Research article

Epilepsy EEG classification method based on supervised locality preserving canonical correlation analysis

Hongming Liu[1,2], Yunyuan Gao[2,4,*], Jianhai Zhang[3,4] and Juanjuan Zhang[5]

[1] Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou, China

[2] College of Automation, Hangzhou Dianzi University, Hangzhou, China

[3] College of Computer & Software, Hangzhou Dianzi University, Hangzhou, China

[4] Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, China

[5] Wenzhou branch, China Mobile Group Zhejiang Co, Ltd

* **Correspondence:** Email: gyy@hdu.edu.cn.

Abstract: Existing epileptic seizure automatic detection systems are often troubled by high-dimensional electroencephalogram (EEG) features. High-dimensional features will not only bring redundant information and noise, but also reduce the response speed of the system. In order to solve this problem, supervised locality preserving canonical correlation analysis (SLPCCA), which can effectively use both sample category information and nonlinear relationships between features, is introduced. And an epileptic signal classification method based on SLPCCA is proposed. Firstly, the power spectral density and the fluctuation index of the frequency slice wavelet transform are extracted as features from the EEG fragments. Next, SLPCCA obtains the optimal projection direction by maximizing the weight correlation between the paired samples in the class and their neighbors. And the projection combination of original features in the optimal direction is the fusion feature. The fusion features are then input into LS-SVM for training and testing. This method is verified on the Bonn dataset and the CHB-MIT dataset and gets good results. On various classification tasks of Bonn data set, the proposed method achieves an average classification accuracy of 99.16%. On the binary classification task of the inter-seizure and seizure epileptic EEG of the CHB-MIT dataset, the proposed method achieves an average accuracy of 97.18%. The experimental results show that the algorithm achieves excellent results compared with several state-of-the-art methods. In addition, the parameter sensitivity of SLPCCA and the relationship between the dimension of the fusion features and the classification results are discussed. Therefore, the stability and effectiveness of the method are further verified.

Keywords: epileptic seizure; electroencephalogram (EEG); feature extraction; feature fusion; supervised locality preserving canonical correlation analysis (SLPCCA)

1. Introduction

Epilepsy is a brain disease with strong suddenness, recurrence and involuntary nature [1]. According to statistics from the World Health Organization, nearly 50 million people worldwide suffer from epilepsy [2]. At present, the clinical diagnosis of epilepsy mainly relies on medical history data and brain examinations, such as visual examination of long-term electroencephalogram(EEG) by experienced doctors. However, this method is time-consuming and the results are subjective. To solve this dilemma, a large number of methods have been proposed to identify epileptic EEG signals through signal processing and machine learning. These methods can reduce the burden of doctors and improve the diagnostic accuracy.

Due to the obvious nonlinearity and nonstationarity of epileptic EEG signals, researchers use multi-channel and high sampling rate EEG acquisition equipment to obtain EEG signals of subjects. In this way, they can ensure the highest spatial and temporal resolution. However, this method directly causes the feature dimension of the multi-channel EEG signal to be too high. On the one hand, information redundancy and noise are increased, which brings interference to accurate identification. On the other hand, the signal processing time is prolonged and the diagnosis efficiency is reduced.

In view of the possible dimension disasters in the feature extraction of multi-channel EEG signals, many feature dimension reduction algorithms are proposed. M. Yildiz et al. [3] used principal component analysis (PCA) to reduce dimensionality, obtaining the best 8-dimensional features and improving the classification accuracy by 9%. A. Matin et al. [4] combined PCA with two independent component analysis (ICA) algorithms and achieved good classification results in a variety of tasks. It have been found that manifold learning can improve the classification effect while reducing the dimension of EEG features. Yang et al. [5] proposed a feature dimension reduction algorithm based on local preserving projection and achieved a classification accuracy of more than 97% in a variety of classification tasks. J.Birjandtalab et al. [6] adopted a non-linear data embedding technique based on random nearest neighbor distance metric and achieved an F-measure of more than 87%. Hou et al. [7] used local linear embedding and random forest algorithms to achieve an average classification accuracy of 95%.

Although there have been many dimension reduction algorithms for a single feature, the relationships among features are not well considered. It may cause redundant information between different features, thereby increasing unnecessary feature dimensions. In order to reduce the feature dimension and the information redundancy between features, SLPCCA is applied to the classification of epilepsy EEG since it can effectively use both sample category information and nonlinear relationships between features. SLPCCA is used to convert the original high-dimensional multi-channel EEG features into low-dimensional fusion features, then the fusion features are input into the classifier. Through experiments on Bonn data set and CHB-MIT data set, it can be proved that SLPCCA can effectively reduce the multi-channel EEG feature dimension. At the same time, the redundant information and irrelevant components in the original features are removed, which improves the classification accuracy and helps to reduce the diagnosis time.

The layout of the paper is as follows. In Section 2, two original features in the experiment are introduced. In Section 3, theories and characteristics of SLPCCA is introduced. The data sets, classifiers and classification evaluation indicators used in the experiment, as well as the classification results under various classification tasks of different data sets are presented in Section 4. Parameter

sensitivity and fusion feature dimension of SLPCCA is analyzed in Section 5. Finally, conclusion are presented in Section 6.

2. Feature extraction

The automatic detection of epileptic seizures is mainly the mining of effective information in EEG through machine learning. Its essence can be attributed to a pattern recognition problem, in which it is important to design features that can effectively distinguish epileptic EEG. Since EEG are significantly different in different frequency bands, power spectral density (PSD) is selected to reflect the changes of EEG energy with frequency. Besides, due to the non-stationary nature of EEG, frequency domain features that require signals to be stationary, such as PSD, have limitations. To overcome the limitations, feature extraction methods based on time-frequency analysis are proposed and fluctuation index of frequency slice wavelet transform (FSWT-FI) is selected here.

2.1. Power spectral density

PSD [8] can describe the energy change of the frequency domain signal. Since the energy of the EEG signal during an epileptic seizure is significantly greater than that of the non-seizure, the PSD can effectively distinguish the seizure and non-seizure EEG. The way to obtain the PSD of signal $f(t)$ is as follows:

$$P(\omega) = \sum_{\tau=-\infty}^{+\infty} E\left[f(t) f^*(t+\tau) e^{-i\omega\tau}\right], \tag{2.1}$$

where $P(\omega)$ means the power spectral density at frequency ω, E means expectation, '$*$' means conjugate and τ is the time delay.

EEG can be roughly divided into five rhythms, namely $\delta(1 \sim 3Hz)$, $\theta(4 \sim 7Hz)$, $\alpha(8 \sim 13Hz)$, $\beta(14 \sim 30Hz)$ and $\gamma(30 \sim 80Hz)$. Each rhythm has different energy characteristics during a seizure. Therefore, the PSD on these rhythms are selected as features of the automatic detection of epileptic seizures.

2.2. Fluctuation index of frequency slice wavelet transform

Frequency slice wavelet transform (FSWT) is a time-frequency analysis method proposed by Yan et al. [9]. The FSWT of signal $f(t)$ is as follows:

$$W_f(t, \omega, k) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(u) \hat{p}^*\left(k\frac{u-\omega}{\omega}\right) e^{iut} du, \tag{2.2}$$

where $W_f(t, \omega, k)$ is the frequency slice wavelet transform form of $f(t)$, '$*$' means conjugate and $\hat{f}$ is the Fourier transform of f. $\hat{p}$ is the Fourier transform of the mother wavelet function, that is, the frequency slice function. The commonly used frequency slicing function is $\hat{p}(\omega) = e^{-0.5\omega^2}$. k is the time-frequency analysis coefficient, which can be obtained as $k = \frac{\Delta\omega_p}{\eta_s}$, where $\Delta\omega_p$ is the width of the frequency window of $\hat{p}(\omega)$ and η_s is the frequency resolution of $f(t)$. The inverse Fourier transform of $W_f(t, \omega, k)$ in the time domain (t_1, t_2) and frequency domain (ω_1, ω_2) is as follows, then the signal

component of the $f(t)$ in the time frequency domain $(t_1, t_2, \omega_1, \omega_2)$ can be obtained,

$$f_{t_1, t_2, \omega_1, \omega_2} = \frac{1}{2\pi} \int_{\omega_1}^{\omega_2} \int_{t_1}^{t_2} W(\tau, \omega, k) e^{i\omega(t-\tau)} d\tau d\omega. \tag{2.3}$$

In real applications, EEG is processed in discrete form. Correspondingly, the discrete FSWT is given by [10]

$$W(\cdot, k) = F^{-1} \left\{ F\{f\} * \hat{P}_k \right\}, \quad k = 0, 1, \cdots, n-1 \tag{2.4}$$

where n is the length of the signal f, F and F^{-1} are the discrete Fourier transform and inverse discrete Fourier transform, respectively. Besides, the frequency slice function $\hat{P}_k$ is a sparse sequence of limited length. Equation (2.4) can be realized by reducing resample in the time domain and using fast Fourier transform for each k. So, the inverse discrete FSWT is as follows:

$$f = \frac{1}{n} F^{-1} \left\{ \sum_{l=0}^{N_\theta - 1} W(l, k) e^{-i(2\pi/n)lk} \right\}, \quad k = 0, 1, \cdots, n-1 \tag{2.5}$$

where N_θ is the number of resample points.

After obtaining the time-frequency distribution of EEG signals through FSWT, the corresponding frequency ranges of rhythms δ, θ, α, β and γ are reconstructed to obtain the time-domain waveforms of each rhythm. Then the fluctuation index (FI) of each rhythm signal is calculated as follows for the subsequent classification task,

$$FI = \sum_{t=1}^{n-1} (f_{t+1} - f_t)^2 \tag{2.6}$$

where n is the total length of the signal f. f_{t+1} and f_t are signals at two adjacent moments.

3. Feature fusion algorithm

Although the above-mentioned PSD and FSWT-FI can well reflect the characteristics of epilepsy EEG, they have the risk of causing a dimension explosion. Because they are extracted on each rhythm of the EEG, when the object is a multi-channel EEG, the feature dimension will increase rapidly. In order to avoid the risk of dimension explosion, feature fusion algorithm is used after feature extraction to fuse information of two features while reducing the dimension.

Canonical correlation analysis (CCA) [11] is a multivariate statistical algorithm dealing with the correlation between two sets of variates. CCA transforms the correlation between two sets of random variables into the correlation between a few pairs of typical variables. It can not only retain the effective information of multiple EEG features participating in the fusion, but also reduce redundant information, thereby reducing the feature dimension.

Although CCA can effectively analyze the correlation between features and has played a role in image recognition tasks [12], it has two shortcomings. On the one hand, CCA is an unsupervised feature fusion algorithm, so it does not contain sample category information. In order to incorporate sample category information into the feature fusion process, discriminant canonical correlation analysis (DCCA) [13] is proposed. The fusion features obtained through the supervised feature fusion algorithm will be better applied to the task of epilepsy EEG classification. On the other hand, as an essentially

linear dimension reduction technique, CCA cannot reveal the nonlinear relationship between features. So, kernel canonical correlation analysis (KCCA) [14] is proposed to converts the nonlinear problem in the original space into a linear problem in the high-dimensional space and uses the kernel trick to solve it. However, it is difficult for KCCA to solve the nonlinear problem embedded in the local space. Therefore, the idea of local preservation is introduced and locality preserving canonical correlation analysis (LPCCA) [15] is proposed to better contain the nonlinear properties of EEG. Inheriting the advantages of DCCA and LPCCA while making up for their shortcomings and adding sample weights, supervised locality preserving canonical correlation analysis (SLPCCA) [16] is proposed.

3.1. Canonical correlation analysis

Suppose there are two random variables ξ and η. CCA is to find the projection direction α and β so that the correlation coefficient of the projection $\alpha^T \xi$ and $\beta^T \eta$ is the largest. α and β are called typical projection directions. $\alpha^T \xi$ and $\beta^T \eta$ are called canonical correlation variables. By analyzing several pairs of unrelated canonical correlation variables, the analysis of the correlation between ξ and η can be completed. Typical projection directions α and β can be obtained by the following criterion function.

$$\arg \max_{\alpha, \beta} \rho = \frac{\alpha^T S_{\xi\eta} \beta}{\sqrt{\alpha^T S_{\xi\xi} \alpha} \sqrt{\beta^T S_{\eta\eta} \beta}}, \tag{3.1}$$

where $S_{\xi\eta}$ represents the cross-covariance matrix of ξ and η. $S_{\xi\xi}$ and $S_{\eta\eta}$ represent the covariance matrix of ξ and η, respectively.

When CCA is used for feature fusion, suppose there are two standardized feature sets $X = [x_1, x_2, \cdots, x_n] \in R^{p*n}$ and $Y = [y_1, y_2, \cdots, y_n] \in R^{q*n}$, which means X and Y respectively contain p-dimensional features and q-dimensional features of n samples. Based on the idea of CCA, the canonical correlation variables between X and Y are extracted first. These canonical correlation variables can be called canonical correlation features, denoted as $\alpha_1^T X$ and $\beta_1^T Y$, $\alpha_2^T X$ and $\beta_2^T Y$, $\cdots$, $\alpha_d^T X$ and $\beta_d^T Y$, a total of d pairs. Then, feature set after projection, namely $X^* = (\alpha_1, \alpha_2, \ldots, \alpha_d)^T X$ and $Y^* = (\beta_1, \beta_2, \ldots, \beta_d)^T Y$, can form the fusion feature $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$ for classification.

3.2. Discriminant canonical correlation analysis

Because the sample category information needs to be considered, DCCA reorders the features in the original feature set X and Y according to categories. Then, the matrix C_w describing the similarity within the feature vector class can be defined as $C_w = XAY^T$, where A is a symmetric positive semi-definite block matrix with rank c and block size n_i. c is the number of sample categories and n_i is the number of samples in ith category. Similarly, the matrix C_b describing the similarity between feature vector classes can be defined as $C_b = \sum_{i=1}^{c} \sum_{j=1 \neq i}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} x_k^{(i)} y_l^{(j)^T} = -XAY^T$.

So, the criterion function of the typical projection direction α and β of DCCA can be defined as

$$\arg \max_{\alpha, \beta} \rho = \frac{\alpha^T XAY^T \beta}{\sqrt{\alpha^T S_{XX} \alpha} \sqrt{\beta^T S_{YY} \beta}}. \tag{3.2}$$

Same as CCA, $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$ can be used as the fusion feature after projection for classification.

3.3. Locality preserving canonical correlation analysis

To analyze the local structure of the data, if $N_X(x_i)$ is the set of local neighbors of the sample x_i, the local correlation of the sample can be expressed as $\alpha^T \cdot \sum_{j \in N_X(x_i)} (x_i - x_j)(y_i - y_j)^T \cdot \beta$. Then, the k neighboring matrices of feature sets X are defined as follows

$$F_X(i, j) = \begin{cases} exp\left(-\frac{\|x_i - x_j\|^2}{t_x^2}\right), & x_i \in N_X(x_j) \text{ or } x_j \in N_X(x_i) \\ 0, & else \end{cases} \quad (3.3)$$

where t_x are the mean value of the distance between all samples in the feature set X. F_Y can be obtained in the same way. So, the local canonical correlation can be expressed as $\alpha^T \cdot \sum_{i=1}^n \sum_{j=1}^n F_X(i, j)(x_i - x_j) F_Y(i, j)(y_i - y_j)^T \cdot \beta$ and the criterion function of α and β of the typical projection direction of LPCCA can be defined as

$$arg \max_{\alpha, \beta} \rho = \frac{\alpha^T X F_X F_Y^T Y^T \beta}{\sqrt{\alpha^T X (D_{F_X} - F_X) X^T \alpha} \sqrt{\beta^T Y (D_{F_Y} - F_Y) Y^T \beta}}, \quad (3.4)$$

where D_{F_X} and D_{F_Y} are two diagonal matrices with diagonal element values $D_{F_X}(i, i) = \sum_{j=1}^n F_X(j, i)$ and $D_{F_Y}(i, i) = \sum_{j=1}^n F_Y(j, i)$. Like CCA and DCCA, $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$ can be used as the fusion feature after projection for classification.

3.4. Supervised locality preserving canonical correlation analysis

Drawing on the idea of LPCCA, SLPCCA constructs the intra-class neighbor set of the sample and defines the intra-class k neighbor matrix of feature sets X as follows,

$$SF_X(i, j) = \begin{cases} exp\left(-\frac{\|x_i - x_j\|^2}{t_x^2}\right), & x_i \in SN_X(x_j) \text{ or } x_j \in SN_X(x_i) \\ 0, & else \end{cases} \quad (3.5)$$

where SN_X is the set of intra-class nearest neighbors of the sample. SF_Y can be obtained in the same way. Similar to LPCCA, the criterion functions of the typical projection directions α and β of SLPCCA are as follows,

$$arg \max_{\alpha, \beta} \rho = \frac{\alpha^T X SF_X SF_Y^T Y^T \beta}{\sqrt{\alpha^T X (SD_{SF_X} - SF_X) X^T \alpha} \sqrt{\beta^T Y (SD_{SF_Y} - SF_Y) Y^T \beta}}, \quad (3.6)$$

where SD_{SF_X} and SD_{SF_Y} are diagonal matrices composed of column elements of SF_X and SF_Y, respectively.

If $\widehat{S_{XY}} = X SF_X SF_Y^T Y^T$, $\widehat{S_{XX}} = X (SD_{SF_X} - SF_X) X^T$ and $\widehat{S_{YY}} = Y (SD_{SF_Y} - SF_Y) Y^T$, the solution of the criterion function is equivalent to the solution of the following optimization problem

$$\max_{\alpha, \beta} \rho = \alpha^T \widehat{S_{XY}} \beta$$
$$s.t. \begin{cases} \alpha^T \widehat{S_{XX}} \alpha = 1 \\ \beta^T \widehat{S_{YY}} \beta = 1 \end{cases} \quad (3.7)$$

In order to solve the above optimization problem, the Lagrange multiplier method is used to obtain the following generalized eigenvalue equation:

$$\begin{pmatrix} 0 & \widehat{S_{XY}} \\ \widehat{S_{YX}} & 0 \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \sigma^2 \begin{pmatrix} \widehat{S_{XX}} & 0 \\ 0 & \widehat{S_{YY}} \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \tag{3.8}$$

where σ is the Lagrange multiplier. Next, the generalized eigenvector $\{\alpha_i, \beta_i\}, i = 1, \cdots, d$, corresponding to the first d maximum generalized eigenvalues can be found. Thus, α and β can be constructed and then $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$ can be obtained.

3.5. Epilepsy EEG feature fusion method based on SLPCCA

After extracting PSD and FSWT-FI from EEG data, in order to reduce the feature dimension and information redundancy, SLPCCA is adopted to merge PSD and FSWT-FI into a fusion feature. Firstly, the k neighbor matrix within the class according to the relationship between samples in the feature set PSD and FSWT-FI is established. Then, samples and neighbor samples in PSD are correlated with the corresponding samples and neighbor samples in FSWT-FI. The local structure of the data is thus preserved. After that, by solving the optimization problem, the projection direction is obtained. Finally, the original features are projected and combined into a fusion vector. The specific process of feature fusion by SLPCCA is as follows

- Construct the intra-class k neighbor matrix SF_X and SF_Y of feature sets X and Y, namely PSD and FSWT-FI;
- Establish criterion function ρ as Eq (3.6);
- Calculate typical projection directions α and β by solving optimization problem as Eq (3.7);
- Get the projection X^* and Y^* of feature set X and Y by α and β;
- Combine X^* and Y^* into a fusion feature $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$;
- Z is the fusion feature set obtained by the fusion of PSD and FSWT-FI, which will be used to train and test the classifier.

If $H = \widehat{S_{XX}}^{-\frac{1}{2}} \widehat{S_{XY}} \widehat{S_{YY}}^{-\frac{1}{2}} = \sum_{i=1}^{r} \lambda_i u_i v_i^T$, where λ is the eigenvalue, then it can be known that the optimal solutions for projection directions α and β are $\widehat{S_{XX}}^{-\frac{1}{2}} u_i$ and $\widehat{S_{YY}}^{-\frac{1}{2}} v_i$ [11], respectively. Under the constraints of the optimization problem shown in Eq (3.7), the typical projection directions, namely α_i and β_i, are at most r pairs $\left(r = rank\left(\widehat{S_{XY}}\right)\right)$. Because the projection directions α_i and β_i corresponding to the larger eigenvalue λ_i contain more information, the projection direction corresponding to the d largest eigenvalues can be used to construct the final d-dimensional projection direction.

Besides, for any α and β that meet the constraints of the optimization problem, there are

$$\begin{cases} \alpha_i^T \widehat{S_{XX}} \alpha_j = \beta_i^T \widehat{S_{YY}} \beta_j = \delta_{ij} \\ \\ \alpha_i^T \widehat{S_{XY}} \beta_j = \lambda_i \delta_{ij} \end{cases} \tag{3.9}$$

where $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. So, if $x_i^* = \alpha_i^T x_i$ and $y_i^* = \beta_i^T y_i$, then for any $i \neq j$ there are

$$cov\left(x_i^*, x_j^*\right) = \alpha_i^T \widehat{S_{XX}} \alpha_j = 0,$$

$$cov\left(y_i^*, y_j^*\right) = \beta_i^T \widehat{S_{YY}} \beta_j = 0, \tag{3.10}$$

$$cov\left(x_i^*, y_i^*\right) = \alpha_i^T \widehat{S_{XY}} \beta_j = 0.$$

Therefore, $cov\left(z_i^*, z_j^*\right) = 0$, which means the components of fusion features are irrelevant [12]. It proves that there is no information redundancy between the fusion features.

In addition, SLPCCA is further analyzed and compared with the above feature fusion algorithm, namely CCA, DCCA and LPCCA.

4. Experiment

4.1. Dataset

The EEG data comes from the German Bonn University dataset [17] and CHB-MIT dataset [18], which have been widely used in epilepsy detection related tasks.

The Bonn dataset contains 5 subsets of A-E. Each subset contains 100 segments of single-channel EEG signals composed of 4097 sampling points and the sampling frequency is 173.61Hz. Subsets A and B are respectively taken from the scalp surface of five healthy volunteers with their eyes open and closed. Subsets C and D are respectively taken from five epilepsy patients outside the lesion area and within the lesion area of the inter-epilepsy. Subset E is taken from the epileptic seizure area of the above five epilepsy patients. Each EEG signal is divided into 4 evenly, that is, each subset is divided into a data set containing 400 single-channel EEG signals composed of 1024 sampling points.

The CHB-MIT dataset consists of multi-channel EEG signals from 24 epilepsy patients aged 3 to 22 years. Most EEG signals adopt 23 EEG channels and their electrode positions and naming adopt the international 10–20 system [19]. A sliding window with a window length of 4 seconds and a step length of 2 seconds is used to extract the inter-seizure segment and the seizure segment from the continuous EEG signal.

In the experiment, 70% of each type of sample is used as the training set and 30% is used as the test set. For the Bonn data set, PSD of each sample and the FSWT-FI of each sample's δ, θ, α, β, and γ rhythms are extracted as features. Then feature fusion algorithms are used to get the fusion feature. For the CHB-MIT data set, 23 channels' PSD of each EEG signal and the FSWT-FI of the five EEG rhythms are extracted as features. After forming the concatenated feature vector, feature fusion algorithms are used to get the fusion feature.

4.2. Pretreatment, classifier and evaluation index

The original EEG will be interfered by noise, such as eye electrical artifact and EMG interference from human body. Therefore, in order to better extract EEG features, wavelet threshold filtering is used for denoising. Besides, the band stop filter is used to remove power-line interference.

Least squares support vector machine (LS-SVM), which has been widely used in various fields such as machine learning, is chosen as the classifier. In order to evaluate the effect of the classifier, accuracy, sensitivity and specificity are introduced as evaluation indicators. They are obtained as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%,$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\%, \qquad (4.1)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\%,$$

where TP (True Positive) is the number of positive samples that are correctly classified, TN (True Negtive) is the number of negative samples that are correctly classified, FP (False Positive) is the number of positive samples that are incorrectly classified and FN (False Negative) is the number of negative samples that are incorrectly classified.

4.3. Experiment result

4.3.1. Experiments on Bonn dataset

In order to consider a wider range of practical clinical situations, such as the classification of multiple types of EEG signals and the classification under the uneven number of samples, 5 subsets of Bonn dataset are used to form 11 classification groups. They include A vs E, B vs E, C vs E, D vs E, AB vs E, AC vs E, AD vs E, ABC vs E, ABCD vs E, A vs C vs E and AB vs CD vs E. These 11 classification groups can be divided into 4 classification tasks: normal EEG vs seizure epileptic EEG, inter-seizure epileptic EEG vs seizure epileptic EEG, non-seizure EEG vs seizure epileptic EEG and normal EEG vs inter-seizure epileptic EEG vs seizure epileptic EEG. The classification results on these four tasks are shown in Table 1, Table 2, Table 3 and Table 4 respectively. The results shown in the tables are the average of all experiments in the corresponding task and the optimal result in each task is bolded.

- normal EEG vs seizure epileptic EEG

Table 1. Result of normal EEG vs seizure epileptic EEG.

Data Sets	Feature	Fusion Algorithm	Feature Dimension	Accuracy	Sensitivity	Specificity
	PSD	\	5	99.16%	98.33%	100%
	FI	\	5	99.16%	98.33%	100%
		series	10	100%	100%	100%
A vs E		PCA	8	99.58%	100%	99.16%
B vs E		CCA	8	99.58%	99.16%	100%
	PSD,FI	DCCA	2	96.67%	95.00%	98.33%
		LPCCA	8	98.75%	97.50%	100%
		SLPCCA	**8**	**100%**	**100%**	**100%**

- inter-seizure epileptic EEG vs seizure epileptic EEG

Table 2. Result of inter-seizure epileptic EEG vs seizure epileptic EEG.

Data Sets	Feature	Fusion Algorithm	Feature Dimension	Accuracy	Sensitivity	Specificity
	PSD	\	5	98.75%	98.33%	99.16%
	FI	\	5	97.91%	97.50%	98.33%
		series	10	99.16%	99.16%	99.16%
C vs E		PCA	8	97.91%	96.67%	99.17%
D vs E		CCA	8	98.75%	99.16%	98.33%
	PSD,FI	DCCA	2	97.08%	98.33%	95.83%
		LPCCA	8	98.75%	98.33%	99.16%
		SLPCCA	**8**	**99.16%**	**98.75%**	**99.58%**

- non-seizure EEG vs seizure epileptic EEG

Table 3. Result of non-seizure EEG vs seizure epileptic EEG.

Data Sets	Feature	Fusion Algorithm	Feature Dimension	Accuracy	Sensitivity	Specificity
	PSD	\	5	98.21%	96.45%	98.67%
	FI	\	5	98.37%	97.29%	98.49%
AC vs E		series	10	98.72%	97.49%	99.18%
AD vs E		PCA	8	97.08%	97.50%	97.66%
ABC vs E		CCA	8	98.19%	97.08%	98.63%
ABCD vs E	**PSD,FI**	DCCA	2	98.11%	96.87%	98.54%
		LPCCA	8	98.41%	97.08%	99.16%
		SLPCCA	**8**	**99.06%**	**98.74%**	**99.16%**

- normal EEG vs inter-seizure epileptic EEG vs seizure epileptic EEG

Table 4. Result of normal EEG vs inter-seizure epileptic EEG vs seizure epileptic EEG.

Data Sets	Feature	Fusion Algorithm	Feature Dimension	Accuracy
	PSD	\	5	96.96%
	FI	\	5	96.83%
		series	10	97.58%
A vs C vs E		PCA	8	96.94%
AB vs CD vs E		CCA	8	97.16%
	PSD,FI	DCCA	2	96.58%
		LPCCA	8	97.16%
		SLPCCA	**8**	**98.44%**

It can be seen from the above results that fusion features that formed by the fusion of PSD and FI using SLPCCA algorithm can achieve the best classification effect in all kinds of classification tasks, with an average of 99.08%. Series features are also superior to single features in all classification tasks, second only to SLPCCA. For other feature fusion algorithms, because the algorithm itself is not suitable for epilepsy EEG classification tasks, the results are even inferior to a single feature on some classification tasks. But because of the excellent performance of SLPCCA, it can be believed that the application of feature fusion algorithm in the classification of epilepsy EEG is feasible and effective.

4.3.2. Experiment on CHB-MIT dataset

Table 5. Result of inter-seizure epileptic EEG vs seizure epileptic EEG.

Feature	Fusion Algorithm	Feature Dimension	Accuracy	Sensitivity	Specificity
PSD	\	115	95.58%	95.35%	95.82%
FI	\	115	94.66%	94.42%	94.89%
	series	230	96.44%	96.21%	96.67%
	PCA	190-210	92.22%	91.11%	93.33%
	CCA	160-180	94.19%	93.96%	94.42%
PSD,FI	DCCA	2	81.47%	80.77%	82.17%
	LPCCA	150-170	95.31%	95.78%	95.02%
	SLPCCA	**50-60**	**97.18%**	**97.10%**	**97.77%**

Although the effectiveness of the feature fusion algorithm has been proven, the difference in the classification effect of each algorithm on the Bonn dataset is small. In addition, because the Bonn dataset is a single-channel EEG dataset with low feature dimensions, the experimental results on the Bonn dataset cannot reflect the advantages of feature fusion algorithm in dimension reduction.

Therefore, the binary classification of inter-seizure epileptic EEG and seizure epileptic EEG on the CHB-MIT dataset are carried out. The results are shown in the Table 5. The optimal result is also bolded.

It can be seen from Table 5 that, in the multi-channel EEG data set, the dimension reduction advantage of the feature fusion algorithm is more obvious. Compared with the series feature, the dimension of each fusion feature has different degrees of reduction. The fusion feature obtained by SLPCCA achieve the highest classification accuracy of 97.18% when the feature dimension is almost one-half of the single feature dimension and one-fourth of the serial feature dimension. It is foreseeable that when the feature dimension or the number of EEG channels is further increased, when the dimensional explosion disaster occurs in the series feature, the fusion feature obtained by SLPCCA can still be in a low dimension. Besides, because the algorithm itself is not suitable for epilepsy EEG classification tasks, the results obtained by other fusion algorithms except SLPCCA are not as good as single feature, just like the experiment on Bonn dataset. However, because CCA and LPCCA consider the correlation between features, their accuracies are still higher than that of single feature dimension reduction methods, such as PCA.

4.3.3. Comparison with other methods

The proposed method is compared with other methods under the same dataset. The comparison results of Bonn dataset and CHB-MIT dataset are shown in the Table 6 and Table 7 respectively. It can be seen from the result that SLPCCA has a higher accuracy in more comprehensive classification tasks.
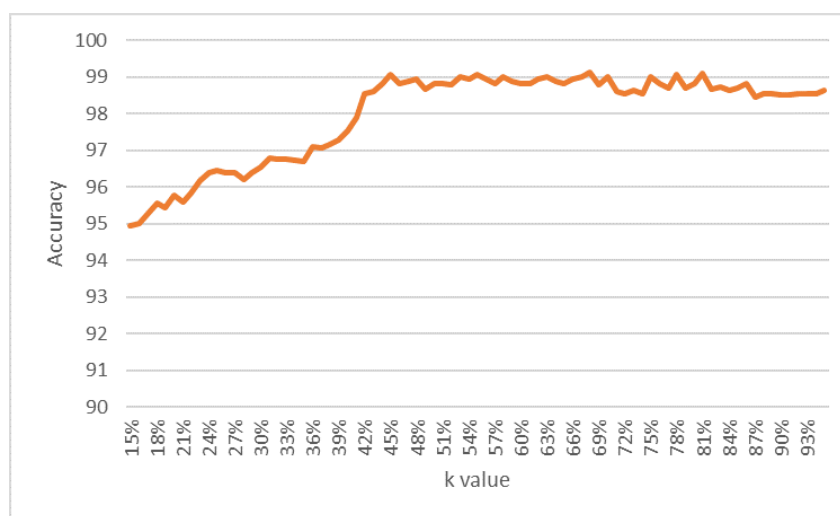
5. Discussion

It can be seen from the results that although CCA, DCCA and LPCCA can achieve dimension reduction, the accuracy of the fusion features obtained by them is negatively affected. CCA and LPCCA are unsupervised algorithms, so they may introduce wrong category information in the feature fusion process, thereby reducing the classification accuracy. The dimension of the fusion features obtained by the DCCA is limited by the number of categories, so DCCA is not suitable for small-category classification tasks. However, SLPCCA can not only achieve feature dimension reduction, but also improve classification accuracy, so it will be further analyzed next.

5.1. Discussion of algorithm parameter sensitivity

In order to add the local structure information of the data, SLPCCA draws on LPCCA to construct the intra-class k neighbor matrix of the feature set. Because the parameter k will affect the effect of the algorithm, sensitivity analysis is performed on k. k ranges from 15% to 90% of the total number of samples in the same category. Accuracy of task AB vs CD vs E on the Bonn dataset is the evaluation index. The result is shown in the Figure 1.

Table 6. Comparison results of Bonn dataset.

Year	Method	Datasets	Accuracy
2013	intrinsic time scale decomposition[20]	A,C,E	95.67%
2014	wavelet feature + statistical feature[21]	A,C,E	84.20%
2015	frequency domain nonlinear features[22]	A,C,E	98.00%
2016	weighted permutation entropy + SVM[23]	A,D,E	97.50%
2016	optimal allocation + logic tree[24]	A,D,E	97.50%
2018	Hankel matrix + Hilbert transform + time-frequency representation[25]	D,E	96.00%
		A,D,E	98.00%
2019	Fourier-Bessel series expansion + weighted multi-scale Renyi permutation entropy[26]	D,E	97.50%
		A,C,E	97.30%
2020	energy feature + approximate entropy + local preserving projection + LS-SV[5]	D,E	93.00%
		C,E	96.50%
		A,D,E	96.69%
2020	statistical features + adaptive LS-SVM[27]	11 groups	99.00%
2021	**proposed method**	**C,E**	**99.16%**
		D,E	**99.16%**
		A,C,E	**98.17%**
		A,D,E	**98.00%**
		11 groups	**99.16%**

Figure 1. Sensitivity analysis result of parameter k.

It can be seen from the figure that when k is bigger than 40%, the accuracy is high and the fluctuation range does not exceed 1%. Therefore, SLPCCA can be considered not sensitive to parameter changes and relatively stable.

Table 7. Comparison results of CHB-MIT dataset.

Year	Method	Patients	Accuracy
2011	discrete wavelet transform[28]	24	80.00%
2012	energy feature + wavelet transform[29]	5	91.80%
2015	Poincare section and phase space[30]	24	93.00%
2016	Convolutional Neural Network[31]	23	85.00%
2017	Mallat's Scattering[32]	24	91.00%
2017	wavelet level selection + SVM[33]	24	92.30%
2018	power spectral density + SVM[34]	24	92.30%
2018	short time Fourier transform (STFT)[35]	24	92.00%
2018	convolutional neural network[36]	21	95.60%
2019	hybrid filter[37]	22	85.80%
2020	complementary ensemble empirical mode decomposition + limit gradient enhancement[38]	24	95.49%
2020	discrete wavelet transform for multiresolution analysis[39]	10	95.30%
2020	non-dominated sorting genetic algorithm[40]	24	97.00%
2020	DFA based on the EMD + teager and instantaneous energy + Higuchi and Petrosian fractal dimension[41]	24	93.00%
2021	**proposed method**	**24**	**97.18%**

5.2. Discussion of feature dimension and classification effect

Among the mentioned various feature fusion algorithms, only the fusion feature dimension of DCCA is limited by the number of classification categories. Except DCCA, in CCA, LPCCA and SLPCCA, the dimension of the projection direction depends on the number of non-zero eigenvalues of the generalized characteristic equation. The dimension of the projection direction will then determine the dimension of the fusion feature and ultimately affect the classification effect. Therefore, the classification accuracies of the fusion features of these three algorithms under different feature dimensions are analyzed. Accuracy on the CHB-MIT dataset is the evaluation index. The result is shown in the Figure 2.
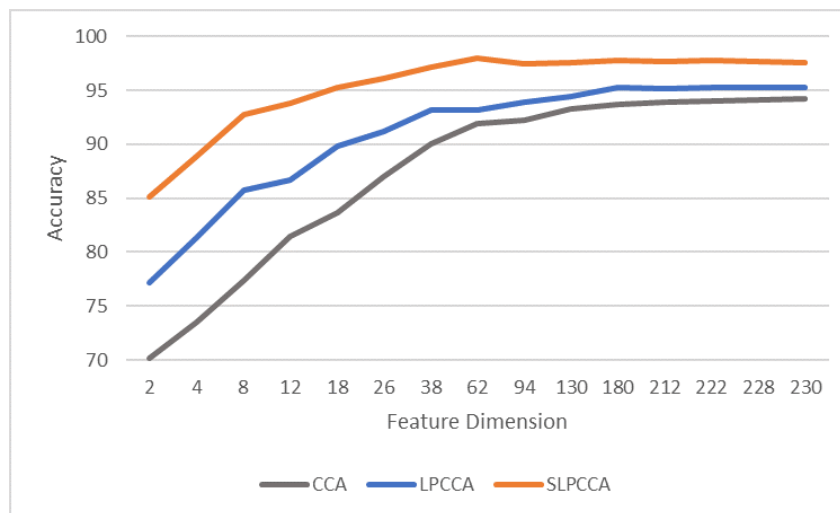
Figure 2. The relationship between feature dimension and classification accuracy.

It can be seen from the figure that as the fusion feature dimension decreases, the accuracy rate of CCA decreases the fastest, followed by LPCCA. And SLPCCA maintains a relatively high accuracy rate when the feature dimension decreases. It is because LPCCA constructs a k-adjacent matrix on the basis of CCA, compared to the fusion features generated by CCA, the fusion features generated by LPCCA have information about the local structure of the data. Besides, SLPCCA borrows from DCCA and changes from an unsupervised feature fusion algorithm to a supervised feature fusion algorithm. The added sample category information enables SLPCCA to contain more information with less dimensions and ultimately achieve a better classification effect. In real applications, EEG features usually have high dimension and contain redundant information. Moreover, because EEG has strong frequency characteristics, features of combined frequency bands are often selected. But some of these frequency bands may be irrelevant to the research, which will lead to the inclusion of irrelevant components. Therefore, it is necessary to use feature fusion methods such as SLPCCA to reduce dimension while connecting multiple features. Thus, redundant information and irrelevant components in features are reduced, and the application effect of EEG, such as classification accuracy, is improved.

6. Conclusions

In order to avoid the dimensional explosion and information redundancy caused by the series connection of multi-channel EEG features, SLPCCA, which can reflect the sample category information and the local structure of the data, is used for feature fusion. Then the epilepsy classification method based on SLPCCA including the following three steps is proposed. First, PSD and FSWT-FI are extracted from EEG as features. Secondly, SLPCCA is used to find a projection direction and merge the two features of one sample into one feature after projection. Compared with other feature fusion algorithms such as CCA, SLPCCA contains not only the local structure of the sample, but also the category information of the sample. Finally, fusion features are used to train and test LS-SVM. The proposed method is verified on the Bonn dataset and the CHB-MIT dataset. The

experiment results show that the proposed method can achieve an average classification accuracy of 99.16%, a sensitivity of 99.06% and a specificity of 99.58% on the 11 classification tasks on the Bonn dataset and the feature dimension is only 80% of the series feature dimension. It can achieve an accuracy of 97.18%, a sensitivity of 97.10% and a specificity of 97.77% on inter-seizure epileptic EEG vs seizure epileptic EEG task on the CHB-MIT dataset and the feature dimension is only 25% of the series feature dimension. It can be seen that the proposed method not only has better classification accuracy than other methods, but also can effectively reduce the feature dimension. What's more, because the proposed method has no special processing, such as selecting specific channels for some patients or selecting different features for different tasks manually, it can be well adapted to other epilepsy EEG dataset.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. A. J. Durnford, W. Rodgers, F. J. Kirkham, M. A. Mullee, A. Whitney, M. Prevett,et al., Very good inter-rater reliability of engel and ilae epilepsy surgery outcome classifications in a series of 76 patients, *Seizure*, **20** (2011), 809–812. doi: 10.1016/j.seizure.2011.08.004.

2. E. Howell, Epilepsy stigma: Moving from a global problem to a global solution, *Seizure-Eur. J. Epilepsy*, **19** (2010), 628–629. doi: 10.1016/j.seizure.2010.10.016.

3. M. Yildiz, E. Bergİl, The investigation of channel selection effects on epileptic analysis of eeg signals, *Balk. J. Electr. Comput. Eng.*, **3** (2015), 236–241.

4. A. Matin, R. A. Bhuiyan, S. R. Shafi, A. K. Kundu, M. U. Islam, A hybrid scheme using pca and ica based statistical feature for epileptic seizure recognition from eeg signal, in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, IEEE, (2019), 301–306. doi: 10.1109/ICIEV.2019.8858573.

5. Y. Liu, B. Jiang, J. Feng, J. Hu, H. Zhang, Classification of EEG signals for epileptic seizures using feature dimension reduction algorithm based on LPP, *Multimedia Tools Appl.*, **80** (2020), 30261-–30282. doi: 10.1007/s11042-020-09135-7.

6. J. Birjandtalab, M. B. Pouyan, M. Nourani, Nonlinear dimension reduction for eeg-based epileptic seizure detection, in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, IEEE, (2016), 595–598. doi: 10.1109/BHI.2016.7455968.

7. Q. Hou, Y. Liu, J. Liu, S. Sun, Epilepsy detection using random forest classification based on locally linear embedding algorithm, in *2020 5th International Conference on Control, Robotics and Cybernetics (CRC)*, IEEE, (2020), 202–206. doi: 10.1109/CRC51253.2020.9253455.

8. K. C. Chua, V. Chandran, R. Acharya, C. Lim, Automatic identification of epilepsy by hos and power spectrum parameters using eeg signals: A comparative study, in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, (2008), 3824–3827. doi: 10.1109/IEMBS.2008.4650043.

9. Z. Yan, A. Miyamoto, Z. Jiang, Frequency slice wavelet transform for transient vibration response analysis, *Mech. Syst. Signal Process.*, **23** (2009), 1474–1489. doi: 10.1016/j.ymssp.2009.01.008.

10. Z. Yan, T. Tao, Z. Jiang, H. Wang, Discrete frequency slice wavelet transform, *Mech. Syst. Signal Process.*, **96** (2017), 385–392. doi: 10.1016/j.ymssp.2017.04.019.

11. H. Hotelling, Relations between two sets of variates, in *Breakthroughs in statistics*, Springer, (1992), 162–190. doi: 10.1007/978-1-4612-4380-9_14.

12. Q. S. Sun, S. G. Zeng, Y. Liu, P. A. Heng, D. S. Xia, A new method of feature fusion and its application in image recognition, *Pattern Recognition*, **38** (2005), 2437–2448. doi: 10.1016/j.patcog.2004.12.013.

13. T. Sun, S. Chen, J. Yang, P. Shi, A supervised combined feature extraction method for recognition, in *Procdings of the IEEE International Conference on Data Mining, Pisa, Italy*, Citeseer, (2008), 1043–1048.

14. T. Melzer, M. Reiter, H. Bischof, Appearance models based on kernel canonical correlation analysis, *Pattern Recognit.*, **36** (2003), 1961–1971. doi: 10.1016/S0031-3203(03)00058-X.

15. T. Sun, S. Chen, Locality preserving cca with applications to data visualization and pose estimation, *Image Vision Comput.*, **25** (2007), 531–543. doi: 10.1016/j.imavis.2006.04.014.

16. S. D. Hou, Q. S. Sun, D. S. Xia, Supervised locality preserving canonical correlation analysis algorithm, *Pattern Recognit. Artif. Intell.*, **2012**, 143–149.

17. R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E*, **2001**, 061907. doi: 10.1103/PhysRevE.64.061907.

18. A. Shoeb, J. Guttat, Application of machine learning To epileptic seizure detection, in *International Conference on Machine Learning*, 2010.

19. R. W. Homan, J. Herman, P. Purdy, Cerebral location of international 10–20 system electrode placement, *Electroencephalogr. Clin. Neurophysiol.*, **66** (1987), 376–382. doi: 10.1016/0013-4694(87)90206-9.

20. R. J. Martis, U. R. Acharya, J. H. Tan, A. Petznick, L. Tong, C. K. Chua, et al., Application of intrinsic time-scale decomposition (itd) to eeg signals for automated seizure prediction, *Int. J. Neural Syst.*, **23** (2013), 1350023. doi: 10.1142/S0129065713500238.

21. N. Ahammad, T. Fathima, P. Joseph, Detection of epileptic seizure event and onset using EEG, *BioMed. Res. Int.*, **2014**. doi: 10.1155/2014/450573.

22. R. J. Martis, J. H. Tan, C. K. Chua, T. C. Loon, S. W. J. YEO, L. Tong, Epileptic eeg classification using nonlinear parameters on different frequency bands, *J. Mech. Med. Biol.*, **15** (2015), 1550040. doi: 10.1142/S0219519415500402.

23. N. S. Tawfik, S. M. Youssef, M. Kholief, A hybrid automated detection of epileptic seizures in eeg records, *Comput. Electr. Eng.*, **53** (2016), 177–190. doi:10.1016/j.compeleceng.2015.09.001.

24. E. Kabir, Y. Zhang, Epileptic seizure detection from eeg signals using logistic model trees, *Brain Inf.*, **3** (2016), 93–100. doi: 10.1007/s40708-015-0030-2.

25. R. R. Sharma, R. B. Pachori, Time-frequency representation using ievdhm-ht with application to classification of epileptic eeg signals, *IET Sci., Meas. Technol.*, **12** (2018), 72–82.

26. V. Gupta, R. B. Pachori, Epileptic seizure identification using entropy of fbse based eeg rhythms, *Biomed. Signal Process. Control*, **53** (2019), 101569. doi: 10.1016/j.bspc.2019.101569.

27. H. Al-Hadeethi, S. Abdulla, M. Diykh, R. C. Deo, J. H. Green, Adaptive boost ls-svm classification approach for time-series signal classification in epileptic seizure diagnosis applications, *Expert Syst. Appl.*, **161** (2020), 113676. doi: 10.1016/j.eswa.2020.113676.

28. N. Rafiuddin, Y. U. Khan, O. Farooq, Feature extraction and classification of eeg for automatic seizure detection, in *2011 International Conference on Multimedia, Signal Processing and Communication Technologies*, IEEE, (2011), 184–187. doi: 10.1109/MSPCT.2011.6150470.

29. Y. U. Khan, N. Rafiuddin, O. Farooq, Automated seizure detection in scalp eeg using multiple wavelet scales, in *2012 IEEE international conference on signal processing, computing and control*, IEEE, (2012), 1–5. doi: 10.1109/ISPCC.2012.6224361.

30. M. Zabihi, S. Kiranyaz, A. B. Rad, A. K. Katsaggelos, M. Gabbouj, T. Ince, Analysis of high-dimensional phase space via poincaré section for patient-specific seizure detection, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **24** (2015), 386–398. doi: 10.1109/TNSRE.2015.2505238.

31. P. Thodoroff, J. Pineau, A. Lim, Learning robust features using deep learning for automatic seizure detection, in *Machine learning for healthcare conference*, PMLR, (2016), 178–190.

32. M. Z. Ahmad, A. M. Kamboh, S. Saleem, A. A. Khan, Mallat's scattering transform based anomaly sensing for detection of seizures in scalp EEG, *IEEE Access*, **5** (2017), 16919–16929. doi: 10.1109/ACCESS.2017.2736014.

33. D. Chen, S. Wan, J. Xiang, F. S. Bao, A high-performance seizure detection algorithm based on discrete wavelet transform (dwt) and EEG, *PloS one*, **12** (2017), e0173138. doi: 10.1371/journal.pone.0173138.

34. J. Bonello, L. Garg, G. Garg, E. E. Audu, Effective data acquisition for machine learning algorithm in eeg signal processing, in *Soft Computing: Theories and Applications*, Springer, (2018), 233–244. doi: 10.1007/978-981-10-5699-4_23.

35. K. M. Tsiouris, S. Markoula, S. Konitsiotis, D. D. Koutsouris, D. I. Fotiadis, A robust unsupervised epileptic seizure detection methodology to accelerate large EEG database evaluation, *Biomed. Signal Process. Control*, **40** (2018), 275–285. doi: 10.1016/j.bspc.2017.09.029.

36. M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, et al., Epileptic seizure detection based on EEG signals and CNN, *Front. Neuroinf.*, **12** (2018), 95. doi: 10.3389/fninf.2018.00095.

37. M. B. Ahmadi, A. Craik, H. F. Azgomi, J. T. Francis, J. L. Contreras-Vidal, R. T. Faghih, Real-time seizure state tracking using two channels: A mixed-filter approach, in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, IEEE, (2019), 2033–2039. doi: 10.1109/IEEECONF44664.2019.9048990.

38. J. Wu, T. Zhou, T. Li, Detecting epileptic seizures in eeg signals with complementary ensemble empirical mode decomposition and extreme gradient boosting, *Entropy*, **22** (2020), 140. doi: 10.3390/e22020140.

39. S. Chakrabarti, A. Swetapadma, A. Ranjan, P. K. Pattnaik, Time domain implementation of pediatric epileptic seizure detection system for enhancing the performance of detection and easy monitoring of pediatric patients, *Biomed. Signal Process. Control*, **59** (2020), 101930. doi: 10.1016/j.bspc.2020.101930.

40. L. A. Moctezuma, M. Molinas, EEG channel-selection method for epileptic-seizure classification based on multi-objective optimization, *Front. Neurosci.*, **14** (2020), 593. doi: 10.3389/fnins.2020.00593.

41. L. A. Moctezuma, M. Molinas, Classification of low-density eeg for epileptic seizures by energy and fractal features based on emd, *J. Biomed. Res.*, **34** (2020), 180–190. doi: 10.7555/JBR.33.20190009.