

Coal and Gas Outbursts Prediction Based on DWT+FICA-LDA Feature Extraction and QPSO-DELM Classifier

Xuning Liu

China University of Mining and Technology - Beijing Campus

Zhixiang Li

Shijiazhuang University

Zixian Zhang

Liaocheng University

Guoying Zhang (✉ zhanglaoshijidian@163.com)

China University of Mining and Technology - Beijing Campus

Research

Keywords: Coal and gas outbursts, Discrete wavelet transform, Fast independent component analysis, Linear discriminant analysis, Deep extreme learning machine, Quantum particle swarm optimization

Posted Date: November 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-111736/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Coal and gas outbursts prediction based on DWT+FICA-LDA feature extraction and QPSO-DELM classifier

Xuning Liu^{1,2}, Zhixiang Li², Zixian Zhang^{1,3}, Guoying Zhang^{1*}

¹ School of Mechanical Electronic & Information Engineering, China University of Mining and Technology, Beijing, China

² Department of Computer Engineering, Shijiazhuang University, Shijiazhuang, China

³ School of Foreign University, Liaocheng University, Liaocheng, China

Corresponding author: E-mail: Guoying Zhang (zhanglaoshijidian@163.com)

Abstract: Due to the severity and great harm of coal and gas outbursts accidents, outbursts prediction becomes very necessary, the paper presents a hybrid prediction model of feature extraction and pattern classification for coal and gas outbursts. First, Discrete wavelet transform (DWT) is utilized as a processing technique to decompose subseries and extract the features with different frequency, and the optimal feature components are retained; Second, in order to eliminate the redundancy between the features and uncorrelation between feature and outbursts, we use the fast independent component analysis(FICA) feature extraction method based on high-order statistics to obtain each independent feature, obtaining the global information in the feature; then the obtained features are input into linear discriminant analysis(LDA), under the guidance of class label, then the local information in features are obtained; Finally, the projected features are input into the deep extreme learning machine(DELM) classifier based on the optimal parameters by quantum particle swarm optimization(QPSO) for training and classification. The experimental results on the data set of coal and gas outbursts show that compared with other models in the current prediction of coal and gas outbursts, this method has significant effect on various indicators such as speed and recognition effect.

Keywords: Coal and gas outbursts; Discrete wavelet transform; Fast independent component analysis; Linear discriminant analysis; Deep extreme learning machine; Quantum particle swarm optimization.

1 Introduction

Coal and gas outbursts is a dynamic process with complex mechanism and many influencing factors, its mechanism and evolution law are not clear up to now, and outbursts disaster cannot be effectively prevented and controlled, which seriously threatens the safety of coal mine production[1]. Therefore, fast and accurate dynamic prediction of coal and gas outbursts has important practical significance for coal mine safety production. The indexes selection of coal and gas outbursts influencing factors play an important part in the effect of outbursts prediction. However, the acquisition of outbursts indexes has noise due to the pollution of measurement accuracy, calculation accuracy, recording and storage process. There exit redundancy between features and irrelevance between feature and outbursts. If the original features are directly used for classifier prediction, the accuracy of classifiers cannot be guaranteed, and the efficiency is very poor, so for the prediction of coal and gas outbursts, feature denoising and feature extraction are very necessary. The accuracy and reliability of outbursts prediction depend on the selected indexes, and the selection of prediction indexes depends on the preprocessing degree of coal and gas outbursts indexes.

Coal and gas outbursts shows the characteristics of random, complexity, non-linear and non-stationary[2], in order to reduce the complexity of coal and gas outbursts prediction, eliminate the influence of noise caused by uncertainty and local random factors on prediction process, recover the whole statistical law of original data as possible regular distribution, as far as possible to maintain the authenticity and integrity of the data based on the interference, it is very necessary to perform feature denoising. However, the traditional feature denoising analysis such as the Fourier transform, Kalman Filter and Particle Filter have some defects, although these methods can achieve better denoising effect, it also cause the loss of some details in the signal, which make it difficult to further

perform the feature extraction and classification, these methods also cannot reflect the key and fundamental characteristics of non-linear and non-stationary time-frequency local characteristics[3]. Wavelet analysis[4] is used to process the characteristic signal, judge the main frequency band range of the signal, obtain the energy distribution and characteristics in different frequency bands, and normalize the relative energy of each frequency band. Wavelet transform signal is decomposed into two kinds: high frequency and low frequency component. The low-frequency part signal contains the main signal performance, and the high-frequency part signal contains more noise signal. Finally, wavelet denoising is realized by smoothing the high-frequency part signal and reconstructing it. The main advantage of DWT is that it can provide more local information by decomposing the signal, so as to solve the above problems effectively[5]. The collected features sample data may be polluted by noise to a certain extent, which interferes with the quality of the data, it is the fluctuation of the data that submerges the original features, which brings difficulties to data analysis. This chapter uses this method to solve the feature denoising and other related problems. The high frequency part of the feature is reconstructed smoothly to realize the denoising. Wavelet theory has the advantages of decorrelation, multi-resolution and base selection flexibility[6]. It can effectively distinguish the noise and mutation components in the sample feature data, remove the noise and obtain the non-stationary characteristics, in this paper DWT is applied in the feature preprocessing.

Many scholars have researched the methods of coal and gas outbursts prediction, the research results mainly focus on the two links of feature extraction and pattern classification. At present, there are many feature extraction methods of coal and gas outbursts influencing factors, which are divided into linear and nonlinear feature extraction. The linear extraction methods include the following: PCA[7-8], ICA[9], LDA[10], FA[11]; The nonlinear extraction methods include the following: LLE[12], KPCA[13-15]and ISOMAP[16]. These feature extraction methods can extract the linear or non-linear feature information which exists in coal and gas outbursts influencing factors, then improve the prediction performance, but every methods have their own merits and demerits. For example, PCA belongs to unsupervised learning method, which can project high-dimensional spatial data into low-dimensional principal component space with the help of linear transformation, it can accurately keep the most expressive features of the original data. However, PCA cannot effectively use the difference information between different classes and obtain the most discriminative features, then cannot use the prior knowledge and extract the nonlinear component information, so the processing of nonlinear data has limitations; LDA can find those dimensions with recognition ability, during the process of linear dimension reduction, it can use prior knowledge and experience of class to extract local discriminant information from data, but it can't extract global information and non-linear feature information like PCA; FA belongs to linear dimension method, it can carry out factor analysis on the influencing factors with a high correlation, find a few factors that can integrate the original variables, so that the few factors can reflect most of the information of the original variables, reduce the information redundancy between the variable factors, improve the prediction performance of the model, but FA can only comprehensively evaluate the characteristic information of original data features; ICA can find components satisfying statistical independence and non-Gaussian. Corresponding high-order statistics analysis not only can analyze the non-Gaussian form of signal, but also can greatly remove the coupling between multi variables and meet the independence requirements, but it is also a linear dimensionality reduction method, which cannot extract the nonlinear components like PCA and LDA; KPCA is global information dimensionality reduction method combining kernel method and PCA, it can extract the global non-linear feature information in data and remove the correlation between data, which can make up for the deficiency that PCA is difficult to extract the non-linear feature information, but it cannot extract the existing local discriminant information and linear components; ISOMAP is a global nonlinear manifold learning algorithm, can reduce dimension, extract feature information and eliminate its correlation, but it is only limited to global feature extraction, ignoring local linear feature extraction; LLE manifold learning method requires that the sample set is dense and uniform, and only retains the local features of the training samples, for the discrete test samples, it cannot extract

more effective features for the prominent classification problem. At the same time, because coal and gas outbursts influencing factors have the characteristics of random and diversity, the relationships between influencing factors, outbursts and influencing factors are also very complex, the single feature extraction method cannot fully mine the internal laws of the outbursts influencing factors. The correlation information between them cannot be extracted, and the information components contained in the influencing factors cannot be integrated, which makes that the prediction effect of the classifier is not high. The literature shows that the sample data of coal and gas outbursts contains quite complex information, including global and local structure information, gaussian and non-gaussian components, which has significant non-linear and non-gaussian characteristics. In order to analyze the feature data better, we need to use the mixed feature fusion method to extract features in multiple feature spaces, any single subspace feature extraction algorithm has its limitations, and cannot be superior to other subspaces in any case. Therefore, this paper focuses on the problems which exist in the single feature extraction of coal and gas outbursts and make full use of the combination of local linear feature and global nonlinear feature extraction methods to make up for the shortcomings of single feature extraction method.

Classifier model selection for outbursts prediction is another issue addressed in this study, the common classifiers are BP Neural Network[17], Support Vector Machine(SVM)[18], Random Forestry(RF)[19], and other intelligent classifier technologies[20-21]. These methods play an important role in improving the prediction of coal and gas outbursts. Among them, PNN and RF all assume the indexes of coal and gas outbursts prediction are independent of each other, and there are some disadvantages such as complex calculation, long learning time and large sample space. ELM is a single hidden layer feedforward neural network, which is characterized by the connection weight matrix between input layer and hidden layer, and the offset of hidden layer nodes only need one-time random initialization, instead of updating by traditional back propagation algorithm. The only thing to be calculated is the weight matrix between hidden layer and output layer, which can be calculated by finding the square of generalized inverse matrix. Compared with other models, ELM has some obvious advantages. From the perspective of classification performance, ELM will achieve classification performance similar to or even better than SVM, and gain many successful experiences in small sample and high-dimensional nonlinear pattern recognition problems. ELM can approach any non-linear objective function with arbitrary precision by giving an excitation function satisfying certain conditions [21], which has strong nonlinear approximation ability and can be used to describe coal and gas outbursts prominent nonlinear evolution of space-time. However, due to the randomness, the parameters of ELM are not optimal, and there are problems in stability and reliability, so parameter optimization is needed. Many kinds of literatures have proposed some optimization methods such as Genetic Algorithm(GA), Ant Colony(AC) and Particle Swarm Algorithm(PSO)[22-27]. To some extent, these intelligent optimization algorithms can solve the problems that the optimal solution of parameters cannot obtain by experience setting, however these optimization algorithms are generally based on the mode of evolutionary algorithm. The disadvantage of the evolutionary algorithm is that the optimization algorithm is easy to trap at local minima, the parameters searched are not optimal, and the efficiency is not high, so the algorithm needs iteration continuously. QPSO is a global searching algorithm, which theoretically guarantees to find the best solution in the searching space. Compared with PSO, the iterative equation of QPSO [28] does not need the particle velocity vector, and it needs less parameters to be adjusted, which can be realized more easily, related experimental results also show that QPSO has better performance than standard PSO algorithm.

These models follow the non-linear characteristics of time and space evolution of coal and gas outbursts from different perspectives, and enhance the prediction ability of coal and gas outbursts. The mechanism of coal and gas outbursts is very complex, and there are many influencing factors, the non-linear relationship between the outbursts risk and influencing factors cannot be expressed by a clear mathematical model, thus increasing the difficulty of prediction. Therefore, it is necessary to establish an effective prediction model of coal and gas outbursts. In order to

improve the detection accuracy and stability, we propose a optimization classifier model based on DWT feature denoising and FICA-LDA feature extraction. The main contributions of this paper are as follows: The first stage is feature decomposition and reconstruction, we use DWT to smooth the coal and gas outburst features, and extract the trend components in the original features, improving the quality of trend term extraction to a certain extent. The second stage is dual space feature extraction, FICA is used to extract high-order information, the accounting information maps the high-dimensional features in the input space to the new low-dimensional feature space, and separates the independent features. Then, LDA is used to extract second-order statistical information, we can find the dimensions with identification ability through data dimensionality reduction, so that the original data can be reflected in these dimensions, and different classes can be distinguished as much as possible to get the optimal identification characteristics. The third stage is machine learning algorithm and parameter optimization, we use the DELM with high learning ability and generalization ability to predict outbursts, and uses the QPSO algorithm with good comprehensive performance to optimize the related parameters of the DELM. Finally, we conduct a comprehensive evaluation using the actual dataset of coal and gas outbursts, this includes a comparison with state-of-the-art feature selection methods in coal and gas outbursts. The remaining part of the article is organized as follows. Section2 offers the materials and methodologies used in the experiments and introduces our proposed approach. The dataset preparation and configuration environment of the proposed schemes are presented, then the experimental results are analyzed in Section3. Conclusions based on this research and future research targets are highlighted in Section4.

2. Theory and method

Coal and gas outbursts is a complex non-linear and random feature set, here we propose a mixed feature extraction and selection model based on DWT+FICA-LDA+QPSO-RELM. Firstly, we use DWT to decompose different features in coal and gas features, and reconstruct them by selecting threshold and threshold functions to form a new feature set of indicators; Second, FICA algorithm is used to extract non Gaussian features, forming independent features to produce a new feature set of space, then LDA is used to extract Gaussian features and local information, FICA-LDA feature extraction method can obtain the discrimination features; Finally, the obtained features are input to the DELM based on QPSO for classification, as shown in the Fig.1:

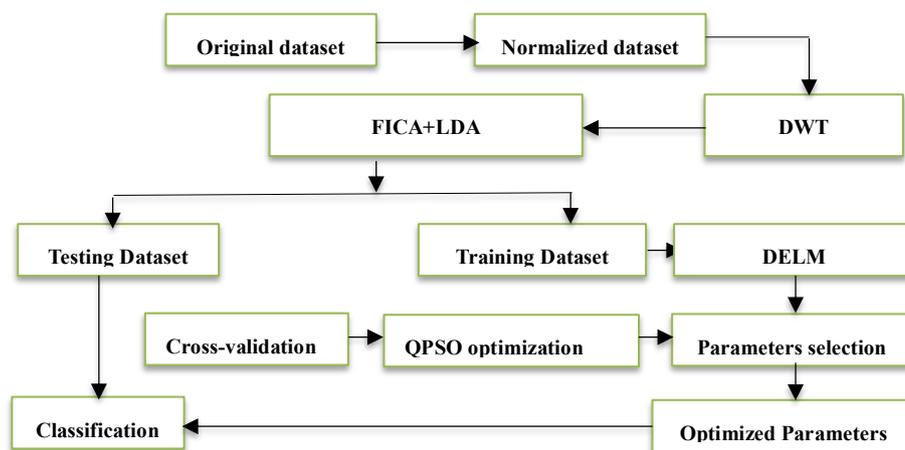


Fig.1. Flowchart of proposed comprehensive prediction model

2.1 Feature preprocessing

2.1.1 Discrete wavelet transform

Wavelet transform[29] is a signal analysis tool for time-frequency analysis, which has unique advantages in time-frequency localization. It can extract information hidden in the original feature data, analyze local

features and non-stationary in the data. Discrete wavelet can be defined as:

$$\mathbf{f}(\mathbf{b}, \mathbf{a}) = \frac{1}{\sqrt{\mathbf{a}}} \int_{-\infty}^{+\infty} \mathbf{f}(t) \varphi\left(\frac{t - \mathbf{b}}{\mathbf{a}}\right) dt \quad (1)$$

$$\mathbf{w}(\mathbf{m}, \mathbf{n}) = 2^{-\left(\frac{\mathbf{m}}{2}\right)} \sum_{t=0}^{T-1} \mathbf{f}(t) \varphi[(t - \mathbf{n} * 2^{\mathbf{m}} / 2^{\mathbf{m}})] \quad (2)$$

Where φ is the parent wavelet, t is the discrete data feature, T is the length of data, integer variables m and n are the scale and conversion factor of discrete wavelet, $\mathbf{a} = 2 * \mathbf{m}$, $\mathbf{b} = \mathbf{n} * 2^{\mathbf{m}}$, the important application in data analysis is data decomposition and reconstruction. $\mathbf{d}\mathbf{i}(\mathbf{i} = 1, 2, \dots, \mathbf{j})$ denotes detail component, $\mathbf{a}\mathbf{i}(\mathbf{i} = 1, 2, \dots, \mathbf{j})$ denotes the approximate component. The series of approximations represents the low frequency components, which contains the trend information of the original signal. The detail series represents the high frequency components, including the characteristic signals of the influencing factors related to the original signal. In this study, a fast discrete wavelet analysis algorithm called Mallat [36] is used, which is based on four filters: low and high pass decomposition filter, low and high pass reconstruction filter. The signal can be decomposed into an approximate value ($\mathbf{a}\mathbf{n}$) and several detail values ($\mathbf{d}\mathbf{n}$) through the multi-level decomposition process based on Mallat. In the decomposition step, the signal is divided into two high frequency and low frequency components, while the low frequency is divided into two parts: high frequency and low frequency. High frequency is called signal detail, low frequency is the approximate value of signal, and the final result is the combination of high frequency and low frequency components. The original signal can be reconstructed by the following wavelet transform:

$$\mathbf{f}(\mathbf{x}) = \int_0^{\infty} \int_{-\infty}^{+\infty} \frac{1}{\mathbf{a}^2} \mathbf{w}\mathbf{f}(\mathbf{b}, \mathbf{a}) \varphi_{(\mathbf{a}, \mathbf{b})}(\mathbf{t}) \mathbf{d}\mathbf{b} \mathbf{d}\mathbf{a} \quad (3)$$

$$\mathbf{s} = \mathbf{a}\mathbf{1} + \mathbf{d}\mathbf{1} = \mathbf{a}\mathbf{2} + \mathbf{d}\mathbf{1} + \mathbf{d}\mathbf{2} = \mathbf{a}\mathbf{3} + \mathbf{d}\mathbf{1} + \mathbf{d}\mathbf{2} + \mathbf{d}\mathbf{3} \quad (4)$$

In wavelet decomposition, a selection of proper wavelet function and the number of decomposition layers are very necessary, many types of wavelet functions such as Daubechies, Symmlet, Gaussian, Mexican hat, Morlet and Shannon wavelets can be used in wavelet decomposition. It is found that Daubechies wavelet function has the advantages of effective local ability in the field of time-frequency. Therefore, Daubechies wavelet function is used as the wavelet base in this paper to balance the wavelength and smoothness. The number of decomposition layers is determined as 3 according to the relevant literature and experience value, which can more effectively separate various feature components.

2.1.2 Selection of threshold and threshold function

The selection rule of threshold and the design of threshold function are the key factors that affect the denoising effect, the threshold function can be divided into two kinds: soft threshold and hard threshold. The hard threshold function has a break point, which simply keeps or removes the signal. While the soft threshold is used to distinguish the threshold value, it also uses the threshold value to attenuate the signal. In this paper, the soft threshold function is used, that is, when the absolute value of the wavelet coefficient is greater than or equal to the given threshold value, the value is the original value minus the threshold value, when it is less than the given threshold value, the value is zero. After the selected threshold function is selected, the threshold needs to be selected. The quality of threshold selection directly determines the performance of noise reduction model, that is, it is very important to select the threshold and carry out threshold quantification to noise reduction [30]. The types of threshold mainly include: fixed threshold, unbiased likelihood estimation threshold, heuristic threshold, minimax threshold, etc. in this paper, the fixed threshold is adopted. In a word, the combination of fixed threshold and soft threshold is used in the selection of wavelet analysis threshold and threshold function.

2.1.3 Wavelet denoising algorithm

DWT is used to decompose the original data, it can compare the signals of different frequency range, and divide

the signals with different characteristics into subspaces of different resolution scales, and then reconstruct the signals with wavelet coefficients. The reconstructed data not only removes the unstable factors, but also maintains good details. The noise of measured data always has high frequency signal. The essence of wavelet de-noising is to extract high frequency part from the signal. The details of wavelet transform process include wavelet decomposition, threshold and threshold function processing and wavelet reconstruction [31]:

- Wavelet decomposition. Select the appropriate wavelet function and the number of multi-scale decomposition layers, then analyze the wavelet transform of the original signal, on this basis, get the lowest approximation coefficient (low frequency) and the discrete detail coefficient (high frequency) of each layer.
- Threshold and threshold function processing. After wavelet processing the threshold method is used to deal with the high frequency coefficients of the discrete detail coefficients of each layer, so as to eliminate the noise.
- Wavelet reconstruction. After threshold processing the low-frequency wavelet coefficients of wavelet decomposition and the high-frequency wavelet coefficients are used to reconstruct, and the smoothed signals are obtained.

2.2 FICA-LDA hybrid feature extraction

2.2.1 ICA

ICA[32-34] model assumes that the observed mixed signal variable $y = (y_1, y_2, \dots)$. These variables are composed of n independent unknown source signals $s = (s_1, s_2, \dots, s_n)$ is a linear mixture, so there is a relationship between the mixed signal y and the source signal s , which can be expressed as $y = ws$ using vectors, and w as the unknown mixed coefficient matrix.

$$y = ws + n \quad (5)$$

Where n is the unknown unpredictable noise, which accords with the Gaussian distribution. ICA recovers the source signal and mixture matrix through y . The main idea is to remove the correlation between the data variables from the non-Gaussian signal of the original data, so that the components of each variable are statistically independent of each other. ICA has a better separation effect on the cross signal, and can more deeply mine the effective information which is covered due to the cross signal. Superior performance in signal separation, redundancy elimination and noise reduction to realize decomposition and fusion of deformation information. In this paper, formula (5) is analyzed by using FastICA algorithm which is more efficient than traditional ICA.

2.2.2 FICA

Because the FICA[35-37] algorithm takes the maximum negative entropy as a search direction, we first discuss the decision criterion of negative entropy, definition of negative entropy is as follows:

$$H(Y) = - \int p(y) \log p(y) dy \quad (6)$$

Where $H(Y)$ is a Gaussian random variable with the same variance, and is the differential entropy of the random variable

$$J(Y) = H(Y_{gauss}) - H(Y) \quad (7)$$

According to information theory, among the random variables with the same variance, the random variables with Gaussian distribution have the maximum differential entropy. When there is a Gaussian distribution, the stronger the non-Gaussian property of f ; is, the smaller the differential entropy is and the larger the value is, so it can

be used as a non-Gaussian measure of random variables. Since it is impractical to calculate the probability density distribution function needed to know for differential entropy according to equation (7), the following approximate formula is adopted:

$$\mathbf{J}(\mathbf{W}) \approx [\mathbf{E}\{\mathbf{G}(\mathbf{W}^T \mathbf{X})\} - \mathbf{E}\{\mathbf{G}(\mathbf{V})\}]^2 \quad (8)$$

Where $G(\cdot)$ is some non-quadratic function and V is a Gaussian variable of standard normal distribution. In this way, the problem of finding the maximum value of $J(W)$ is transformed into the problem of solving the maximum value of $\mathbf{E}\{\mathbf{G}(\mathbf{W}^T \mathbf{X})\}$, that is, when solving the value of mixed matrix W , each independent component has the strongest non Gaussian property and the best separation effect. The iteration formula (9) is used for calculation

$$\mathbf{w}(k+1) = \mathbf{E}\{\mathbf{X}g(\mathbf{w}(k)^T \mathbf{X})\} - \mathbf{E}\{\mathbf{X}g(\mathbf{w}(k)^T \mathbf{X})\}\mathbf{w}(k)^T \quad (9)$$

Where $g(\cdot)$ is the derivative of function $G(\cdot)$. At the same time, $w(K+1)$ was normalized by (10)

$$\mathbf{w}(k+1) = \mathbf{w}(k+1) / \|\mathbf{w}(k+1)\|_2 \quad (10)$$

Therefore, the specific iterative steps are as follows:

- Step 1: Select the initial value w ;
- Step 2: Calculate $w(k+1)$ according to (9);
- Step 3: Normalize $w(k+1)$ according to (10);
- Step 4: If the adjacent w change is less than the given value, the iteration stops; otherwise, it turns to step 2

2.2.3 LDA

LDA[38-39] is a supervised linear dimensionality reduction method based on Fisher's discriminant criteria, the key goal is to find a projection matrix G to maximize the Fisher criterion after sample projection. The optimal projection direction can be found by eigenvalue decomposition of the scattering matrix. The specific steps are as follows:

Step 1: Let x be a matrix of k class data $[x_1, x_2, \dots, x_k]$, each x_k includes n_i samples, then calculate the mean vector of all kinds of samples x_i' and the mean vector of all samples x' , which are expressed as follows:

$$x_i' = \frac{1}{n_i \sum_{i \in X_i} x_i} \quad (11)$$

$$x' = \frac{1}{n \sum_{i=1}^k \sum_{i \in X_i} x_i} \quad (12)$$

Step 2: Calculate the inter-class scatter matrix S_b and intra-class scatter matrix of the sample S_w , which are respectively expressed as follows:

$$S_w = \frac{1}{n \sum_{i=1}^k \sum_{i \in X_i} (x_i - x_i')(x_i - x_i')^T} \quad (13)$$

$$S_b = \frac{1}{n \sum_{i=1}^k \sum_{i \in X_i} (x_i' - x')(x_i' - x')^T} \quad (14)$$

Step 3: Find the projection matrix G and define fisher criterion function, in order to obtain the maximum value of G , it can be solved in the following ways:

$$G = G^T S_b G / G^T S_w G \quad (15)$$

$$S_w^{-1} S_b g_i = \lambda g_i, i = 1, 2, 3 \dots w, G = [g_1^T, g_2^T, g_3^T, \dots, g_w^T] \quad (16)$$

Where, g_i is the eigenvector corresponding to the eigenvalue λ of matrix $S_w^{-1} S_b$.

Step 4: The Lagrange multiplier method is used to solve all samples with the most optimal projection direction, then

can achieve the purpose of dimension reduction:

$$y = G^T X \quad (17)$$

2.2.4 FICA-LDA feature extraction algorithm

The FICA is to extract the high-order statistics information of non-Gaussian from original features, which can provide a means for many processing problems that cannot be solved by the second-order statistics method. From the non-Gaussian feature, we can find nonlinear expression which makes the components become statistically independent or as independent as possible. FICA can effectively extract non-Gaussian components of process data and reduce the complexity of detection model, but FICA is an unsupervised learning method, and it is unable to extract all kinds of correlation information between class label and various features, so it reduce its accuracy and sensitivity in feature extraction [40]. LDA is mainly to find an optimal projection matrix, so that the feature space can get a new feature space after the matrix mapping. In the new feature space, each category is easier to be recognized, and the dimension can also be reduced. LDA only depends on the second-order statistics information, and the main variables obtained are uncorrelated, and the data is Gaussian distribution. For random variables or processes that do not obey Gaussian distribution, the second-order statistics cannot fully represent their statistical characteristics [41-42].

In view of this situation, we propose a dual space feature extraction fusion algorithm based on the combination of FICA and LDA, the basic idea is that we use the FICA of high-order statistical information to extract the useful information of coal and gas outbursts, and separate the relevant and unrelated features in the new feature space, then use LDA supervised learning to extract the second-order statistics method. In this way, we can make full use of the complementarity of the two feature spaces, the difficult recognition patterns in one feature subspace are easy to recognize in the other feature subspace, FICA can extract complete useful information and separate useful features and noise features, LDA can reduce the dimension of features, improving the classification accuracy. In the two feature spaces, we combine the advantages of FICA and LDA method to mine all kinds of hidden information in the features of coal and gas outbursts to the greatest extent, which contribute to improve the prediction accuracy of classifier. The core of the algorithm is to combine the advantages of FICA and LDA dual space feature extraction algorithm, in view of the complexity and diversity of coal and gas outbursts, considering the characteristics of existing small outburst sample data, we design a dual space feature extraction algorithm combining FICA and LDA, and the implementation steps of the fusion method are as follows:

| |
|--|
| <p>Algorithm 2 FICA-LDA algorithm</p> <p>Input: original data set D, feature subset k1, k2.</p> <p>Output: feature set s.</p> <ol style="list-style-type: none">1. Normalization the original data, and produce the new data set D;2. Run the FICA algorithm on dataset d to centralize and whiten the dataset, the independent component signal k1 and noise component k2 are separated from the data set;4. Run LDA algorithm on dataset k15. Input data set to LDA training, computer matrix intra class scatter matrix and inter class scatter matrix.6. Get the best projection matrix and eigenvector, the dimension of training samples and test samples are reduced by mapping matrix, and the feature subset k2 of low dimension is obtained. |
|--|

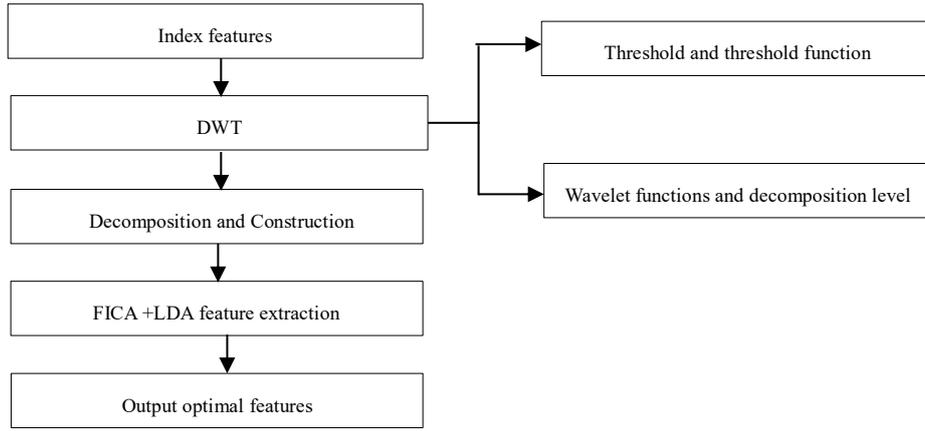


Fig.2 Flowchart of proposed DWT+FICA-LDA feature extraction model

2.3 DELM

2.3.1 ELM

ELM[43-45] is a single hidden layer feedforward neural network, this method is mainly composed of three layers: input layer, single hidden layer and output layer. With the advantages of parameter selection, fast learning speed and good generalization performance, the ELM is fast, simple, easy to realize and has strong generalization. In small samples, logical regression and neural network are obviously affected, while ELM function maintains good performance, which makes it have obvious advantages under the condition that samples are difficult to obtain. Suppose that there are N discrete training samples $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}$. Single layer feedforward neural network model can be expressed as follows:

$$\mathbf{y}_i = \sum_{i=1}^L \beta_i g(\omega_i * \mathbf{x}_i + \mathbf{b}_i) \quad (8)$$

Where L is the number of neurons in the hidden layer, G (x) is the activation function, ω_i is the weight value of the ith neuron in the input layer and the hidden layer, \mathbf{b}_i is the bias value of the ith neuron in the hidden layer, β_i is the weight value of the first neuron and the output layer in the hidden layer, y IUI is the actual output value. The matrix form can be expressed as: in $t = h$ β , the hidden layer output matrix H can be expressed as:

$$\square = \square \square \quad (9)$$

The output matrix H of the hidden layer can be expressed as follows:

$$\mathbf{H}(\omega_1, \dots, \omega_L, \mathbf{b}_1, \dots, \mathbf{b}_L, \mathbf{x}_1, \dots, \mathbf{x}_L) = \begin{bmatrix} g(\omega_1 * \mathbf{x}_1 + \mathbf{b}_1), \dots, g(\omega_L * \mathbf{x}_1 + \mathbf{b}_L) \\ \dots \dots \dots \\ g(\omega_1 * \mathbf{x}_N + \mathbf{b}_1), \dots, g(\omega_L * \mathbf{x}_N + \mathbf{b}_L) \end{bmatrix}_{N \times L} \quad (10)$$

The output weight matrix of the hidden layer can be expressed as: $\beta = [\beta_1, \beta_1, \beta_L]^T$, the output matrix of the fitting is as follows: $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_1, \mathbf{T}_N]^T$, its biggest feature is that the connection weight matrix between the input layer and the hidden layer, and the offset of the nodes in the hidden layer only need to be initialized randomly once, without updating through the traditional back propagation algorithm. The only calculated one is the weight matrix between the hidden layer and the output layer, which can be obtained by the method of finding the generalized inverse matrix. Algorithm flow is as follows:

Step 1: Given the training data set, randomly determine the hidden layer node parameters and input weights

and thresholds.

$$\beta = \{\beta_1, \beta_2, \beta_3, \dots, \beta_n\} \quad (11)$$

Step 2: Select the activation function to calculate the output matrix H of the hidden layer.

Step 3: H is the generalized inverse matrix of the hidden layer output matrix, the least square method is used to calculate the output layer weight B.

$$B = (H^T H + \lambda I)^{-1} H^T T \quad (12)$$

2.3.2 DELM

Although ELM has more advantages, it also has some problems. ELM is not suitable for constructing deeper network structure, it reduces the amount of computation and machine overhead. The weight β from the hidden layer node to the output layer node of ELM is solved by the generalized inverse matrix of the hidden layer output, and the robustness of the model is not strong. ELM only considers the empirical risk, not is the structural risk, so the data may be over-fitting, this problem can be solved by using the DELM. DELM[46-50] is a multiple-layer neural network, its learning procedure is highly efficient in learning time and has good generalization capabilities. Compared with ELM, DELM adds the restriction of regular term to prevent over-fitting. In this method, the regularized terms are constructed in the linear equations to solve the problem of insufficient computational stability and over-fitting, and the structural risk and empirical risk of ELM are balanced. DELM is mainly to obtain the output weight by evaluating the loss function with the weighted least squares.

$$\min_{\beta} L = \frac{1}{2} \|\beta\|^2 + \frac{c}{2} \|T - H\beta\|^2 \quad (13)$$

C is used to balance the empirical risk and structural risk of the learning machine. For the solution of the above-mentioned conditional extremum problem, the partial derivative is used after the Lagrange equation is converted to the unconditional extremum.

$$\beta - CH^T(T - H\beta) = 0 \quad (14)$$

Therefore, for the DELM, the output weight can be expressed as follows:

$$\beta = \begin{cases} \frac{1}{c} + H^T H^{-1} H^T H, N \gg n \\ H^T (\frac{1}{c} + H^T H)^{-1} T, N < n \end{cases} \quad (15)$$

The objective function is

$$y = \sum_{i=1}^k \beta_i g(\omega_i x + b_i) = \sum_{i=1}^k (H^T H + \frac{1}{c})^{-1} H^T T g(\omega_i x + b_i) \quad (16)$$

2.4 QPSO

2.4.1 PSO

Particle swarm optimization (PSO) [51] is a global optimization algorithm based on iteration, which is easy to implement and has no many parameters to adjust. In particle swarm optimization, a group of particles is composed of M particles in d-dimensional space. The position of the ith particle is $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iD})$, and the velocity is $V_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{iD})$. Each iteration of particles in particle swarm optimization algorithm updates its own velocity and position through individual extremum pbest and global extremum gbest. The updating

formula of velocity and position are as follows.

$$v_{id}^{j+1} = wv_{id}^j + c_1r_1(p_{id}^j - x_{id}^j) + c_2r_2(p_{gd}^j - x_{id}^j) \quad (17)$$

$$x_{id}^{j+1} = x_{id}^j + v_{id}^{j+1} \quad (18)$$

Where $\mathbf{i} = 1, 2, \dots, \mathbf{m}$; $\mathbf{d} = 1, 2, \dots, \mathbf{n}$, \mathbf{m} is the particle size; p_{id}^j is the D dimension of the individual extreme value pbest of the i th particle in the j iteration, p_{gd}^j is the D dimension of the global extreme value gbest of all particles in the j iteration, W is the inertia weight factor, c_1 and c_2 are the particle acceleration learning factors, R_1 and R_2 are the random numbers between 0 and 1. Inertia weight W is used to balance the global and local exploration ability of particle swarm, which is very important for particle swarm. A large inertia weight is conducive to the exploration ability but weak particle generalization ability. On the contrary, a small inertia weight is conducive to the fast generalization ability and lead to local optimization. In the current work, we propose a nonlinear decreasing inertia weight method to adjust the W value, taking further performance improvement as the equation:

$$W = W_{max} - \frac{(W_{max} - W_{min}) \times (t-1)^2}{(t_{max}-1)^2} \quad (19)$$

Where w_{max} and w_{min} are the maximum and minimum values of w respectively, t is the current number of iterations and t_{max} is the maximum number of iterations.

2.4.2 QPSO

The main defect of PSO is that the global generalization ability cannot be guaranteed, and there are too many parameters to be set, which is not conducive to find the optimal parameters of the model to be optimized. The change of particle position is lack of randomness and easy to fall into the trap of local optimum. In order to solve this problem, Sunetal proposed the quantum particle swarm optimization algorithm(QPSO)[52], is inspired by the particle swarm optimization algorithm trace analysis and quantum mechanics. The algorithm cancels the particle's moving direction attribute, and the update of particle's position has nothing to do with the particle's previous motion. In this way, the random particles of particle's position are added, and the particles are moved according to the following iterative formula,

$$mbest_j(t) = \frac{1}{m} pbest_{ij}(t) \quad (20)$$

Where pbest and ij are the best positions of all pbests in particle swarm, and mbest represents the average best position of the particle history of pbest. Where X_{ij} represents the position of the i th particle and M represents the size of the particle swarm.

$$p_{ij}(t) = \phi_{ij} pbest_{ij}(t) + (1 - \phi_{ij}) gbest_j(t) \quad (21)$$

$p_{ij}(t)$ represents the update of the i th particle position. U , ϕ_{ij} are the random value generated by the unified probability distribution in the range of $[0,1]$, α parameter is the contraction expansion factor, it is the only parameter controlling the particle generalization speed in the quantum particle swarm algorithm, and it is very sensitive to the population number and the maximum number of iterations.

$$X_{ij}(t+1) = p_{ij}(t) \pm \alpha |mbest_j(t) - x_{ij}(t)| * \ln(1/u) \quad (22)$$

Here we set the value range of α as (0.5,1) by searching, and the combination of particle swarm optimization and equation(21) is quantum particle swarm optimization.

| Algorithm1 QPSO algorithm |
|---|
| <p>The search space and fitness function f are defined, and the population number m and position x_{ij} are initialized and the dimension of particles</p> <p>For $t=1$-Maximum Iteration T do</p> <p style="padding-left: 2em;">Compute the mean best position $mbest$</p> <p style="padding-left: 2em;">$B=(1.0-0.5) \dots (T-t)/T+0.5$</p> <p style="padding-left: 2em;">For $i=1$---population size M do</p> <p style="padding-left: 4em;">If $f(X_i) < f(P_i)$ then</p> <p style="padding-left: 6em;">$P_i = X_i$</p> <p style="padding-left: 4em;">END IF</p> <p style="padding-left: 2em;">$P_g = \min(P_{p_i})$</p> <p style="padding-left: 2em;">For $j=1$-dimension D do</p> <p style="padding-left: 4em;">$\emptyset = \text{rand}(0,1); u = \text{rand}(0,1);$</p> <p style="padding-left: 4em;">$P_{ij} = \emptyset * P_{ij} + (1-\emptyset) * P_{gj}$</p> <p style="padding-left: 4em;">If $(\text{rand}(0,1)) > 0.5$ then</p> <p style="padding-left: 6em;">$X_{ij} = P_{ij} + B$</p> <p style="padding-left: 4em;">Else</p> <p style="padding-left: 6em;">$X_{ij} = P_{ij} - B * (mbest - X_{ij}) * \log(1/u)$</p> <p style="padding-left: 4em;">End if</p> <p style="padding-left: 2em;">End for</p> <p style="padding-left: 2em;">End for</p> <p>End for</p> |

2.5 QPSO-DELM

Coal and gas outbursts is a small sample data, the common logic regression and neural network are obviously affected, and the DELM function maintains good performance, which makes it have obvious advantages under the condition that the sample is difficult to obtain. Because DELM randomly generates input weights and hidden layer thresholds, the generalization ability of the model is poor and the prediction accuracy is not ideal. In this paper, QPSO is used to optimize the input weights and hidden layer thresholds of the DELM, which greatly improves the prediction accuracy and efficiency of the elm model. The algorithm is as follows :

| Algorithm3 QPSO-DELM algorithm |
|--------------------------------|
| |

Input: Search space N, Iteration number G, population M, fitness function.

Output: Regularize parameter C, input the optimal value of weight and hidden layer threshold.

Step1: Define related parameters: Search space N, Iteration number G, population M, fitness function.

Step2: Quantum code the cascade of input weights and hidden layer thresholds of DELM model to be optimized, and initialize pbest & gbest of particles, initialize the parameters of position and velocity.

Step3: for each particle to get the best solution mbest, N_{rad} is between [0,1], and it is randomly generated for each qubit. If $N_{rad} \leq |\alpha^2|$ then the qubit is 0, otherwise it is 1.

Step4: Calculate the fitness function of each particle.

Step5: Update pbest & gbest according to fitness function

$$P_g = \min(P_{p_i}) \quad P_{ij} = \emptyset * P_{ij} + (1 - \emptyset) * P_{gj}$$

Step6: If ($N_{rad} > 0.5$) then $X_{ij} = P_{ij} + (1.0 - 0.5) * \dots * (T - t) / T + 0.5$
Else $X_{ij} = P_{ij} - (1.0 - 0.5) * \dots * (T - t) / T + 0.5 * (mbest - X_{ij}) * \log(1/u)$

Step7: Determine whether the most conditional or the maximum number of iterations are met, output the optimal solution regularization parameters, input the weights and hidden layer thresholds, and return to step 3 for execution.

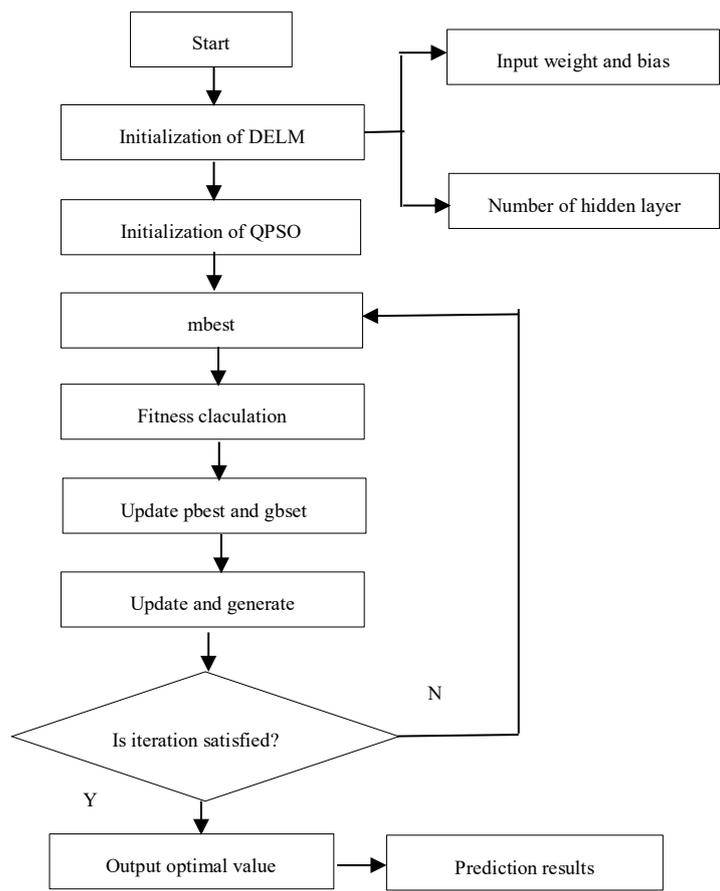


Fig.3 Flowchart of proposed QPSO-DELM classifier model

3. Experiment

3.1 Dataset description and preprocessing

The influencing factors of coal and gas outbursts include geological stress, gas and physical properties of coal seam. The experimental data come from the historical data of Pingdingshan No.8 mine in Henan Province. With

reference to previous studies[53] the coal and gas outbursts influencing indexes include the following: gas pressure(1),initial velocity of gas output(2), initial velocity of gas emission(3), coal seam firmness coefficient(4),structural coal thickness(5), fault structure complexity(6). These indexes are operability, extensive and applicable in engineering practice. Gas outbursts is a bipartition problem, namely 1 for the outburst type and 0 for the non-outburst type, as listed in column7. Table 1 gives a part of coal and gas outbursts sample data.

Table 1 Raw data for coal and gas outbursts

| 1 (MPa) | 2 (m ³ /t) | 3 (m ³ /t) | 4 (m) | 5 (m) | 6 (m) | 7 |
|----------|-----------------------|-----------------------|--------|-------|-------|---|
| 0.25 | 2.08 | 22.4 | 0.2291 | 0.22 | 0 | 0 |
| 0.16 | 2.02 | 18.9 | 0.1928 | 0.38 | 0 | 0 |
| 0.36 | 1.13 | 18.8 | 0.202 | 0.26 | 54 | 0 |
| 0.35 | 0.73 | 20.5 | 0.2222 | 0.54 | 254.3 | 0 |
| 0.2 | 4.03 | 15.8 | 0.25 | 0.36 | 90 | 1 |
| 0.43 | 0.69 | 15.8 | 0.4174 | 0.27 | 371.1 | 1 |
| 0.25 | 1.67 | 17.7 | 0.2696 | 0.48 | 371.1 | 1 |
| 0.3 | 1.41 | 17 | 0.2901 | 0.34 | 190.7 | 1 |
| 0.34 | 1.93 | 18.5 | 0.2381 | 0.27 | 190.7 | 1 |
| 0.25 | 2.26 | 16.6 | 0.2264 | 0.43 | 87 | 1 |

In order to make the experimental results more objective, the method of ten-fold cross-validation is used to verify the classification effect, that is, the data set is randomly divided into ten parts, one of which is taken as the testing set in turn, the other nine parts are taken as the training set, and then the corresponding classification method is run to classify and learn the data in the training set, and the testing set is used to test, and the corresponding 10 times testing results are obtained. In this paper, coal and gas outbursts prediction is a two-classification problem (outburst or normal),according to the data distribution characteristics of coal and gas outbursts, due to the use of different units in the acquisition, there are great differences in the order of magnitude, which affects the convergence speed and accuracy of the algorithm. Before training the model, this paper uses the normalization method to process these data into new data within the range of [0,1], and removes the dimension to make the indexes more comparable.

3.2 Experimental environment and parameter setting

The validity of the algorithm is further verified by comparing experiments on real datasets. All experiments are completed on a CPU with the main frequency of 2.8GhZ and a PC with 4GB memory. The operating system is windows7, and the algorithm is implemented in Python 3.7 environment.

True TP indicates the number of samples correctly marked for application, true negative TN indicates the number of samples correctly marked for application, false negative FN indicates the number of samples incorrectly marked for application, and false negative FP indicates the number of samples incorrectly marked for application. According to the above evaluation parameters, the overall accuracy, precision, sensitivity, specificity are obtained according to reference. In order to eliminate the random factors, we perform ten times with different random seeds on each experiment. The average performance of these ten repetitions is regarded as the final result. In order to ensure that the fair comparison between the methods, we use RF as a filter feature section classifier to select the best subset, the proposed computational method such as (SVM, KNN, RF, DT)might be sensitive to the values of its main controlling parameter, we use grid search optimization to tune all kinds of parameters. The wavelet basis function is DB4, the number of decomposition layers is 3, the parameters of the DELM and SVM are selected by QPSO and cross validation.

3.3 Experimental results and analysis

In this paper, the following experimental steps are used to verify the comprehensive performance of our proposed model: (1) Combing different feature extraction and feature selection methods, we use ELM and SVM as classifiers in different stages to compare the advantages of our proposed feature extraction method. (2) Comparing the design of the optimal classifiers, using different optimization algorithm, the comprehensive performance of the SVM, ELM and the DELM before and after the optimization are compared and analyzed to verify the advantages of the DELM proposed by us. (3) Experiment is used to verify the advantages of our comprehensive model compared with the prominent prediction models in the literature. In the following experiments, only the accuracy, feature dimension and execution time are compared, and other indicators are shown in the Table.

3.3.1 Performance comparison of FICA and DWT+FICA using SVM and ELM

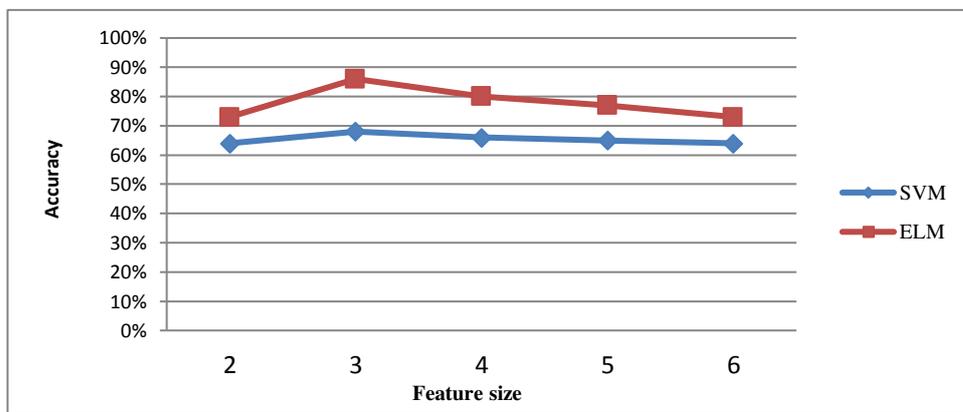


Fig.2 Comparison of FICA feature extraction method

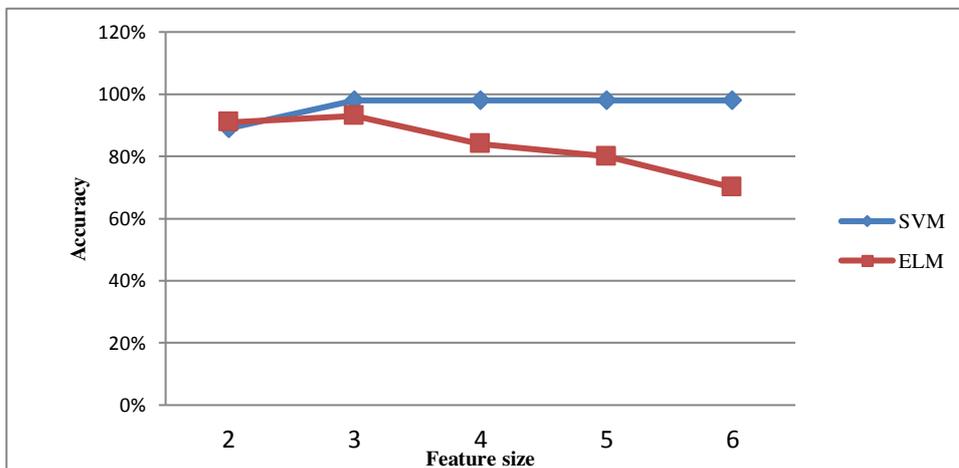


Fig.3 Comparison of DWT+FICA method

Table1 Accuracy comparison of different ICS using different classifiers

| | IC2 | IC3 | IC4 | IC5 | IC6 |
|---------------|------|------|------|------|------|
| FICA- SVM | 0.64 | 0.68 | 0.66 | 0.65 | 0.64 |
| FICA- ELM | 0.73 | 0.86 | 0.80 | 0.77 | 0.73 |
| DWT+FICA- SVM | 0.89 | 0.98 | 0.98 | 0.98 | 0.98 |
| DWT+FICA- ELM | 0.91 | 0.93 | 0.84 | 0.80 | 0.70 |

As can be seen from Fig.2-3 and Table 1, the accuracy of the four algorithms are shown with the increase of IC independent number, the discrimination accuracy of DWT+FICA with different classifiers are higher

than that of FICA. At the same time we also find out that the prediction accuracy of different feature extraction with different classifiers is highest when the independent component is IC3.

The results show that when the number of independent components is too low, it is difficult to completely separate the independent information carried in the feature, and the independent components decomposed cannot well express the sample information. When the number of independent components gradually rises, the independent information in the detection signal can be separated one by one to fully express the sample information it represents, which leads to the improvement of the final discrimination accuracy. When the number of independent components reaches a certain level, the prediction accuracy of the discriminant model is stable. The accuracy does not continue to increase with the increase of independent number. At the same time, a certain feature of the sample may be reflected in multiple features. In order to avoid the complexity brought by feature redundancy and interference to the stability of the model, it is necessary to eliminate the feature describing the same sample property in different dimension space and compress the data to a certain extent while mining the sample difference features. The characteristics of DWT+FICA can solve this series of problems very well, which is very helpful to improve the signal-to-noise, enhance the statistical independence and non-Gaussian of the feature, and separate the components as independent components with non-Gaussian as independence measure.

3.3.2 Performance comparison of DWT+FICA combining with different feature extraction methods

Table2 Comparison of different feature extraction methods using SVM

| | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DWT+FICA-LDA | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+KPCA-LDA | 0.98 | 0.92 | 0.91 | 0.91 | 0.93 | 0.92 |
| DWT+PCA-LDA | 0.97 | 0.95 | 0.97 | 0.97 | 0.97 | 0.91 |
| DWT+LLE-LDA | 0.96 | 0.96 | 0.94 | 0.96 | 0.96 | 0.91 |
| DWT+KPCA | 0.95 | 0.95 | 0.92 | 0.91 | 0.91 | 0.92 |
| DWT+LLE | 0.94 | 0.94 | 0.91 | 0.92 | 0.92 | 0.92 |
| DWT+PCA | 0.92 | 0.93 | 0.91 | 0.91 | 0.91 | 0.81 |
| DWT+LDA | 0.94 | 0.92 | 0.92 | 0.92 | 0.91 | 0.83 |

Table3 Comparison of different feature extraction methods using ELM

| | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| DWT+FICA-LDA | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+KPCA-LDA | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| DWT+PCA-LDA | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 |
| DWT+LLE-LDA | 0.98 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| DWT+KPCA | 0.95 | 0.96 | 0.95 | 0.90 | 0.95 | 0.96 |
| DWT+LLE | 0.93 | 0.96 | 0.95 | 0.90 | 0.95 | 0.96 |
| DWT+PCA | 0.93 | 0.96 | 0.95 | 0.90 | 0.92 | 0.83 |
| DWT+LDA | 0.94 | 0.96 | 0.95 | 0.90 | 0.92 | 0.83 |

In order to verify the accuracy and validity of the FICA-LDA feature extraction proposed in this paper, all kinds of feature information such as linear feature, Gaussian feature, non-Gaussian feature, nonlinear feature and any two features are extracted to form new feature vector respectively. SVM and ELM classifiers are used to classify, and the classification results are compared with the methods proposed in this paper. Here we mainly

compare the effects of several different feature combination. From Table2-3, it can be seen that after FICA, KPCA, PCA, LLE and LDA are combined on the two classifiers, the accuracy of algorithms are improved. Among them, FICA-LDA feature extraction is the best way to extract features on ELM, with the accuracy 0.99, which is higher than that of KPCA-LDA, PCA-LDA and LLE-LDA, and the execution time is the lowest. The indexes of SVM can be slightly lower than that of ELM, this is because after the original data is decomposed by FICA, the feature points of the original feature is compressed, and the redundant information in the feature is also removed, the relatively independent information in the original feature is mined by using the characteristics of high-order moments of the dataset, so that the feature information covered by the original cross feature can be effectively separated from each other, which can improve the modeling effect. The LDA removes the redundancy and obtains the characteristics of the optimal discrimination, so we can establish a discrimination model based on this kind of information through LDA on the basis of FICA, which has a better discrimination effect and improves the classification performance. KPCA, PCA and LLE are linear and nonlinear feature extraction methods respectively, and their effects are close to FICA.

The coal and gas outbursts influencing factors has the characteristics of non-Gaussian distribution and high nonlinearity. The traditional feature extraction methods only consider the information maximization while neglects the cluster structure information of the sample data, which leads to the incomplete feature extraction. The features are linear and each feature has non-Gaussian characteristics, which are the complex of Gaussian signal and non-Gaussian signal. FICA-LDA feature extraction is very suitable for the prediction of outbursts in this paper, it can effectively select feature information and ensure the integrity of the information to achieve better classification results. From the different indicators of the two classifiers, we can see that the effect of various indicators obtained by FICA-LDA feature extraction is not only higher than that of the single feature method, but also is the highest in various feature combinations. It shows that the combination of FICA-LDA feature extraction multiple features can extract more features and improve the classification accuracy.

3.3.3 Comparison of different kernels function using ELM and DELM

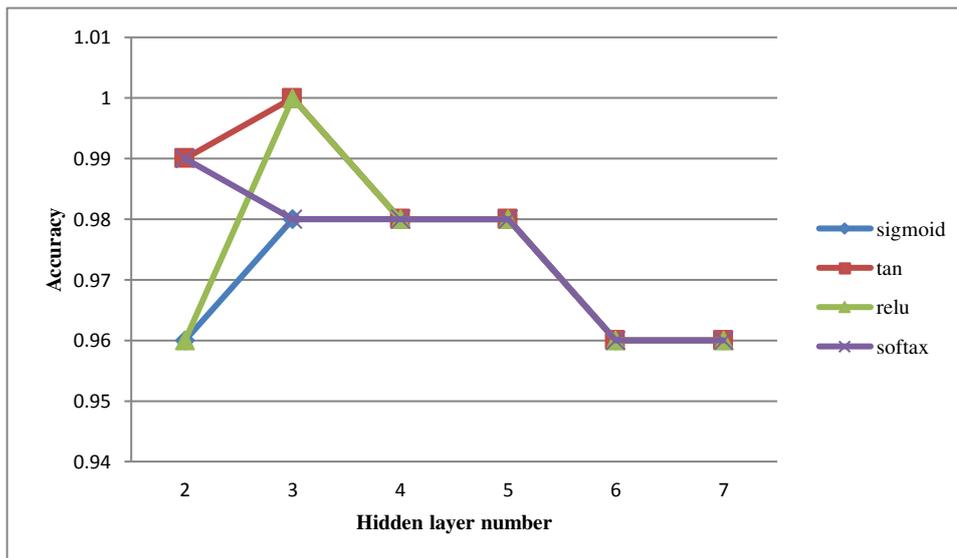


Fig.4 Comparison of different functions using ELM

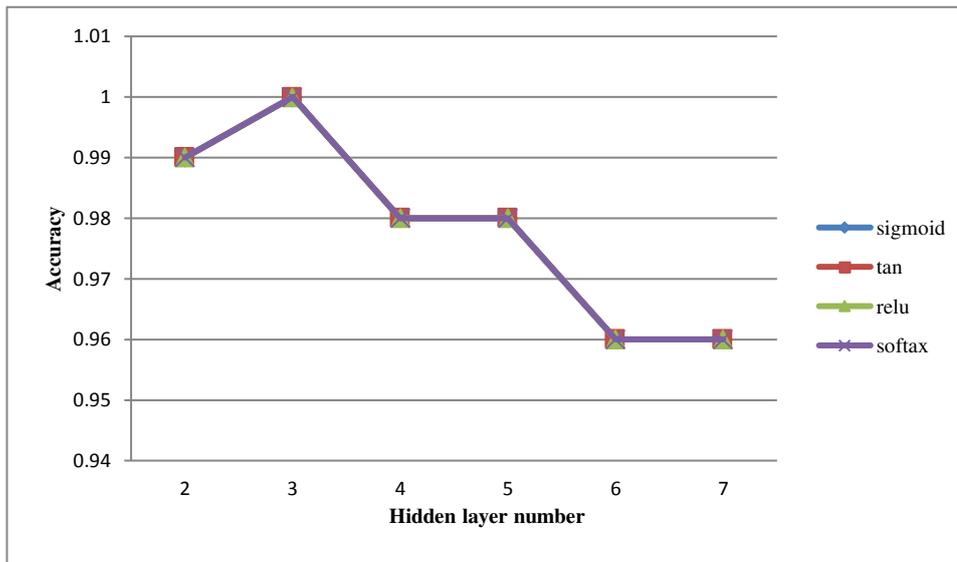


Fig.5 Comparison of different functions using DELM

Table4 Comparison of different functions using ELM

| Methods | Number | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS | TIME |
|---------|--------|----------|-----------|-------------|-------------|------|------|-------|
| Sigmoid | 3 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.006 |
| Tan | 3 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.006 |
| Relu | 3 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.008 |
| Softwax | 3 | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.006 |

Table5 Comparison of different functions using DELM

| Methods | Number | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS | TIME |
|---------|--------|----------|-----------|-------------|-------------|------|------|--------|
| Sigmoid | 3 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.002 |
| Tan | 3 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.0019 |
| Relu | 3 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.003 |
| Softwax | 3 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.003 |

From Table 4-5, we can see that the accuracy and execution time of different kernel functions of ELM and DELM algorithms with different hidden layers are compared. Among the four kernel functions, the accuracy of Tan and Relu kernel functions with ELM and DELM are the best, which are 0.98 and 0.99 respectively, and the execution time are 0.006 and 0.0019 seconds respectively. The overall performance is the best with DELM, for the model established by the algorithm, the training accuracy of Softwax and Sigmoid kernel functions are about 95% and 98% for outbursts datasets, which are lower than that of DELM model, and the execution time is long. Sigmoid kernel function obtain the worst effect with ELM classifier, only is 0.95 accuracy, and the time is 0.006s. The effect of kernel functions with DELM is better than that of ELM. From Fig. 4-5, we can also see that when the number of hidden layer nodes in ELM changes, the accuracy also changes. When the number of hidden layer nodes in ELM increases from 2 to 7, the accuracy increases slowly. When the number of hidden layer nodes is 3, the accuracy is the highest. When the number of hidden layer nodes is higher than 3, the accuracy begins significantly less.

Compared with the SVM and ELM learning models, DELM model realizes the function of adding regularization to prevent over-fitting, and it can still maintain the characteristics of rapid learning. The training classification accuracy of DELM model is higher than that of ELM and SVM, and the testing time is lower, which shows that

optimization can make the model classification result more stable. Compared with all kinds of kernel functions, the model with Tan function as excitation function has higher accuracy, shorter execution time and more stable. Therefore, we choose the DELM model as the classifier after considering the relationship between the execution time and the accuracy. The number of the best hidden layer nodes is set to 3, and the best kernel function is Tan kernel function. Under the premise of less training time, we can get higher training accuracy.

3.3.4 Performance comparison of different process of ours using different classifiers

Table6 Performance comparison of different process using SVM

| Method | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS |
|--------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| NONE | 0.88 | 0.88 | 0.92 | 0.83 | 0.92 | 0.83 |
| DWT | 0.92 | 0.96 | 0.95 | 0.90 | 0.95 | 0.87 |
| DWT+FICA | 0.98 | 0.96 | 0.95 | 1.0 | 0.95 | 0.96 |
| DWT+LDA | 0.98 | 0.96 | 0.90 | 1.0 | 0.97 | 0.89 |
| DWT+FICA-LDA | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+PSO | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+QPSO | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |

Table7 Performance comparison of different process using ELM

| Methods | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS |
|--------------------------|----------|-----------|-------------|-------------|----------|----------|
| NONE | 0.88 | 0.88 | 0.92 | 0.83 | 0.83 | 0.83 |
| DWT | 0.93 | 0.90 | 0.80 | 0.90 | 0.90 | 0.85 |
| DWT+FICA | 0.95 | 0.90 | 0.88 | 1.0 | 0.97 | 0.89 |
| DWT+LDA | 0.98 | 0.92 | 0.90 | 1.0 | 0.97 | 0.89 |
| DWT+FICA-LDA | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+PSO | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+QPSO | 1 | 1 | 1 | 1 | 1 | 1 |

Table8 Performance comparison of different process using DELM

| Methods | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS |
|--------------------------|----------|-----------|-------------|-------------|----------|----------|
| NONE | 0.88 | 0.88 | 0.92 | 0.83 | 0.83 | 0.83 |
| DWT | 0.93 | 0.90 | 0.80 | 0.90 | 0.90 | 0.85 |
| DWT+FICA | 0.99 | 0.90 | 0.88 | 1.0 | 0.97 | 0.89 |
| DWT+LDA | 0.99 | 0.92 | 0.90 | 1.0 | 0.97 | 0.89 |
| DWT+FICA-LDA | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+PSO | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| DWT+FICA-LDA+QPSO | 1 | 1 | 1 | 1 | 1 | 1 |

From Table 6-8, we can see the experimental comparison of various indicators on two different classifiers in different stages of ours. Only the classification accuracy and execution time are compared here, and other indicators are shown in the Table. For SVM classifier, the classification accuracy of DWT+FICA -LDA is 0.99, which is higher than that of DWT+FICA and DWT+LDA. DWT+FICA-LDA with ELM and DELM classifier all have the highest accuracy, this shows that the combination of DWT+FICA-LDA can produce good prediction accuracy. FICA only relies on high-order statistics information, which can effectively extract non-Gaussian components and global information of process data and reduce the complexity of detection model.

But FICA is an unsupervised learning method, it can't extract all kinds of correlation information between class label and feature, which reduces its accuracy and sensitivity in feature extraction. LDA is mainly to find an optimal projection matrix, so that the feature space can get a new feature space after this matrix mapping, and each class in the new feature space is easier to be identified, and the dimension can also be reduced. LDA only relies on the second-order statistics information, and obtains the local information. The main variables obtained are uncorrelated and cannot fully represent their statistical characteristics. In this way, FICA and LDA feature extraction methods are combined with each other, with the complementary advantages, and the feature information of coal and gas outbursts extraction are complete. From Table6-8 we can also see that the QPSO and PSO algorithm are compared with the classifiers of SVM, ELM and DELM. In terms of accuracy and other indicators, the QPSO algorithm can achieve 1 accuracy. Compared with PSO, the accuracy is improved and the running time is less, but the accuracy of SVM is no change in the accuracy. This shows that the combination of different classification algorithms and appropriate optimization algorithms is very important for the prediction effect.

3.3.5 Comparison with existing methods in literatures

Table9 Comparison with different algorithms in existing literatures

| Methods | Accuracy | Precision | Specificity | Sensitivity | AUC | MCSS | TIME |
|-------------------|----------|-----------|-------------|-------------|----------|----------|------------|
| Ours | 1 | 1 | 1 | 1 | 1 | 1 | 1.4 |
| KPCA+CS-ELM[11] | 0.84 | 0.90 | 0.80 | 0.90 | 0.90 | 0.85 | 2.0118 |
| KPCA+PSO-PNN[18] | 0.89 | 0.86 | 0.96 | 0.90 | 0.95 | 0.85 | 2.0679 |
| PCA+BP[3] | 0.67 | 0.67 | 0.80 | 0.85 | 0.67 | 0.80 | 2.35 |
| ISOMAP-WLSVM [10] | 0.87 | 0.89 | 0.88 | 0.89 | 0.88 | 0.89 | 2.55 |
| LLE+FOA-BP[12] | 0.70 | 0.70 | 0.83 | 0.88 | 0.70 | 0.83 | 2.35 |
| FA+RF[13] | 0.85 | 0.88 | 0.85 | 0.87 | 0.85 | 0.85 | 0.95 |

It can be seen from Table9 that compared with the models in the literatures, the method proposed in this paper has excellence results, the accuracy is 1. Compared with KPCA + CS-ELM, KPCA +PSO-PNN and LDA, this proposed method adopts DWT to perform the feature denoising, and FICA-LDA is used in feature extraction, then the strong nonlinear fitting ability of DELM is used to achieve better prediction results, with the accuracy increased by 0.16, 0.15 and 0.11 respectively; Compared with ISOMAP-WLSVM, PCA+SVM, FA+RF, the effect of proposed method in this paper is more effective, with the accuracy increased by 0.17, 0.15 respectively, it can achieve less consumption of calculation time and lower feature dimensionality on the premise of a high accuracy. Consequently, we can conclude that the proposed method has an advantage in combination of property in training time and testing accuracy than the other compared methods.

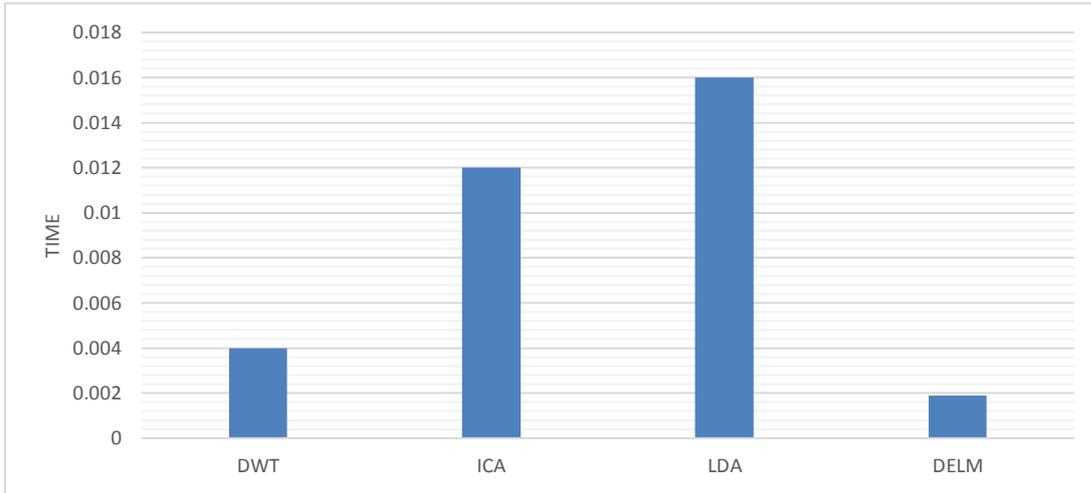


Fig.5 Average computation time at different stages of the proposed method

At the same time we also compare their performances in terms of computation complexity, which reveals their running time for real data. The running time is obtained in the following mode, and we run all the methods and employ RF as classifier to select the best subset. It can be seen from Table9 and Fig.5that the running time of ours is 1.40s, which is higher than other feature extraction methods and is lower than most feature selection methods. Moreover, it can predict outbursts according to the distribution characteristics of coal and gas outbursts sample data, which not only has better classification performance, but also has a low running time. The feature selection methods are between 0.95s-2.55s. Because these feature selection methods have the defect of high complexity, which make them more time-consuming. The advantages of our proposed methods are significant compared with most feature selection methods. Therefore, Comprehensive analyses demonstrate that the proposed feature selection method is effective in feature selection and classification performance for outbursts.

To sum up, our proposed method has higher prediction accuracy and average computation cost over the compared prediction methods in prediction of coal and gas outbursts. However, the proposed model has some defects. The proposed model has only been validated on one available dataset which come from pingdingshan no8 mine, but dataset of different mines can be tested in order to achieve better generalization performance. The current work deals with solving a two-class classification problem, however solving a multi-class outbursts classification problem is highly in demand. Further, ours requires more time cost to obtain the high accuracy, so a model with less computation time cost based optimization scheme can be investigated in future.

3.3.8 Statistical significance test

Table10 Comparison of traditional feature selection methods and ours using binomial test

| Methods compared | E-values | Significant (Y/N) |
|-------------------|----------|-------------------|
| KPCA+CS-ELM[13] | 1.53e-05 | Y |
| KPCA+PSO-PNN[15] | 9.33e-09 | Y |
| PCA+BP[7] | 3.12e-06 | Y |
| ISOMAP-WLSVM [11] | 1.53e-05 | Y |
| LLE+FOA+BP[12] | 1.93e-05 | Y |
| FA+RF[16] | 6.57e-05 | Y |

In order to evaluate the effectiveness of our proposed method, the statistical analysis of the proposed method is compared with other feature selection models mentioned in the literatures. For some filter feature selection methods

such as MIFS, MRMR, MIFS-ND, FCBF, GRA, FAT, AHP and EW, we use RF to find out the optimal features, at the same KNN is used as a classifier to prediction outbursts. In this paper, a statistical significance test called a binomial test is conducted between ours and other methods. A brief description of the binomial test is given below according to reference[56]. We compare the accuracy of the proposed with the accuracy of others in literatures. The results are achieved and the confidence interval is set to 0.95. Table 10 shows the E-value of independent sample binomial test. If E-values is a very small number, then the proposed algorithm is better than the other methods with high confidence, it means that there is a significant difference between the two methods. From the result, it shows the proposed method is better than the other methods with high confidence for this problem. The values less than 0.05, it proves that the performance of ours are better than others. According to the results of Table10, there are not values more than 0.05, which shows the performance of ours have much different with other methods, the performance of ours can obtain the highest accuracy compared with other methods in literatures.

4 Conclusion

According to the redundancy and uncorrelated characters in coal and gas outbursts, this paper introduces the DWT into the feature preprocessing of coal and gas outbursts, and reasonably eliminates the random fluctuation and noise feature information, which contributes to improve the accuracy of coal and gas outbursts.

Due to the complexity of coal and gas outbursts, the components obtained by DWT has redundant and uncorrelated information, which can't get better noise reduction effect, the coal and gas outbursts have non-linear and non-Gaussian, on the basis of the above, this paper deeply analyzes the combination of FICA and LDA algorithm, which can capture the essential information contained in it as much as possible, and uses the non-Gaussian to separate the components as independent components. In view of the incompleteness of information extracted by FICA, we input the features and class labels separated by FICA into the LDA supervision layer, projects the features and makes the features have class differentiation, carry out feature fusion and dimensionality reduction, extract Gaussian feature components, and eliminate the redundancy and correlation between features, then new features are formed; In view of the randomness of the parameter setting of DELM algorithm, the QPSO algorithm is applied into the optimization of DELM parameters, and the coal and gas outbursts prediction model based on QPSO-DELM is proposed, then the prediction accuracy of the model is significantly improved. Through example verification and comparative analysis, it is proved that the model has high identification, strong generalization ability, and high prediction accuracy, and it can be applied in the field of coal and gas outbursts prediction.

In the future, we will verify the proposed methods on different data sets of coal and gas outbursts, improve the generalization ability and adaptability of the model, provide new ideas and effective technical means for coal and gas outbursts prediction.

Availability of data and material

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

Declaration of Competing Interest

The authors declare no competing financial interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (U1704242).

Reference

- [1] Chen K P. A new mechanistic model for prediction of instantaneous coal outbursts Dedicated to the memory of Prof. Daniel D. Joseph[J]. *International Journal of Coal Geology*, 2011, 87(2):72-79.
- [2] Li H Y. Major and minor structural features of a bedding shear zone along a coal seam and related gas outburst, Pingdingshan coalfield, Northern China[J]. *International Journal of Coal Geology*, 2001,47(2):101-113.
- [3] NAN Cunquan,FENG Xiating. Coal and gas outburst area prediction using SVM. *Chinese Journal of Rock Mechanics and Engineering*, 2005,24(2):263-267.
- [4] Wang W, Ding J. Wavelet network model and its application to the prediction of hydrology. *Nat Sci*, 2003, 1:67-71.
- [5] Z. Tan, J. Zhang, J. Wang, and J. Xu, Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models, *Applied Energy*,2010,87(11):3606–3610.
- [6] Daubechiesi, *Ten lectures on wavelets*,1992,61. SIAM, Philadelphia.
- [7] Zhu, Zj., Zhang HW., Han J., Song, WH. Prediction of Coal and Gas Outburst Based on PCA-BP Neural Network. *China Safety Science Journal*,2013, 23:45-51.
- [8] Fu, H., Wang, X., Wang, Z. Research on the Soft Sensor of Coal and Gas Outburst Based on PCA and PSO-ELM. *Chinese Journal of Sensors and Actuators*,2014,27(12):1710-1715.
- [9] Li, Dafeng. Research on prediction of coal and gas outburst based on ICA-SVM. *Industry and Mine Automation*,2009,10:36-39.
- [10] Gao,W. The Application of Fisher Discriminant in Predicting Dangerous Level of Coal and Gas Outburst. *China Safety Science Journal*,2010,20(10):26-31.
- [11] Wen, T., Zhang, B., Shao, L. Prediction of coal and gas outburst based on random forest model. *Computer Engineering and Applications*,2014, 50(10):233-238.
- [12] Long NZ., Yi, AM., A XJ., Zhang, XL. Prediction of coal and gas outburst intensity based on LLE-FOA-BP model. *Industry and Mine Automation*,2019,45(10):68-74.
- [13] Wang, Y., Wang., Fu, H., Zhang, Y. The identification Model of Coal and Gas Outburst Intensity Based on KPCA and CIPSO-PNN. *Chinese Journal of Sensors and Actuators*,2015, 28(2):272-279.
- [14] Li, S., Hu, H. Risk identification of coal and gas outburst based on KPCA and improved extreme learning machine model. *Application Research of Computers*,2017, 35(1):172-177.
- [15] Kang, F., Xu, W., Zhang, S. A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing*,2014,18:178-184.
- [16] Fu, H., Feng, S., Gao, Z., Yang Y. Study on double coupling algorithm Based model for coal and gas outburst prediction. *China Safety Science Journal*,2018,28(3):84-90.
- [17] Fu, H., Feng, S., Gao, Z., Yang Y., Study on double coupling algorithm based model for coal and gas outburst prediction. *China Safety Science Journal*,2018, 28(3):84-90.
- [18] Wen, T., Zhang, B., Shao, L., Prediction of coal and gas outburst based on random forest model. *Computer Engineering and Application*,2014, 50(10):233-238.
- [19] Fu, H., Wang, X., Wang, Z., Research on the Soft Sensor of Coal and Gas Outburst Based on PCA and PSO-ELM. *Chinese Journal of Sensors and Actuators*,2014, 27(12):1710-1715.
- [20] Liang, B.,Qin, B.,Sun, W.,Wang, S.,Shi, Y., The application of intelligent weighting grey target decision model in the assessment of coal-gas outburst. *Journal of China Coal Society*,2013, 38(7):1611-1615.
- [21] Zhang, Y.,Cui, J.,Jiao,X., Study of the multi-index coupling forecasting model of coal and gas outburst and its application. *Chinese Journal of Engineering*,2018, 40(11):1309-1316.
- [22] Kuang, F., Zhang, S., Jin, Z., Xu, W.A. A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection, *Soft Computing*,2005, 19:1187–1199.

- [23] Kuang, F., Jin, Z., Xu, W. A novel chaotic artificial bee colony algorithm based on Tent map. *IEEE transactions on Evolutionary Computation*.2014.
- [24] Yan, X., Fu, H., Chen, W. Prediction of coal mine gas emission based on Markov chain improving IGA-BP mode. *Computer Modeling & New Technologies*,2014, 10:2-24.
- [25] Jaouher, BenAli., Lotfi, Saidi., Aymen, Mouelhi., Brigitte, Chebel-Morello.,Farhat, Fnaiecha.,(2015)Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. *Engineering Applications of Artificial Intelligence*42:67-81.
- [26] An J., You Z., Chen X. Robust and accurate prediction of protein self-interactions from amino acids sequence using evolutionary information. *Molecular Biosystems*,2016,12(12):3702-3710.
- [27] Hoang, ND., Bui,DT. A Novel Relevance Vector Machine Classifier with Cuckoo Search Optimization for Spatial Prediction of Landslides. *Journal of Computing in Civil Engineering*,2016,30(5):1943-5487.
- [28] J.Sun,W Xu,B.Feng.A global search strategy of quantum-behaved particle swarm optimization,in:2004IEEEConference on Cybernetics and Intelligent Systems,vol1,IEEE:2004,pp.111-116
- [29] Santoso S, Grady WM, Powers EJ, Lamoure J, Bhatt SC. Characterization of distribution power quality events with Fourier and Wavelet transforms. *IEEE Trans Power Deliv* 2000;15(1):245–7.
- [30] Gaouda M, Salama MMA, Sultan MR, Chikhani AY. Power quality detection and classification using wavelet multiresolution signal decomposition. *IEEE Trans Power Deliv* 1999;14(4):1469–76.
- [31] Oleskovicz Mário, Coury Denis V, Felho Odilon Delmont, Usida Wesley F,Adriano AFM, Carneiro, et al. Power quality analysis applying a hybrid methodology with wavelet transforms and neural networks. *Int J Electr Power Energy Syst*, 2009;31(5):206–12.
- [32] HYVARINEN A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*,1999, 10(3): 626 – 634.
- [33] Subasi A, Gursoy M I. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications*, 2010,37 (12) : 8659-8666
- [34] C.J. Lu. Integrating independent component analysis-based denoising scheme with neural network for stock price prediction, *Expert Syst. Appl.* 37(10)1016-1023.
- [35] Lu C J, Lee T S, Chiu C C. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 2009, 47(2):115–126.
- [36] Kao, L.-J., et al. Integration of Nonlinear Independent Component Analysis and Support Vector Regression for Stock Price Forecasting. *Neurocomputing*,2013, 99, 534-542
- [37] BELL B A, SEJNOWSKI T. An information maximization approach to blind source separation and blind deconvolution. *Neural Computation*, 1995, 7(6): 1129 -1159.
- [38] Zhong Wei, Huang Yuanliang, Hao Zhenzhen, et al. Research on implementation method of LDA algorithm based on Particle Swarm Optimization. *Computer Engineering and Applications*,2017,53(1):39-43.
- [39] Liu, L.,Yang, H., Xiao,G. Time-frequency domain feature extraction algorithm based on linear discriminant analysis. *Systems Engineering and Electronic*,2019, 2(7):14-19.
- [40] Ding Sijing, Sun Zhong,Yang Sun, Yanku. Combining principal component analysis network with linear discriminantanalysis for the classification of retinal optical coherence tomography images [J];*Journal of Image and Graphic*, 2019,24(01):0115-0123.
- [41] TANG Yongbo. Transformer Fault Diagnosis Model Based on PCA and KICA Feature Extraction [J], *High Voltage Engineering*,2014,40(2): 557-563.
- [42] WANG Weidong. Dual-space Feature Extraction Algorithm Based on PCA and ICA [J]. *Journal of Image and Graphics*, 2008,13(11):2163-2169
- [43] Huang G B, Wang D H, Lan Y. Extreme learning machines: a survey. *International Journal of Machine Learning*

- and Cybernetics, 2011, 2(2): 107-122
- [44] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and applications. *Neurocomputing*, 2006, 70(13):489–501
- [45] DING SF, XUX Z, NIE R. Extreme Learning Machine and Its Applications. *Neural Computing and Applications*, 2014,25(34): 549-556.
- [46] UZAIR M, SHAFAIT F, GHANEM B, et al. Representation learning with deep extreme learning machines for efficient image set classification. *MACINTYRE J. Neural Computing and Applications*. London: Springer, 2015:1-13.
- [47] XU SuHui, MU XiaoDong, CHAI Dong, LUO Chang. Domain Adaption Algorithm with ELM Parameter Transfer, *Acta automatic sinica*, 2018, 44(2): 311-317.
- [48] Huang G B, Zhou H M, Ding X J, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, 42(2):345-367
- [49] XU JiaMing, ZHANG WeiQiang. Manifold Regularized Extreme Learning Machine for Language Recognition. *Acta automatic sinica*, 2015,41(9):234-250.
- [50] SHAN Dong, XU Xinzheng. Multi-label Learning Model Based on Multi-label radial Basis Function Neural Network and Regularized Extreme Learning Machine. *PR&AI*,2017,30(9):833-840.
- [51] Peng Liying, Guosheng Fan. Hybridizing DEMD and Quantum PSO with SVR in Electric Load Forecasting. *Energies*,2016,22(1):1-20
- [52] Sudheer Ch, Nitin Anand, BK, Panigrahi, Shashi Mathur, Streamflow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputin*,2013,101:18-23.
- [53] Zhang, Y.,Cui, J.,Jiao,X., Study of the multi-index coupling forecasting model of coal and gas outburst and its application. *Chinese Journal of Engineering*,2018, 40(11):1309-1316.

Figures

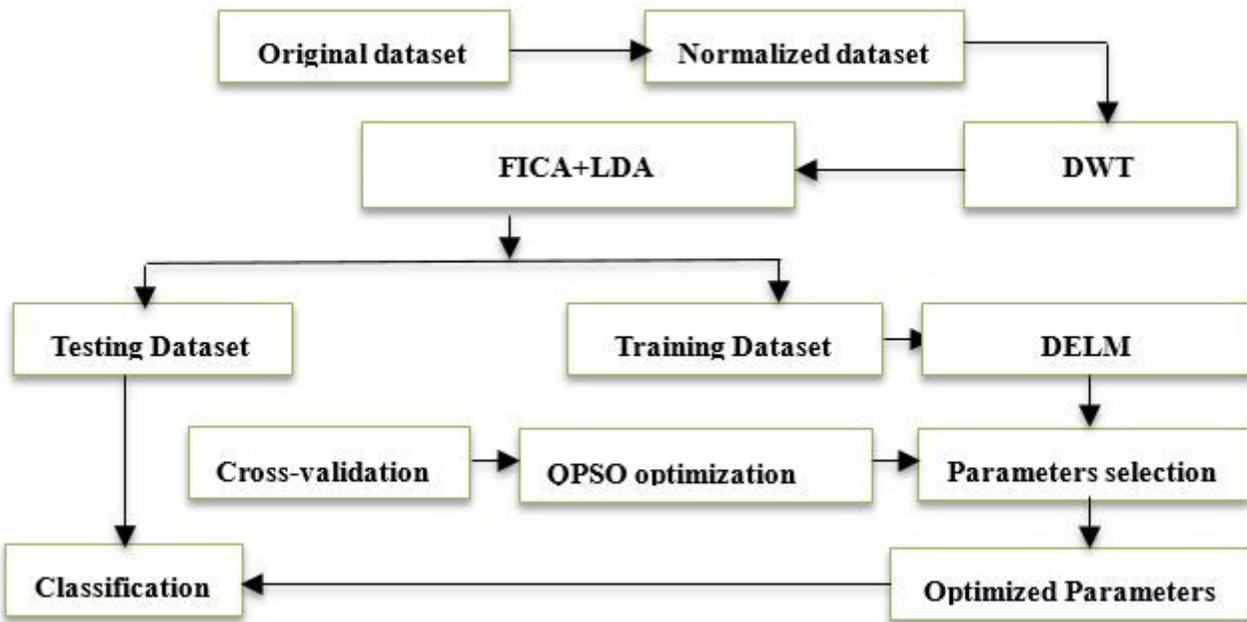


Figure 1

Flowchart of proposed comprehensive prediction model

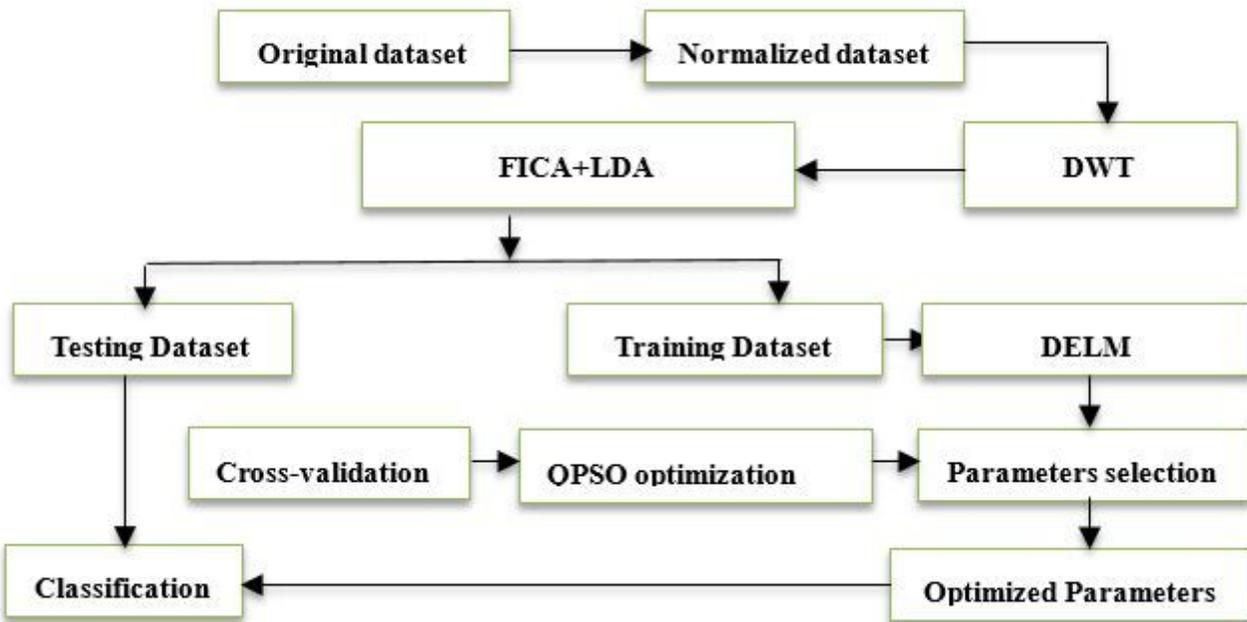


Figure 1

Flowchart of proposed comprehensive prediction model

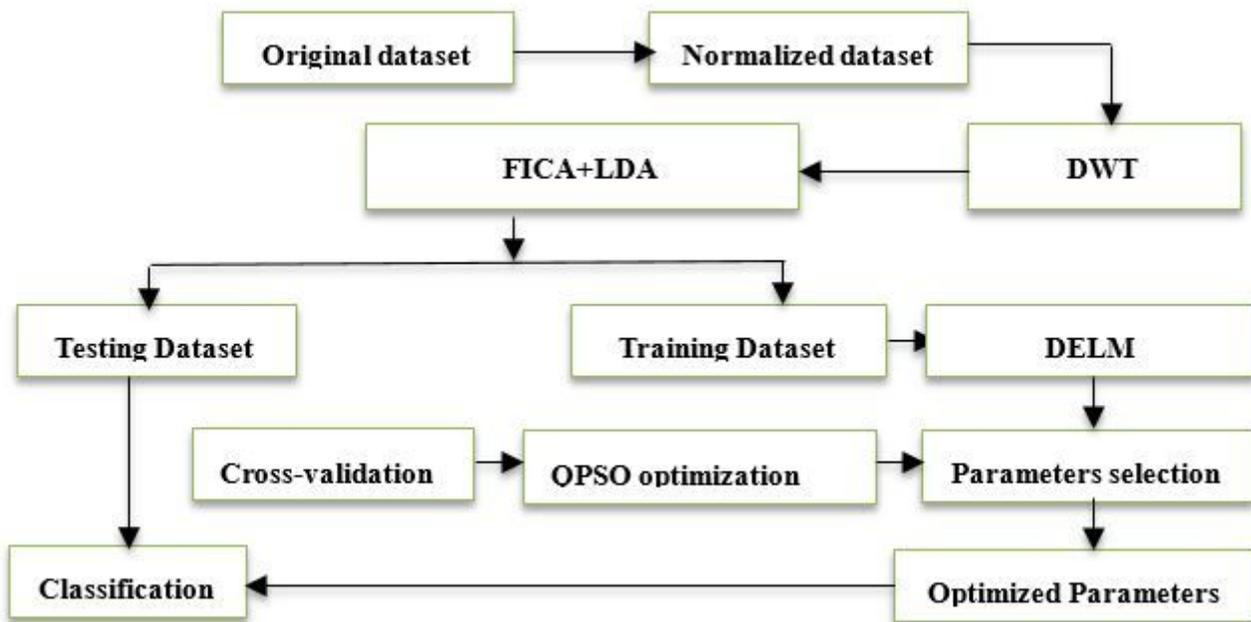


Figure 1

Flowchart of proposed comprehensive prediction model

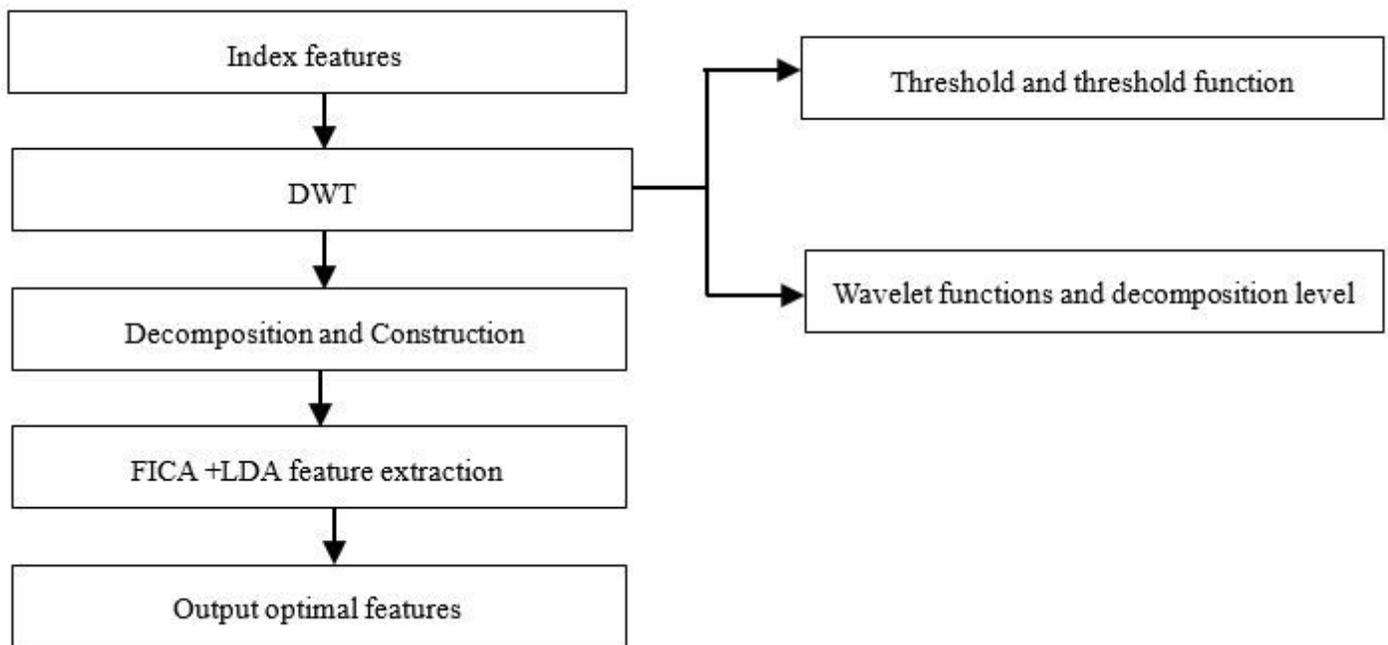


Figure 2

Flowchart of proposed DWT+FICA-LDA feature extraction model

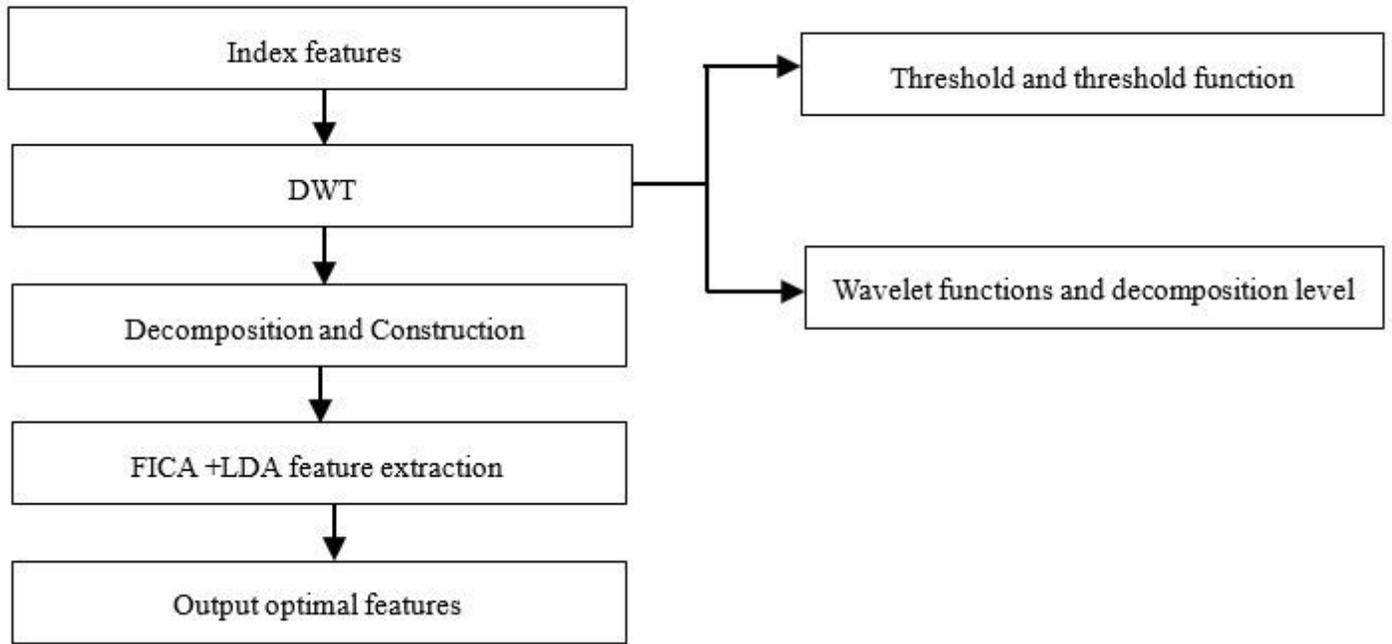


Figure 2

Flowchart of proposed DWT+FICA-LDA feature extraction model

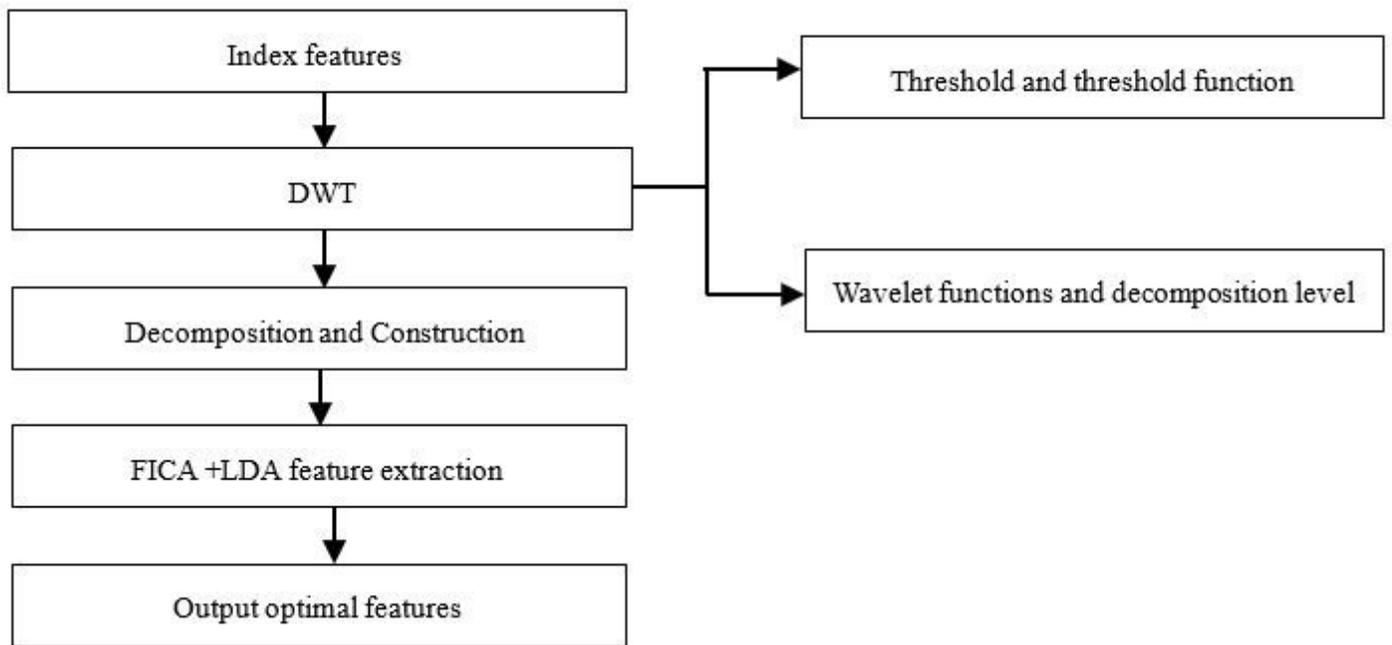


Figure 2

Flowchart of proposed DWT+FICA-LDA feature extraction model

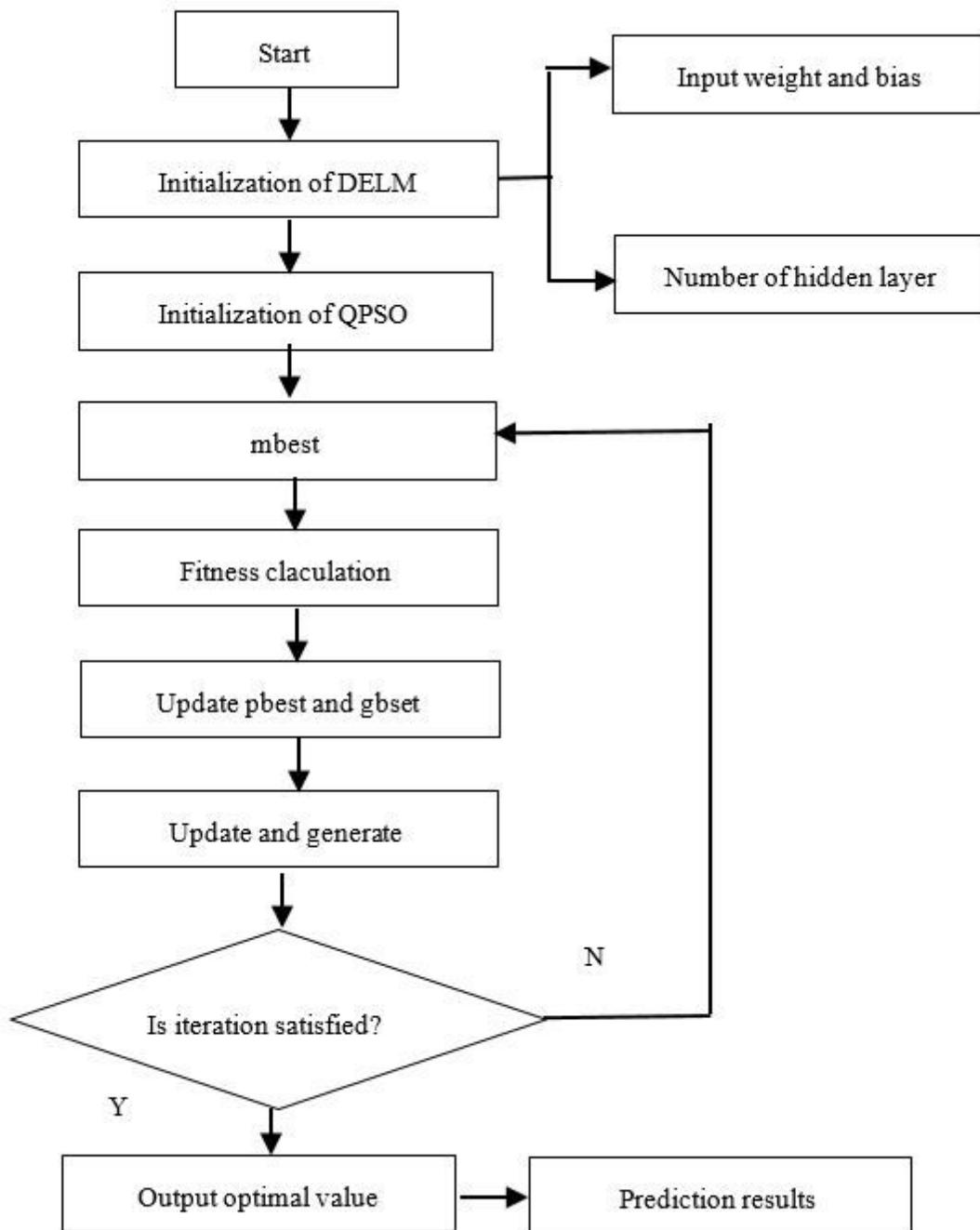


Figure 3

Flowchart of proposed QPSO-DELM classifier model

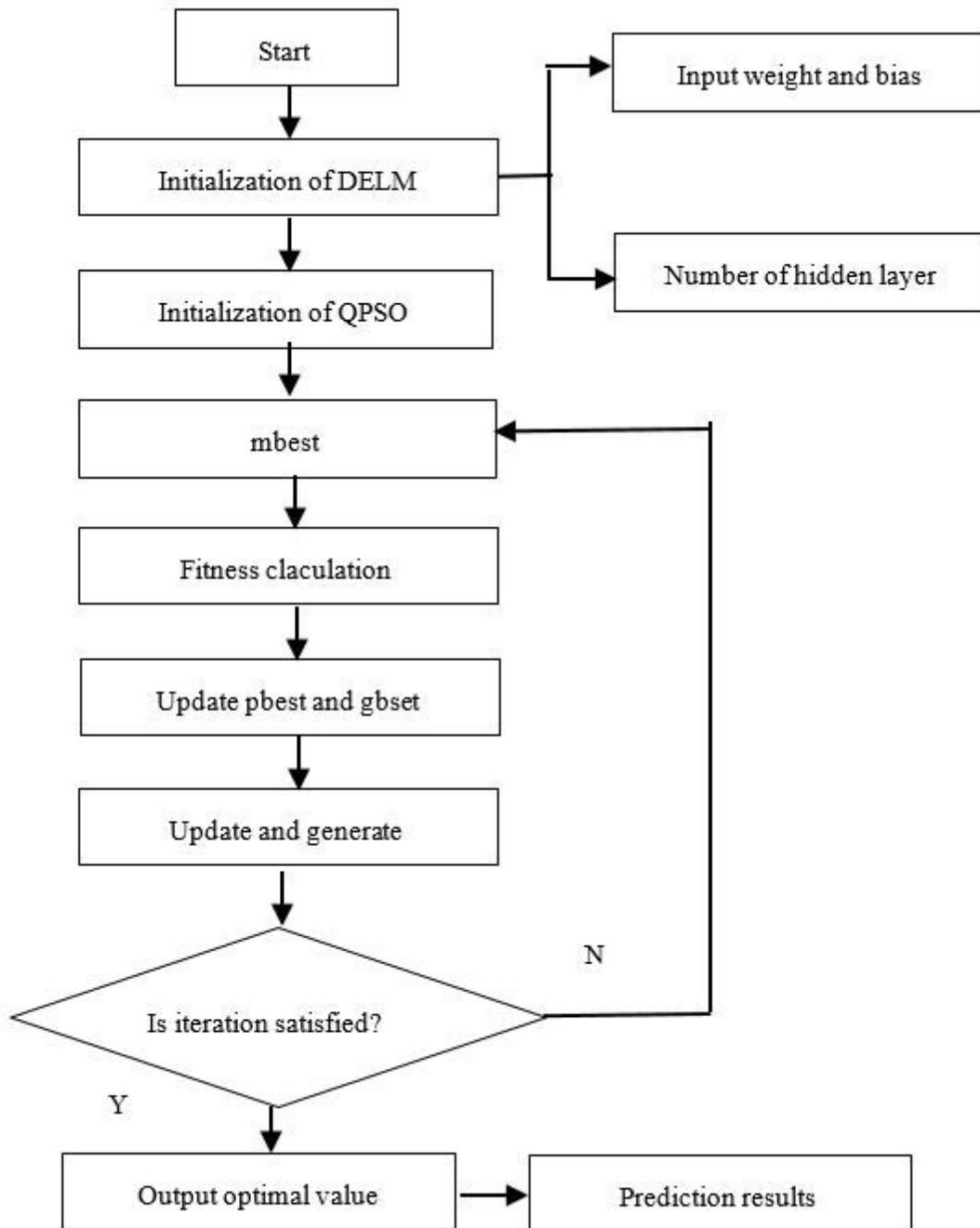


Figure 3

Flowchart of proposed QPSO-DELM classifier model

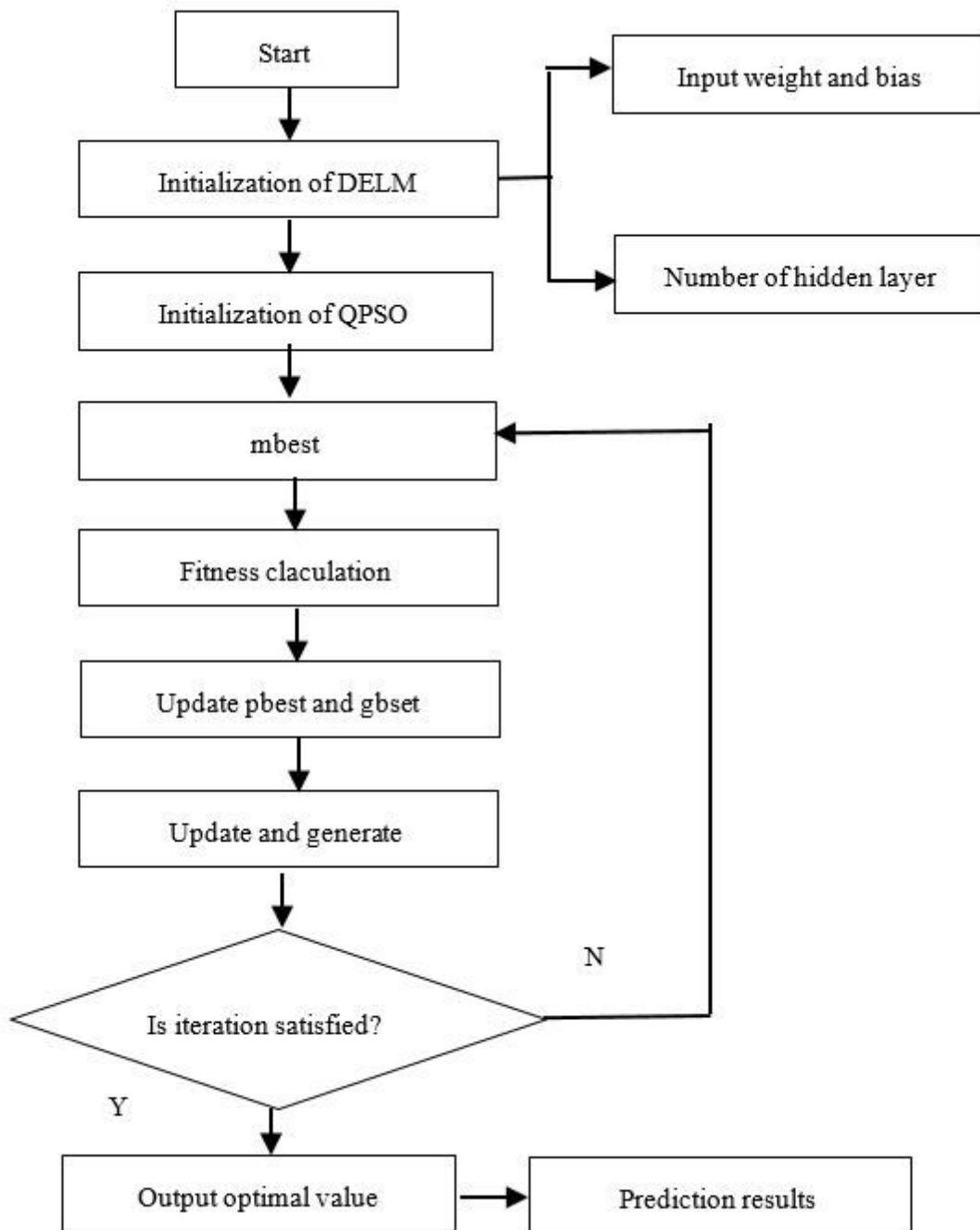


Figure 3

Flowchart of proposed QPSO-DELM classifier model

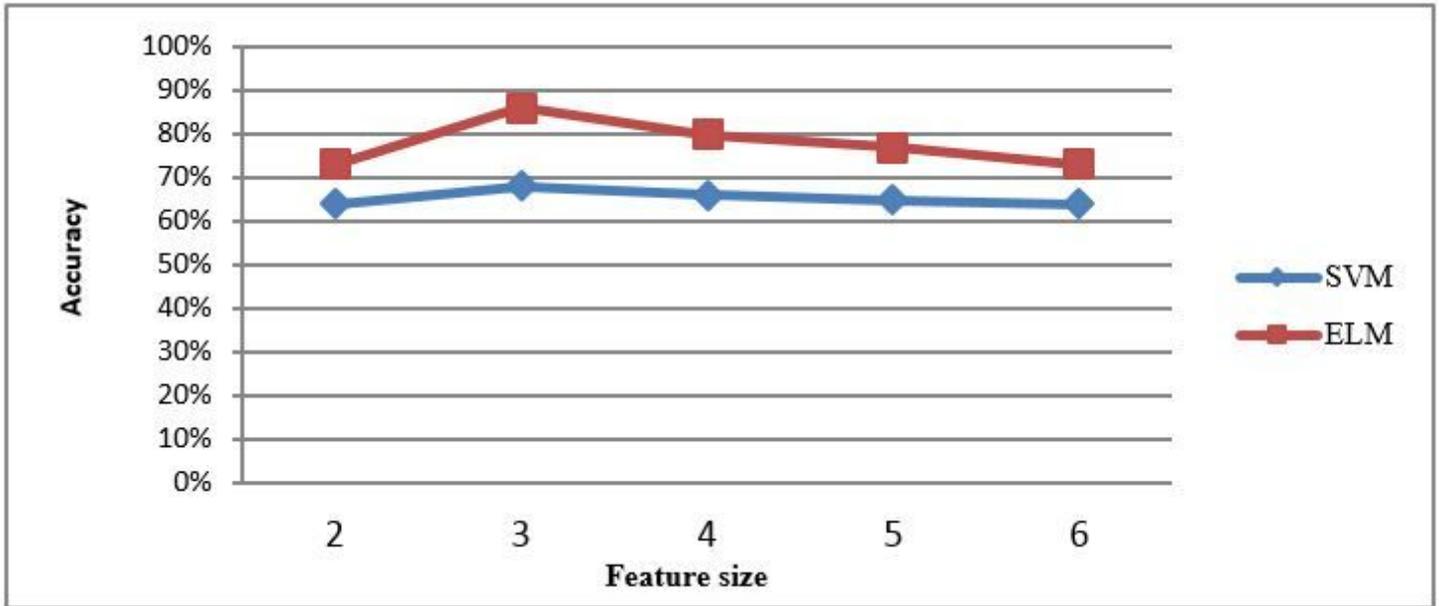


Figure 4

Comparison of FICA feature extraction method

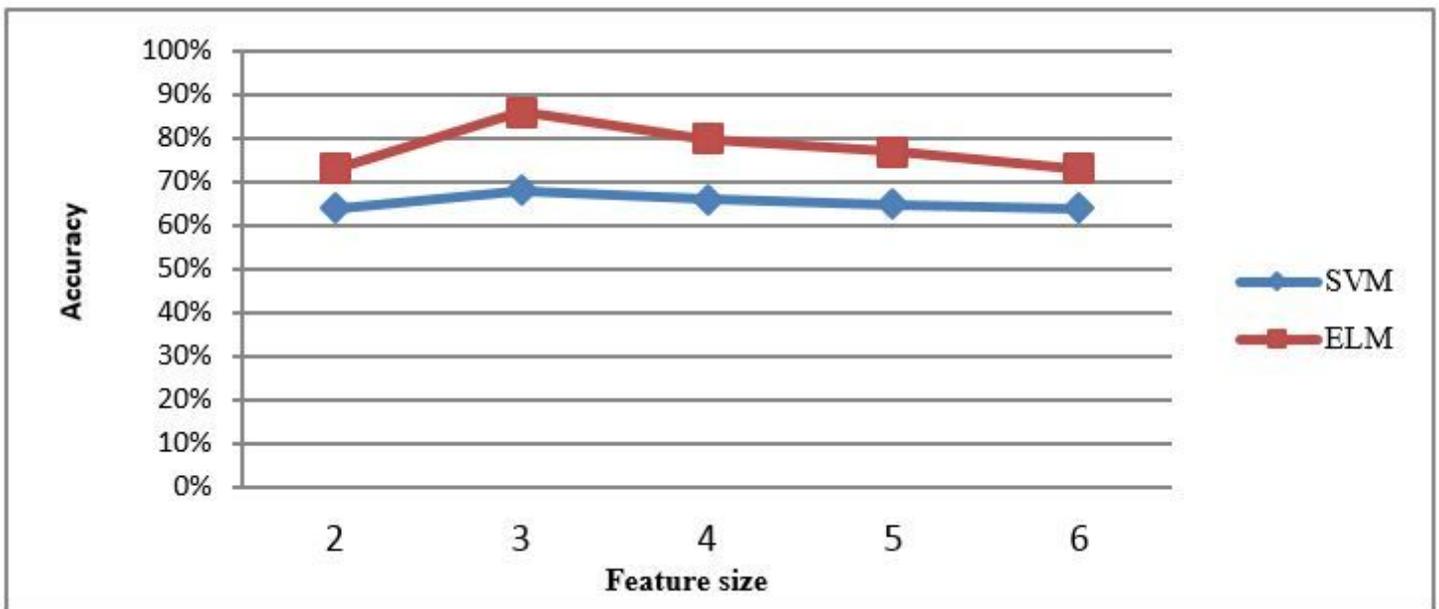


Figure 4

Comparison of FICA feature extraction method

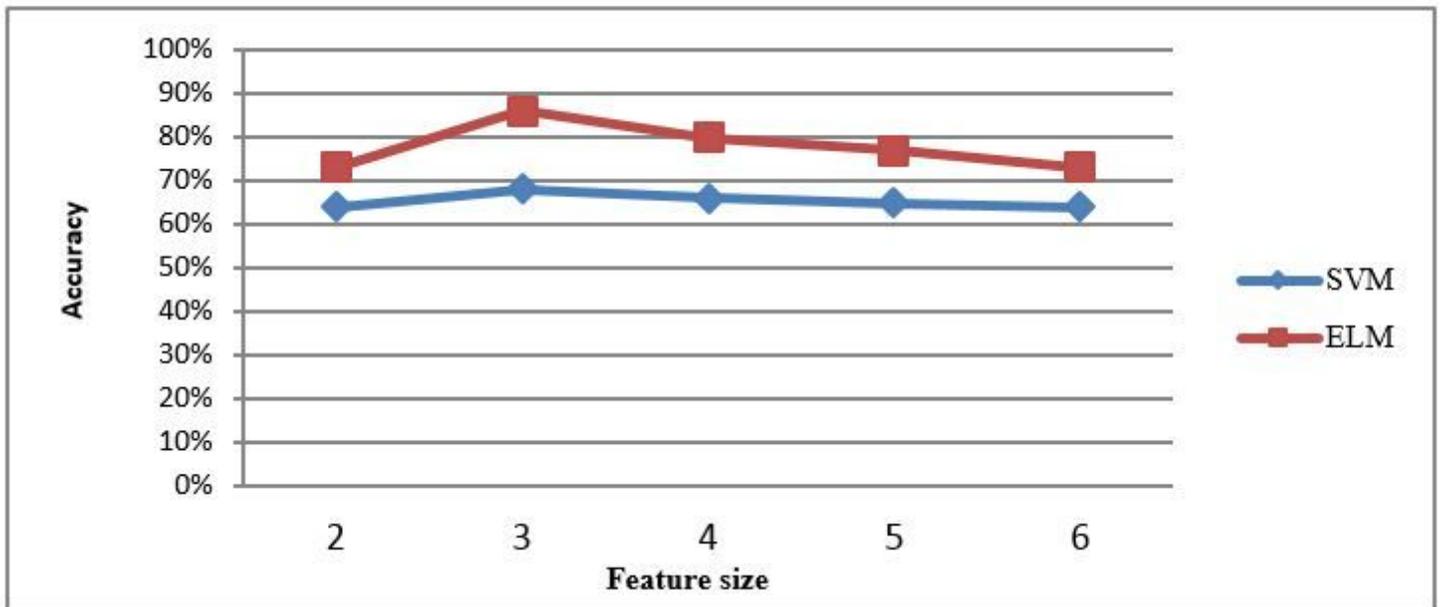


Figure 4

Comparison of FICA feature extraction method

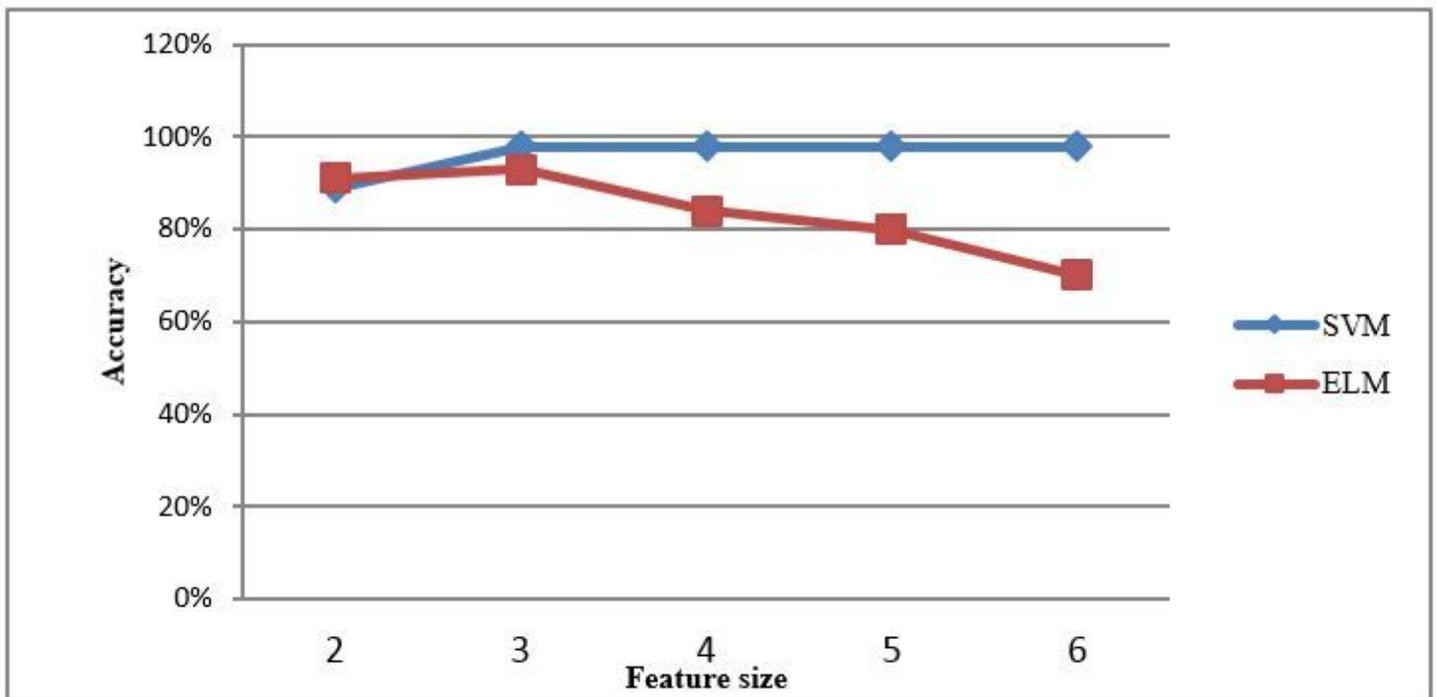


Figure 5

Comparison of DWT+FICA method

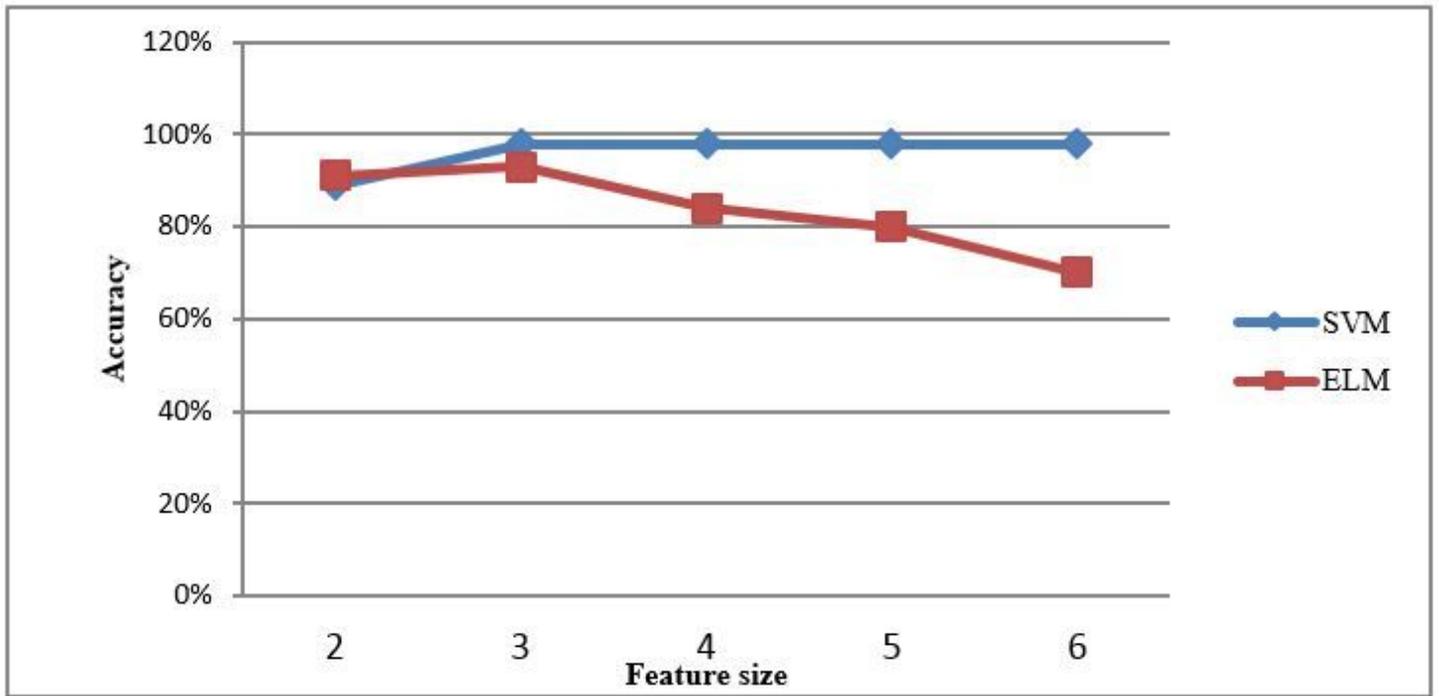


Figure 5

Comparison of DWT+FICA method

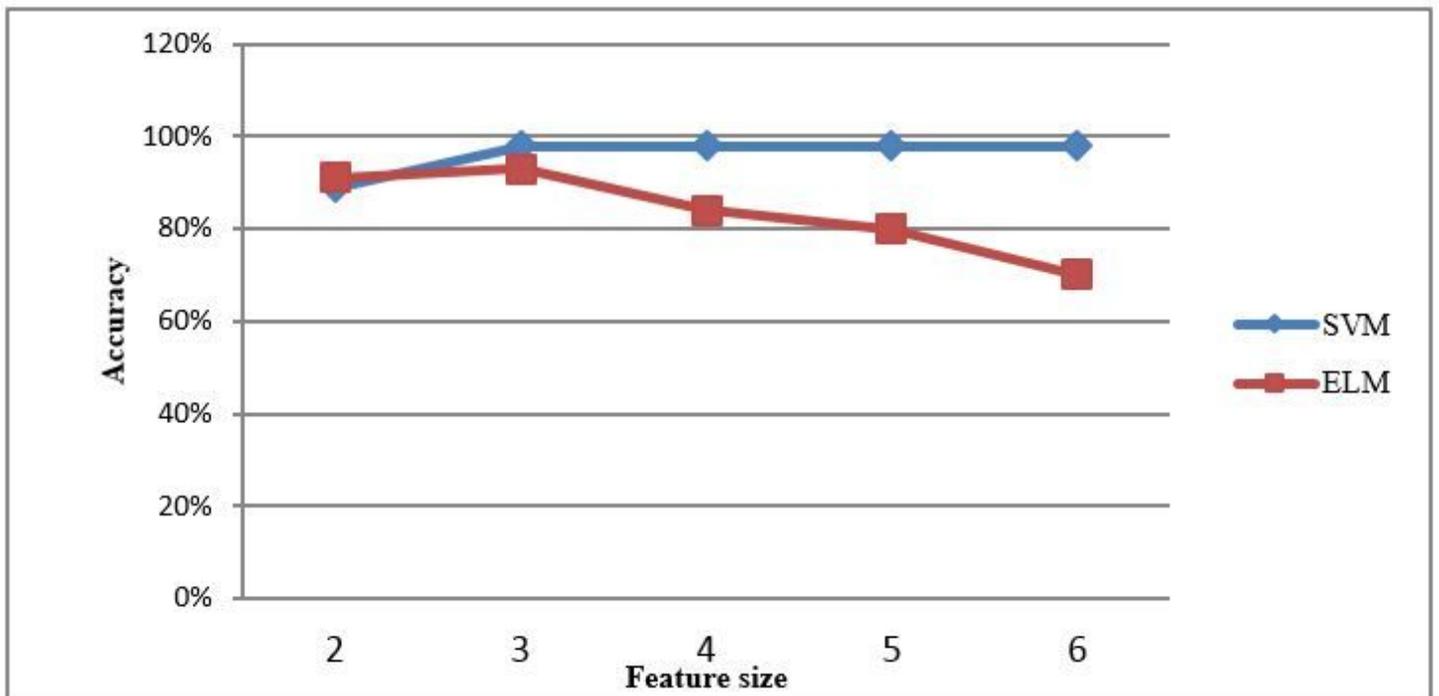


Figure 5

Comparison of DWT+FICA method

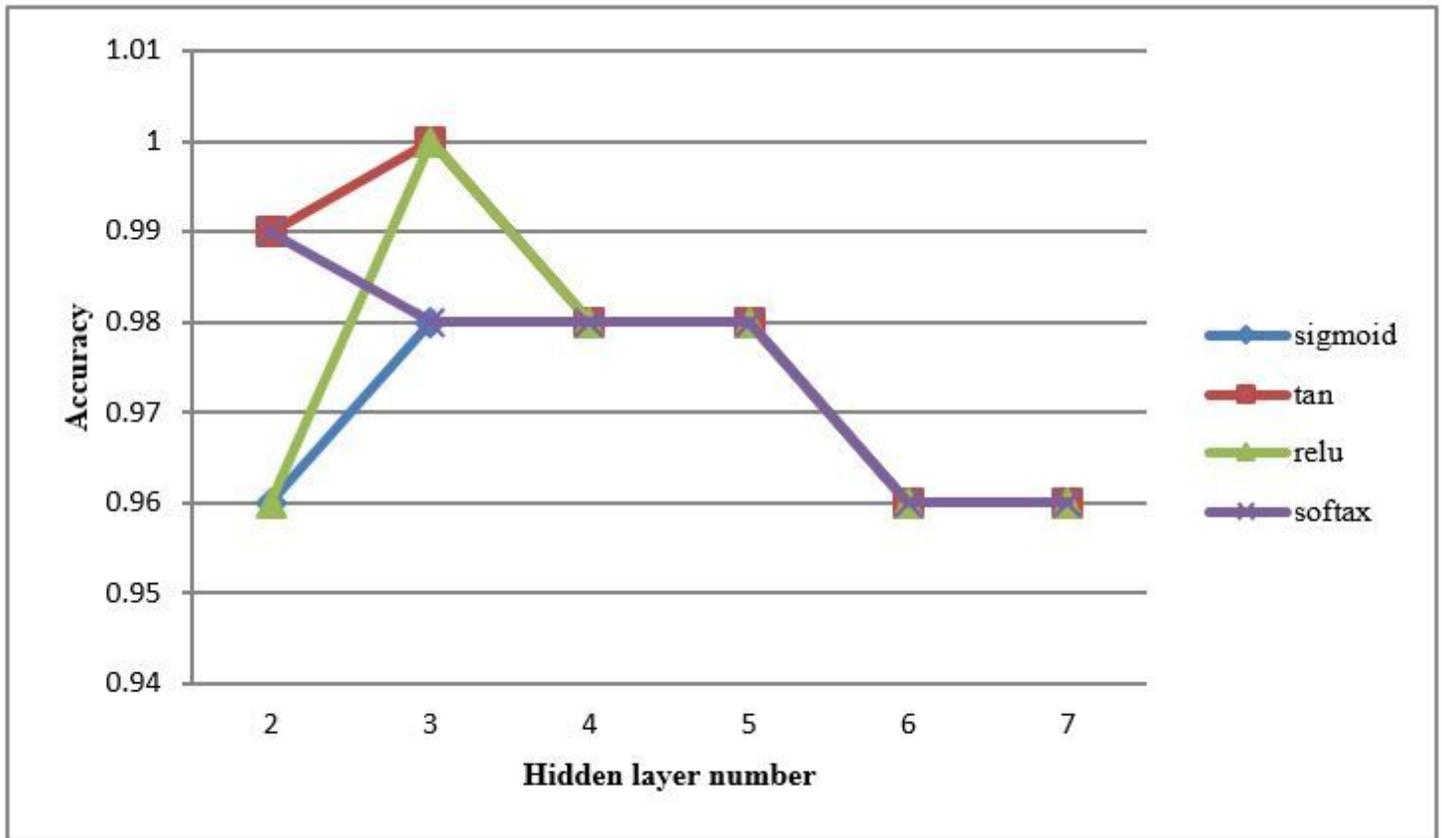


Figure 6

Comparison of different functions using ELM

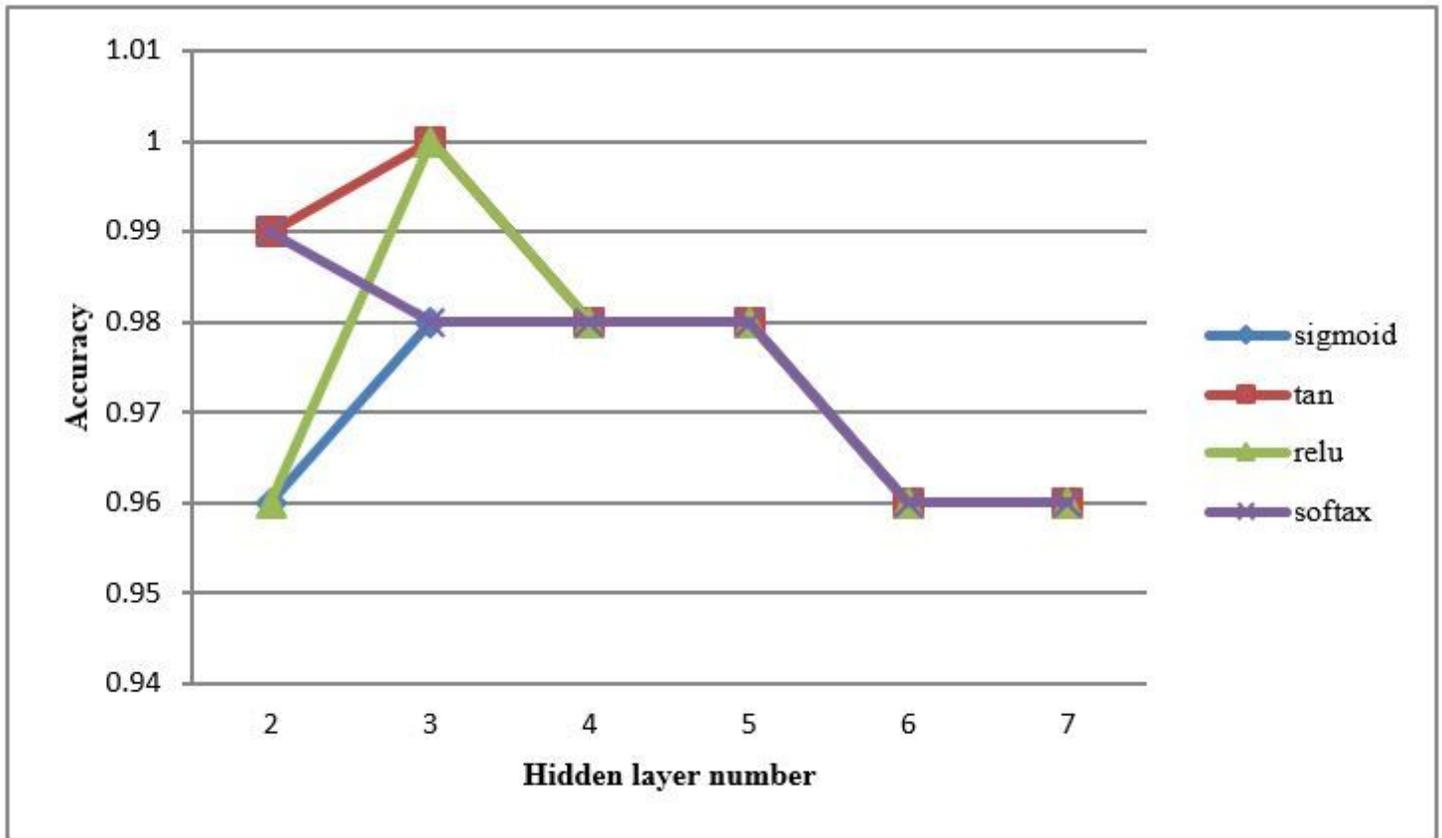


Figure 6

Comparison of different functions using ELM

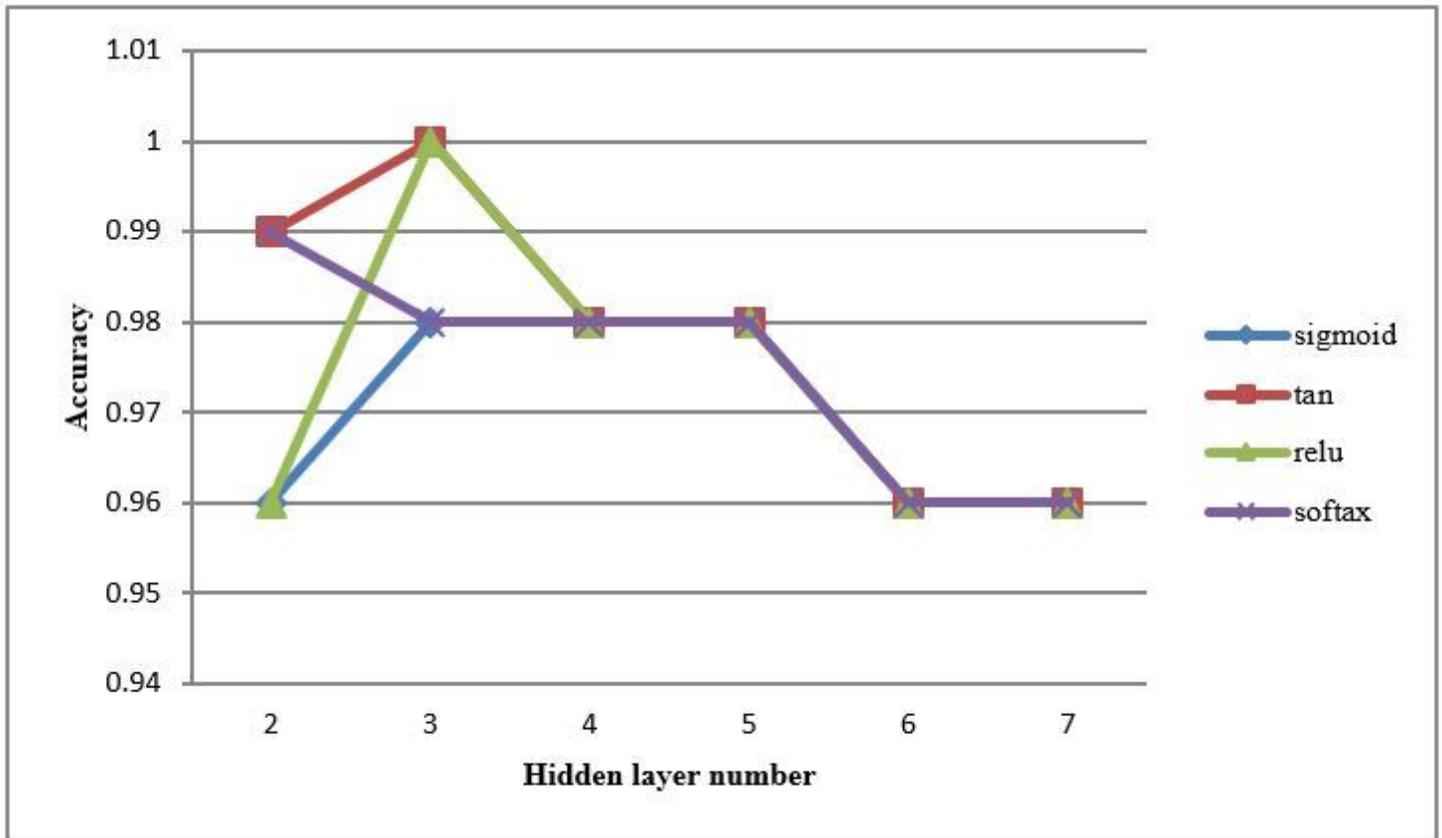


Figure 6

Comparison of different functions using ELM

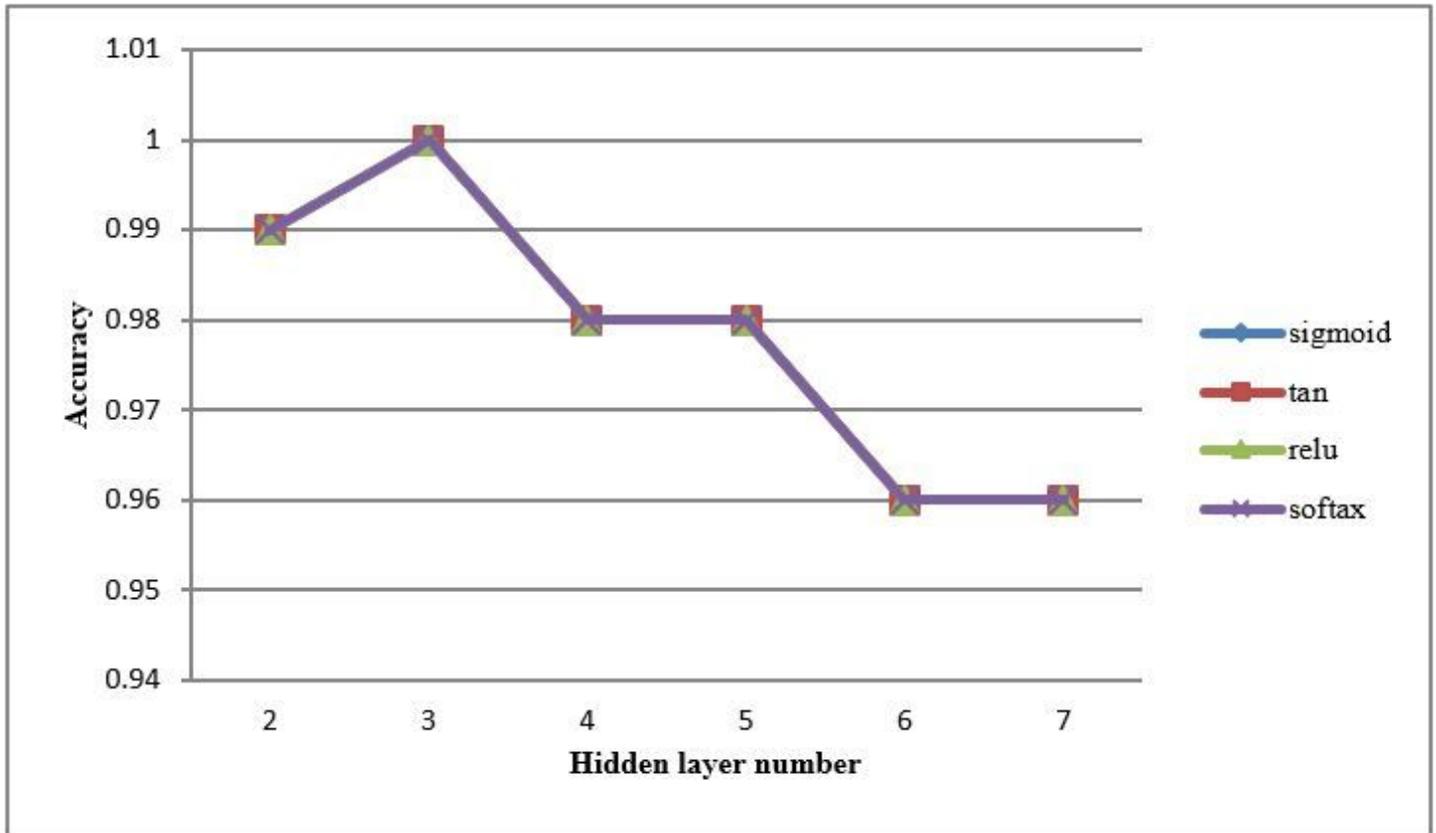


Figure 7

Comparison of different functions using DELM

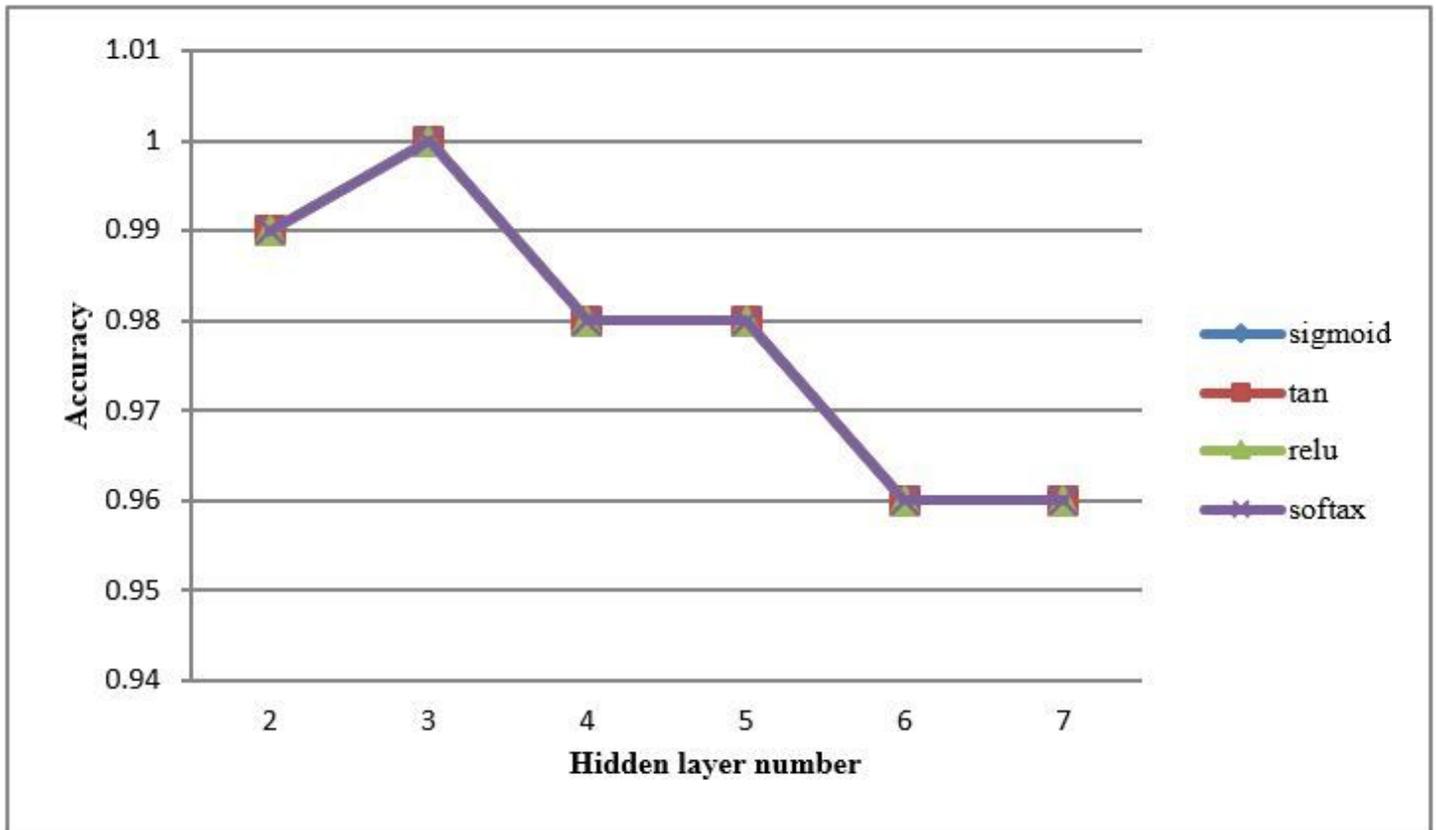


Figure 7

Comparison of different functions using DELM

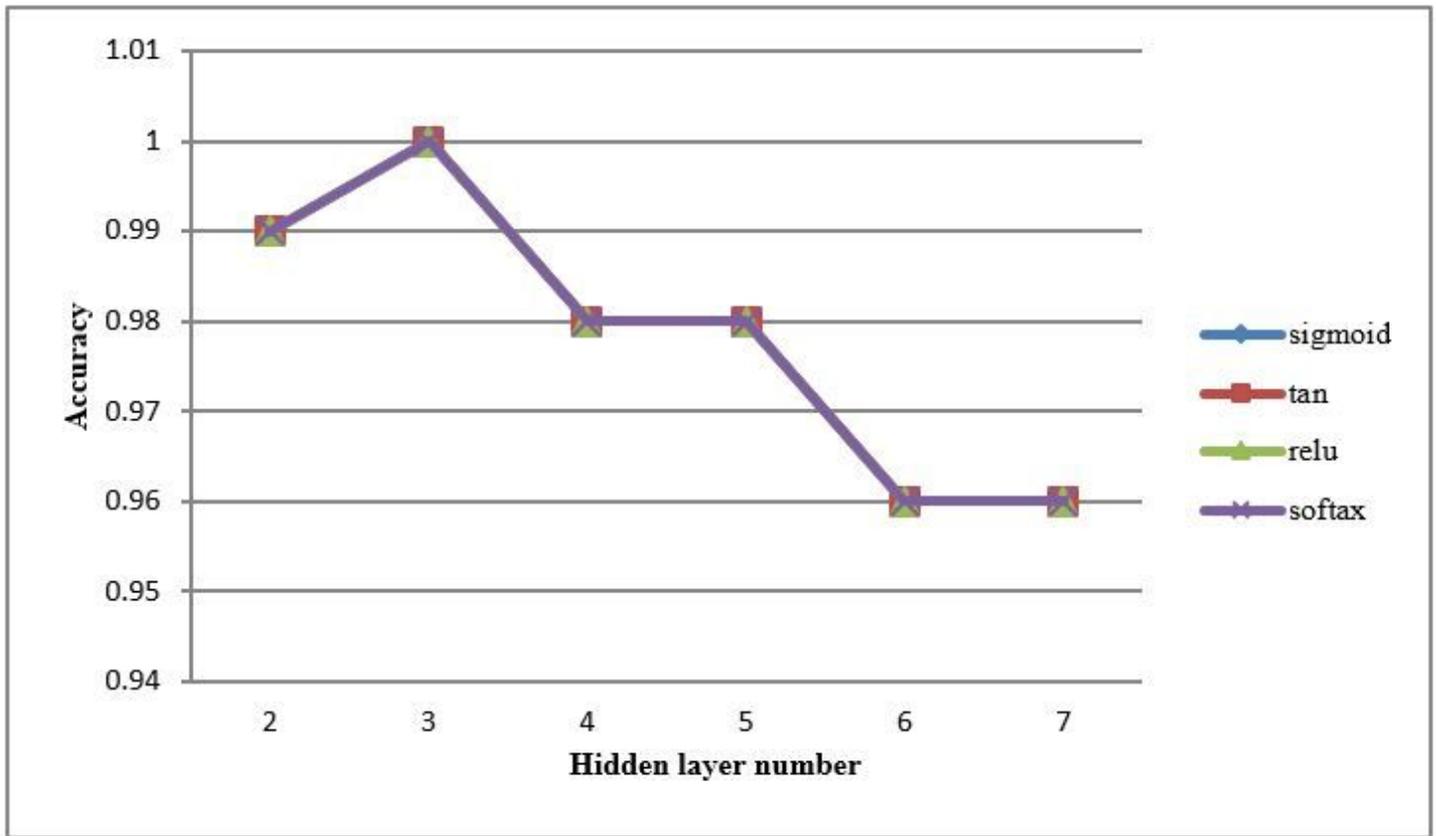


Figure 7

Comparison of different functions using DELM

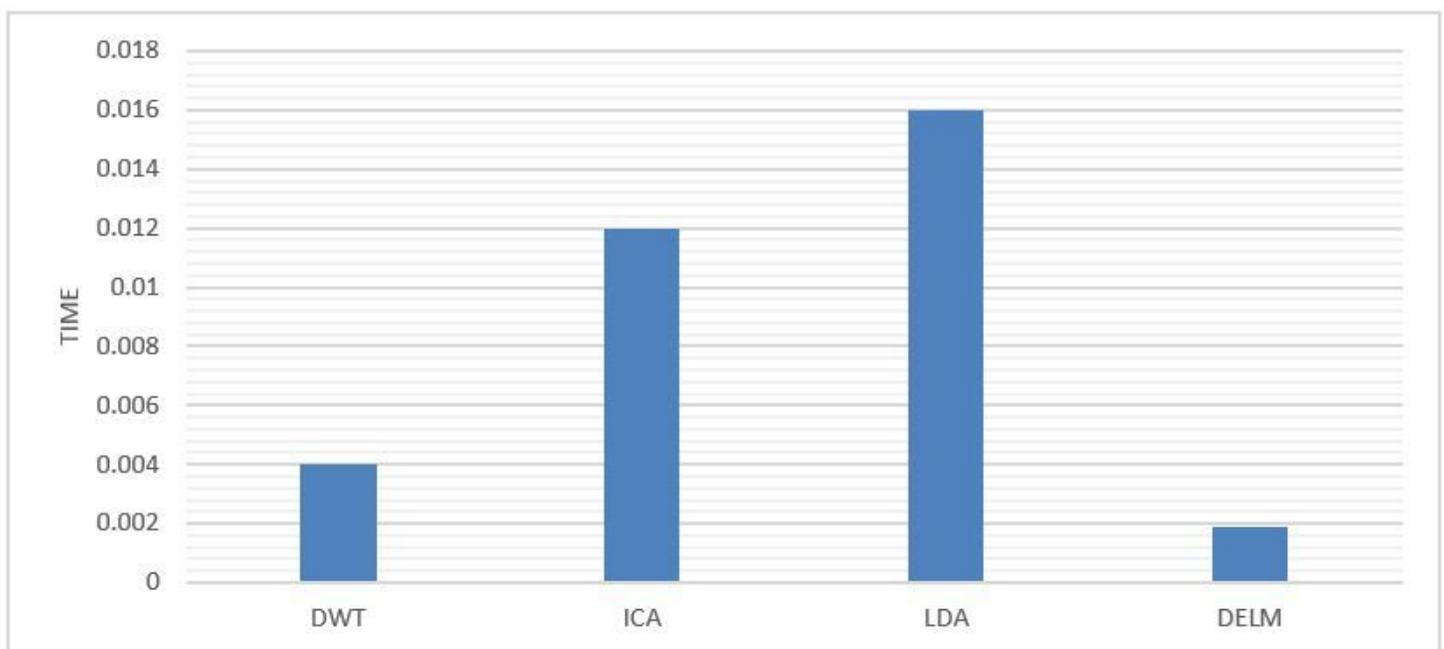


Figure 8

Average computation time at different stages of the proposed method

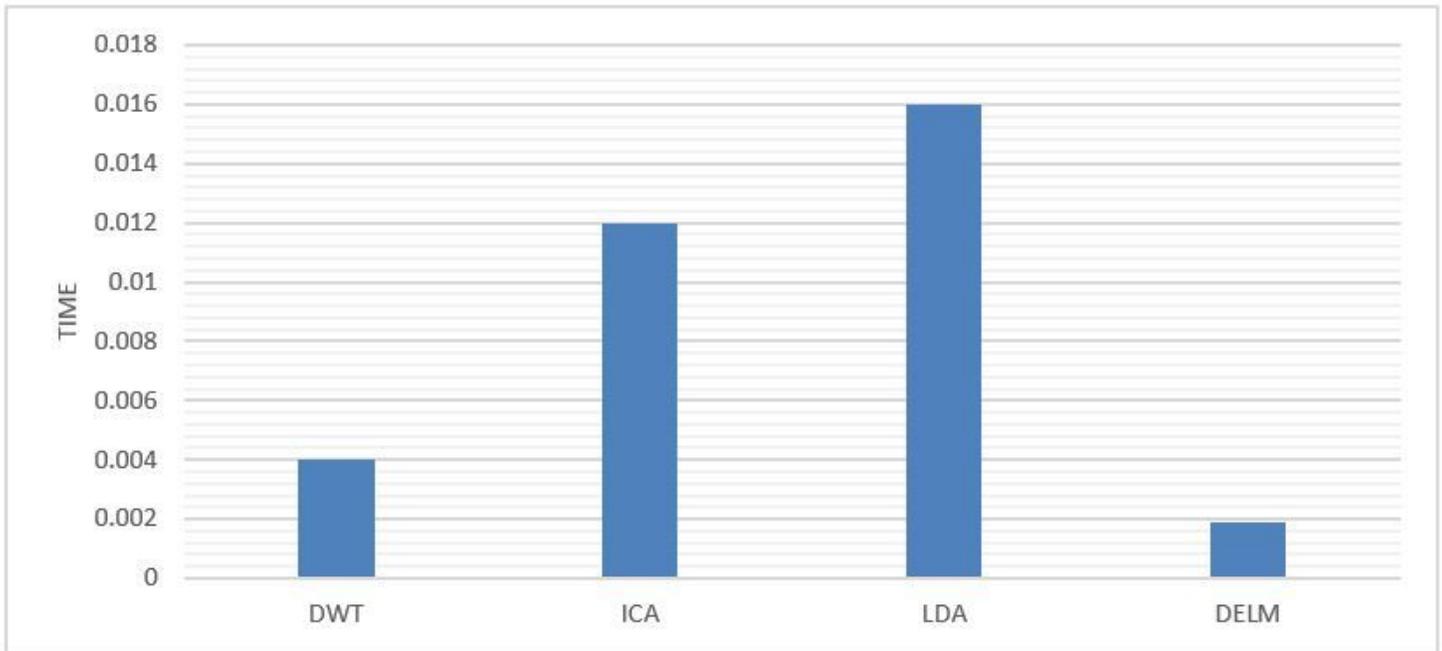


Figure 8

Average computation time at different stages of the proposed method

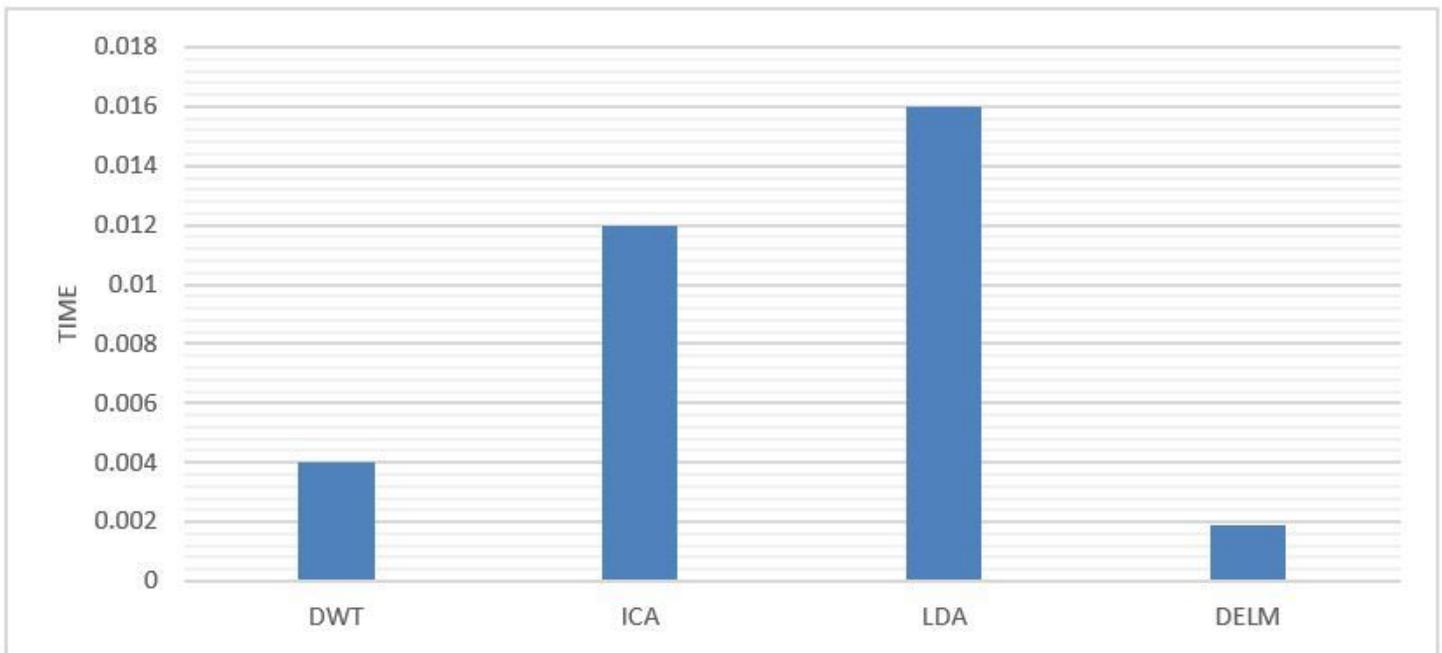


Figure 8

Average computation time at different stages of the proposed method