Chin. Quart. J. of Math.
2022, 37(3): 281–300

EEG Feature Learning Model Based on Intrinsic Time-Scale Decomposition and Adaptive Huber Loss

YANG Li-jun[1,2], *JIANG Shu-yue*[1], *WEI Xiao-ge*[1], *XIAO Yun-hai*[1,2]

(1. School of Mathematics and Statistics, Henan Engineering Research Center for Artificial Intelligence Theory and Algorithms, Henan University, Kaifeng 475004, China; 2. Center for Applied Mathematics of Henan Province, Henan University, Zhengzhou 450046, China)

Abstract: According to the World Health Organization, about 50 million people worldwide suffer from epilepsy. The detection and treatment of epilepsy face great challenges. Electroencephalogram (EEG) is a significant research object widely used in diagnosis and treatment of epilepsy. In this paper, an adaptive feature learning model for EEG signals is proposed, which combines Huber loss function with adaptive weight penalty term. Firstly, each EEG signal is decomposed by intrinsic time-scale decomposition. Secondly, the statistical index values are calculated from the instantaneous amplitude and frequency of every component and fed into the proposed model. Finally, the discriminative features learned by the proposed model are used to detect seizures. Our main innovation is to consider a highly flexible penalization based on Huber loss function, which can set different weights according to the influence of different features on epilepsy detection. Besides, the new model can be solved by proximal alternating direction multiplier method, which can effectively ensure the convergence of the algorithm. The performance of the proposed method is evaluated on three public EEG datasets provided by the Bonn University, Childrens Hospital Boston-Massachusetts Institute of Technology, and Neurological and Sleep Center at Hauz Khas, New Delhi(New Delhi Epilepsy data). The recognition accuracy on these two datasets is 98% and 99.05%, respectively, indicating the application value of the new model.

Keywords: Epilepsy; EEG signals; Intrinsic time-scale decomposition; Huber loss function

2000 MR Subject Classification: 00A69

CLC number: O29 **Document code**: A

Article ID: 1002-0462 (2022) 03-0281-20

Received date: 2022-07-12

Foundation item: Supported by National Natural Science Foundation of China (Grant Nos. 11701144, 11971149); Henan Province Key and Promotion Special (Science and Technology) Project (Grant No. 212102310305).

Biographies: YANG Li-jun (1979-), female, native of Ruzhou, Henan, associate professor of Henan University, engages in intelligent information processing; XIAO Yun-hai (1978-), male, native of Puyang, Henan, professor of Henan University, engages in numerical optimization. Corresponding author YANG Li-jun: yanglijun@henu.edu.cn.

DOI: 10.13371/j.cnki.chin.q.j.m.2022.03.006

§1. Introduction

Epilepsy is a complex disease of human brain, which has a bad effect on patients' lives. The World Health Organization reports that about 50 million people worldwide suffer from epilepsy [55]. Fortunately, up to 70 percent of patients can survive if they are diagnosed and treated timely. Long-term irregular epileptic seizures not only bring physical pain to the patient, but also cause serious burden to the patient's family. Electroencephalogram (EEG) can indicate the brain's electrical activities associated with epilepsies [47], and it can accurately record the scattered slow spike or irregular spike signals during epileptic seizure. Hence, EEG can help doctors diagnose epilepsies and rationally use antiepileptic drugs. However, reading continuous EEG signals is a time-consuming and tedious process, and data analyzing is not efficient. As the number of channels increases, this problem becomes more pronounced. Therefore, it has become an urgent and vital task to develop a rapid and reliable method to assist neuroscientists in identifying epilepsies from EEG recordings. With the development of artificial intelligence, rapid progress has been made in the study of EEG signals, and the automatic detection of EEG signal has been deeply concerned by researchers.

EEG signal is non-stationary and nonlinear, which makes the automatic detection and analysis still a challenging problem. Time domain analysis was applied to directly extract feature waves in EEG signals that are similar to the epilepsy signal [34]. However, the information obtained from the time domain is hard to get the expected recognition accuracy. Therefore, some researchers used time-frequency analysis to extract the features of EEG signals [5, 31]. In recent years, some scholars have found new ways from the perspective of establishing or improving optimization models. Some optimization models can learn EEG features adaptively [26, 27]. And many algorithms have been developed to provide ideas for solving these optimization models [6, 28, 36, 44]. The extracted features from the optimization models usually show better performance than features extracted manually. By introducing different regularization terms to the optimization model, the EEG features corresponding to each dataset can be extracted. The regularization terms used popularly are L_1 norm, L_2 norm, or a combination of both. After obtaining EEG features through the optimization model, the automatic detection of epileptic seizures can be achieved by choosing appropriate classifiers, such as random forest (RF), support vector machine (SVM) and feedforward neural network (FNN) [7, 9, 22]. Deep learning methods are increasingly applied to EEG signals [10, 50]. The main advantage of deep learning models is that original data can be directly input into the network without feature extraction and the output is the detection result, that is, feature extraction and epilepsy detection process are integrated into a network. Deep convolutional neural networks (CNN), deep belief networks (DBN) and long short term memory networks (LSTM) have been applied to EEG signal recognition [18, 20, 51]. However, the training process of deep learning requires a large amount of training data, which is difficult to achieve in epilepsy detection. In order to obtain sufficient training data, some researchers have tried to use sliding windows to segment EEG signals, but the results were not satisfactory. In this paper, we propose an EEG feature learning model called Adaptive Huber(A-Huber), which combines adaptive weight with Huber loss function, and the experimental results show that the A-Huber model can obtain preferable recognition accuracy on different epilepsy datasets.

The organization of this study is as follows. Section 2 describes the datasets used in the experiments and introduces the relevant literature. Section 3 gives the proposed adaptive feature learning model and the algorithm for solving the model. Section 4 applies the proposed model to three public EEG datasets and analyzes the experimental results. Section 5 discuss the effects of adaptive weights. Section 6 gives the conclusion and future work.

§2. Dataset and literature review

2.1. Dataset

In this study, three EEG datasets are used to test the performance of the proposed model. They are all available public and come from Bonn University, Childrens Hospital Boston-Massachusetts Institute of Technology (CHB-MIT), and Neurological and Sleep Center at Hauz Khas(New Delhi Epilepsy data), and are commonly used in the recognition of epilepsy.

2.1.1. Bonn dataset Bonn dataset is collected through an amplifier system, which has 128 channels and digitized on 12-bit A/D converter. It consists of five subsets from different classes, denoted as A, B, C, D, and E. Each subset contains 100 single-channel EEG signals, the duration of every recording is 23.6 seconds and the sampling frequency is 173.61 Hz. A summary description of the dataset is provided in Table1. Please refer to [4] for more detailed description of the Bonn dataset.

Table 1 Description of Bonn dataset.

	Set A	**Set B**	**Set C**	**Set D**	**Set E**
Subjects	Healthy	Healthy	Epileptic	Epileptic	Epileptic
State	Eyes opened	Eyes closed	Interictal	Interictal	Ictal
Electrode placement	International 10-20 system	International 10-20 system	Opposite to epileptogenic zone	Within epileptogenic zone	Within epileptogeniz zone
Number	100	100	100	100	100
Time duration	23.6s	23.6s	23.6s	23.6s	23.6s
Sample rate	173.61 Hz	173.61 Hz	173.61 Hz	173.61 Hz	173.61 Hz

2.1.2. CHB-MIT dataset The CHB-MIT dataset[1] consists of 24 EEG signal sets from two groups of subjects. The first group recorded EEG signals from 22 subjects (5 males, ages

[1]https://physionet.org/content/chbmit/1.0.0/

3-22; 17 females, ages 1.5-19), and a total of 23 EEG signal sets were obtained, where chb21 was the data of chb01 after 1.5 years. The second group recorded the EEG signals of one subject. Each subset contained between 9 and 42 consecutive edf files. The sampling rate of all signals is 256Hz and the resolution is 16 bits. This data was recorded by using an international 10-20 EEG electrode location and naming system.

2.1.3. New Delhi epilepsy dataset The dataset[2] consists of the EEG signals of 10 epilepsy patients from Neurology and Sleep Centre, Hauz Khas, New Delhi, and was obtained by using Grass Telefactor Comet AS40 Amplification System. The signals were filtered by a bandpass filter with a frequency range of 0.5-70 Hz and then segmented into preictal, interictal and ictal stages. Every folder contains fifty MAT-files of EEG time series signals, and each MAT-file consists of 1024 samples of one EEG time series data with a duration of 5.12 seconds. A summary description of the dataset is provided in Table 2.

Table 2 Description of new Delhi epilepsy data.

	Interictal	**Preictal**	**Ictal**
Number	50	50	50
Time duration	5.12s	5.12s	5.12s
Sample rate	200 Hz	200 Hz	200 Hz
Placement	10-20 system	10-20 system	10-20 system

2.2. Literature review

Decomposition is a technique widely used in EEG feature extraction. Gotman [16] proposed to decompose all EEG signals into several half-wave signals. Then the average amplitude, average duration and coefficient of the decomposed half-wave signals are extracted as the features for epilepsy recognition. Polat [38] applied fast Fourier transform to EEG signals and calculated power spectral density (PSDs) as the features of EEG signals. They considered a binary classification problem (AvsE on Bonn dataset) and chose decision tree (DT) classifier. Finally, the accuracy reached 98.72%. Li [32] used wavelet transform and envelope analysis to extract EEG features. The neural network ensemble was used as the classifier and the accuracy of 98.78% was obtained. In recent years, Gupta and Pachori [19] proposed the Fourier-Bessel series expansion method (FBSE), used to extract the rhythm features of EEG signals based on weighted multiscale Renyi permutation entropy (WMRPE). The highest accuracy of the binary classification problems of AvsE, BvsE, CvsE and DvsE on Bonn dataset were 99.5%, 99.5%, 99.5% and 97.5%, respectively. Besides, Empirical mode decomposition (EMD) and intrinsic

[2]https://www.researchgate.net/publication/308719109

time-scale decomposition (ITD) were also applied to EEG characteristic extraction [14,23]. Both EMD and ITD can decompose EEG signals into multiple components of different frequencies. For every frequency component, its instantaneous amplitude and instantaneous frequency are suitable to further extract features since they contain a lot of physiological and pathological information [33, 40, 56]. However, these decomposition techniques have some shortcomings, such as end-effects, mode mixing, and non-robustness to noise, which limit their applications [21,57,58].

The performance of classification based on decomposition methods depends on manual extraction of features. However, the features extracted manually have subjective consciousness which limit its robustness. Therefore, researchers begin to consider automatic feature learning from EEG data. Hussein et al. [27] applied LASSO model to EEG feature learning, and conducted experiments to test the performance of the model. Moreover, they noticed that the LASSO model can not maintain high levels of performance with severe noise. The reason is that when there are outliers in the samples, the L_2 loss function will give higher weights to the outliers, which reduces the feature learning ability of the model. Hence, Hussein et al. [26] proposed a robust feature extraction model, which replaced the L_2 loss function in LASSO model with L_1 loss function. The experimental results showed that the model can perform well even under severe noise levels. L_1 loss function can effectively alleviate the influence of sample outliers. However, all components of the regression coefficient vector are punished to the same extent in the two models mentioned above, so that the contributions of different features are ignored. Based on this, we construct a feature learning model called adaptive Huber (A-Huber) in this paper. On the one hand, the adaptive weights of the regression coefficients are considered in the new model. On the other hand, the A-Huber model employs the Huber loss function to combine the advantages of L_1 and L_2 loss functions. Experiments show that the A-Huber model has good performance.

§3. Methodology

3.1. The Huber loss function

In regression analysis, the L_2 loss function is susceptible to outliers, which will affect the performance of regression prediction. In order to weaken the influence of outliers on regression results, Edgeworth et al. [13] proposed the regression estimation based on L_1 loss. Further, Huber proposed a robust loss function [24], which can effectively balance L_2 loss and L_1 loss. For an ginven positive real number c, the Huber loss function $H(u)$ is as follows:

$$H(\mathrm{u})=\begin{cases}\frac{1}{2}\mathrm{u}_\mathrm{i}^2, & |\,\mathrm{u}_\mathrm{i}\,|\le \mathrm{c},\\ c\,|\,\mathrm{u}_\mathrm{i}\,|-\frac{1}{2}\mathrm{c}^2, & |\,\mathrm{u}_\mathrm{i}\,|> \mathrm{c}.\end{cases} \tag{3.1}$$

From the definition (3.1), it can be seen that the Huber function is quadratic when u is less than or equal to the parameter c, and it grows linearly when u is greater than parameter c. The parameter c describes where the transition from quadratic function to linear function takes place. In 1973, Huber proved the large sample properties of Huber's regression coefficient estimators [25]. Huber loss function integrates the advantages of L_2 loss and L_1 loss, which is more robust to outliers.

3.2. Adaptive Huber EEG feature learning model

Hui Zou proposed adaptive Lasso to solve the bias problem by adding adaptive weights to L_1 penalty term of the coefficients and gave the following criterion in 2006 [63].

$$Q^{adl}(\beta)=\left\|y-\sum_{j=1}^{d}x_j\beta_j\right\|^2+\lambda\sum_{j=1}^{d}\omega_j^{adl}|\beta_j|, \tag{3.2}$$

where y is the response value, $\{x_j\}_{j=1}^d$ is the sample, $\lambda\geq 0$ is a regularization parameter, $\beta=\{\beta_j\}_j$ is the calculated regression coefficient, and ω^{adl} is the weight vector of the regression coefficients. $\omega_j^{adl}(j=1,2,\cdots,d)$ are usually set the reciprocal of the least square coefficients. The loss function in adaptive LASSO model is L_2 norm function. Lacroix et al. [48] combined Huber criterion with adaptive LASSO and proposed Huber adaptive LASSO (H-AD-LASSO) criterion in 2011, as shown below:

$$Q^{Hadl}(\alpha,\beta,s)=L_H(\alpha,\beta,s)+\lambda\sum_{j=1}^{d}\omega_j^{Hadl}|\beta_j|, \tag{3.3}$$

and

$$L_H(\alpha,\beta,s)=\begin{cases} ns+\displaystyle\sum_{i=1}^{n}H_M\left(\frac{y_i-\alpha-x_i^T\beta}{s}\right), & if\quad s>0,\\ 2M\displaystyle\sum_{i=1}^{n}|y_i-\alpha-x_i^T\beta|, & if\quad s=0,\\ +\infty, & if\quad s<0, \end{cases} \tag{3.4}$$

where, $H_M(\cdot)$ represents the Huber loss function, s is the data measurement scale, and α represents the regression intercept from the data. ω^{Hadl},β have the same meanings as ω^{adl},β in criterion (3.2). In the H-AD-LASSO criterion, Huber loss function is combined with adaptive LASSO, and the advantages of L_2 loss and L_1 loss are combined to make β sparse. Therefore, the optimization model based on the H-AD-LASSO criterion has strong robustness. Although the optimization model has several good properties, it is complex and hard to be applied in applications. Based on this, we propose a feature learning model of EEG signals in this study, called A-Huber. Our innovation is to improve the model proposed by Hussion [26], simplify the model (3.3), and apply it to seizure detection. Experimental results show that Huber loss function combined with adaptive weight method is effective in the analysis of epileptic EEG signals. The flow chart of the experiment in this paper is presented in Figure 1, including preprocessing of EEG signals, feature learning and epilepsy detection based on A-Huber model. We will describe each part in detail in the following.

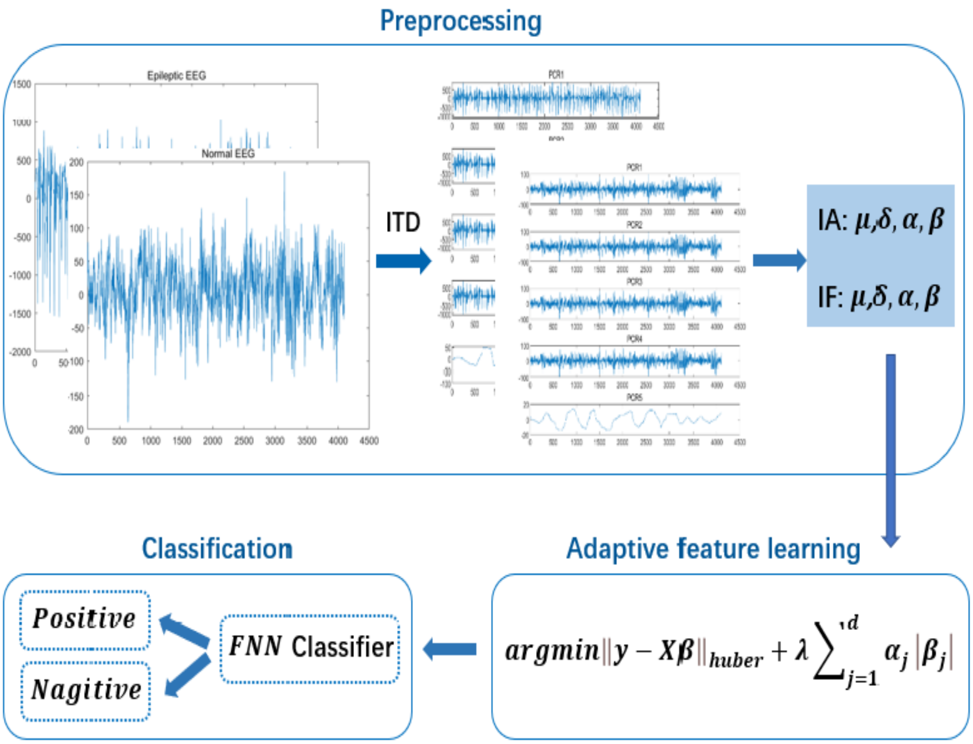

Fig. 1 The flow chart of epilepsy recognition proposed in this study, which includes three step: preprocessing of EEG signals, A-Huber model for feature learning, and epilepsy detection.

3.2.1. Preprocessing EEG signals are non-stationary and no-nlinear, and they fluctuate greatly under the influence of time. ITD is a time-frequency analysis method commonly used in the field of signal processing, and it is an effective tool for processing non-stationary and nonlinear signals [14]. It considers linear operator and single-layer iteration, and can process data in real time and analyze data effectively. In this study, we use ITD to decompose each EEG signal into a set of proper rotation components (PRCs) in descending order of frequency. The PRCs contain a large amount of pathological and physiological information of the original EEG signals. Then, we calculate the mean, variance, skewness, and kurtosis of each PRC instantaneous amplitude and frequency as the EEG features. These indexes are defined as follows:

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i, \qquad \delta^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \mu)^2,$$

$$s = \frac{1}{\delta^3}\sum_{i=0}^{N-1}(x_i - \mu)^3, \quad k = \frac{1}{\delta^4}\sum_{i=0}^{N-1}(x_i - \mu)^4,$$

where x_i corresponds to the value at the ith point of instantaneous frequency or instantaneous amplitude of each PRC, and N is the length of each PRC. EEG signals contains different oscillation modes. Since the EEG signals are usually decomposed into five bandwidths based on frequency range [52], we choose the first five PRCs for further feature learning. Figure 2 shows an example, where an EEG signal is decomposed into five PRCs by ITD.

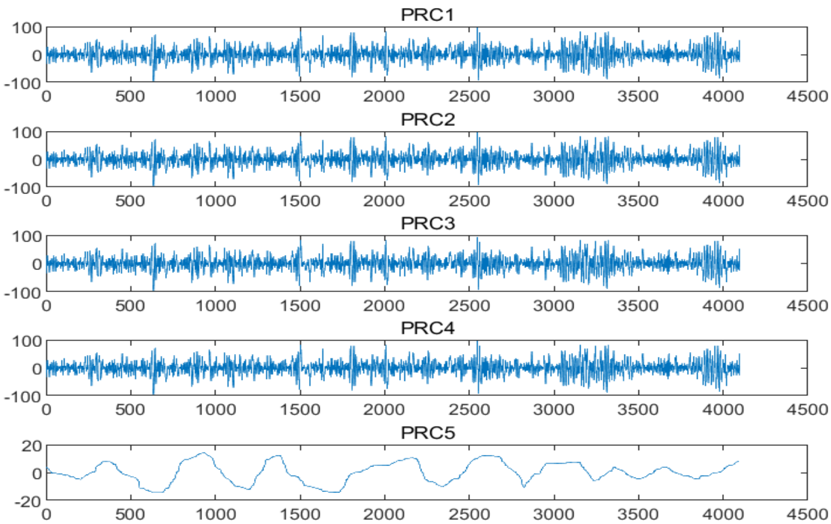

Fig. 2 The PRCs of an EEG signal decomposed by ITD.

3.2.2. A-Huber In this study, we propose the A-Huber feature learning model. The A-Huber model is an unconstrained optimization problem which is defined as follows:

$$\hat{\beta}(\lambda)=\arg\min_{\beta}\|y-X\beta\|_{huber}+\lambda\sum_{j=1}^{d}\alpha_j|\beta_j|, \tag{3.5}$$

where $X\in R^{n\times d}$ and can be expressed as $X=[x_1;x_2;\ldots;x_n]$. Each x_i is a sample which has been preprocessed by ITD. n represents the number of EEG signals in X, and d represents the dimension of x_i. Since x_i is a stack of four statistical indicators of the amplitudes and frequencies of all five PRCs, $d=40$. $\lambda\geq 0$ is the regularization parameter, and a sparse solution can be obtained when the value of λ is large.

$$\alpha_j=\frac{1}{|\hat{\beta}_j^1|^{\sigma}}$$

is the weight of the jth regression coefficient, $\sigma>0$ is an adjustable parameter, we take $\sigma=1$ in this paper. $\hat{\beta}_j^1$ is the solution of $\arg\min_\beta\|y-X\beta\|_2^2$. The vector y is defined as the label vector corresponding to X. We write X and y as:

$$X=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \quad \& \quad y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}. \tag{3.6}$$

The optimization model (3.5) can be simplified as follows.

$$\begin{aligned} \hat{\beta}(\lambda) &= \arg\min_{\beta} \|y-X\beta\|_{huber}+\lambda\sum_{j=1}^{d}\alpha_j|\beta_j| \\ &= \arg\min_{\beta} \|y-X\beta\|_{huber}+\lambda\sum_{j=1}^{d}|\alpha_j\beta_j| \\ &= \arg\min_{\beta^*} \|y-X^*\beta^*\|_{huber}+\lambda\|\beta^*\|, \end{aligned} \tag{3.7}$$

and

$$X^*=\begin{pmatrix} \alpha_1^{-1}x_{11} & \alpha_2^{-1}x_{12} & \cdots & \alpha_d^{-1}x_{1d} \\ \alpha_1^{-1}x_{21} & \alpha_2^{-1}x_{22} & \cdots & \alpha_d^{-1}x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{-1}x_{n1} & \alpha_2^{-1}x_{n2} & \cdots & \alpha_d^{-1}x_{nd} \end{pmatrix} \& \beta^*=(\alpha_1\beta_1,\alpha_2\beta_2,\cdots,\alpha_d\beta_d)^T. \tag{3.8}$$

Then the model (3.7) is equivalent to the following form:

$$\hat{\beta}^*=\arg\min_{\beta^*} H(y-X^*\beta^*)+\lambda\|\beta^*\|_1, \tag{3.9}$$

where $H(\cdot)$ is the Huber function, $X^*\in R^{n\times d}$ and $y\in R^{n\times 1}$. If the solution $\hat{\beta}^*$ of the optimization problem (3.9) is obtained, then the solution of (3.5) is $\hat{\beta}:=\hat{\beta}^*/\alpha$.

Solution

The optimization problem (3.9) can be reformulated as the following constrained optimization problem:

$$\begin{aligned} &\hat{\beta}^*=\arg\min_{u,\beta} H(u)+\lambda\|\beta\|_1, \\ &s.t. \quad u=y-X^*\beta. \end{aligned} \tag{3.10}$$

Note that the Huber loss function is differentiable, while the L_1 norm is not differentiable. Therefore, this problem cannot be solved by the gradient descent method, Newton method and so on. We can employ the methods such as proximal alternating direction multiplier method (P-ADMM), proximal gradient methods [6] and block coordinate descent [36]. P-ADMM is an improvement of ADMM which converges faster and has smaller errors. Therefore, P-ADMM is used to solve problem (3.9) in this paper.

Consider the augmented lagrange function of (3.10):

$$L_A(\beta,u,z)=\lambda\|\beta\|_1+H(u)-z^T(X^*\beta+u-y)+\frac{\eta}{2}\|X^*\beta+u-y\|_2^2. \tag{3.11}$$

where $z\in R^{n\times 1}$ is the lagrange multiplier, and $\eta>0$ is the penalty parameter.

(1) Firstly, fix variables $u=u_k, z=z_k$, solve the problem (3.11) about β, we have:

$$\begin{aligned}\beta_{k+1}&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1-z_k^T(X^*\beta+u_k-y)+\frac{\eta}{2}\|X^*\beta+u_k-y\|_2^2\\&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1-z_k^T(X^*\beta+u_k-y)+\frac{\eta}{2}\|X^*\beta+u_k-y\|_2^2+\frac{1}{2\eta}z_k^Tz_k\\&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1+\frac{\eta}{2}\|X^*\beta+u_k-y-\frac{z_k}{\eta}\|_2^2.\end{aligned}\tag{3.12}$$

Since the analytic solution of (3.12) is hard to be obtained, we can compute the first-order Taylor expansion of the second term at β_k as an approximation. Suppose g_k is the gradient of $\frac{1}{2}\|X^*\beta+u_k-y-\frac{z_k}{\eta}\|_2^2$ at $\beta=\beta_k$, i.e.,

$$g_k=X^{*T}(X^*\beta_k+u_k-y-\frac{z_k}{\eta}).$$

The formula (3.12) can be reformed in the following form:

$$\begin{aligned}\beta_{k+1}&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1+\frac{\eta}{2}\|X^*\beta+u_k-y-\frac{z_k}{\eta}\|_2^2\\&\approx\arg\min_{\beta\in R^d}\lambda\|\beta\|_1+\eta(g_k^T(\beta-\beta_k)+\frac{1}{2\tau}\|\beta-\beta_k\|_2^2)\\&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1+\frac{\eta}{2\tau}(2\tau g_k^T(\beta-\beta_k)+\|\beta-\beta_k\|_2^2+\tau^2g_k^Tg_k)\\&=\arg\min_{\beta\in R^d}\lambda\|\beta\|_1+\frac{\eta}{2\tau}\|\beta-(\beta_k-\tau g_k)\|_2^2\\&=\mathrm{Shrink}(\beta_k-\tau g_k,\frac{\lambda\tau}{\eta})\\&=max(|\beta_k-\tau g_k|-\frac{\lambda\tau}{\eta},0)\frac{\beta_k-\tau g_k}{|\beta_k-\tau g_k|}.\end{aligned}\tag{3.13}$$

where $|\cdot|$ and max represent the absolute value and maximum, respectively. We define $0\cdot 0/0=0$.

(2) Secondly, fix variables $\beta=\beta_{k+1}$, $z=z_k$, minimize problem (3.11) about u, we can obtain the following formula:

$$\begin{aligned}u_{k+1}&=\arg\min_{u\in R^n}H(u)-z_k^T(X^*\beta_{k+1}+u-y)+\frac{\eta}{2}\|X^*\beta_{k+1}+u-y\|_2^2\\&=\arg\min_{u\in R^n}H(u)+\frac{\eta}{2}\|X^*\beta_{k+1}+u-y-\frac{z_k}{\eta}\|_2^2.\end{aligned}\tag{3.14}$$

Variables in (3.14)can be separated and each element of u can be obtained as:

$$u_{k+1}^i=\arg\min_{u^i\in R}H(u^i)+\frac{\eta}{2}(u^i-(y^i+\frac{z_k^i}{\eta}-(X^*\beta_{k+1})^i))^2.\tag{3.15}$$

To give a solution for (3.15), define an optimization operator as shown in Formula (3.16), where $\alpha>0$ is a constant.

$$\mathrm{Prox}_\rho(\xi,\alpha)=\arg\min_{\lambda\in R}H(\lambda)+\frac{1}{2}\alpha(\lambda-\xi)^2.\tag{3.16}$$

By simple calculation we can get the following formula.

1.When $|\lambda|\le c$, we have

$$\arg\min_{\lambda\in R} H(\lambda)+\frac{1}{2}\alpha(\lambda-\xi)^2=\arg\min_{\lambda\in R}(\frac{1}{2}+\frac{1}{2}\alpha)\lambda^2-\alpha\xi\lambda$$

$$\mathrm{Prox}_\rho(\xi,\alpha)=\begin{cases}-c, & \frac{\alpha\xi}{1+\alpha}<-c,\\ \frac{\alpha\xi}{1+\alpha}, & -c\le\frac{\alpha\xi}{1+\alpha}\le c,\\ c, & \frac{\alpha\xi}{1+\alpha}>c.\end{cases}\tag{3.17}$$

2.When $\lambda>c$, we have

$$\arg\min_{\lambda\in R} H(\lambda)+\frac{1}{2}\alpha(\lambda-\xi)^2=\arg\min_{\lambda\in R}(\frac{\alpha}{2}\lambda^2+c\lambda-\alpha\xi\lambda)$$

$$\mathrm{Prox}_\rho(\xi,\alpha)=\begin{cases}\frac{\alpha\xi-c}{\alpha}, & \frac{\alpha\xi-c}{\alpha}\ge c,\\ c, & \frac{\alpha\xi-c}{\alpha}<c.\end{cases}\tag{3.18}$$

3.When $\lambda<-c$, we have

$$\arg\min_{\lambda\in R} H(\lambda)+\frac{1}{2}\alpha(\lambda-\xi)^2=\arg\min_{\lambda\in R}(\frac{\alpha}{2}\lambda^2-c\lambda-\alpha\xi\lambda)$$

$$\mathrm{Prox}_\rho(\xi,\alpha)=\begin{cases}\frac{\alpha\xi+c}{\alpha}, & \frac{\alpha\xi+c}{\alpha}\le -c,\\ -c, & \frac{\alpha\xi-c}{\alpha}>-c.\end{cases}\tag{3.19}$$

Then, we can obtain the expression of the optimization operator $\mathrm{Prox}_\rho(\xi,\alpha)$ explicitly through equations (3.17), (3.18) and (3.19).

$$\mathrm{Prox}_\rho(\xi,\alpha)=\begin{cases}\frac{\alpha\xi+c}{\alpha}, & \frac{\alpha\xi+c}{\alpha}<-c,\\ \frac{\alpha\xi}{1+\alpha}, & -c\le\frac{\alpha\xi}{1+\alpha}\le c,\\ \frac{\alpha\xi-c}{\alpha}, & \frac{\alpha\xi-c}{\alpha}>c.\end{cases}\tag{3.20}$$

Let

$$\xi=y^i+\frac{z_k^i}{\eta}-(X\beta_k)^i,\quad \alpha=\eta,\quad \lambda=u_i,$$

and then we can get

$$\begin{aligned}u_{k+1}^i&=\mathrm{Prox}_\rho(\xi,\alpha)\\ &=Prox_\rho(y^i+\frac{z_k^i}{\eta}-(X^*\beta_k)^i,\eta),\quad i\epsilon[n].\end{aligned}\tag{3.21}$$

(3) Finally, fix variables $u=u_{k+1}$, $\beta=\beta_{k+1}$, update the lagrange multiplier z_{k+1}

$$z_{k+1}=z_k-\gamma\eta(X^*\beta_{k+1}+u_{k+1}-y). \tag{3.22}$$

where $\gamma>0$ is a constant. Therefore, the P-ADMM algorithm for solving the problem of Huber-L_1 is as follows:

Table 3 Iterative process.

P-ADMM algorithm
Input: X; y; λ; η; τ; max; $\epsilon=1e-6$;
Step 1. Initialization: Given initial value $u=u_0$; $z=z_0$; $k=0$;
Step 2. While not converge, do
$k\leq max$ & $\|\beta_{k+1}-\beta_k\|^2>\epsilon$;
$\beta_{k+1}\leftarrow Shrink(\beta_k-\tau g_k,\frac{\lambda\tau}{\eta})$;
$u^i_{k+1}\leftarrow Prox_\rho(y^i+\frac{z^i_k}{\eta}-(X^*\beta_k)^i,\eta)$;
$z_{k+1}\leftarrow z_k-\gamma\eta(X^*\beta_{k+1}+u_{k+1}-y)$;
End while
Step 3: Output iteration result: β_k.

3.3. EEG classification

After obtaining the EEG features learned from the A-Hubr model, RF, SVM, and FNN can be used as classifiers for detection tasks. In subsection 4.1, the A-Hubr model is combined with three classifiers to detect epilepsy, and we found that FNN the best performance among the three classifiers (please refer to Table 4 for detailed experimental results). As consequence, we choose FNN as the classifier and compare the experimental results with the published studies in other experiments.

§4. Experimental studies

In this section, three kinds of experiments will be conducted to test the performance of the A-Huber model. The representation ability of the A-Huber model is tested through different classifiers. Moreover, we also compare the proposed method with other methods in the literature. In the experiments, we use the well-known performance metrics: sensitivity (Sen), specificity (Spe), and accuracy (Acc) [2, 26, 43], which are defined as follows:

$$Sen=\frac{TP}{FN+TP}\times 100\%, \tag{4.1}$$

$$Spe=\frac{TN}{TN+FP}\times 100\%, \tag{4.2}$$

$$Acc = \frac{TP+TN}{FN+TP+TN+FP} \times 100\%, \tag{4.3}$$

where TP represents the number of true positive samples whose predicted results by the classifier are also true positive. Similarly, TN, FP and FN represent the number of true negative, false positive and false negative samples, respectively. Besides, the ten fold cross validation technique is used in the experiments [54]. We randomly divide the initial sample dataset D into 10 disjoint subsets, i.e.,

$$D = D_1 \cup D_2 \cup \cdots \cup D_{10}, D_i \cap D_j = \varnothing (i \neq j),$$

where $\varnothing$ is an empty set. Then the model is trained by 9 subsets and tested by the remaining one subset. As a result, we obtain 10 classification results, whose average will be employed to evaluate the performance.

4.1. Experiment 1: Performance test of the feature representation

To test the feature representation performance of the A-Huber model, we combine the A-Huber features with three classifiers (RF, SVM, and FNN), and perform the two-classification task of epileptic EEG signals. The sensitivities, specificities, and accuracies of the experimental results are shown in Table 4. It can be seen from the experimental results that all accuracies are higher than 80%, which shows that the A-Huber model can effectively learn the discriminant features for epilepsy detection. In addition, we also find that the FNN classifier has the highest accuracy among the listing classifiers. Therefore, we will only use the combination of the proposed model and FNN to compare with the methods in the literature.

Table 4 Classification results with different classifier(Sen-Spe-Acc)(%).

Problems	SVM	RF	FNN
A vs E	100-98-99	99-98-98.5	**100-100-100**
C vs E	95-99-97	96-97-96.5	**99-99-99**
B vs E	94-100-97	98-98-88	**99-100-99.5**
D vs E	88-90-89	91-85-88	**98.5-99.0-98.0**
CHB-MIT 1	98.81-99.76-99.29	98.33-99.76-99.05	**100-100-100**
Ictal vs Interictal	97 -94 -100	96-97-96.5	**100-100-100**
Preictal vs Ictal	96 - 98- 97	98-96-97	**98-99-98.5**
Interictal vs Preictal	84- 90- 82	84- 90-85	**86-94-90**

4.2. Experiment 2: Comparison with the state-of-the-art methods

In this section, we compare the proposed epilepsy detection methods with the state-of-the-art methods on three public datasets.

In the Bonn dataset, we perform four experiments, which are A vs E, B vs E, C vs E and D vs E. Table 5 shows the comparison results with different methods. According to the results, we can see that the proposed A-Huber model can learn the discriminative features and has better classification performance.

Table 5 Comparison of Bonn two classification results.

Classes	Method	Classifier	Sen (%)	Spe (%)	Acc (%)
AvsE	Kumar2014Machine [30]	SVM	98.00	96.00	97.5
	Admin2016A [3]	RUSBoost	-	-	97.87
	Diykh2017Classify [12]	SVM	-	-	100
	Raghu2019A [39]	MLP	-	-	99.45
	Hussein2019Robust [26]	RF	100	99.0	99.50
	Sameer2020Wireless [42]	RF	-	-	98
	A-Huber	**FNN**	**100**	**100**	**100**
BvsE	Supriya2016Weighted [41]	SVM	-	-	97.25
	Diykh2017Classify [12]	SVM	-	-	99.76
	Peng2019novel [37]	DLWH	100	95	97.50
	Raghu2019A [39]	MLP	-	-	96.06
	Sameer2020Wireless [42]	RF	-	-	96
	A-Huber	**FNN**	**99.00**	**100.00**	**99.50**
CvsE	Samiee2014Rational [41]	MLP	98.5	99.3	97.7
	Diykh2017Classify [12]	SVM	-	-	96
	Raghu2019A [39]	MLP	-	-	97.60
	Peng2019Tnovel [37]	DLWH	99	98	100
	A-Huber	**FNN**	**99.00**	**99.00**	**99.00**
DvsE	Samiee2014Rational [41]	MLP	94	100	88
	Diykh2017Classify [12]	SVM	-	-	93.7
	Raghu2019A [39]	MLP	-	-	97.60
	Sameer2020Wireless [42]	Adaboost	-	-	95.5
	A-Huber	**FNN**	**98.50**	**99.00**	**98.00**

For CHB-MIT dataset, the data amount of two classes are highly unbalanced since there has significant difference in the length of EEG signals recorded in epileptic period and non-epileptic period. Hence, the continuous EEG signals are cut into some segments with the length of 256 sampling points, and some non-epileptic data samples are discarded so that the epileptic sample size is the same as the size of non-epileptic samples. The experimental results presented in Table 6 show that the average accuracy, sensitivity, and specificity of the proposed method are 99.05%, 99.11%, and 98.97%, respectively. Compared with the state-of-the-art methods, the proposed methods has better performance.

Table 6 Comparison of CHB-MIT two classification results.

Method	Number of subjects	Classifier	Acc (%)	Sen (%)	Spe (%)
Ahammad2014Detection [35]	24	Linear classifier	84.2	-	98.5
Gill2015Analysis [15]	12	GMM	86.93	86.26	87.58
Khan et al2017 [29]	/	WT and CNN	/	87.8	/
Chen et al2017 [11]	18	DWT	/	91.71	92.89
Truong et al2018 [53]	13	STFT and CNN	/	81.2	/
Peng2019A [37]	24	DLWH	95.06	94.33	95.06
Zhang2021EEG [60]	23	STFT+CNN	98.46	-	-
A-Huber	24	**FNN**	**99.05**	**99.11**	**98.97**

Experimental results show that the proposed model can further improve recognition performance and obtain the best the accuracy, sensitivity and specificity, and the comparison results are shown in Table 7.

Table 7 Comparison of New Delhi Epilepsy Dataset two classification results.

Classes	Method	Classifier	Acc (%)	Sen (%)	Spe (%)
Ictal vs Interictal	Sameer2020Wireless [42]	SVM	98	-	-
	P.Swami2019Selection [49]	ENE,STD	98.12	97.49	98.74
	A-Huber	**FNN**	**100**	**100**	**100**
Preictal vs Ictal	Sameer2020Wireless [42]	SVM	97	-	-
	A-Huber	**FNN**	**98.5**	**98**	**99**
Interictal vs Preictal	Sameer2020Wireless [42]	SVM	71	-	-
	A-Huber	**FNN**	**90**	**86**	**94**

4.3. Experiment 3: Comparison with other optimization models

The comparison methods listed in Table 8 are feature learning methods for epilepsy detection, which are all considered from the perspective of optimization model. Yuan et al. [59] proposed the kernel sparse representation classification (KSRC) model, which was an improvement on the traditional SRC model. They performed two-class experiments (AvsE and DvsE) on Bonn dataset to test the validity of their model. The accuracy is 98.63% by using Gaussian kernel. Peng et al. [37] combined dictionary learning with sparse representation (DLSR) for epilepsy detection. They also considered the two-class classification and the accuracies were varying from 97.5% to 100%. Hussein et al. [26] extracted EEG features by using L_1-penalized robust regression and classified by using the RF classifier. On the two-class problem AvsE, they reached the accuracy of 99.50% with 10-fold cross-validation method. Compared with these optimization models, the A-Huber shows better performance, which obtains the accuracies varying from 98% to 100% in two-class problems.

Table 8 Comparison the A-Huber model with other optimization models (Acc)(%).

Optimization model	AvsE	BvsE	CvsE	DvsE	CHB-MIT
KSRC [59]	98.63	-	-	98.63	-
L_1PR [26]	99.5	-	-	-	-
DLSR [37]	100	97.5	99	99.5	-
A-Huber	**100**	**99.5**	**99**	**98**	**99.05**

§5. Discussion

Experimental results illustrate that the A-Huber model can perform well in EEG-based epilepsy detection. By using A-Huber model to learn an adaptive weight for feature learning, we find that this model has good classification results and can learn discriminant features for detecting epilepsy. The effectiveness of A-Huber model is verified by comparing with recent published recognition results in different EEG datasets.

The A-Huber model is similar to previous studies by using some optimization models to learn the features of epilepsy. Adaptive weight plays a key role in model structure. Hussein et al. [26] selected features from the perspective of sparsity, while the proposed model in this study further considered the advantages of L_1 loss function and L_2 loss function. Therefore, the A-Huber model proposed in this paper can make full use of the structural information of EEG signals, which combines the advantages of adaptive weight and Huber loss function.

§6. Conclusion and future work

A feature learning model (A-Huber) for seizure detection is proposed in this paper, which is based on adaptive weight and Huber loss function. The A-Huber model learns an adaptive weight by using ITD preprocessed EEG signals, so as to carry out feature selection of EEG signals. In addition, the proposed model can be solved by the P-ADMM algorithm, which effectively ensures the convergence. The model extracts discriminant features from EEG signals and performs well when using different classifiers.

The limitation of this method is that, although P-ADMM algorithm can ensure convergence and high operation efficiency, it has many adjustable parameters, and the fluctuation of parameters has a great impact on the model solution. Therefore, other methods can be considered to optimize the algorithm or further optimize the model. In addition, this paper mainly carries out the classification experiments of two classes, and multi-classification experiments can be studied in the future.

Acknowledgements

We would like to express our sincere gratitude to anonymous reviewers for their valuable suggestions and comments. The authors would like to thank the journal editor and anonymous reviewers for their constructive comments. This work was supported by the National Natural Science Foundation of China (grant numbers 11701144, 11971149), and Programs for Science and Technology Development of Henan Province (grant numbers 212102310305).

[References]

[1] ACHARYA U R, SREE S V, ALVIN A P C, SURI J S. Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework[J]. Expert Syst. Appl., 2012, 39(10): 9072-9078.

[2] ALTMAN D G, BLAND J M. Diagnostic tests 1: Sensitivity and specificity[J]. BMJ: British Medical Journal (International Edition), 1994, 308(6943): 1552-1552.

[3] AMIN S, KAMBOH A M. A robust approach towards epileptic seizure detection[C]. In Proceedings of the Twenty-sixth IEEE International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2016, 1-6.

[4] ANDRZEJAK R G, LEHNERTZ K, MORMANN F, RIEKE C, DAVID P, ELGER C E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state[J]. Phys. Rev. E Stat. Nonlin. Soft Matter Phys., 2001, 64(6): 061907.

[5] BIJU K S, HAKKIM H A, JIBUKUMAR M G. Ictal EEG classification based on amplitude and frequency contours of IMFs[J]. Biocybern. Biomed. Eng., 2017, 37(1): 172-183.

[6] BOTTOU L. Large-scale machine learning with stochastic gradient descent[J]. Proceedings of COMPSTAT' 2010, 2010, 16: 177-186.

[7] BREIMAN L. Random forest[J]. Mach. Learn., 2001, 45: 5-32.

[8] BUGEJA S, GARG L, ELISHA A E. A novel method of EEG data acquisition, feature extraction and feature space creation for early detection of epileptic seizures[C]. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, 837-840.

[9] CHANDAKA S, CHATTERJEE A, MUNSHI S. Cross-correlation aided support vector machine classifier for classification of eeg signals[J]. Expert Syst. Appl., 2009, 36(2, Part 1): 1329-1336.

[10] CHAO H, ZHI H L, DONG L, LIU Y L. Recognition of emotions using multichannel EEG data and DBN-GC-based ensemble deep learning framework[J]. Comput. Intel. Neurosc., 2018, 2018: 1-11.

[11] CHEN D, WAN S R, XIANG J, BAO F S. A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG[J]. Plos One, 2017, 12(3): e0173138.

[12] DIYKH M, LI Y, WEN P. Classify epileptic EEG signals using weighted complex networks based community structure detection[J]. Expert Syst. Appl., 2017, 90(dec.30): 87-100.

[13] EDGEWORTH F Y. On observations relating to several quantities[J]. Hermathena, 1887, 6: 279-285.

[14] FREI M G, OSORIO I. Intrinsic time-scale decompositon: time-frequency-energy analysis and real-time filtering of non-stationary signals[J]. P. Roy. Soc. A-Math. Phy. 2007, 463(2): 321-342.

[15] GILL A F, FATIMA S A, AKRAM M U, KHAWAJA S G, AWAN S E. Analysis of EEG signals for

detection of epileptic seizure using hybrid feature set[J]. Lecture Notes in Electrical Engineering, 2015, 344: 49-57.

[16] GOTMAN J. Automatic recognition of epileptic seizures in the EEG[J]. Electroencephalogr. Clin. Neurophysiol., 1982, 54(5): 530-540.

[17] GOTMAN J, FLANAGAN D, ZHANG J, ROSENBLATT B. Automatic seizure detection in the newborn: methods and initial evaluation[J]. Electroencephalogr. Clinical Neurophysiol., 1997, 103(3): 356-362.

[18] GRAVES A. Long short-term memory[J]. Springer Berlin Heidelberg, 2012, 385: 37-45.

[19] GUPTA V, PACHORI R B. Epileptic seizure identification using entropy of FBSE based EEG rhythms[J], Biomed. Signal Proces., 2019, 53: 101569.

[20] HINTON G E. Deep belief networks[J]. Scholarpedia, 2009, 4(6): 5947.

[21] HU X Y, PENG S L, HUANG W L. EMD revisited: A new understanding of the envelope and resolving the mode-mixing problem in am-fm signals[J]. IEEE T. Signal Proces., 2012, 60(3): 1075-1086.

[22] HUANG D S, MA S D. Linear and nonlinear feedforward neural network classifiers: a comprehensive understanding[J]. J. Intell. Syst., 1999, 9(1): 1-38.

[23] HUANG N E, SHEN Z, LONG S R, WU M C, SHIH H H, ZHENG Q N, YEN N, TUNG C C, LIU H H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. P. Roy. Soc. A-Math. Phy., 1998, 454(1971): 903-995.

[24] HUBER P J. Robust estimation of a location parameter[J]. The Annals of Mathematical Statistics, 1964, 35(1): 73-101.

[25] HUBER P J. Robust regression: asymptotics, conjectures and monte Carlo[J]. Ann. Stat., 1973, 1(5): 799-821.

[26] HUSSEIN R, ELGENDI M, WANG Z J, WARD R K. Robust detection of epileptic seizures based on L_1-penalized robust regression of EEG signals[J]. Expert Sys. Appl., 2018, 104: 153-167.

[27] HUSSEIN R, WANG Z J, WARD R. L1-regularization based eeg feature learning for detecting epileptic seizure[C]. In Proceedings of the 2016 IEEE global conference on signal and information processing (GlobalSIP), 2016, 1171-1175.

[28] JIN Z F, WANG D, SHANG Y L, LV J M. An efficient algorithm for low rank matrix restoration problem with unknown noise level[J]. Chin. Quart. J. of Math., 2021, 36(4): 356-368.

[29] KHAN H, MARCUSE L, FIELDS M, SWANN K, YENNER B. Focal onset seizure prediction using convolutional networks[J]. IEEE Trans. Biomed. Eng., 2017, 65(9): 2109-2118.

[30] KUMAR A, KOLEKAR M H. Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals[C]. In Proceedings of the 2014 International Conference on Medical Imaging, M-health and Emerging Communication Systems (MedCom) IEEE, 2014, 139: 412-416.

[31] LANG M, GUO H, ODEGARD J E, BURRUS C S, WELLS R O. Noise reduction using an undecimated discrete wavelet transform[J]. IEEE Signal Proc. Let., 2002, 3(1): 10-12.

[32] LI M Y, CHEN W Z, ZHANG T. Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble[J]. Biomed. Signal Proces., 2017, 31: 357-365.

[33] MARTIS R J, ACHARYA U R, TAN J H, PETZNICK A, TONG L, CHUA C K, NG E Y K. Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction[J]. Int. J. Neural Syst., 2013, 23(5): 1350023.

[34] MINASYAN G R, CHATTEN J B, CHATTEN M J, HARNER R N. Patient-specific early seizure detection from scalp EEG[J]. J. Clin. Neurophysiol., 2010, 27(3): 163-178.

[35] NABEEL A, THASNEEM F, PAUL J. Detection of epileptic seizure event and oneset using EEG[J]. Biomed Res. Int., 2014, 2014: 450573.

[36] NESTEROV Y. Efficiency of coordinate descent methods on huge-scale optimization problems[J]. Siam J. Optimiz., 2012, 22(2): 341-362.

[37] PENG H, LI C, CHAO J, WANG T, HU B. A novel automatic classification detection for epileptic seizure based on dictionary learning and sparse representation[J]. Neurocomputing, 2019, page in press.

[38] POLAT K, GUNES S. Classification of epileptiform eeg using a hybrid system based on decision tree classifier and fast fourier transform[J]. Appl. Math. Comput., 2007, 187(2): 1017-1026.

[39] RAGHU S, SRRAAM N, HEGDE A S, KUBBEN P L. A novel approach for classification of epileptic seizures using matrix determinant[J]. Expert Syst. Appl., 2019, 127(AUG.): 323-341.

[40] RIAZ F, HASSAN A, REHMAN S, NIAZI I K, DREMSTRUP K. EMD-based temporal and spectral features for the classification of EEG signals using supervised learning[J]. IEEE T. Neur. Sys. Reh., 2016, 24(1): 28-35.

[41] SAMIEE K, KOVACS P, GABBOUJ M. Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform[J]. IEEE T. Bio-Med. Eng., 2014, 62(2): 541-552.

[42] SAMEER M, GUPTA B. Detection of epileptical seizures based on alpha band statistical features[J]. Wireless Pers. Commun., 2020, 115(5): 909-925.

[43] SHARMA M, PACHORI R B, ACHARYA U R. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension[J]. Pattern Recogn. Lett., 2017, 94(jul.15): 172-179.

[44] SHEN N, JIN Z F, WANG A Y. Nested alternating direction method of multipliers to low-rank and sparse-column matrices recovery[J]. Chin. Quart. J. of Math., 2021, 36(1): 90-110.

[45] SIULY, LI Y, WEN P. Clustering technique-based least square support vector machine for EEG signal classification[J]. Comput. Meth. Prog. Bio., 2011, 104(3): 358-372.

[46] SIULY N A, LI Y, WEN P. EEG Signal classification based on simple random sampling technique with least square support vector machine[J]. Int. J. Biomed. Eng. Tec., 2011, 7(4): 390.

[47] SMITH S J M. EEG in the diagnosis, classification, and management of patients with epilepsy[J]. J. Neurol. Neurosur. Ps., 2005, 76(Suppl 2): ii2-7.

[48] SOPHIE L L, ZWALD L. Robust regression through the Huber's criterion and adaptive lasso penalty[J]. Electron. J. Stat., 2011, 5: 1015-1053.

[49] SWAMI P,BHATIA M, TRIPATHI M, CHANDRA P S, GANDHI T K. Selection of optimum frequency bands for detection of epileptiform patterns[J]. Healthc. Technol. Lett., 2019, 6(5): 126-131.

[50] SUWICHA J, SETHA P N, PASIN I. EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation[J]. The Scientific World J., 2014, 2014: 627892.

[51] TABAR Y R, HALICI U. A novel deep learning approach for classification of EEG motor imagery signals[J]. J. Neural Eng., 2017, 14(1): 016003.

[52] TATUM W O. ELLEN R. Grass lecture: extraordinary EEG[J]. The Neurodiagnostic Journal, 2014, 54(1): 3-21.

[53] TRUONG N D, NGUYEN A, KUHLMANN L, BONYADI M R, YANG J, IPPOLITO S, KAVEHEI O. Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram[J]. Neural Networks, 2018, 105: 104-111.

[54] WIENS T S, DALE B C, BOYCE M S, KERSHAW G P. Three way k-fold cross-validation of resource selection functions[J]. Ecol. Model., 2008, 212(3-4): 244-255.

[55] WORLD HEALTH ORGINAZTION (2022, February, 9). Around 50 million people worldwide have epilepsy, making it one of the most common neurological diseases globally. Retrieved from https://www.who.int/news-room/fact-sheets/detail/epilepsy.

[56] YANG L J, DING S J, ZHOU H M, YANG X H. A strategy combining intrinsic time-scale decomposition and feedforward neural network for automatic seizure detection[J]. Physiol. Meas., 2019, 40(9): 095004.

[57] YANG L J, YANG Z H, YANG L H, Zhang P. An improved envelope algorithm for eliminating undershoots[J]. Digit. Signal Process., 2013, 23: 401-411.

[58] YANG L J, YANG Z H, ZHOU F, YANG L H. A novel envelope model based on convex constrained optimization[J]. Digit. Signal Process., 2014, 29: 138-146.

[59] YUAN Q, ZHOU W D, YUAN S S, LI X L, WANG J W, JIA G J. Epileptic EEG classification based on kernal sparse representation[J]. Int. J. Neural Syst., 2014, 24(4): 1450015.

[60] ZHANG J W, XU H, LONG H C, HOU K K, LI Z W, YANG Z Y, WANG Y N, LI Q J. EEG signal epilepsy detection system based on convolutional neural network and memristor array[C]. 2021 IEEE 4th International Conference on Electronics Technology (ICET) IEEE, 2021.

[61] ZHOU M N, TIAN C, CAO R, WANG B, NIU Y, HU T, GUO H, XIANG J. Epileptic seizure detection based on EEG signals and CNN[J]. Front Neuroinform, 2018, 12: 95.

[62] ZHOU W, LIU Y, LI X. Epileptic seizure detection using lacunarity and Bayesian linear discriminant analysis in intracranial EEG[J]. IEEE T. Bio-Med. Eng., 2013, 60(12): 3375-3381.

[63] ZOU H. The adaptive Lasso and its oracle properties[J]. J. Am. Stat. Assoc., 2006, 101(476): 1418-1429.