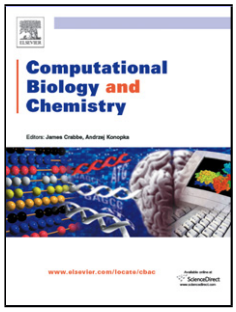


Title: Chaos enhanced Grey Wolf Optimization wrapped ELM for diagnosis of paraquat-poisoned patients

Authors: Xuehua Zhao, Xiang Zhang, Zhennao Cai, Xin Tian, Xianqin Wang, Ying Huang, Huiling Chen, Lufeng Hu

PII: S1476-9271(18)30796-5
DOI: https://doi.org/10.1016/j.compbiolchem.2018.11.017
Reference: CBAC 6961

To appear in: *Computational Biology and Chemistry*

Received date: 28 October 2018
Accepted date: 18 November 2018

Please cite this article as: Zhao X, Zhang X, Cai Z, Tian X, Wang X, Huang Y, Chen H, Hu L, Chaos enhanced Grey Wolf Optimization wrapped ELM for diagnosis of paraquat-poisoned patients, *Computational Biology and Chemistry* (2018), https://doi.org/10.1016/j.compbiolchem.2018.11.017

Chaos enhanced Grey Wolf Optimization wrapped ELM for diagnosis of paraquat-poisoned patients

Xuehua Zhao, Xiang Zhang, Zhennao Cai, Xin Tian, Xianqin Wang, Ying Huang, Huiling Chen*, Lufeng Hu*

***Abstract*—Paraquat (PQ) poisoning seriously harms the health of humanity. An effective diagnostic method for paraquat poisoned patients is a crucial concern. Nevertheless, it's difficult to identify the patients with low intake of PQ or delayed treatment. Here, a new efficient diagnostic approach to integrate machine learning and gas chromatography-mass spectrometry (GC-MS), named GEE, is proposed to identify the PQ poisoned patients. First, GC-MS provides the original data that efficiently identified the paraquat-poisoned patients. According to the high dimensionality of the original data, in the second stage, the chaos enhanced grey wolf optimization (EGWO) is adopted to search the optimal feature sets to improve the accuracy of identification. Finally, the extreme learning machine (ELM) is used to identify the PQ poisoned patients. To efficiently evaluate the proposed method, four measures were used in our experiments and comparisons were made with six other methods. The PQ-poisoned patients and robust volunteers can be well identified by GEE and the values of AUC, accuracy, sensitivity and specificity were 95.14%, 93.89%, 94.44% and 95.83%, respectively. Our experimental results demonstrated that GEE had better performance and might serve as a novel candidate diagnosis of PQ-poisoned patients.**

***Index Terms*—diagnosis, paraquat, chaos, grey wolf optimization, extreme learning machine**

This research is supported by the Science and Technology Plan Project of Wenzhou, China (ZG2017019), Science and Technology Committee of Shanghai Municipality of China (KF1405), Zhejiang Provincial Natural Science Foundation of China (LY17F020012, LY14H230001, LY14F020035), Guangdong Natural Science Foundation (2018A030313339), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (17YJCZH261), Science Research Cultivation Project of Shenzhen Institute of Information Technology (ZY201718), Special Innovation Project of Guangdong Education Department (2017GKTSCX063) and Special Funds for the Cultivation of Scientific, Technological Innovation for College Students in Guangdong (pdjh2018b0861) and The 13th Five-Year Plan Project of Philosophy and Social Sciences in Shenzhen (SZ2018D017).

*Corresponding Author: Huiling Chen and Lufeng Hu.

Xuehua Zhao is with School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China. (e-mail: lcrlc@sina.com)

Xiang Zhang is Wenzhou Mingcheng Construction Investment Group Co., Ltd.(zhxan@126.com)

Zhennao Cai is with College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, China. (e-mail: cznao@163.com)

Xin Tian is with Cancer Hospital Chinese Academy of Medical Sciences (Shenzhen Hospital), Shenzhen, 518000; China. (e-mail: 947952187@qq.com)

Xianqin Wang is with Analytical and Testing Center, Wenzhou Medical University, Wenzhou 325000, China. (e-mail: lankywang@163.com)

Ying Huang is with School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China. (e-mail: 1296634883@qq.com)

Huiling Chen is with College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou, 325035, China. (e-mail: chenhuiling.jlu@gmail.com).

Lufeng Hu is with Department of Pharmacy, the First Affiliated Hospital of Wenzhou Medical University, Nanbaixiang street, Ouhai District, Wenzhou City 325000, China. (e-Mail: hulufeng79@sina.com).

I. INTRODUCTION

Paraquat is the widely used herbicides because of its extraordinary weeding effect and harmless to the environment [1, 2]. Nevertheless, for humanity, PQ can lead to severe kidney, liver, heart, lung failure and Parkinson's disease risk [3, 4]. Statistics have shown lots of people died every year because of PQ poisoning [5]. As an example, more than 500 individuals die from PQ poisoning every year in Korea [6]. The treatments of PQ poisoning mainly are removing PQ from the body [1, 7]. And the earlier the treatment start, the lower the mortality is [5]. Therefore, timely and accurate diagnosis is the key for PQ-poisoned patients.

Currently, the diagnosis of PQ-poisoned patients mainly depends on the PQ concentration which is in blood. Since PQ is spread out all organs in the body quickly [1, 8], testing PQ in blood still faces difficulties, especially when the patients intake of low PQ or they hadn’t gone to hospital in time (some of patients went to hospital at several days later). Therefore, developing an efficient diagnosis approach for PQ poisoning has become an important topic in medicine.

So far, classification has been widely researched in the domain of medical diagnosis since many medical diagnoses can model as the classification tasks. For example, Krawczyk and Schaefer proposed a hybrid classifier with neural networks (PNN) and support vector machines (SVM) to recognize the patients with breast cancer [9]. Alex et al proposed a medical diagnosis method using the radial basis

function neural network (RBFNN) [10]. Kotu et al adopted two classifiers, which respectively are *k*-nearest neighbor (*k*-NN) classifier and SVM classifier, to identify post-myocardial infarction patients [11]. Alickovic and Subasi used random forests for the diagnosis of heart arrhythmia [12]. Extreme learning machine (ELM) is an effective classification method which is based on the single hidden layer feed forward neural networks (SLFNs) [13], and has been applied to lots of fields, for example, text classification, image recognition and fault diagnosis [14]. As a consequence, in our previous study, we first propose a new method based on the ELM to efficiently identify PQ patients [15]. The experiments demonstrate the proposed method has the excellent performance with the accuracy of identification, 91%. However, the previous method does not consider the quality of feature set and the original high dimensional feature set is directly used to identify the patients. Therefore, it is possible to lead to the low accuracy due to the redundant features contained in the feature sets. Additionally, the previous method cannot also provide the favorable information of the key features affecting the diagnosis accuracy. Feature selection (FS) is an effective approach to figure out the high dimensional space of features. Due to it efficiently reducing the redundant features to improve the accuracy of identification, FS has been applied to the wide range of fields such as text classification [16], emotion recognition [17], medical diagnosis [18, 19] and so on.

In this study, we present an efficiently and effective diagnosis framework based on gas chromatography coupled with mass spectrometry (GC-MS), Enhanced grey wolf optimization (EGWO) and ELM together, namely GEE. To evaluate the methods, four measures, accuracy (ACC), the area under the receiver operating characteristic curve (AUC), specificity (SPE) and sensitivity (SEN), are taken in our experiments. Comparisons are made with other methods. The experimental results indicate that the proposed method, GEE, not only can provide the efficient diagnosis accuracy of PQ patients but also can help to analyze the key features affecting the performance of diagnosis from the metabolomics perspective.

In summary, there mainly are three contributions in our work: (1) a novel perspective for identifying the PQ-poisoned patients using GC-MS technology is presented; (2) a novel feature selection method based on chaos theory enhanced GWO is proposed; (3) the most key features can be found by FS.

The remainder of this paper is as follows. Section 2 introduces FS and GWO. The GEE is described in Section 3. The experiments are described in Section 4. The discussion is in Section 5 and conclusions in Section 6.

II. PRELIMINARY METHODS

A. *Feature selection*

The high dimensionality of data usually is an important question for learning tasks. For training model, the high dimensional data can load to overfitting and become less comprehensive because of the irrelevant and redundant features. FS is one of effective approaches for dimensionality reduction by identifying and discarding irrelevant and redundant features. Various studies have illustrated that FS can boost the performance of the classification models [20].

The relevance and redundancy are two key terms in FS. The prediction ability of feature set is decreased since one feature is removed, that is to say, the feature is relevant. The prediction ability of feature set is not decreased because one feature is removed, that is to say, the feature is redundant. Consequently, FS is to search the minimum subset from the original set, which has maximal relevance and minimum redundancy as possible. Usually, FS process consists of four steps, which respectively are as follows: 1) The candidate subsets are produced. 2) The candidate subsets are evaluated. The subset with a best value of evaluation remained. 3) According to the given stopping criterion, the subset generation and evaluation are stopped. 4) The final selected subset is validated by given learning algorithms. Obviously, the exhaustive search is the way to find the optimal subset, but this way is NP-completeness. To resolve the problem, the heuristic search strategies have been proposed [20, 21].

B. *Grey wolf optimization* (*GWO*)

In recent years, the GWO is proposed to seek the good solution of combination question by Mirjalili et al. [22]. Like other swarm intelligence algorithms [23-31], GWO has the strong search capability of finding the global optimum, GWO mainly mimics the three keys which are respectively encircling prey, hunting and attacking prey. In GWO, *alpha* represents the best solution, *beta* represents the second best solution represents, *delta* represents the third best solution, and the rest of the solutions are named as *omega*.

The equations (1), (2), (3) and (4) are used to simulate the encircling behavior of grey wolves.

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_{prey}(t) - \vec{X}_{wolf}(t) \right| \tag{1}$$

$$\vec{X}_{wolf}(t+1) = \vec{X}_{prey}(t) - \vec{A} \cdot \vec{D} \tag{2}$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2\vec{r}_2 \tag{4}$$

where *t* is the current iteration, $\vec{A}$ and $\vec{C}$ are coefficiencies, $\vec{X}_{prey}$ and $\vec{X}_{wolf}$ respectively are the position of the prey

and a grey wolf. $\vec{a}$ is linearly decreased from 2 to 0, and $\vec{r}_1$, $\vec{r}_2$ are random vectors in the interval of [0, 1].

The three first best solutions, *alpha*, *beta* and *delta*, are saved, the *omega* is obliged to update their positions according to the following equations.

$$\vec{D}_{alpha} = \left|\vec{C}_1 \cdot \vec{X}_{alpha} - \vec{X}\right| \quad (5)$$

$$\vec{D}_{beta} = \left|\vec{C}_2 \cdot \vec{X}_{beta} - \vec{X}\right| \quad (6)$$

$$\vec{D}_{delta} = \left|\vec{C}_3 \cdot \vec{X}_{delta} - \vec{X}\right| \quad (7)$$

$$\vec{X}_1 = \vec{X}_{alpha} - \vec{A}_1 \cdot \vec{D}_{alpha} \quad (8)$$

$$\vec{X}_2 = \vec{X}_{beta} - \vec{A}_2 \cdot \vec{D}_{beta} \quad (9)$$

$$\vec{X}_3 = \vec{X}_{delta} - \vec{A}_3 \cdot \vec{D}_{delta} \quad (10)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (11)$$

C. Chaotic mapping

Chaos, as a widespread nonlinear phenomenon in nature, has the characteristics of randomness, ergodicity, sensitivity to initial conditions and so on [32]. Due to the characteristics of ergodicity and randomness, chaotic motions can traverse all the states in a certain range according to their own laws without repetition. Therefore, if we use chaos variables to search optimally, we will undoubtedly have more advantages than random search. Chaos ergodicity features can be used to optimize the search and avoid falling into the local minima; therefore, chaos optimization search method has become a novel optimization technique. Chaotic sequences generated by different mappings can be used. In this paper, chaotic sequences are generated by using logistic mapping as follows.

$$x_{i+1} = ux_i(1 - x_i) \quad (12)$$

u is the control parameter and let u = 4. When u = 4, the logistic mapping comes into a thorough chaotic state. Let $x_i \in (0,1)$ and $x_i \neq 0.25, 0.5, 0.75$.

The initial bacterial population θ is mapped to the chaotic sequence that has been generated according to Eq. (12), resulting in a corresponding chaotic bacterial population *pch*.

$$pch = x_i * \theta \quad (13)$$

D. Extreme learning machine (*ELM*)

The ELM is an efficient learning method for the SLFNs [33]. Compared to the traditional learning algorithms, the ELM learns the model by minimizing training error and the norm of output weights.

Given the training set $D = \{x_i, t_i \mid i = 1,2,...,N\}$, where x_i is the input feature and t_i is the label. The SLFNs with an activation function and the number of hidden neurons are modeled as follows

$$\sum_{j=1}^{\tilde{N}} \beta_i \, g(w_i \bullet x_j + b_j) = o_j, \; j = 1,2,...,N \quad (14)$$

where $\tilde{N}$ and $g(x)$ respectively denote the number of hidden neurons and the activation function, β_i and w_i are the weight coefficiencies, b_i and o_j respectively denotes the bias of the *i*-th hidden neuron and the output. If SLFNs can make these *N* samples with zero error, there exist β_i, w_i, b_j such that $\sum_{j=1}^{\tilde{N}} \beta_i \, g(w_i \bullet x_j + b_j) = t_j, \; j = 1,2,...,N$. The above relation can be formulated as follows:

$$\mathrm{H}\beta = \mathrm{T} \quad (15)$$

where

$$\mathrm{H}(w_1,...,w_{\tilde{N}},b_1,...,b_{\tilde{N}},x_1,...,x_{\tilde{N}}) = \begin{pmatrix} g(w_1 \bullet x_1 + b_1) & \cdots & g(w_{\tilde{N}} \bullet x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \bullet x_N + b_1) & \cdots & g(w_{\tilde{N}} \bullet x_N + b_{\tilde{N}}) \end{pmatrix}_{N \times \tilde{N}} \quad (16)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}_{N \times m} \quad (17)$$

For ELM, w, and b are arbitrarily value. β can be calculated by seeking the least square solution:

$$\| \mathrm{H}(w_1,\cdots,w_{\tilde{N}},b_1,\cdots,b_{\tilde{N}})\hat{\beta} - \mathrm{T} \| = \min_{\beta} \| \mathrm{H}(w_1,\cdots,w_{\tilde{N}},b_1,\cdots,b_{\tilde{N}})\beta - \mathrm{T} \| \quad (18)$$

We can use the Moor-Penrose (MP) generalized inverse of H to easily accomplish the above equation

$$\hat{\beta} = \mathrm{H}^{\dagger}\mathrm{T} \quad (19)$$

where $\mathrm{H}^{\dagger}$ is the MP generalized inverse.

III. PROPOSED METHOD FOR DIAGNOSIS OF PQ PATIENT

Then, we mainly introduce our diagnosis method for the PQ patients, namely GEE. The basic idea of GEE is described as follows.

To efficiently identify the PQ patients, three technologies, which are respectively GC-MS, FS and classification, fuse together in GEE. GC-MS can be regarded as the data generator and provides the data of original high dimension that is used to build classification model and identify PQ patients. FS is used to improve the performance of diagnosis of PQ by obtaining the high distinguishing feature sets, and the classification is used to identify the PQ patients. In GEE, we adopt EGWO to select the optimal feature sets and use the ELM with good generalization as the classifier of diagnosis of the PQ patients. The entire diagnosis of PQ patients consists of four steps: (1) generating

the original high dimension data of PQ patients by GC-MS; (2) based on the original data, obtaining the optimal feature subsets by EGWO; (3) rebuilding the dataset according to the optimal feature sets; (4) training the ELM model and identifying the PQ patients. The detailed process of GEE is shown in Fig.1. Compared with the previous method only using single ELM, the GEE can further improve the accuracy of diagnosis of PQ patients.

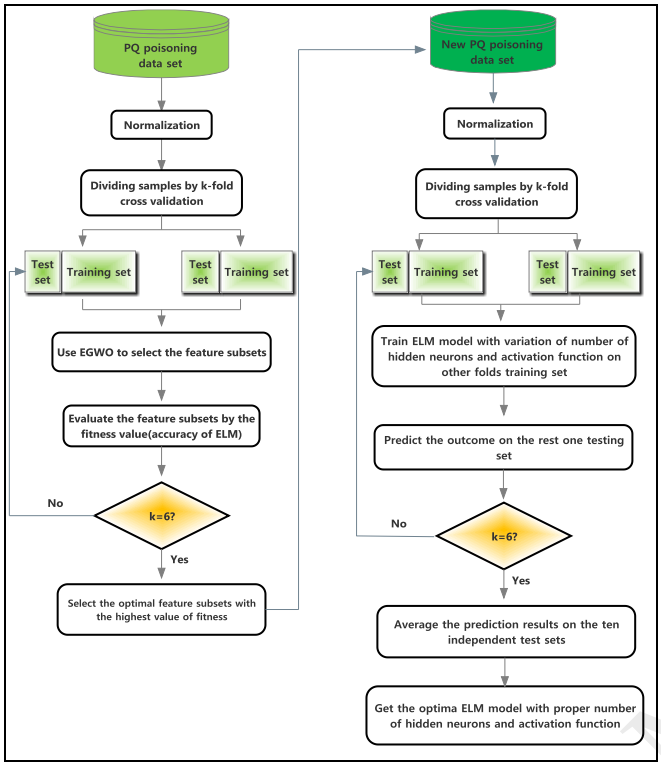


Fig.1 Flowchart of the GEE methodology

A. Data collection

In our study, the Medical Ethics Committee of The First Affiliated Hospital of Wenzhou Medical University provides us data of PQ-poisoned patients. The data only can be used by the investigators. Our work obeys the Declaration of Helsinki.

The data of 15 patients, which had a history of direct contact with PQ poisoning, had not been received any drug treatment. We have collected their hemoperfusion (HP) or hematodialysis (HD) treatment in our work. Their plasma PQ concentration is from 100 to 1500 ng/mL. We also selected 16 healthy volunteers as control. Consequently, the blood samples of 31 patients are analyzed by the GC-MS method, for more detailed information please refer to our previous article [34]. For each sample, 119 peaks are detected, in which 20 peaks are identified in terms of retention time (RT) and mass spectra data. Consequently, each sample of patients consists of the 119-dimensional feature vector.

B. EGWO for feature selection

To search in a discrete space, updating the grey wolf positions is according to the following equation:

$$flag_{i,j} = \begin{cases} 1 & X_{i,j} > 0.5 \\ 0 & otherwise \end{cases} \quad (20)$$

Where $X_{i,j}$ indicates the *j*th position of the *i*th grey wolf.

In the EGWO, chaotic theory was embedded into GWO for generating the more diversified population to achieve a more suitable balance between the exploitation and exploration. Currently, many researchers also used the chaos theory to improve the performance of GWO. For example, Emary and Zawbaa applied the chaos to adjust the exploration rate parameter [28], Kohli and Arora mapped the chaos with the algorithm along with the initialization of its first chaotic number and a variable [29], and Mehrotra and Pal replaced control parameters or random variables by chaotic variables to avoid to get caught in a local optimum [30]. Different from the above approaches, we use the chaos theory in the process of initialization. The classification accuracy obtained by the ELM classifier is used as the fitness function for evaluating the selected feature sets. The the EGWO algorithm is presented in Table 1.

Table 1. the EGWO algorithm

EGWO Algorithm

```
Begin
  Initialize the parameters: n, maxiter, dim, pos, a, A, and C;
  Generate the initial populations;
  for i = 1 : n
    for j = 1 : dim
      if pos(i, j) > 0.5
        flag(j) = 1;
      else
        flag(j) = 0;
      end if
    end for
  end for
  Calculate the fitness of grey wolves with selected features;
  while k < maxiter
    for i = 1 : n
      Update the position of the current grey wolf;
    end for
    for i = 1 : n
      for j = 1 : dim
        if pos(i, j) > 0.5
          flag(j) = 1;
        else
          flag(j) = 0;
        end if
      end for
    end for
  Update a, A, and C;
  Calculate the fitness of grey wolves with selected features;
  Update alpha, beta, and delta;
  k = k + 1;
  end while
  Return the selected features of alpha as the optimal feature
    subset;
End
```

C. *Data rebuilding*

In the data rebuilding step, the data of PQ patients was rebuilt based on the optimal subset obtained by EGWO. The data items corresponding to the optimal subset remained and the other items corresponding to the redundant features are discarded. Finally, the new data of PQ patients is reproduced. After data rebuilt, PQ patients can be presented by the new feature vector consisting of the optimal feature subset.

D. *ELM for identifying PQ patients*

The classification model is the key component to identify PQ patients in our method. In our study, we select the ELM as the classification model due to its good generalization ability and high efficiency. In this step, an optimal ELM model is constructed as follows:

Step1: Input the training set $D=\{x_i,t_i \mid i=1,2,...,N\}$.

Step2: Model the standard SLFNs with the function $g(x)$ and the number of hidden neurons $\tilde{N}$ according to Eq. (14).

Step3: Give the equation $\mathrm{H}\beta = \mathrm{T}$.

Step4: Randomly give the input weights and the biases of SLFNs.

Step5: Determine the output weights.

Step6: Calculate the $\hat{\beta}$.

IV. EXPERIMENTS

For 119 peaks, each of them is regarded as a feature to form the original feature set of PQ patients. That means the original data of each patient can be represented by a 119-dimensional vector. Since the original feature set with 119 features could include the redundant features, we need to search the optimal feature sets from the original feature set. Considering the fantastic global search ability of EGWO, the EGWO in GEE to reduce the redundant features and rebuild the feature set.

To validate the proposed algorithm, we compare GEE with ELM, *k*-NN [35], SVM [35], BPNN [36], RBFNN [37] and PNN [38].

A. *Experimental setup*

The experiments are conducted on Intel(R) Core(TM) i7-4770 CPU @ 3.40GHZ with 16GB of RAM on the Windows 7 operation system. All the algorithms are coded and run in MATLAB 2010b. In our experiments, the values of parameters are set according to the experience. For EGWO, the detailed parameters are set according to Table 2.

Table 2. Parameters setting of EGWO

Parameters	Value
Control parameter ($\vec{a}$)	[2, 0]
The number of wolves	50
The number of generations	30

For the reasonableness of the evaluation, the *k*-fold cross-validation (CV) [39] is adopted in our experiments. CV is a common way to estimate the performance of classifiers. The initial sample is divided into *K* subsamples, and one of subsamples is reserved as the data for the validation model, and the other *K*-1 samples are used to train model. As a special form of cross validation, leave one out CV (LOOCV) only use one of all the samples as validation sample, while the rest as training samples. This step is repeated until each sample has been used as a validation data [39].

Before the data is split, the data is normalized using the following Eq. (21).

$$x' = \frac{x - \mu}{\sigma} \tag{21}$$

where x denotes the old value, x' denotes the scaled value, μ, and σ respectively denote the mean and standard deviation of feature.

B. Evaluation measures

In our experiments, four measures are used to evaluating the performance of the algorithms. They are respectively the ACC, AUC [40], SEN and SPE, which are written as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{22}$$

$$SEN = \frac{TP}{TP + FN} \tag{23}$$

$$SPE = \frac{TN}{FP + TN} \tag{24}$$

where TP denotes the number of cases which are correctly identified as PQ poisoned patients; FN is the number of PQ-poisoned cases which are identified as healthy controls; TN is the number of healthy controls cases which are correctly classified as healthy controls; and FP is the number of healthy controls cases which are identified as PQ-poisoned patients. AUC is one common way to evaluate the performance of the binary classifier. The larger the value of AUC is, the better the performance of the classifier.

C. Results

The activation functions are closely related to the performance of ELM. So, we firstly determine the best activation function for ELM model based on our data. There are five activation functions tested in our experiments, which are Triangular basis function (tribas), Sigmoid function (sig), , Radial basis function (radbas) Hard-limit function (hardlim) and Sine function (sin).

1) Determination of activation function and number of neurons

In Fig. 2, the relationship between the activation functions and the accuracy is shown. As we can see, the Sigmoid function outperforms the other activation functions. Consequently, the Sigmoid function is used by the ELM model in our experiments.

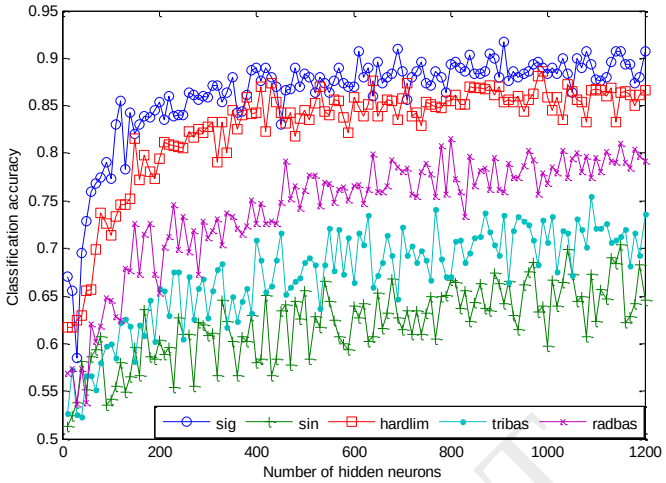


Fig. 2 Classification accuracy of ELM with different activation functions

We can also see in Fig. 2, the accuracy of ELM is related to the number of hidden neurons. To set the number of hidden neurons for ELM with Sigmoid function, different models with the number of hidden neurons varying from 10 to 1200 with the interval 10 are built. The average classification accuracy of 10 runs of k-fold CV is shown in Fig. 3. As we can see in Fig. 3, when the number of hidden neurons is increased from 10 to 200, there is a sharp rise for the performance of ELM. But when the value is more than 200, the performance of ELM became stable. Additionally, we can also see that the 6-fold CV has the better performance than another k-fold CV. Therefore, the ELM model with 800 hidden neurons is used in the following experiments and 6-fold CV is used to validate the performance of algorithms.

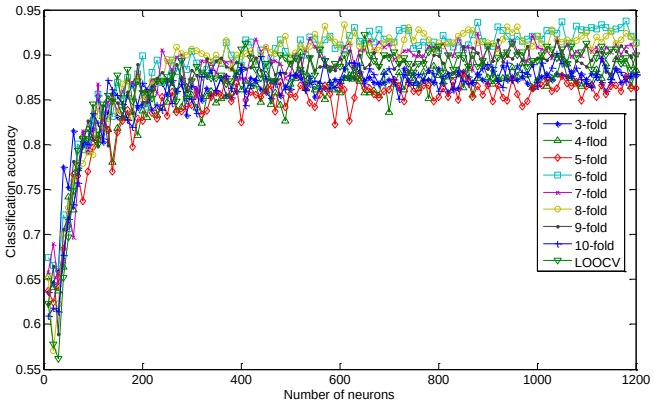


Fig. 3 Classification accuracy of ELM with different number of hidden neurons

2) Performance evaluation

For the GEE, the optimal feature set is selected from the original feature set with 119 features, after that based on the selected optimal subset, the model of ELM is learned and used to predict. Fig. 4 showed the relation of fitness and iterations of each fold in the process of feature selection. As we can see, in the beginning stage, the fitness of feature sets

had the low value, with the increasing of iteration times, the fitness value is higher and higher. This denotes the selected subsets are continually adjusted, until the optimal feature subsets appeared. Fig. 5 showed the change of size of feature sets in the process of FS. We can see that the features of subsets are adjusted until the optimal feature subsets are selected.

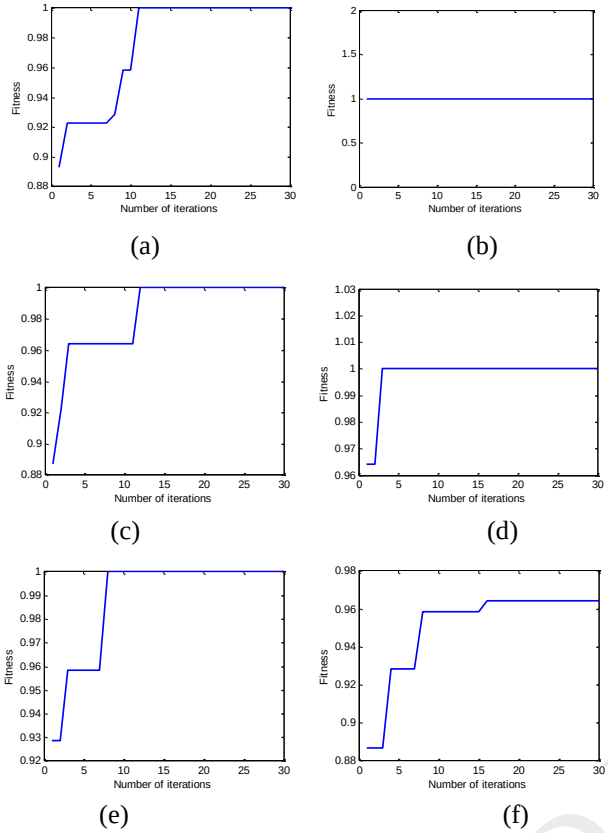


Fig. 4 Relation of fitness value and iteration times of each fold in the process of feature selection

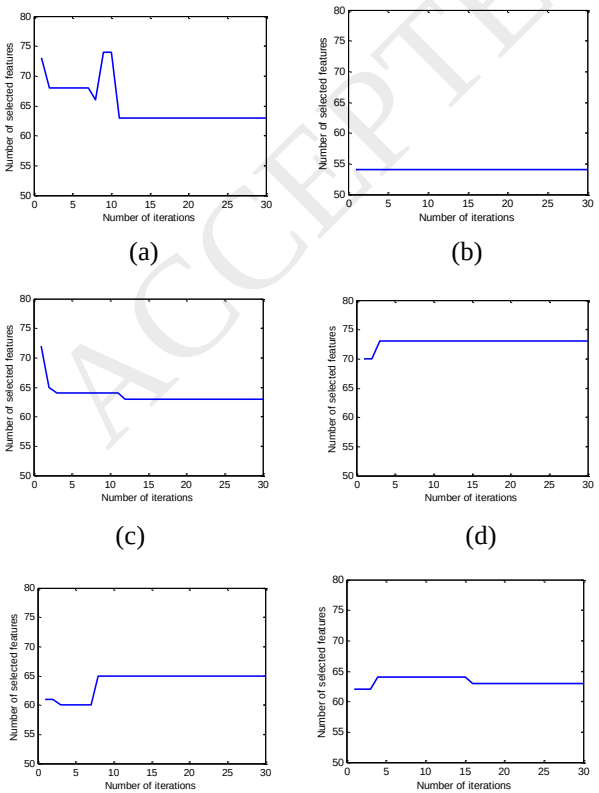


Fig. 5 Relationship of number of features and iteration in each fold.

Based on the selected feature set, the GEE built the model of ELM for diagnosis of PQ patients. To validate the ELM model, we test the model in the 6 folds CV via 10 runs and evaluate it by ACC, AUC, sensitivity and specificity measure. The mean and standard deviation of four measures are listed in Table 3. As we can see in Table 3, the GEE has the good performance, the ACC, AUC, SEN and SPE are respectively 0.9389, 0.9514, 0.9444 and 0.9583. Table 4 showed the confusion matrix. As shown in Table 4, we can see that the only two samples are incorrectly predicted.

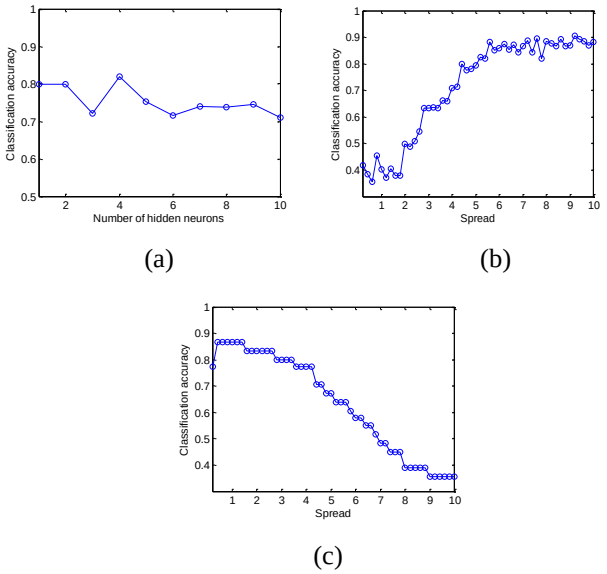
Table 3 The performance of the GEE via 10 runs of 6-fold CV

ACC	AUC	SEN	SPE
0.9389	0.9514	0.9444	0.9583
[0.022]	[0.0234]	[0.0234]	[0.0351]

Table 4 The confusion matrix via 10 runs of 6-fold CV

GEE	Predicted PQ patients	Predicted healthy controls
PQ patients	14	1
Healthy controls	1	15

Then, we made comparisons with ELM, *k*-NN, SVM, BPNN, RBFNN and PNN. For *k*-NN, we respectively set *k* to 1, 2 and 3. For SVM, the nonlinear SVM with Gaussian kernel and grid-search method [41] is used, and 4-fold CV is used to find the optimal parameters. The parameters C and γ vary between $C=\{2^{-5}, 2^{-3},\ldots, 2^{15}\}$ and $\gamma=\{2^{-15},2^{-13},\ldots,2^{1}\}$. The parameters are set according to the combinations of (C, γ) with the best CV accuracy. For BPNN, an important parameter is the number of hidden neurons. Fig. 6 (a) shows the relationship of ACC and the number of hidden neurons. As we can see, when the number of hidden neurons is 4, the ELM has the highest accuracy. So, for BPNN, the number of neurons is set to 4 for comparison. For RBFNN, the spread is the key parameter, Fig. 6 (b) shows the relation of accuracy and spread value in the dataset of PQ patients. We can see that the accuracy became stable when the value of spread is more than 5. So, we set the spread to 9 for comparison in our experiments. For PNN, Fig. 6 (c) shows the relation of accuracy and spread value. We can see that there is the highest accuracy for PNN when the spread value is set in the range from 0.2 to 1.4, so we set spread value to 1 for PNN.



(a) (b) (c)

Fig. 6 Relation of accuracy and parameters for three algorithms. (a) Relation of accuracy and number of hidden neurons for BP. (b) Relation of accuracy and spread for RBF (c) Relation of accuracy and spread for PNN.

All these algorithms run 10 times in 6-folds CV. The mean and standard deviation of four measures are listed in Table 5, where the standard deviation is in the brackets. As showed in Table 5, the proposed algorithm has the highest mean value of ACC, AUC, sensitivity and specificity among nine algorithms, and has the smallest standard deviation value of ACC, AUC and Specificity. These indicate the proposed algorithm has the best performance among nine algorithms.

Table 5 The results of nine methods

Methods	ACC	AUC	Sensitivity	Specificity
BP	0.8194	0.8076	0.8042	0.8111
	[0.0817]	[0.1003]	[0.0963]	[0.1272]
RBF	0.9044	0.9158	0.9197	0.9119
	[0.0446]	[0.0429]	[0.0393]	[0.0739]
PNN	0.8667	0.8781	0.9139	0.8422
	[0.0312]	[0.0267]	[0.0570]	[0.0485]
KNN-1	0.8811	0.8994	0.9361	0.8628
	[0.0337]	[0.0351]	[0.0320]	[0.0496]
KNN-2	0.7567	0.7875	0.8806	0.5944
	[0.0339]	[0.0263]	[0.0271]	[0.0550]
KNN-3	0.8322	0.8514	0.8972	0.8056
	[0.0265]	[0.0243]	[0.0288]	[0.0687]
SVM	0.8639	0.8907	0.8681	0.9133
	[0.0607]	[0.0399]	[0.0302]	[0.0775]
ELM	0.9072	0.8986	0.9361	0.8611
	[0.0280]	[0.0242]	[0.0132]	[0.0403]
GEE	0.9389	0.9514	0.9444	0.9583
	[0.0222]	[0.0234]	[0.0234]	[0.0351]

3) Features analysis

The proposed algorithm, GEE, not only can improve the accuracy of diagnosis of PQ, but also can help us find these key features and feature combination so that further study PQ. Here, we select ten groups of feature subsets obtained by the proposed method to analyze the characteristic of features. Table 6 lists the performance and size of ten feature subsets obtained by the proposed algorithm. We can see that (1) ten groups of feature subset have a good classification performance, their ACC is higher than 0.92, (2) the number of features in these subsets is much less than that in the original set. For these subsets, the number of the smallest feature subsets is 56 and the number of the largest feature subsets is only 73. This indicates the original set contains the many redundant features which seriously affect the classification of feature set, and FS can efficiently improve the performance of classification algorithms.

Table 6 Ten feature subsets obtained by GEE

No.	Features	ACC	AUC	Sensitivity	Specificity
No.1	63	0.9344	0.9396	0.9611	0.9180
		[0.0319]	[0.0274]	[0.0268]	[0.0361]
No.2	52	0.9239	0.9417	0.9722	0.9111
		[0.0277]	[0.0218]	[0.0293]	[0.0264]
No.3	57	0.9300	0.9250	0.9778	0.8722
		[0.0367]	[0.0366]	[0.0287]	[0.0657]
No.4	59	0.9289	0.9194	0.9667	0.8722
		[0.0285]	[0.0369]	[0.0430]	[0.0527]
No.5	74	0.9306	0.9410	0.9944	0.8875
		[0.0320]	[0.0321]	[0.0176]	[0.0546]
No.6	63	0.9278	0.9236	0.9667	0.8806
		[0.0257]	[0.0212]	[0.0287]	[0.0377]
No.7	53	0.9367	0.9535	0.9944	0.9125
		[0.0246]	[0.0232]	[0.0176]	[0.0414]
No.8	74	0.9339	0.9236	0.9500	0.8972
		[0.0301]	[0.0343]	[0.0583]	[0.0568]
No.9	52	0.9267	0.9382	0.9028	0.9736
		[0.0439]	[0.0346]	[0.0532]	[0.0324]
No.10	63	0.9367	0.9208	0.9250	0.9167
		[0.0105]	[0.0132]	[0.0264]	[0]

Then we analyze the characteristic of features of ten feature subsets. Fig. 7 shows the frequency of each feature appears in ten subsets. As we can see, there are two features all of which appeared in ten subsets, these features are respectively No.3 and No.87, i.e. butanoic acid and uric acid. There are two features all of which appeared in nine subsets, these features are respectively No.20 and No.39. There are

seven features all of which appeared in eight subsets, these features are respectively No.6, No.7, No.17, No.48, No.49, No.66 and No.107. This indicates these features are the key features of diagnosis of PQ which can help to effectively recognize the PQ patients, especially, butanoic acid and uric acid.

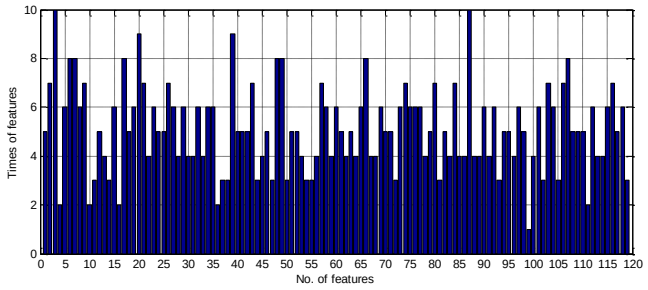


Fig. 7 Frequency of the selected features in ten subsets

V. DISCUSSION

Metabolomics or metabolite profiling is a recently developed technology for systems biology, which contains the systematic and comprehensive profiling of metabolites, cellular and systemic changes in response to disease and environmental influences. When integrating high-throughput analytical technologies, the metabolomics is the excellent irreplaceable technology to assess global changes in small molecular signatures. Due to separation efficiency, good reproducibility and efficient identification of metabolites, GC-MS becomes an efficient analytical tool in metabolomics. It has begun to apply to the domain of medical diagnosis and shows important value. Since GC-MS will generate a plenty of data, we need a high efficient data process method, such as classification, FS.

In our proposed diagnosis method for PQ-poisoned patients, the GC-MS technology is the key factor identifying the PQ patients poisoned. Most existing diagnosis method deeply depending on determination of PQ concentration, so it is difficult to identify PQ patients with low concentration of PQ in blood. The metabolomics data resources from GC-MS present the metabolic changes in blood after PQ poisoning. This is the main reason that our method is more effective than current methods.

The data based on GC-MS is the basis of the proposed method, however, the data usually is high-dimensional data, which not only loads to the long training time and identifying time but also affects the accuracy of identification due to including many irrelevant and redundant features. So, the FS fuse into the proposed method to help to select the feature set with high classification accuracy and to analyze the key feature identifying PQ patients. In this study, EGWO is adopted to find the good feature subset due to its fantastic global search ability. Finally, the ELM with the excellent generalization performance is used to identify the PQ-poisoned patients. Though other classifiers can be used to identify PQ-poisoned patients instead of ELM, the ELM is more efficient in the data of PQ-poisoned patients.

In PQ-poisoned diagnosis, finding the sample of PQ-poisoned patients and hadn't received any treatment is very difficult. This is the reason why we only have 15 data of PQ-poisoned patients. Generally, if there is no definite history of exposure to PQ poisoning, it is difficult due to the ambiguous clinical manifestations of PQ-poisoned. Based on GC-MS, we can obtain plenty of mass spectra data from blood samples of PQ-poisoned patients and don't consider how many PQ the patient ingested or how long they went to hospital. Because the retention time of each compound in GC-MS was different, and, their molecular mass was unique, the results of GC-MS analysis are credible. The proposed method can be used to find the PQ-poisoned patients. In the current condition that the PQ concentration in the body determined by high-performance liquid chromatography or mass spectrum is the only approach for diagnosis of PQ poisoning, the proposed method could be a valuable diagnosis method.

VI. CONCLUSIONS

This work presents an efficient framework for diagnosing PQ by fusing GC-MS, EGWO and ELM together. As the core of the framework, EGWO integrates the distinctive inherent sensitivity, ergodicity and randomness of chaos theory, which is able to explore much more searching space, therefore selecting the most relevant features for PQ diagnosis. The results of empirical experiments have shown the advantage of the proposed method in terms of four measures which are ACC, AUC, SEN and SPEC, respectively. Additionally, the most important feature for PQ-poisoned diagnosis has been found. As a consequence, the proposed method will provide a practical way for medical decision support. In future works, the other mechanisms such as orthogonal learning or quadratic interpolation will be investigated to further enhance the search capability of GWO. In addition, the core EGWO can be fused into the computer-aided system for much easier decision making.

VII. COMPETING INTERESTS

The authors declare that they do not have any conflicts of interest (financial or otherwise) related to the data presented in this manuscript.

Reference:

[1] S. C. Yoon, "Clinical outcome of paraquat poisoning," *Korean J Intern Med,* vol. 24, pp. 93-4, Jun 2009.

[2] J. L. Lin, M. L. Leu, Y. C. Liu, and G. H. Chen, "A prospective clinical trial of pulse therapy with glucocorticoid and cyclophosphamide in moderate to severe paraquat-poisoned patients," *Am J Respir Crit Care Med,* vol. 159, pp. 357-60, Feb 1999.

[3] Y. Zou, Y. Shi, Y. Bai, J. Tang, Y. Chen, and L. Wang, "An improved approach for extraction and high-performance liquid chromatography analysis of paraquat in human plasma," *J Chromatogr B Analyt Technol Biomed Life Sci,* vol. 879, pp. 1809-12, Jun 15 2011.

[4] M. J. Rio and C. Velez-Pardo, "Paraquat induces apoptosis in human lymphocytes: protective and rescue effects of glucose, cannabinoids and insulin-like growth factor-1," *Growth Factors,* vol. 26, pp. 49-60, Feb 2008.

[5] C. W. Hsu, J. L. Lin, D. T. Lin-Tan, K. H. Chen, T. H. Yen, M. S. Wu*, et al.*, "Early Hemoperfusion May Improve Survival of Severely Paraquat-Poisoned Patients," *Plos One,* vol. 7, Oct 29 2012.

[6] J. R. Koo, J. C. Kim, J. W. Yoon, G. H. Kim, R. W. Jeon, H. J. Kim*, et al.*, "Failure of continuous venovenous hemofiltration to prevent death in paraquat poisoning," *American Journal of Kidney Diseases,* vol. 39, pp. 55-59, Jan 2002.

[7] W. P. Wu, M. N. Lai, C. H. Lin, Y. F. Li, C. Y. Lin, and M. J. Wu, "Addition of Immunosuppressive Treatment to Hemoperfusion Is Associated with Improved Survival after Paraquat Poisoning: A Nationwide Study," *Plos One,* vol. 9, Jan 27 2014.

[8] P. Houze, F. J. Baud, R. Mouy, C. Bismuth, R. Bourdon, and J. M. Scherrmann, "Toxicokinetics of paraquat in humans," *Hum Exp Toxicol,* vol. 9, pp. 5-12, Jan 1990.

[9] B. Krawczyk and G. Schaefer, "A hybrid classifier committee for analysing asymmetry features in breast thermograms," *Applied Soft Computing,* vol. 20, pp. 112-118, Jul 2014.

[10] A. Alexandridis and E. Chondrodima, "A medical diagnostic tool based on radial basis function classifiers and evolutionary simulated annealing," *Journal of Biomedical Informatics,* vol. 49, pp. 61-72, Jun 2014.

[11] L. P. Kotu, K. Engan, R. Borhani, A. K. Katsaggelos, S. Orn, L. Woie*, et al.*, "Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients," *Artificial Intelligence in Medicine,* vol. 64, pp. 205-215, Jul 2015.

[12] E. Alickovic and A. Subasi, "Medical Decision Support System for Diagnosis of Heart Arrhythmia using DWT and Random Forests Classifier," *Journal of Medical Systems,* vol. 40, Apr 2016.

[13] S. F. Ding, H. Zhao, Y. N. Zhang, X. Z. Xu, and R. Nie, "Extreme learning machine: algorithm, theory and applications," *Artificial Intelligence Review,* vol. 44, pp. 103-115, Jun 2015.

[14] X. G. Zhao, X. Bi, G. R. Wang, Z. Zhang, and H. B. Yang, "Uncertain XML documents classification using Extreme Learning Machine," *Neurocomputing,* vol. 174, pp. 375-382, Jan 22 2016.

[15] L. F. Hu, G. L. Hong, J. S. Ma, X. Q. Wang, and H. L. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine,* vol. 59, pp. 116-124, Apr 1 2015.

[16] A. S. Ghareb, A. Abu Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications,* vol. 49, pp. 31-47, May 1 2016.

[17] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert Systems with Applications,* vol. 47, pp. 35-41, Apr 1 2016.

[18] J. Y. Choi, D. H. Kim, K. N. Plataniotis, and Y. M. Ro, "Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography," *Expert Systems with Applications,* vol. 46, pp. 106-121, Mar 15 2016.

[19] R. Sheikhpour, M. A. Sarram, and R. Sheikhpour, "Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer," *Applied Soft Computing,* vol. 40, pp. 113-131, Mar 2016.

[20] X. H. Zhao, D. L. Li, B. Yang, H. L. Chen, X. B. Yang, C. L. Yu*, et al.*, "A two-stage feature selection method with its application," *Computers & Electrical Engineering,* vol. 47, pp. 114-125, Oct 2015.

[21] X. H. Zhao, D. L. Li, B. Yang, C. Ma, Y. G. Zhu, and H. L. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing,* vol. 24, pp.

585-596, Nov 2014.

[22] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software,* vol. 69, pp. 46-61, 2014.

[23] Q. Zhang, H. Chen, J. Luo, Y. Xu, C. Wu, and C. Li, "Chaos Enhanced Bacterial Foraging Optimization for Global Optimization," *IEEE Access,* pp. 1-1, 2018.

[24] X. Wang, Z. Wang, J. Weng, C. Wen, H. Chen, and X. Wang, "A new effective machine learning framework for sepsis diagnosis," *IEEE Access,* vol. 10.1109/ACCESS.2018.2867728, 2018.

[25] J. Luo, H. Chen, Q. zhang, Y. Xu, H. Huang, and X. Zhao, "An improved grasshopper optimization algorithm with application to financial stress prediction," *Applied Mathematical Modelling,* vol. 64, pp. 654-668, 2018/12/01/ 2018.

[26] Z. Cai, J. Gu, C. Wen, D. Zhao, C. Huang, H. Huang*, et al.*, "An Intelligent Parkinsons' Disease Diagnostic System Based on a Chaotic Bacterial Foraging Optimization Enhanced Fuzzy KNN Approach," *Computational and Mathematical Methods in Medicine,* vol. 2018, p. 24, 2018.

[27] M. Wang, H. Chen, H. Li, Z. Cai, X. Zhao, C. Tong*, et al.*, "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Engineering Applications of Artificial Intelligence,* vol. 63, pp. 54-68, 8// 2017.

[28] M. Wang, H. Chen, B. Yang, X. Zhao, L. Hu, Z. Cai*, et al.*, "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing,* vol. 267, pp. 69-84, 2017/12/06/ 2017.

[29] W. Deng, H. Zhao, X. Yang, J. Xiong, M. Sun, and B. Li, "Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment," *Applied Soft Computing,* vol. 59, 2017.

[30] W. Deng, H. Zhao, L. Zou, G. Li, X. Yang, and D. Wu, "A novel collaborative optimization algorithm in solving complex optimization problems," *Soft Computing,* vol. 21, pp. 4387-4398, 2017/08/01 2017.

[31] W. Deng, R. Yao, H. Zhao, X. Yang, and G. Li, "A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm," *Soft Computing,* pp. 1-18, 2017.

[32] T. Kapitaniak, "Continuous control and synchronization in chaotic systems," *Chaos Solitons & Fractals,* vol. 6, pp. 237-244, 1995.

[33] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics,* vol. 2, pp. 107-122, Jun 2011.

[34] X. Wang, M. Zhang, J. Ma, Y. Zhang, G. Hong, F. Sun*, et al.*, "Metabolic changes in paraquat poisoned patients and support vector machine model of discrimination," *Biol Pharm Bull,* vol. 38, pp. 470-5, 2015.

[35] X. D. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda*, et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems,* vol. 14, pp. 1-37, Jan 2008.

[36] Y. L. Yu, W. Li, D. R. Sheng, and J. H. Chen, "A hybrid short-term load forecasting method based on improved ensemble empirical mode decomposition and back propagation neural network," *Journal of Zhejiang University-Science A,* vol. 17, pp. 101-114, Feb 2016.

[37] G. A. Montazer and D. Giveki, "An improved radial basis function neural network for object image retrieval," *Neurocomputing,* vol. 168, pp. 221-233, Nov 30 2015.

[38] D. F. Specht, "Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification," *IEEE Trans Neural Netw,* vol. 1, pp. 111-21, 1990.

[39] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition,* vol. 48, pp. 2839-2846, Sep 2015.

[40] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, pp. 861-874, Jun 2006.

[41] S. J. Chen, C. Y. Chang, K. Y. Chang, J. E. Tzeng, Y. T. Chen, C. W. Lin*, et al.*, "Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines," *Ultrasound Med Biol,* vol. 36, pp. 2018-26, Dec 2010.