

Effective and extensible feature extraction method using genetic algorithm-based frequency-domain feature search for epileptic EEG multiclassification

Tingxi Wen, MS, Zhongnan Zhang, PhD*

Abstract

In this paper, genetic algorithm-based frequency-domain feature search (GAFDS) method is proposed for the electroencephalogram (EEG) analysis of epilepsy. In this method, frequency-domain features are first searched and then combined with nonlinear features. Subsequently, these features are selected and optimized to classify EEG signals. The extracted features are analyzed experimentally. The features extracted by GAFDS show remarkable independence, and they are superior to the nonlinear features in terms of the ratio of interclass distance and intraclass distance. Moreover, the proposed feature search method can search for features of instantaneous frequency in a signal after Hilbert transformation. The classification results achieved using these features are reasonable; thus, GAFDS exhibits good extensibility. Multiple classical classifiers (i.e., k-nearest neighbor, linear discriminant analysis, decision tree, AdaBoost, multilayer perceptron, and Naïve Bayes) achieve satisfactory classification accuracies by using the features generated by the GAFDS method and the optimized feature selection. The accuracies for 2-classification and 3-classification problems may reach up to 99% and 97%, respectively. Results of several cross-validation experiments illustrate that GAFDS is effective in the extraction of effective features for EEG classification. Therefore, the proposed feature selection and optimization model can improve classification accuracy.

Abbreviations: AB = AdaBoost, ANN = artificial neural network, DT = decision tree, EEG = electroencephalogram, EER = extreme energy ratio, FFT = fast Fourier transformation, GA = genetic algorithm, GAFDS = genetic algorithm-based frequency-domain feature search, k-NN = k-nearest neighbor, LDA = linear discriminant analysis, MFDFA = multifractal detrended fluctuation analysis, MLP = multilayer perceptron, NB = Naïve Bayes, SEER = semisupervised extreme energy ratio, SVM = support vector machine.

Keywords: EEG classification, epilepsy, GAFDS, nonlinear features

1. Introduction

Epilepsy is a chronic disease characterized by a sudden abnormal discharge of brain neurons. In 2013, over 50 million patients were afflicted with epilepsy worldwide, with most patients originating from developing countries.^[1] Approximately 9 million epileptic patients were recorded in China in 2011. Every year, 600,000 new epileptic patients are recorded.^[2] In China,

epilepsy has become the second most common nerve disease, coming in second to headache. Therefore, the accurate diagnosis and prediction of epilepsy are significant. Electroencephalogram (EEG) signals are often used to evaluate the neural activities of the brain. These signals, which are acquired by electrodes placed on the scalp, can reflect the state of brain neurons at a specific time. The recorded EEG signals are complex, nonlinear, unstable, and random because of the complex interconnection among billions of neurons. Several scholars have focused on EEG signal analysis and processing to aid in the diagnosis and treatment of epilepsy.

The first step in EEG signal analysis is to extract and select relevant features. The major signal feature extraction methods are based on time-domain, frequency-domain, time–frequency domain, and nonlinear signal analyses.^[1] Altunay et al^[3] presented a method for epileptic EEG detection based on time-domain features. Chen et al^[4] extracted features of EEG signals by using Gabor transform and empirical mode decomposition, which involves frequency-domain and time–frequency domain technologies. For nonlinear signal analysis, Zhang and Chen^[5] extracted 6 energy features and 6 sample entropy features of EEG signals. In feature extraction, researchers have often mixed multiple methods and obtained new features by various models. Zhang et al^[6] combined an autoregressive model and sample entropy to extract features, and results showed that the combined strategy can effectively improve the classification of EEG signals. Geng et al^[7] used correlation dimension and Hurst exponent to extract nonlinear features. Ren and Wu^[8] used convolutional

Editor: Mark Stecker.

This work was supported by the Science and Technology Guiding Project of Fujian Province, China (2016H0035); the Enterprise Technology Innovation Project of Fujian Province, China; and the Science and Technology Project of Xiamen, China (3502Z20153026).

The authors have no conflicts of interest to disclose.

Software School, Xiamen University, Xiamen, Fujian, China.

* Correspondence: Zhongnan Zhang, Software School, Xiamen University, Xiamen, Fujian 361005, China (e-mail: zhongnan_zhang@xmu.edu.cn).

Copyright © 2017 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

Medicine (2017) 96:19(e6879)

Received: 12 January 2017 / Received in final form: 11 March 2017 / Accepted: 15 April 2017

<http://dx.doi.org/10.1097/MD.00000000000006879>

deep belief networks to extract EEG features. By contrast, other researchers have extracted fixed features. Chen et al^[9] extracted dynamic features by recurrence quantification analysis. Tu and Sun^[10] proposed a semisupervised feature extractor called semisupervised extreme energy ratio (SEER). Improving on this work, they further proposed 2 methods for feature extraction, namely, semisupervised temporally smooth extreme energy ratio (EER) and semisupervised importance weighted EER.^[11] Both methods presented better classification capabilities than SEER. Rafiuddin et al^[12] conducted wavelet-based feature extraction and used statistical features, interquartile range, and median absolute deviation to form the feature vector. Wang et al^[13] extracted EEG features by wavelet packet decomposition.

After feature extraction, the selected features should be classified to recognize different EEG signals. Classifiers for EEG classification can be grouped into 5, namely, linear classifiers, neural networks, nonlinear Bayesian classifiers, nearest neighbor classifiers, and combinations of classifiers.^[14] Li et al^[15] used a multiple kernel learning support vector machine (SVM) to classify EEG signals. Murugavel et al^[16] used an adaptive multiclass SVM. Zou et al^[17] classified EEG signals by Fisher linear discriminant analysis (LDA). Djemili et al^[18] fed the feature vector to a multilayer perceptron (MLP) neural network classifier. The classification capacity of a single classification method is limited; thus, an increasing number of researchers have attempted to combine 2 or more methods to improve classification accuracy. For example, Subasi and Erçelebi^[19] adopted an artificial neural network (ANN) and logistic regression to classify EEG signals. Wang et al^[20] combined cross-validation with a k-nearest neighbor (k-NN) classifier to construct a hierarchical knowledge base for detecting epilepsy. Murugavel and Ramakrishnan^[21] proposed a novel hierarchical multiclass SVM integrated with an extreme learning machine as kernel to classify EEG signals. To classify multisubject EEG signals, Choi^[22] used multitask learning, which treats subjects as tasks to capture intersubject relatedness in the Bayesian treatment of probabilistic common spatial patterns.

Researchers have also studied the application of machine learning and optimization algorithms to improve the accuracy of epilepsy detection. Amin et al^[23] compared the classification

accuracy rates of SVM, MLP, k-NN, and Naïve Bayes (NB) classifiers for epilepsy detection. Nunes et al^[24] used an optimum-path forest classifier for seizure identification. Moreover, artificial bee colony^[25] and particle swarm optimization^[26] algorithms were also used to optimize neural networks for EEG data classification. Nevertheless, the study of the application of machine learning and optimization algorithms to epilepsy detection is currently insufficient.

In this paper, genetic algorithm-based frequency-domain feature search (GAFDS) method is proposed. This method searches for effective classification features in the frequency spectrum rather than using the maximum, minimum, and mean values of the frequency spectrum as features. This method can be easily extended to the feature extraction of other spectra. For multiclassification problems, a high classification accuracy can be achieved using the features selected by the GAFDS method. The accuracy can be further improved by combining other nonlinear features. An optimization algorithm is used to optimize feature selection.

2. Methodology

The EEG data used in this study are based on a previous published dataset,^[39] thus ethics approval is not required for this study. As shown in Fig. 1, numerous features of the EEG signals are first extracted by various feature extraction methods. Subsequently, the best feature combination subset is selected by a feature selection method. Finally, the feature combination subset is used by a classifier for the classification of the EEG signals.

2.1. Feature extraction

2.1.1. GAFDS method. Genetic algorithm (GA) is a random search method that simulates the biological laws of evolution. GA is a probability optimization method, which exhibits global optimization capability. The following standard parallel GA^[27] is used in this study to search for features in the frequency domain:

$$GA = \{C, E, P_0, M, \Phi, \Gamma, \Psi, T\}, \quad (1)$$

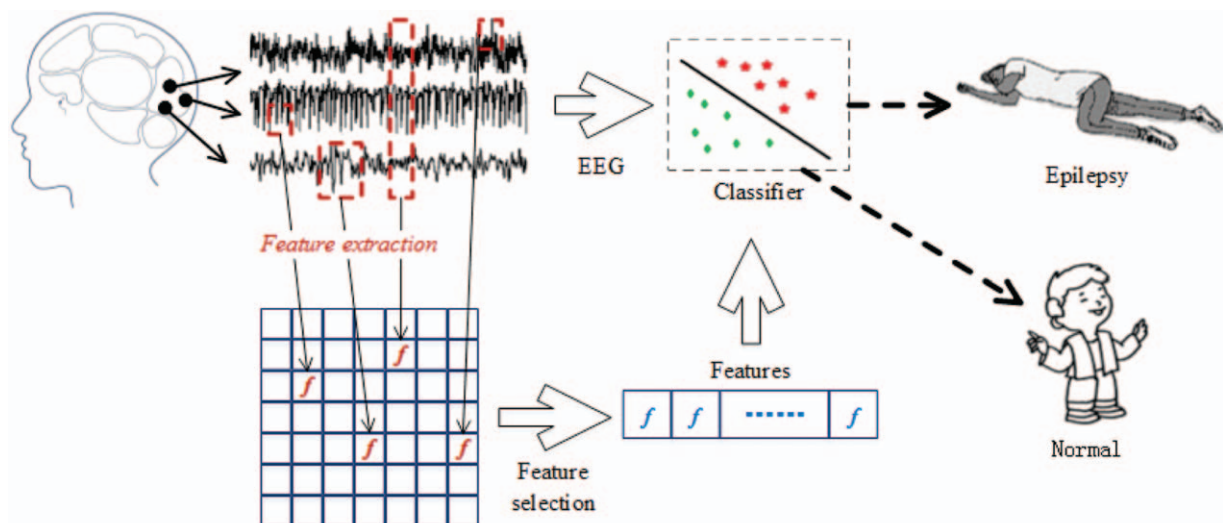


Figure 1. Overall process of EEG signal classification. In feature extraction, GAFDS and extraction methods are used for nonlinear features; a genetic algorithm is used to optimize the selection of the features. The classifiers used for analysis include the k-nearest neighbor, linear discriminant analysis, decision tree, AdaBoost, multilayer perceptron, and Naïve Bayes. EEG = electroencephalogram, GAFDS = genetic algorithm-based frequency-domain feature search.

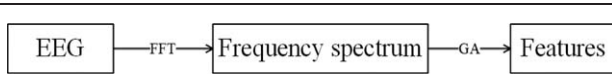


Figure 2. Flowchart of genetic algorithm-based frequency-domain feature search.

where C is the chromosome coding in GA, E is the individual fitness function, P_0 is the initial population, M is the size of the initial population, Φ is the selection operator, Γ is the crossover operator, Ψ is the mutation operator, and T is the given termination condition.

In signal processing, the frequency domain is a coordinate system that describes the frequency features of the signals. Often used to analyze signal features, a frequency spectrogram reflects the relationship between the frequency and amplitude of a signal. The GAFDS method adopts GA to search for a set of classification-suitable features in the frequency spectrum (Fig. 2).

Figure 3 shows the frequency spectrograms of 5 classes of signals (i.e., A, B, C, D, and E) after fast Fourier transformation (FFT). The x-axis represents the frequency, whereas the y-axis represents the amplitude. A significant variation occurs in each class at a certain frequency, such as the amplitudes enclosed in red boxes. By contrast, the amplitudes enclosed in green boxes are difficult to distinguish. The proposed feature extraction method searches for several superior frequency spaces in the frequency spectrogram. Subsequently, the mean values of the amplitudes in the spaces are used as the features. Then, the GA with global search capability is employed to search for the optimal frequency spaces.

A time series $X\{x_1, x_2, \dots, x_n\}$ with a length of n is formed after a signal is sampled. Then, a series $Y\{y_1, y_2, \dots, y_m\}$ with a length of m is obtained by applying FFT to X . For $i, j \in [1, \dots, m]$ and $i < j$,

$$f_{ij} = \frac{1}{j-i+1} \sum_{k=1}^j y_k. \quad (2)$$

The f_{ij} in Eq. (2) is the feature in the frequency interval $[i, j]$.

The main process of using GA in frequency intervals involves obtaining several frequency features with high distinguishing capabilities. Details of this process follow.

1. Individual encoding

On the assumption that the total number of features to be searched for is α , the length of the individual coding array C is 2α . The value of each element in C is between 0 and the highest frequency. Both C_{2i} and C_{2i+1} ($0 \leq i < \alpha$) from C are taken as the frequency range to calculate the features. As shown in Fig. 4, C_3 and C_4 can be used to calculate the feature f_{C_3, C_4} .

However, the constraint $i < j$ should be applied to calculate feature $f_{i,j}$. When $i \geq j$, the feature makes no sense. Therefore, a negative slack variable β (when $i \geq j$, $\beta = 0$) is adopted to implement the constraint.

2. Fitness function

Traversing C to calculate the features yields α features $\{f_g^1, f_g^2, \dots, f_g^\alpha\}$ and α slack variables $\{\beta_1, \beta_2, \dots, \beta_\alpha\}$. For the optimization of the features, the samples should ideally present larger interclass distances and smaller intraclass distances in the feature space. LDA is employed to evaluate these features. The calculation involves a large number of iterations; thus, LDA is used because of its high calculation speed. The fitness value is calculated by

$$\text{Fitness}(C) = \text{LDA}(f_g^1, f_g^2, \dots, f_g^\alpha) - \sum_{i=1}^{\alpha} \beta_i. \quad (3)$$

3. Operators

In GAFDS, Γ is a multipoint crossover operator, Ψ is a Gaussian mutation operator, and Φ is a roulette wheel selection operator.

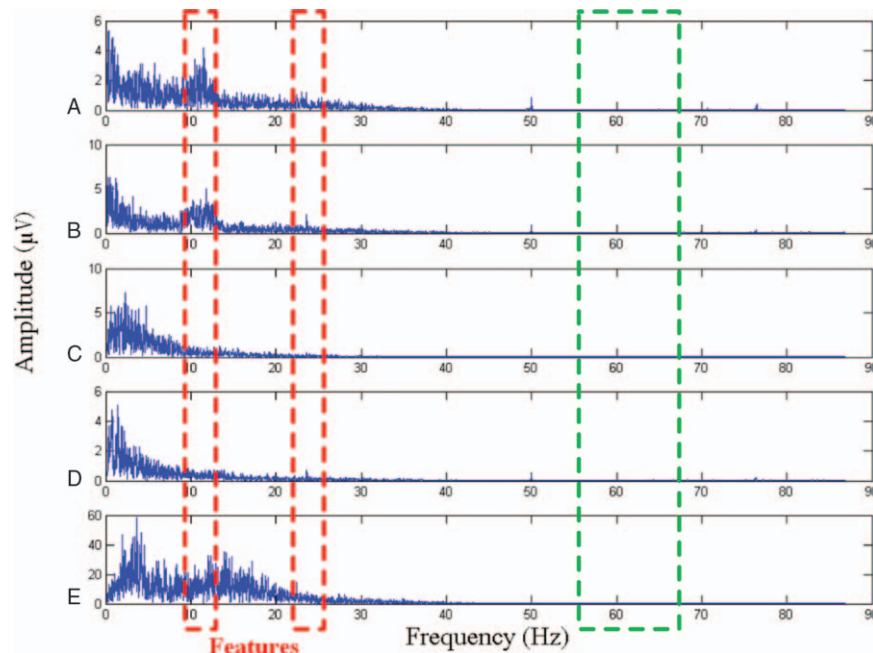


Figure 3. Frequency-domain feature extraction. The features in the red boxes are suitable for classification, whereas those enclosed in green boxes are ineffective. The goal of the genetic algorithm-based frequency-domain feature search method is to find a number of effective features.

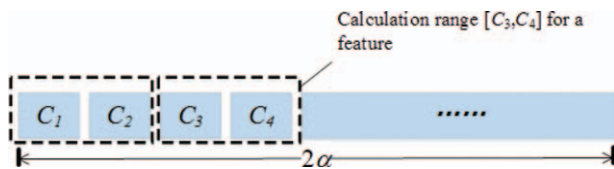


Figure 4. Structure of the individual encoding.

2.1.2. Nonlinear features. An EEG signal is random and unstable; thus, using FFT alone cannot effectively distinguish EEG signals. For this reason, other nonlinear methods, namely, sample entropy, Hurst exponent, Lyapunov exponent, and multifractal detrended fluctuation analysis (MFDFA), are used in this study to extract the effective features.

Proposed by Richman and Moorman^[28] in 2000, sample entropy, which improves on Pincus approximate entropy,^[29] is a measure of regularity to quantify the levels of complexity of a time series. Sample entropy is often used to extract the features of EEG signals.^[5,6] A feature based on sample entropy is defined by

$$f_{se} = \text{Sample Entropy}(X, sn, sm, sr). \quad (4)$$

Sample entropy requires 3 parameters: signal length, sn ; embedding dimension, sm ; and similar tolerance, sr . The value of sr is the standard deviation of X multiplied by parameter χ .

The Hurst exponent was first proposed by England hydrologist H.E. Hurst.^[30] Often used in the chaos–fractal analysis of a time series, it is an index for judging whether the time series data are random walk or biased random walk. In a previous study,^[7] the Hurst exponent was adopted as the main feature for EEG classification and defined by

$$f_{hurst} = \text{Hurst Exponent}(X). \quad (5)$$

The Lyapunov exponent is used for computing how fast nearby trajectories in a dynamic system diverge. This exponent is one of the features used to recognize chaotic motions.^[31] In this study, the largest Lyapunov exponent is used as a feature of EEG and given by

$$f_{lle} = \text{LLE}(Y). \quad (6)$$

In physiology, fractal structures exist in physiological signals. Multifractals can reveal the complexity and inhomogeneity of fractals. MFDFA is the algorithm used for analyzing the multifractal spectrum of a biomedical time series.^[32]

Figure 5 shows the MFDFA-based multifractal spectra of the 5 classes of signals (i.e., A, B, C, D, and E). When the q -order moments of the wave function are $-8, -6, -4, -2, 0, 2, 4, 6$, and 8 , several multifractal spectra are formed. Three points from each multifractal spectrum, namely, p_1 , p_2 , and p_3 , are selected, where p_1 is the point with the minimum h_q , p_2 with the maximum h_q , and p_3 with the maximum D_q . Each point has 2 coordinate values; thus, 6 features are obtained. The 6 features include f_{mf}^1, f_{mf}^2 (i.e., the h_q and D_q values of p_1); f_{mf}^3, f_{mf}^4 (i.e., the h_q and D_q values of p_2); and f_{mf}^5, f_{mf}^6 (i.e., the h_q and D_q values of p_3). The maximum value of D_q is always equal to 1; accordingly, f_{mf}^6 can be removed. Finally, the feature set $\{f_{mf}^1, f_{mf}^2, \dots, f_{mf}^5\}$ can be obtained using MFDFA. At the same

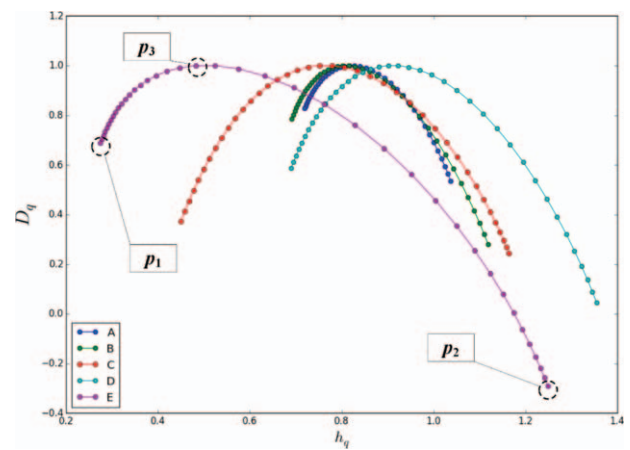


Figure 5. Multifractal detrended fluctuation analysis-based multifractal spectra of the 5 classes of signals. For each sample, the coordinates of p_1 , p_2 , and p_3 in the fractal spectrum are taken as the features of the sample.

time, the detrended fluctuation analysis value of the signal is also a feature:

$$f_{dfa} = \text{DFA}(Y). \quad (7)$$

2.1.3. Feature selection and optimization. Feature extraction is useful in data visualization and comprehension. It reduces the requirement for data calculation and storage as well as the time for training and application. Numerous signal feature extraction algorithms are used in practice. Researchers often combine several feature extraction algorithms to analyze data. However, the use of multiple algorithms usually results in feature dimension expansion and feature redundancy. Feature selection reduces the dimension of a feature space and thus facilitates data training and application.

The selection of an optimal feature subset is a nondeterministic polynomial time problem; therefore, GA is used to search for the optimal feature subset. The algorithm codes individuals in the population in a binary array whose length is the number of features. In the array, 1 means the feature is selected, whereas 0 indicates otherwise. The object function of the algorithm is

$$\min(\text{FPR} - (1 - \text{TPR})), \quad (8)$$

where FPR is the fall-out or false positive rate and TPR is the sensitivity or true positive rate.

2.2. Classification model

After feature extraction, multiple models, including k -NN, LDA, decision tree (DT), AdaBoost (AB), MLP, and NB, are used to classify EEG signals.

Cover and Hart^[33] first proposed k -NN. The main idea of k -NN is as follows: if most of the k samples most similar (nearest in feature space) to a sample belong to a class, then the sample also belongs to that class.

LDA was introduced into pattern recognition and artificial intelligence by Belhumeur et al.^[34] The basic functional concept of LDA is projecting high-dimensional pattern onto the optimal discriminant vector space to extract classification information and reduce feature space dimension. After projection, the pattern samples exhibit the maximum interclass distances and the minimum intraclass distances in the new subspace; that is, the pattern presents the best separability in the space.

DT^[35] implements a group of classification rules represented by a tree structure to minimize the loss function on the basis of the known occurrence probability of each situation. This model is a graphical method that intuitively uses probability analysis. The decision nodes resemble branches of a tree, and thus, the model is called a DT.

AB^[36] is an iterative algorithm that trains different classifiers (weak classifiers) with the same training set and combines these weak classifiers of different weights to construct a stronger classifier (strong classifier).

MLP^[37] is a feed-forward ANN consisting of multiple layers of nodes in a directed graph, with each layer fully connected to the next. Each node is a processing element with an activation function. MLP uses backpropagation for training the network to distinguish data.

NB^[38] is the classification method based on Bayes theorem and feature conditional independence assumption. NB originates from classical mathematics and offers solid mathematical basis and stable classification accuracy. In addition, this model only requires few parameters. It is also insensitive to missing data. In theory, the NB model presents minimal error compared with other classification methods.

2.3. Method flow

Figure 6 shows the process of combining the optimization algorithm and classification model. The original dataset is divided into a training dataset and a testing dataset. After feature extraction and selection, a feature subset is acquired from the training dataset. Subsequently, on the basis of the feature subset, features are separately extracted from the training dataset and

testing dataset to obtain the new training set and testing set. Finally, a new training set is used to train the classifier and test the classifier on the new testing set.

3. Experiments and results

3.1. Dataset

The dataset is obtained from the work of Andrzejak et al.^[39] It includes 5 classes of data (i.e., A, B, C, D, and E). Each class has 100 single-channel EEG samples, each with a length of 23.6 s. The sampling frequency is 173.61 Hz; thus, each sample is a time series with 4097 numbers. Sets A and B are collected when healthy volunteers open and close their eyes. Sets C, D, and E are from epileptics. The samples in set D are recorded from the epileptogenic zone, whereas those in set C are obtained from the hippocampal formation of the opposite hemisphere of the brain. Sets C and D contain only the activities measured during seizure-free intervals, whereas set E only contains seizure activities. The relation among these 5 classes of EEGs is shown in Fig. 7. Numerous previous studies focused on A, E classification; {C, D}, E classification; A, D, E classification; and A, B, C, D, E classification. This paper examines A, E classification; {C, D}, E classification; and A, D, E classification. Figure 8 illustrates the sample data of the 5 classes.

3.2. Environment and parameter

All the algorithms are written using Python programming language and run on a computer (Ubuntu 14.04 LTS, Core i7-6850K CPU, 3.6 GHz, 128 GB memory space). The algorithms are implemented using Pyevolve^[40] and Scikit-learn^[41] libraries.

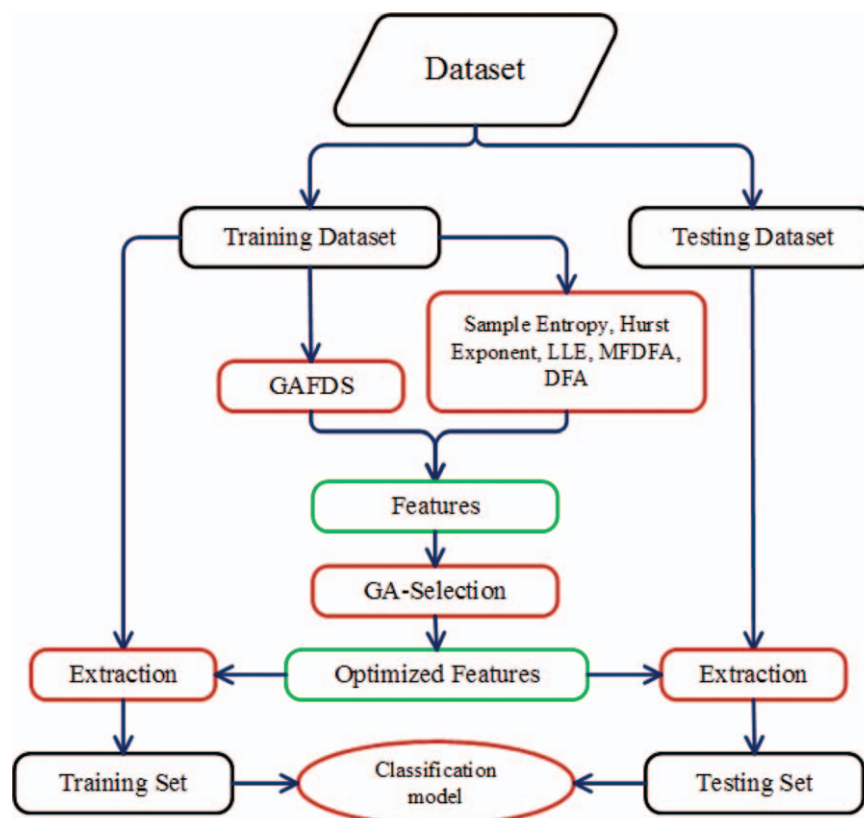


Figure 6. Combination of the optimization algorithm and classification model.

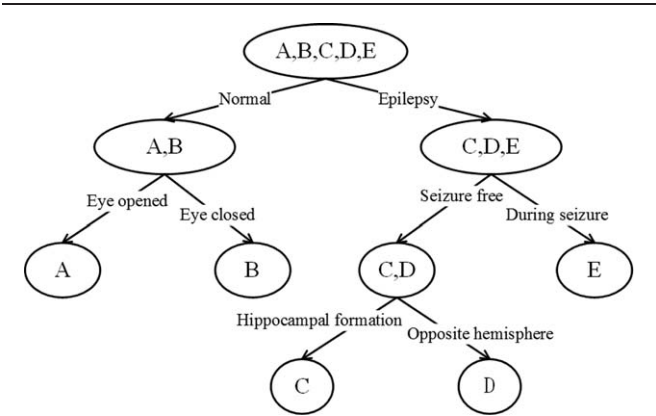


Figure 7. Relation graph of the 5 classes of electroencephalogram signals. A and B are captured from the scalp, whereas C, D, and E come from the intracranial electrodes.

Table 1 lists the parameters used in the experiments. These parameters are based on the general guidelines given in the literature and the authors’ computational experiments on the proposed algorithms.

3.3. Results

3.3.1. Feature extraction and selection. The results of the feature selection inevitably influence the classification results. The nonlinear features of signals do not change across different classification problems. However, the optimization variables of the object function of GAFDS vary across different classification problems, and the selected features are different. In this study, the object function of GAFDS distinguishes the 5 classes. When $\alpha = 4$, 4 features (i.e., f_1, f_2, f_3 , and f_4) are extracted by GAFDS, and 5 features (i.e., f_5, f_6, f_7, f_8 , and f_9) are extracted by MFDFA. In addition, sample entropy f_{10} , Hurst exponent f_{11} , largest Lyapunov exponent f_{12} , and DFA feature f_{13} are obtained.

Table 1		
Parameters and their values.		
N	Description	Value
1	Population size	100
2	Maximum number of generations	1000
3	Crossover probability	0.90 per pair
4	Mutation probability	0.03 per integer
5	α	4
6	β	0.05
7	χ	2
8	sm	2

Figure 9 shows the distributions of the 5 classes of samples in feature space f_1 . Most feature values of class A are below 3, whereas most feature values of class E exceed 3. Therefore, classes A and E can be distinguished by feature f_1 . Approximately 50% of the samples of classes B and C can be classified by f_1 , but the other samples are mixed. Most class C and class D samples are mixed and difficult to distinguish. Overall, classes C and D present outliers, and class E is discrete. Figure 10 shows the distributions of each class in different feature spaces. The median increases from class A to class E for f_4 but decreases for f_6 . Feature f_{13} can effectively distinguish {B, C} and {B, D}. This analysis shows that a single feature usually distinguishes 2 classes at most. Therefore, further analysis of the feature combination is necessary.

As shown in Fig. 11, features $f_1, f_4, f_6, f_7, f_9, f_{10}, f_{11}$, and f_{13} are extracted to construct different 2-dimensional feature spaces. A point in the space represents a sample of a class. In space $\{f_1, f_{11}\}$, all samples of the 5 classes are concentrated on the left side, and they are difficult to separate. In space $\{f_4, f_9\}$, the outlines of classes A, B, E, and {C, D} are clear, but classes C and D are mixed. Every class is discrete in space $\{f_6, f_7\}$ and $\{f_{10}, f_{13}\}$; however, in space $\{f_{10}, f_{13}\}$, each class occupies a certain distribution area, and each class crosses only at the edges.

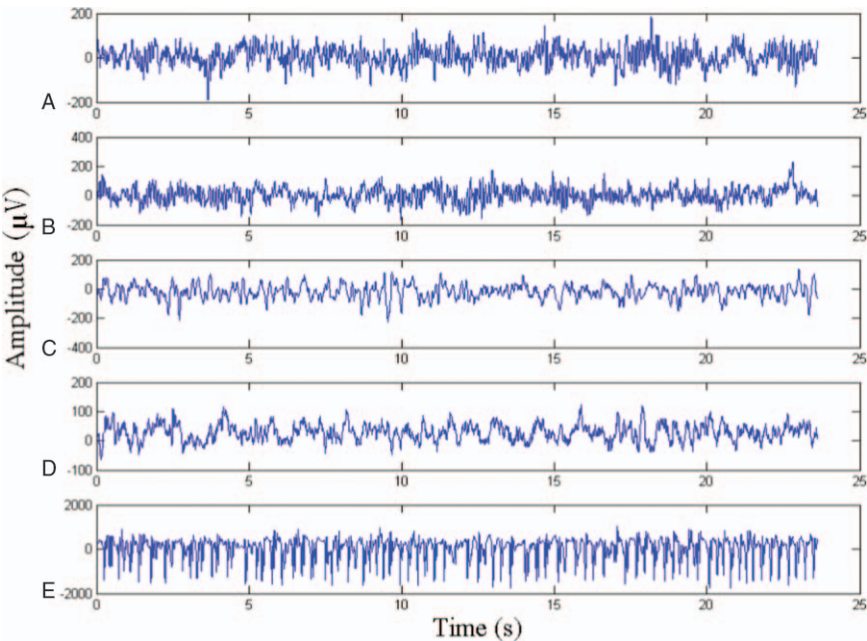


Figure 8. Sample data of the 5 classes (i.e., A, B, C, D, and E).

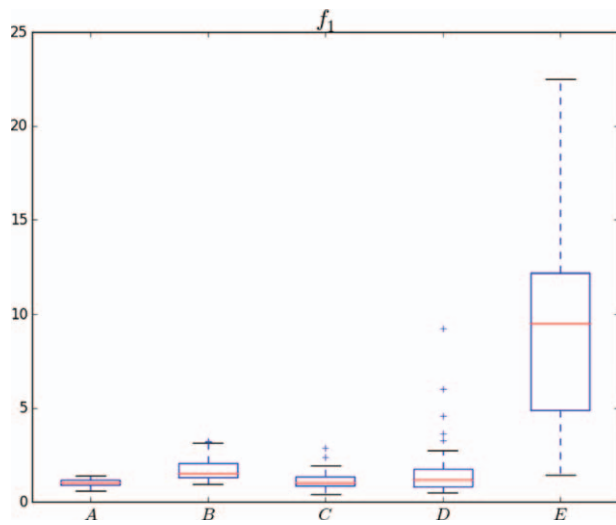


Figure 9. Distributions of the 5 classes of samples in f_1 . A and B as well as E and {A, B, C, D} are distinguishable.

In 1- or 2-dimensional feature spaces, features can be directly observed. However, as the dimension increases, feature evaluation based on distances is the most direct method, regardless of the classifier. In this study, the features $\{f_1, f_2, f_3, f_4\}$ extracted by GAFDS are evaluated by comparing the ratios of the interclass distance and intraclass distance of all the classes with those of the nonlinear features $\{f_{10}, f_{11}, f_{12}, f_{13}\}$. The interclass distance between 2 classes is their distance in a feature space. With n samples in class A, n feature vectors $\{v_1^A, \dots, v_i^A, \dots, v_j^A, \dots, v_n^A\}$ are generated after feature extraction. Vector v_i^A represents sample $i \in A$. Accordingly, the intraclass distance of class A is calculated by

$$\text{dist1} = \frac{1}{n \cdot n} \sum_{i=1}^n \sum_{j=1}^n (v_i^A - v_j^A)^2. \quad (9)$$

The distance between sample i and itself is 0. The interclass distance between n samples in class A and m samples in class B is calculated by

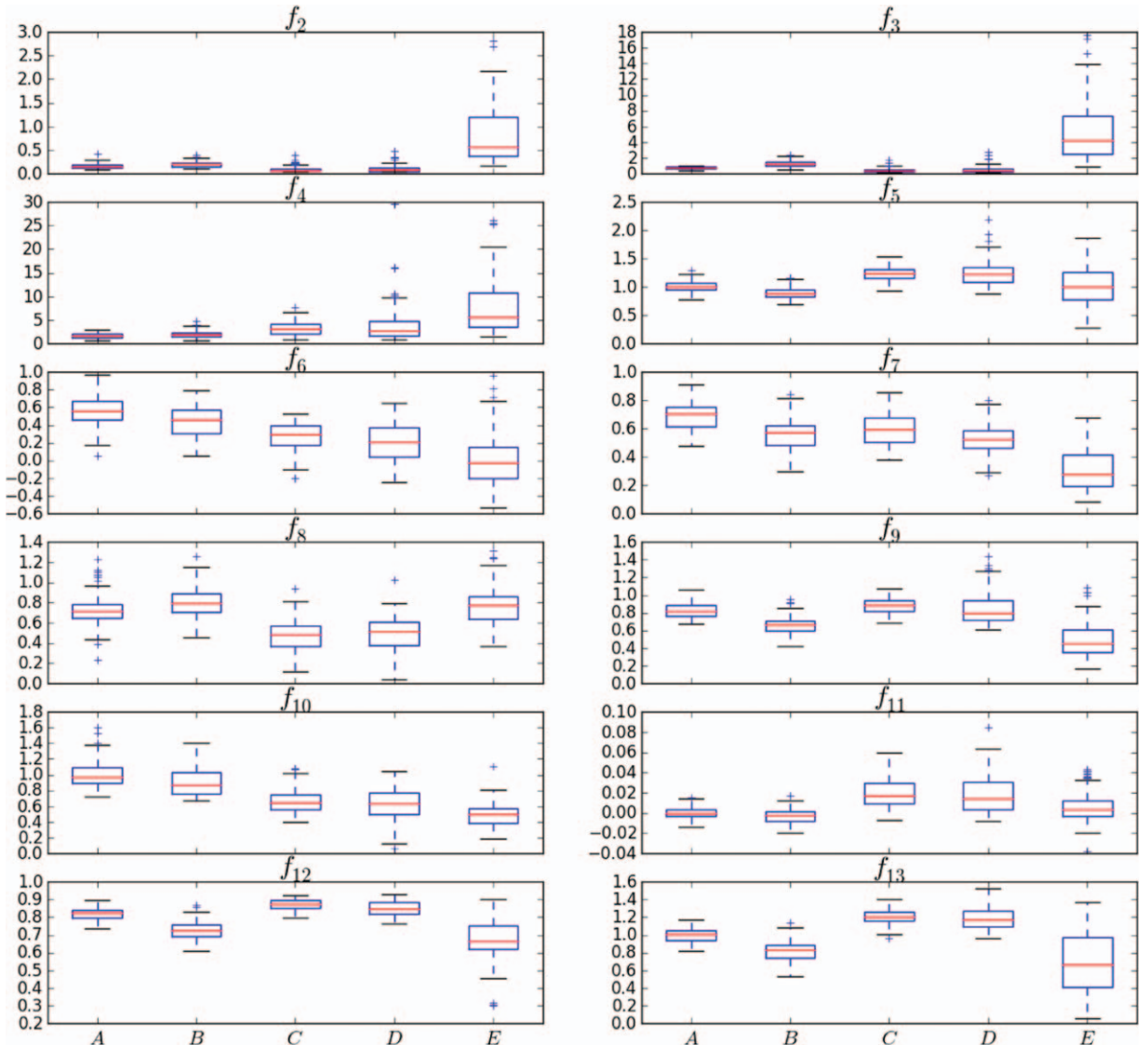


Figure 10. Distributions in feature spaces f_2 – f_{13} of the 5 classes of samples.

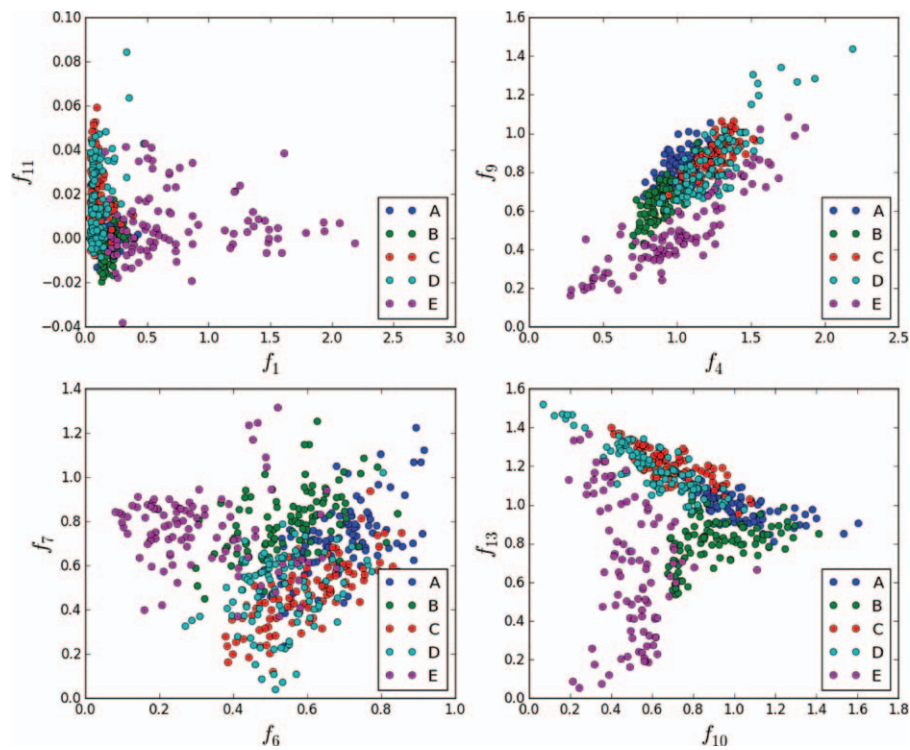


Figure 11. Distributions of the 5 classes in different 2-dimensional feature spaces. In the 2-dimensional combination space (f_{10}, f_{13}) , classes E, B, and {A, C, D} are evidently divided into 3 parts.

Table 2

Ratios of the interclass and intraclass distances of the 5 classes in feature spaces $\{f_1, f_2, f_3, f_4\}$ and $\{f_{10}, f_{11}, f_{12}, f_{13}\}$.

	f_1, f_2, f_3, f_4					$f_{10}, f_{11}, f_{12}, f_{13}$				
	A	B	C	D	E	A	B	C	D	E
A	1	0.745	0.826	2.996	0.743	1	0.914	1.952	2.355	1.642
B	1.793	1	1.318	7.067	1.110	1.238	1	2.616	3.296	2.061
C	2.434	1.613	1	8.405	1.009	1.980	1.958	1	2.125	1.233
D	1.176	1.153	1.120	1	1.198	1.368	1.413	1.217	1	1.521
E	4.250	2.639	1.959	17.460	1	2.166	2.006	1.603	3.452	11

The bold-emphasized value illustrates that the ratio of the interclass and intraclass distances in feature space $\{f_1, f_2, f_3, f_4\}$ is greater than that in feature space $\{f_{10}, f_{11}, f_{12}, f_{13}\}$. For example, r_{BA} is 1.793 in feature space $\{f_1, f_2, f_3, f_4\}$, but its value in feature space $\{f_{10}, f_{11}, f_{12}, f_{13}\}$ is 1.238.

$$\text{dist2} = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m (v_i^A - v_j^B)^2. \quad (10)$$

Therefore, the ratio of the interclass distance and intraclass distance between class A and class B is

$$r_{AB} = \text{dist2} / \text{dist1}. \quad (11)$$

The ratio of the interclass distance and intraclass distance between a class and itself is 1; that is, $r_{AA} = r_{BB} = r_{CC} = r_{DD} = r_{EE} = 1$.

Table 2 shows the ratios of the interclass distance and intraclass distance of the 5 classes in feature spaces $\{f_1, f_2, f_3, f_4\}$ and $\{f_{10}, f_{11}, f_{12}, f_{13}\}$. The features retrieved by GAFDS are comparable to the nonlinear features.

The 5 classes of samples in $\{f_1, f_2, f_3, f_4\}$ and $\{f_{10}, f_{11}, f_{12}, f_{13}\}$ feature spaces are simultaneously classified by numerous classical classifiers. The classifiers used, namely, k-NN, LDA, DT, AB, MLP, and NB, are included in the Scikit-learn library. For the k-NN classifier, $k=3$. The maximum depth of DT is 5. For the

MLP, $\alpha=1$. Other parameters use the default values in the library.

A, E classification involves classifying the EEG signals produced by healthy people and epileptics. As shown in Table 3, numerous classifiers can achieve high accuracies by using the listed features.

The classification of {C, D} and E means classifying EEG signals produced during seizure-free intervals and during

Table 3

Classification accuracies of the common classifiers for classes A and E in feature spaces $\{f_1, f_2, f_3, f_4\}$ and $\{f_{10}, f_{11}, f_{12}, f_{13}\}$ (using k-fold cross-validation, $k=5$).

A, E	k-NN	LDA	DT	MPL	AB	NB
$\{f_1, f_2, f_3, f_4\}$	0.995	0.89	0.995	0.995	0.995	0.995
$\{f_{10}, f_{11}, f_{12}, f_{13}\}$	0.995	0.995	0.975	0.985	0.97	0.98

AB=AdaBoost, DT=decision tree, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 4

Classification accuracies of common classifiers for {C, D} and E classes in feature spaces $\{f_1, f_2, f_4, f_5, f_7, f_8, f_9, f_{10}, f_{12}\}$ (using k-fold cross-validation, k=5).

{C, D}, E	k-NN	LDA	DT	MPL	AB	NB
$\{f_1, f_2, f_4, f_5, f_7, f_8, f_9, f_{10}, f_{12}\}$	0.973	0.983	0.957	0.98	0.987	0.977

AB=AdaBoost, DT=decision tree, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 5

Classification accuracies of the common classifiers for A, D, E in feature spaces $\{f_1, f_2, f_4, f_5, f_6, f_9, f_{10}, f_{11}, f_{12}, f_{13}\}$ (using k-fold cross-validation, k=5).

A, D, E	k-NN	LDA	DT	MPL	AB	NB
$\{f_1, f_2, f_4, f_5, f_6, f_9, f_{10}, f_{11}, f_{12}, f_{13}\}$	0.967	0.917	0.967	0.933	0.967	0.917

AB=AdaBoost, DT=decision tree, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 6

Accuracies for {C, D} and E classification based on the features extracted by GAFDS.

Cross-validation	k-NN	LDA	DT	MPL	AB	NB
2-fold	0.963	0.910	0.947	0.943	0.963	0.947
5-fold	0.967	0.907	0.963	0.953	0.973	0.950
10-fold	0.950	0.923	0.973	0.96	0.970	0.950

AB=AdaBoost, DT=decision tree, GAFDS=genetic algorithm-based frequency-domain feature search, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 7

Accuracies for {C, D} and E classification based on the features optimized by GA selection.

Cross-validation	k-NN	LDA	DT	MPL	AB	NB
2-fold	0.977	0.983	0.960	0.963	0.973	0.95
5-fold	0.983	0.983	0.973	0.97	0.98	0.957
10-fold	0.98	0.983	0.96	0.976	0.99	0.967

AB=AdaBoost, DT=decision tree, GA=genetic algorithm, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 8

Accuracies for A, D, E classification based on features extracted by GAFDS.

Cross-validation	k-NN	LDA	DT	MPL	AB	NB
2-fold	0.93	0.857	0.917	0.9	0.693	0.81
5-fold	0.973	0.887	0.897	0.9	0.67	0.823
10-fold	0.937	0.886	0.92	0.913	0.71	0.817

AB=AdaBoost, DT=decision tree, GAFDS=genetic algorithm-based frequency-domain feature search, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 9

Accuracies for A, D, E classification based on the features optimized by GA selection.

Cross-validation	k-NN	LDA	DT	MPL	AB	NB
2-fold	0.93	0.94	0.913	0.916	0.68	0.897
5-fold	0.96	0.953	0.927	0.94	0.587	0.907
10-fold	0.943	0.953	0.92	0.933	0.71	0.893

AB=AdaBoost, DT=decision tree, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

Table 10

Classification accuracies of the common classifiers for A and E classes in feature spaces $\{f_1^*, f_2^*, f_3^*, f_4^*\}$ (using k-fold cross validation, k=5).

A, E	k-NN	LDA	DT	AB	MPL	NB
$\{f_1^*, f_2^*, f_3^*, f_4^*\}$	0.995	0.89	0.99	0.925	0.99	0.995

AB=AdaBoost, DT=decision tree, k-NN=k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes.

seizure. GA is used to select the features for this classification. The results are shown in Table 4. A, D, E classification is the classification of EEG signals acquired from healthy people, seizure-free epileptics, and epileptics during seizure. The results are shown in Table 5.

3.3.2. Classification results. In the previous section, the features extracted by GAFDS are optimized for the classification problems of the 5 classes. However, for real binary or 3-classification problems, the optimization object should be set according to the requirement; that is, the object function in Eq. (3) should be adjusted.

Table 3 shows good results for A, E classification. For the {C, D} and E classification problem, Table 6 lists the results based on the features extracted by GAFDS, whereas Table 7 illustrates the results based on the features optimized by GA selection from the GAFDS-obtained features and other features.

For a multiclassification problem, Tables 8 and 9 show that GAFDS and the feature selection method can obtain good results for A, D, E classification.

4. Discussion

4.1. GAFDS method

EEG signals are nonlinear, time varying, and unbalanced. FFT is a global linear method. However, a frequency spectrum does not reflect the frequency changes in the time domain; thus, FFT has certain limitations when applied to nonstationary signal analysis. As shown in Table 2, the features extracted by GAFDS present poor cohesiveness compared with other features. For example, class E presents a wide distribution in f_6 , whereas class D has numerous outliers. Thus, features $\{f_1, f_2, f_3, f_4\}$ in Figs. 9 and 10 are standardized to obtain new features $\{f_1^*, f_2^*, f_3^*, f_4^*\}$ within the range [0, 1]. After feature standardization, the accuracy of the AB classifier presents minimal reduction, whereas those of other classifiers remain unchanged, as shown in Table 10. These results indicate that the features extracted by GAFDS have better independence.

Table 2 presents a comparison based on Eq. (11) between the features extracted by GAFDS with nonlinear features. As shown in the table, r_{BA} is good even though the features extracted by GAFDS present poor cohesiveness. Therefore, the features extracted by GAFDS are superior to the nonlinear features in A, E classification. Furthermore, GAFDS presents great extensibility. GAFDS selects features by searching a frequency spectrum; however, it can also search for new features in a Hilbert spectrum and several other signal spectra. Figure 12 shows the instantaneous frequency change in the samples in classes A and E in the time domain after Hilbert transformation. A period along the time axis can also be searched using GAFDS. Subsequently, the average value of the positive instantaneous frequencies in this period can be used as a feature. With this feature, the accuracy of the LDA classifier

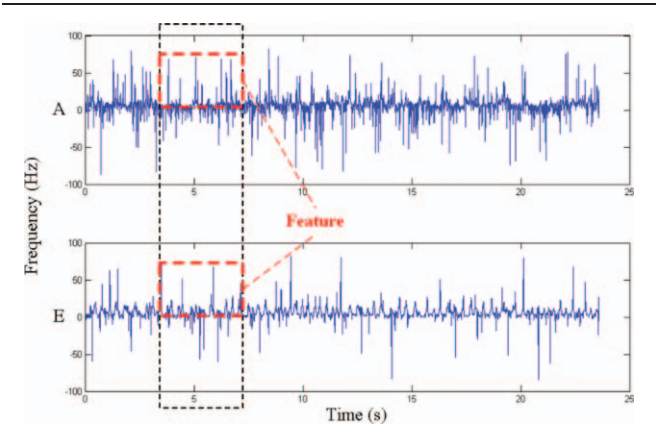


Figure 12. Instantaneous frequencies of the samples in classes A and E whose mean value *t* in the red box is taken as the feature of the sample.

can reach up to 74.5% when classifying samples in classes A and E.

4.2. Analysis and comparison of classification results

As shown in Tables 3–5, the classifiers have different accuracies in different feature spaces. This study uses few features and small searching space. The classification results show that the GA-based feature selection can obtain superior feature combination.

For the A, E classification problem, the features extracted by GAFDS can effectively facilitate classification. For the {C, D} and E classification, Tables 6 and 7 show that the classification accuracy increases after combining new features with the features extracted by GAFDS and feature selection. However, when the complexity of the problem increases, such as in the A, D, E classification (Tables 8 and 9), the classification accuracies of the classifiers using the features extracted by GA selection are not significantly higher than those of the classifiers using the features only generated by GAFDS. Furthermore, the AB classifier performs well in the 2-classification problem but poorly in the multiclassification problem because the parameters of the classifiers are not optimized.

Table 11 shows a comparison of the classification results between recent classifiers and the classification scheme proposed in this paper. Using wavelet transform-based statistical features,

largest Lyapunov exponent, and approximate entropy features, Murugavel et al^[21] developed an ANN and hierarchical multiclass SVM with a new kernel classifier to improve the accuracy for A, D, E classification to 96%. Sharma and Pachori^[46] used the features based on 2- and 3-dimensional phase space representation of intrinsic mode functions as well as an SVM classifier to classify {C, D}, E. Their work achieved a classification accuracy of 98.67%. The scheme proposed in the present study exhibits better classification results with the use of several classifiers based on the GAFDS-selected features and nonlinear features.

5. Conclusion

EEG provides important information for epilepsy detection. Feature extraction, selection, and optimization methods exert significant influence in EEG classification. In this study, a GA-based frequency feature search method is proposed for EEG classification. The method presents global searching capability to search for classification-suitable features in EEG frequency spectra and combine these features with nonlinear features. Finally, GA is used to select effective features from the feature combination to classify EEG signals.

The experimental results show that the standardization and normalization of the features extracted by GAFDS do not affect the accuracy of the classification results and thus indicate that the features extracted by GAFDS have good independence. Compared with nonlinear features, GAFDS-based features allow for high classification accuracy. Furthermore, GAFDS can effectively extract features of instantaneous frequency in the signal after Hilbert transformation; thus, GAFDS presents good extensibility.

For the A, E and {C, D}, E 2-classification problems and the A, D, E 3-classification problem, the GAFDS-based features and optimized features are used by several classifiers (i.e., k-NN, LDA, DT, AB, MLP, and NB). The classification accuracies achieved are better than those by previous classification models.

In our future work, we will use GAFDS to extract new features and use time-domain, frequency-domain, or time–frequency domain features in feature optimization and selection to achieve improved classification accuracy. The parameters and performance of GAFDS also need further improvement. The precision, complexity, and dimension of EEG data increase; thus, we need to continuously improve the extraction method and conduct further research on feature selection optimization to meet the challenging requirements of EEG analyses.

Table 11
Comparison of the results of existing models for EEG classification and the scheme proposed in this paper.

Work	Year	Feature extraction	Classification	Set	Accuracy, %
Chandaka et al ^[42]	2009	Cross-relation	Support vector machine	A, E	99
Guo et al ^[43]	2010	Line length feature	Artificial neural network	A, E	98
Liang et al ^[44]	2010	Wavelet transform and line length feature	Artificial neural network	A, E	97
Pachori et al ^[45]	2014	SODP of IMFs	Artificial neural network	{C, D}, E	97.75
Sharma et al ^[46]	2015	2D PSR of IMFs and IQR of the Euclidian distances of the 3D PSR of IMFs	Support vector machine	{C, D}, E	98.67
Murugavel et al ^[21]	2016	Wavelet transform-based statistical features, largest Lyapunov exponent, and approximate entropy	Artificial neural network, hierarchical multi-class SVM with new kernel	A, E	99
Present work	2016	GAFDS and nonlinear features	k-NN, LDA, DT, AB, MPL, and NB	A, D, E	96
				A, E	99
				{C, D}, E	99
				A, D, E	97

AB=AdaBoost, DT=decision tree, EEG = electroencephalogram, GAFDS=genetic algorithm-based frequency-domain feature search, IMFs=intrinsic mode functions, IQR=interquartile range, k-NN= k-nearest neighbor, LDA=linear discriminant analysis, MLP=multilayer perceptron, NB=Naïve Bayes, PSR=phase space representation, SODP=second-order difference plot, SVM=support vector machine.

References

- [1] Acharya UR, Vinitha Sree S, Swapna G, et al. Automated EEG analysis of epilepsy: a review. *Knowledge Based Syst* 2013;45:147–65.
- [2] Yu PM, Zhu GX, Ding D, et al. Treatment of epilepsy in adults: expert opinion in China. *Epilepsy Behav* 2011;23:36–40.
- [3] Altunay S, Telatar Z, Eroglu O. Epileptic EEG detection using the linear prediction error energy. *Expert Syst Appl* 2010;37:5661–5.
- [4] Chen L, Zhao E, Wang D, et al. Feature extraction of EEG signals from epilepsy patients based on Gabor Transform and EMD decomposition. *Int Conf Nat Comput* 2010;3:1243–7.
- [5] Zhang A, Chen Y. EEG feature extraction and analysis under drowsy state based on energy and sample entropy. Paper presented at: 2012 5th International Conference on BioMedical Engineering and Informatics; 2012:501–5.
- [6] Zhang Y, Ji X, Liu B, et al. Combined feature extraction method for classification of EEG signals. *Neural Comput Appl* 2016; DOI: 10.1007/s00521-016-2230-y. Epub ahead of print.
- [7] Geng S, Zhou W, Yuan Q, et al. EEG non-linear feature extraction using correlation dimension and Hurst exponent. *Neurol Res* 2013;33:908–12.
- [8] Ren Y, Wu Y. Convolutional deep belief networks for feature extraction of EEG signal. *Int Joint Conf Neural Netw* 2014;2850–3.
- [9] Chen L, Zou J, Zhang J. Dynamic feature extraction of epileptic EEG using recurrence quantification analysis. Paper presented at: Proceedings of the 10th World Congress on Intelligent Control and Automation, Beijing; 2012:5019–22.
- [10] Tu W, Sun S. Semi-supervised feature extraction with local temporal regularization for EEG classification. *Int Joint Conf Neural Netw* 2011; 75–80.
- [11] Tu W, Sun S. Semi-supervised feature extraction for EEG classification. *Pattern Anal Appl* 2013;16:213–22.
- [12] Rafiuddin N, Uzzaman Khan Y, Farooq O. Feature extraction and classification of EEG for automatic seizure detection. Paper presented at: International Conference on Multimedia, Signal Processing and Communication Technologies, IEEE; 2011:184–187.
- [13] Wang D, Miao DQ, Wang RZ. A new method of EEG classification with feature extraction based on wavelet packet decomposition. *Tien Tzu Hsueh Pao/Acta Electron Sin* 2013;41:193–8.
- [14] Lotte F, Congedo M, Lécuyer A, et al. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng* 2007;4:R1–3.
- [15] Li X, Chen X, Yan Y, et al. Classification of EEG signals using a multiple kernel learning support vector machine. *Sensors* 2014;14:12784–802.
- [16] Murugavel ASM, Ramakrishnan S, Maheswari U, Sabetha BS. Combined seizure index with adaptive multi-class SVM for epileptic EEG classification. Paper presented at: International Conference on Emerging Trends in Vlsi, Embedded System, Nano Electronics and Telecommunication System, IEEE; 2013:1–5.
- [17] Zou L, Wang X, Shi G, Ma Z. EEG feature extraction and pattern classification based on motor imagery in brain-computer interface. Paper presented at: IEEE International Conference on Cognitive Informatics, IEEE; 2010:536–541.
- [18] Djemili R, Bourouba H, Korba MCA. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals. *Biocybern Biomed Eng* 2015;36:285–91.
- [19] Subasi A, Erçelebi E. Classification of EEG signals using neural network and logistic regression. *Comput Methods Prog Biomed* 2005;78:87–99.
- [20] Wang D, Miao D, Xie C. Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Syst Appl* 2011;38:14314–20.
- [21] Murugavel ASM, Ramakrishnan S. Hierarchical multi-class SVM with elm kernel for epileptic EEG signal classification. *Med Biol Eng Comput* 2016;54:149–61.
- [22] Choi S. Multi-subject EEG classification: Bayesian nonparametrics and multi-task learning. Paper presented at: International Winter Conference on Brain-Computer Interface, IEEE; 2015:1.
- [23] Amin HU, Malik AS, Ahmad RF, et al. Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australas Phys Eng Sci Med* 2015;38:1–1.
- [24] Nunes TM, Coelho ALV, Lima CAM, et al. EEG signal classification for epilepsy diagnosis via optimum path forest—a systematic assessment. *Neurocomputing* 2014;136:103–23.
- [25] Satapathy SK, Dehuri S, Jagadev AK. ABC optimized RBF network for classification of EEG signal for epileptic seizure identification. *Egypt Inform J* 2017;18:55–66.
- [26] Lin CJ, Hsieh MH. Classification of mental task from EEG data using neural networks based on particle swarm optimization. *Neurocomputing* 2009;72:1121–30.
- [27] Goldberg DE. *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.; 1989.
- [28] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. *AJP Heart Circ Physiol* 2000;278:H2039–49.
- [29] Pincus SM. Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 1991;88:2297–301.
- [30] Hurst HE. Long term storage capacity of reservoirs. *Trans Am Soc Civil Eng* 1951;116:776–808.
- [31] Wolf A, Swift JB, Swinney HL, et al. Determining Lyapunov exponents from a time series. *Phys D Nonlin Phenomena* 1985;16:285–317.
- [32] Ihlen EA. Introduction to multifractal detrended fluctuation analysis in Matlab. *Front Physiol* 2012;3:141.
- [33] Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inform Theory* 1967;13:21–7.
- [34] Belhumeur PN, Hespanha JP, Vs KME. Fisherfaces: Recognition Using Class Specific Linear Projection. *Computer Vision—ECCV '96*. Berlin: Springer; 1996.
- [35] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 1991;21:660–74.
- [36] Zhu J, Zou H, Rosset S, et al. Multi-class AdaBoost. *Stat Its Interface* 2006;2:349–60.
- [37] Hassanaitlaasri E, Akhouayri ES, Agliz D, et al. Seismic signal classification using multi-layer perceptron neural network. *Int J Comput Appl* 2013;79:35–43.
- [38] Cheeseman P, Stutz J, Fayyad UM, Piatetsky-Shapiro G, Uthurusamy R. Bayesian classification (AutoClass): theory and results. *Advances in knowledge discovery and data mining American Association for Artificial Intelligence, Menlo Park, CA*; 1996:153–80.
- [39] Andrzejak RG, Lehnertz K, Mormann F, et al. 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;64:116–26.
- [40] Perone CS. Pyevolve: a python open-source framework for genetic algorithms. *ACM SIGEVolution* 2009;4:12–20.
- [41] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2012;12:2825–30.
- [42] Chandaka S, Chatterjee A, Munshi S. Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Syst Appl* 2009;36:1329–36.
- [43] Guo L, Rivero D, Dorado J, et al. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J Neurosci Methods* 2010;191:101–9.
- [44] Liang SF, Wang HC, Chang WL. Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP J Adv Signal Process* 2010;2010:1–5.
- [45] Pachori RB, Patidar S. Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions. *Comput Methods Programs Biomed* 2014;113:494–502.
- [46] Sharma R, Pachori RB. Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions. *Expert Syst Appl* 2015;42:1106–17.