuchi, H., and Wang, S. (2018). Deep takagi–sugeno–kang fuzzy classifier with shared linguistic fuzzy rules. *IEEE Trans. Fuzzy Syst.* 26, 1535–1549. doi: 10.1109/tfuzz.2017.2729507
- Zhang, Y., Zhou, Z., Bai, H., Liu, W., and Wang, L. (2020). Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization. *Front. Neurosci.* 11:496. doi: 10.3389/fnins.2020.00496
- Zheng, W. (2017). Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Trans. Cogn. Dev. Syst.* 9, 281–290. doi: 10.1109/tcds.2016.2587290
- Zhu, Z., Wang, Z., Li, D., Zhu, Y., and Du, W. (2020). Geometric structural ensemble learning for imbalanced problems. *IEEE Trans. Cybernet.* 50, 1617–1629. doi: 10.1109/tcyb.2018.2877663

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

Copyright © 2020 Jiang, Tang, Wang, Qiu, Zhang and Lin. 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.