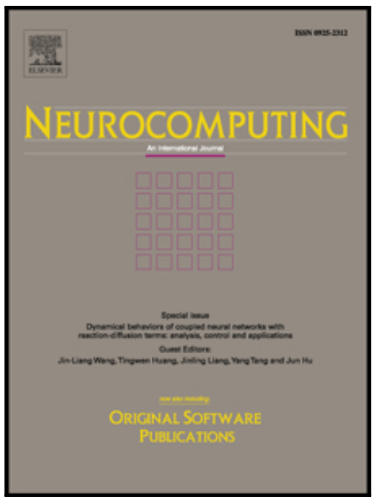


Towards an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses

Mingjing Wang , Huiling Chen , Bo Yang , Xuehua Zhao , Lufeng Hu , ZhenNao Cai , Hui Huang , Changfei Tong

PII: S0925-2312(17)30830-5
DOI: 10.1016/j.neucom.2017.04.060
Reference: NEUCOM 18427

To appear in: *Neurocomputing*

Received date: 14 December 2016
Revised date: 18 March 2017
Accepted date: 27 April 2017

Please cite this article as: Mingjing Wang , Huiling Chen , Bo Yang , Xuehua Zhao , Lufeng Hu , ZhenNao Cai , Hui Huang , Changfei Tong , Towards an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.04.060

Towards an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses

Mingjing Wang[1], Huiling Chen[1,3]*, Bo Yang[2,3], Xuehua Zhao[4], Lufeng Hu[5], ZhenNao Cai[1], Hui Huang[1], Changfei Tong[1]

[1](College of Physics and Electronic Information Engineering, Wenzhou University, 325035, Wenzhou, China)
[2](College of Computer Science and Technology, Jilin University, Changchun 130012, China)
[3](Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China)
[4](School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China)
[5](Department of Pharmacy, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, China)

*Corresponding author at: Huiling Chen (chenhuiling.jlu@gmail.com)

Abstract

This study proposes a novel learning scheme for the kernel extreme learning machine (*KELM*) based on the chaotic moth-flame optimization (*CMFO*) strategy. In the proposed scheme, *CMFO* simultaneously performs parameter optimization and feature selection. The proposed methodology is rigorously compared to several other competitive *KELM* models that are based on the original moth-flame optimization, particle swarm optimization, and genetic algorithms. The comparison is made using the medical diagnosis problems of Parkinson's disease and breast cancer. And the proposed method has successfully been applied to practical medical diagnosis cases. The experimental results demonstrate that, compared to the alternative methods, the proposed method offers significantly better classification performance and also obtains a smaller feature subset. Promisingly, the proposed *CMFOFS-KELM*, can serve as an effective and efficient computer aided tool for medical diagnosis in the field of medical decision making.

Keywords: Kernel extreme learning machine; Parameter optimization; Feature selection; Improved moth-flame optimization; Medical diagnosis

1 Introduction

As a new learning algorithm for single hidden layer feed-forward neural networks, the extreme learning machine (*ELM*) [1] has the ability to learn rapidly, has very few tuning parameters, and does not require the selection of input weights and hidden biases. Since its introduction, it has been used to address a variety of practical problems including cancer diagnosis [2], paraquat poisoning diagnosis [3], prediction of overweight status [4], face recognition [5], image quality

assessment [6], image classification [7], excavation equipment recognition [8], land cover classification [9], and hyperspectral image classification [10]. Despite its many good properties, *ELM* may not be stable or sensitive in some cases because of the random choice of the input weights and biases. In order to overcome this drawback, Huang et al. recently [11] expanded on the original *ELM*, and proposed the theory of kernel extreme learning machine (*KELM*), which does not require randomness in assigning connection weights between input and hidden layers. *KELM* can perform comparatively, or better, than popular machine learning paradigms like the support vector machine (*SVM*), in many scenarios like hyperspectral remote-sensing image classification [12, 13], disease diagnosis [14-17], bankruptcy prediction [18], activity recognition [19], 2-D profiles reconstruction [20], recognition of foreign fibers in cotton [21], and fault diagnosis [22].

However, our previous works show that when *KELM* is applied to practical applications like disease diagnosis [23] and bankruptcy prediction [24], its performance is heavily influenced by the two key parameters of kernel bandwidth γ and penalty parameter *C*. Parameter γ refers to the non-linear mapping from the input space to a high-dimensional feature space and *C* controls the trade-off between the model complexity and the fitting error minimization. Meta-heuristic algorithms have been proposed to tackle the issue of *KELM* parameter setting. These algorithms include particle swarm optimization (*PSO*) [25] and genetic algorithm (*GA*) [20], among others. It is also important to note that there are always many irrelevant or redundant features in the data as described in [26], which can jeopardize the classifier used in *KELM*. Therefore, it is essential to first select features prior to constructing the classifier.

In this study, we attempt to simultaneously address both the issues of parameter setting and feature selection for *KELM*. In order to achieve this goal, we use one of the latest meta-heuristic algorithms, moth-flame optimization [27] (*MFO*), to construct a new learning scheme for *KELM*. To the best of our knowledge, this study is the first to investigate the ability of *MFO* to train the *KELM* model for classification. *MFO* is based on the behavior of moths, which seek out flames in the night through a navigation mechanism called transverse orientation. *MFO* has a good balance between exploitation and exploration, because it updates its position through a special transverse orientation method. Due to its excellent characteristics, *MFO* has attracted the attention of researchers a lot recently. Allam et al. [28] has successfully applied *MFO* to extract the parameters of the three diode model for the multi-crystalline solar cell/module. The experimental results showed that *MFO* algorithm was obviously superior to other algorithms for the parameter extraction process of the three tested models. Emary et al. [29] has proposed an improved moth-flame optimization (*IMFO*) under controlling of exploration/exploitation rates in the domain of machine learning for feature selection. The improvement was that exploration/exploitation rate of *MFO* was controlled by chaotic operation. A comparative study exhibited that the proposed IMFO could select the features commendably. Li et al. [30] has also put forward an improved version of *MFO* based on Lévy-flight (*LMFO*) and has successfully solved the problem of engineering optimization. Recently, Li et al. [31] constructed an efficient and novel least squares support vector machine based on moth-flame optimization (*MFOLSSVM*) for annual power load forecasting. The forecasting results of China's annual electricity consumption revealed that the proposed *MFOLSSVM* model shows much better forecasting performance than other models optimized by other meta-heuristic algorithms. Sayed et al. [32] has used *MFO* to tackle segmentation biomedical imaging problem for thermogram breast cancer detection. And the

proposed scheme has classified breast cancer images into normal or non-normal well. Zhang et al. [33] proposed a moth-firefly optimization based on the logarithmic spiral search capability of *MFO* to increase local exploitation of the fireflies for the intelligent facial emotion recognition. After being evaluated on the various images, the proposed method significantly outperformed other state-of-the-art feature optimization methods and related facial expression recognition models. From the above MFO-related works, the effectiveness of *MFO* for parameters optimization for machine learning algorithms has been proven. However, it has to be admitted that the property of the *MFO* has much room to improve as described by Li et al. [30]. Therefore, Chaos theory is introduced in order to exploit MFO's potential and enhance its performance in this study, Chaos refers to a chaotic dynamic system that is non-repetitive and has ergodicity [34]. It has been confirmed that chaos can search at a high speed with good distribution [35]. Due to these characteristics, chaos has garnered much attention in recent years [36, 37].

In this study, two new chaotic strategies, chaos population initialization and the chaotic disturbance mechanism, are both introduced into the original *MFO*. The resultant *CMFO* strategy is used to identify the two key parameters and the optimal feature subset for KELM. The resultant *KELM* model, *CMFOFS-KELM*, is rigorously validated in terms of classification accuracy, on medical diagnosis problems via cross validation analysis, area under the ROC cure (AUC), sensitivity, and specificity. In order to illustrate the effectiveness of the proposed method, meta-heuristic-based *KELM* methods including *MFO*, *PSO*, *GA* and other three competitive approaches including original *ELM, KELM*, and logit model are investigated for comparison with *CMFOFS-KELM*. Further, a statistical t-test is used to evaluate the significance of the proposed method over other methods. The proposed method has also been successfully applied to diagnosis of paraquat-poisoned patients and overweight statuses in the practical application. As the experimental results show, the proposed learning scheme can enable a more stable *KELM* model with higher predictive accuracy compared to others. The main contributions of this study:

a) First, in order to further balance exploration and exploitation for *MFO*, we introduce two chaotic mechanisms (chaotic population initialization and chaotic disturbance) that are non-repetitive, and have the properties of ergodicity and regularity. This makes *MFO* have a faster convergence rate and also avoids it getting trapped into the local optimum.
b) The improved *MFO* strategy based chaos is successful in simultaneously tackling the optimization parameters and feature selection issues for *KELM*. The resulting model, CMFOFS-KELM, is not only verified on the two medical datasets strictly, but also successfully applied to diagnose of paraquat-poisoned patients and predict overweight statuses.
c) The proposed *CMFOFS-KELM* manages to achieve significantly better classification performance, and offers more stable and robust results when compared to several other meta-heuristic-based *KELM* models.

The remainder of this paper is structured as follows: Section 2 contains a brief description of the background of *KELM*, *MFO*, and chaos strategy. Section 3 explains the proposed methodology in detail. The experimental design is presented in Section 4. Section 5 presents the experimental results. The application of the proposed methodology is presented in Section 6. A discussion is given in Section 7. Section 8 provides the conclusion.

2 Background

2.1 Kernel extreme learning machine

KELM is one of the most popular learning techniques. It is derived from ELM; however, it is better at making generalizations, and has extensive practical applications. A brief de scription of the KELM method is given as follows：

A training set with $n\times1$ input feature vector of x_i and $m\times1$ of t_i target vector is given as $A=\{(x_i,t_i)\,|\,x_i\in R^n,t_i\in R^m,k=1,2,...,N\}$. The conventional SLFNs with $\tilde{N}$ hidden neurons and an activation function $h(\mathrm{x})$ can be modeled as follows:

$$\sum_{i=1}^{\tilde{N}}\beta_i h(w_i\cdot x_j+b_i)=o_j,\ j=1,2,...,N \tag{1}$$

The w_i is the vector of weight between the *i*th neuron in the hidden layer and the input layer, the bias of the *i*th neuron in the hidden layer is termed b_i, the vector of the weight between the *i*th hidden neuron and the output layer is β_i; o_j is the *j*th input data's target vector. $w_i\cdot x_j$ is the result of the inner product of w_i and x_j.

In order to accurately evaluate these samples, $\sum_{j=1}^{N}\|o_j-t_j\|=0,$ is given here. There are $\beta_i,w_i,$ b_i such that $\sum_{i=1}^{\tilde{N}}\beta_i h(w_i\cdot x_j+b_i)=t_j,\ j=1,2,...N$. The equation can been described as follows:

$$H\beta=T \tag{2}$$

$$\text{Where } H(w_1,...,w_{\tilde{N}},b_1,...,b_{\tilde{N}},x_1,...x_N)=\begin{pmatrix} h(w\cdot x_1+b_1) & \cdots & h(w_{\tilde{N}}\cdot x_1+b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ h(w\cdot x_N+b_1) & \cdots & h(w_{\tilde{N}}\cdot x_N+b_{\tilde{N}}) \end{pmatrix}_{N\times\tilde{N}} \tag{3}$$

$$\beta=\begin{bmatrix}\beta_1^T\\ \vdots\\ \beta_{\tilde{N}}^T\end{bmatrix}_{\tilde{N}\times m}\ \text{and}\ T=\begin{bmatrix}t_1^T\\ \vdots\\ t_N^T\end{bmatrix}_{N\times m} \tag{4}$$

H is the output matrix of the neural network of the hidden layer, with the *i*th column of *H* being

the *i*th hidden neuron output with respect to $x_1, x_2, ..., x_N$. Huang et al. [1] have suggested that input weights and the hidden layer bias vector do not need to be tuned. Based on this assumption, the output weights can be mathematically offered by the least square solution $\hat{\beta}$ of the linear system $H\beta = T$:

$$\| H(w_1,...,w_{\tilde{N}},b_1,...,b_{\tilde{N}})\hat{\beta} - T \| = min_{\beta} \| H(w_1,...,w_{\tilde{N}},b_1,...,b_{\tilde{N}})\beta - T \| \quad (5)$$

According to the Moor-Penrose (MP) generalized inverse, the MP generalized inverse of matrix *H* can be given as follows:

$$H\beta = T \Rightarrow \hat{\beta} = H^{\dagger}T \quad (6)$$

$H^{\dagger}$ is the generalized inverse of the matrix H from the above equation. In order to obtain the minimum normal least-squares (LS) solution the MP generalized inverse mechanism is used, which can keep the minimum criterion in all of the LS solutions. When the user does not know the feature mapping, a kernel matrix, which is called the kernel mapping function for the ELM, can be used through the following equation:

$$\Omega_{ELM} = HH^T : \Omega_{ELM_{i,j}} = h(x_i) \cdot h(x_j) = K(x_i, x_j) \quad (7)$$

The data can be mapped from the input space to ensure that it is linearly separable in the hidden-layer feature space H , using $h(x)$. An orthogonal projection approach, $H^{\dagger} = H^T (HH^T)^{-1}$, is adopted to calculate the MP generalized matrix's inverse, and a positive constant C is proposed to diagnose HH^T. In general, the output function of *ELM* is given as follows:

$$F(\mathrm{x}) = \mathrm{h}\,\beta = h(\mathrm{x})\,\mathrm{H}^{\dagger}(\frac{1}{C} + \mathrm{HH}^{\dagger})^{-1}T = \begin{bmatrix} K(\mathrm{x}, \mathrm{x}_1) \\ \vdots \\ K(\mathrm{x}, \mathrm{x}_N) \end{bmatrix}^T \left(\frac{1}{C} + \Omega_{ELM}\right)^{-1} T \quad (8)$$

In many practical applications, the radial basis function kernel (RBF) is adopted. RBF kernel function is always defined as $K(x, x_i) = \exp(-\gamma \| x - x_i \|^2)$. Based on the above descriptions, the two key parameters are the penalty parameter *C* and kernel parameter gamma γ. The first parameter *C* can fulfill a better trade-off between the fitting error minimization and the model complexity. The second parameter, γ, defines the non-linear mapping from the input space to a high-dimensional feature space.

2.2 Moth-Flame Optimization (MFO)

The MFO is one of the newest swarm intelligence optimization techniques [27]. The MFO

algorithm vividly simulates the behavior of moths, which at nighttime navigate around flames using a mechanism called transverse orientation. The MFO makes an improved trade-off between exploration and exploitation of the search space using a specific flame which is assigned to each moth. In this study, we adopted multi-dimensional spaces in order to identify the penalty parameter C and kernel bandwidth γ and feature subset size in KELM. In order to model the spiral of the moths, a matrix is used to represent a set of n moths:

$$M = \begin{bmatrix} m_{1,1} & \cdots & m_{1,d} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,d} \end{bmatrix} \tag{9}$$

where n is the number of moths and d is the number of dimensions. For all of the moths, an assumption is made that there is an array for storing the corresponding fitness values as in Eq. (10).

$$OM = \begin{bmatrix} OM_1 \\ OM_2 \\ \vdots \\ OM_n \end{bmatrix} \tag{10}$$

In this proposed algorithm, the other key components are flames. A similar matrix to the moths' matrix is created:

$$F = \begin{bmatrix} f_{1,1} & \cdots & f_{1,d} \\ \vdots & \ddots & \vdots \\ f_{n,1} & \cdots & f_{n,d} \end{bmatrix} \tag{11}$$

Assume that there exists an array for sorting the flames based on the value of the objective function:

$$OF = \begin{bmatrix} OF_1 \\ OF_2 \\ \vdots \\ OF_n \end{bmatrix} \tag{12}$$

The search agents move around the search area, while flames are the optimal position of moths.

The general framework of the MFO algorithm contains three tuple of approximation function that is described as:

$$MFO = (I, P, T) \tag{13}$$

As can be seen from Eq. (13), I is a function that generates a random population of moths and their corresponding fitness values. The function I is shown as follows:

$$I : \varphi \rightarrow \{M, OM\} \tag{14}$$

The P function is an essential component in the MFO algorithm that moves the moths around the search space. A matrix of M is received and a new M is returned by this function.

$$P : M \rightarrow M \tag{15}$$

The *T* function returns an answer of either true or false depending on whether the stop criterion is satisfied or not. The function returns true if the termination criterion is satisfied and false if the termination criterion is not satisfied. It can be defined as follows:

$$T : M \rightarrow \{true, false\} \tag{16}$$

In the *P* function, the position of each moth is updated with respect to the corresponding flame using Eqs. (17-19):

$$M_i = S(M_i, F_j) \tag{17}$$

$$S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j \tag{18}$$

$$D_i = | F_j - M_i | \tag{19}$$

where *S* is the spiral function, M_i refers to l^{th} moth and F_j indicates j^{th} flame. Where D_i means the distance between the l^{th} moth and j^{th} flame, *b* is a constant in order to define the spiral function, and *t* is derived randomly between -1 and 1.

Eq. (18) is the mathematical model that simulates the spiral-flying path of moths. The spiral equation is the key component of the MFO algorithm because it describes how the moths update their positions around the flames but not necessarily in the space between them. Therefore, the trade-off between exploration and exploitation of the search space in this algorithm can be reasonably guaranteed. The general framework of MFO is as follows:

Moth-flame optimization

Begin

Step 1: Parameter Initialization.

Initialize the generation counter G, the population size P, the boundary of search, and the dimensionality of the space.

Step 2: Population Initialization Randomly.

Set the initialization parameters to initialize the positions randomly.

Step 3: Calculate the fitness of each moth.

Calculate the fitness of each moth and save them as a vector.

Step 4: Update the position of each moth and flame.

Update the position and flame according the corresponding mechanism;

Step 5: Control the positions in the search boundary.

Control the position in the search boundary range in the dimensionality of the space.

Step 6: Iterative optimization.

Enter to the iterative optimization to repeat steps 3-5.When the iterative number reaches the maximal iterative number, the process of optimization stops.

End

2.3 Chaotic operation for population initiation

As we all know, most of the traditional heuristic algorithms work under an assumption of uniform distribution, and generate initiation positions randomly. This may lead to each agent being far away from the optimum point and could increase uncertainty. Therefore, the *MFO* properties might be influenced, to a certain extent, by the initial position of this algorithm. Having good initial population positions ensures fast convergence and a good final solution for swarm intelligence optimization algorithms [38, 39]. Chaos is an unstable dynamic behavior that is sensitive and dependent on initial conditions that have ergodicity and stochasticity. Logistic maps are one of the most commonly used chaotic sequences. Therefore, this chaos map has also been chosen in this study in order to initialize the population in *MFO* (*CMFO*). Logistic maps are shown in Eq. (20). Where μ is the parameter that controls the degree of chaos, there is a completely chaotic state when μ has a value of 4 as shown in Fig. 1.

$$X_{n+1} = \mu \times X_n \times (1 - X_n) \quad \mu \in [0,4] \, , X_n \in (0,1) \tag{20}$$

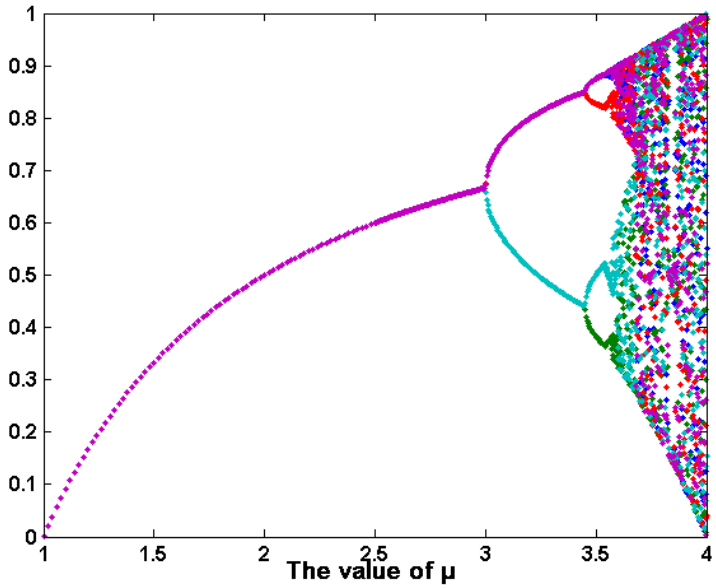


Fig. 1. The bifurcation diagram of the logistic model with different values of μ

Since chaotic sequences have been successfully applied in a wide span of practical problem domains, it is reasonable to believe that chaotic sequences have good potential for optimizing initialization. The initialization steps for the *MFO* are as follows:

Step 1: Randomly *set the position of the first moth;*

Step 2: *Adopt the Logistic Maps to map the remaining moth positions, taking the first position as the initial value;*

Step 3: *Calculate the fitness of each moth;*

Step 4: *Sort the moths in terms of the fitness;*

Step 5: *Save the predetermined number of moths in front of the sorted sequence as the initial population;*

Step 6: *End initialization.*

2.4 Chaotic disturbance mechanism

MFO is not able to avoid falling into the local optimum on some practical optimization problems, although it is able to maintain an excellent balance between exploitation and exploration as described by [27]. In order to bridge this gap, a chaotic disturbance mechanism was proposed to ensure that the local optimum did not appear in this study. In this mechanism, we adopted the same chaotic mapping used in the population initiation procedure. The disturbance mechanism can be described as follows:

$$m'_{best} = m_{best} + \lambda \cdot chaos \tag{21}$$

where m_{best} is the position vector of the current optimal solution obtained thus far, λ decreases from 1 to 0 with the increase of iterations in order to control the degree of disturbance, and λ can be adaptively obtained. In the early steps, the value of λ must be much bigger so that the agent can fly in a wider space in the global search. In the later steps, it should be smaller to allow for the local search. This improves the balance between exploration and exploitation during the disturbance. The parameter *Chaos* indicates the chaos value. As long as the current optimal solution is not improved after *k* iterations (with *k* set as 5, determined by trial and error), the position of the current optimal agent can be updated according to Eq. (21).

3 Proposed CMFOFS-KELM methodology

This study proposed a new learning strategy that introduces chaos theory into *MFO*, which we named *CMFO*, in order to identify the key parameters and discriminative features at the same time as obtaining an optimal *KELM* model. The flowchart of the proposed *CMFOFS-KELM* is shown in Fig. 2. The proposed methodology consists of two main parts. The first part involves optimizing inner parameters and selecting the sub-features. The second part evaluates the outer classification performance. During the selection of the inner parameters and features, the optimal parameters and representative features for the training set are dynamically adjusted by the *CMFO* strategy via the 5-fold cross validation (CV) analysis. Then, the obtained optimal parameters and feature subset are fed into the *KELM* model to conduct the classification task for medical diagnosis in the outer loop via 10-fold CV analysis.

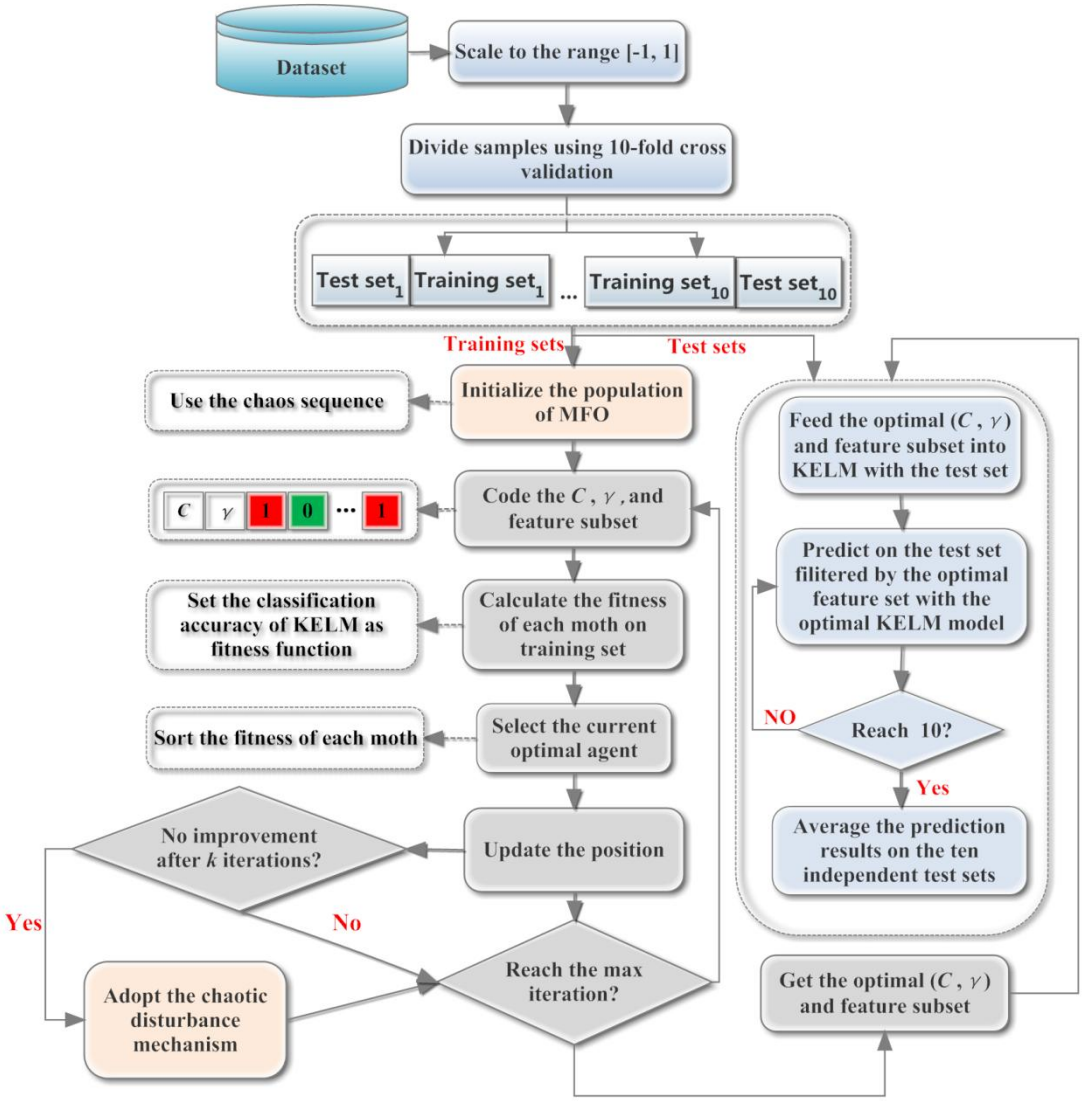


Fig. 2. Flowchart of the proposed CMFOFS-KELM

It is assumed that the features will be selected from a total of *n*-2 features. Detailed description of the coding of each agent in *MFO* is given in Fig. 3. The first two dimensions of each agent are coded as the two key parameters *C* and γ. The coding strategy of the features is as follows: each feature is set as a value x_{ith} in the range [0, 1]. Namely, if the value of the function Logsig (x_{ith}) (as in Eq. (22)) is less than or equal to 0.5, then the corresponding feature is not selected and is set as 0. Conversely, if the value of the function Logsig (x_{ith}) is greater than 0.5, it means that the feature is chosen and set as 1.

$$Logsig(x_{ith}) = 1/(1+\exp(-x_{ith})) \tag{22}$$

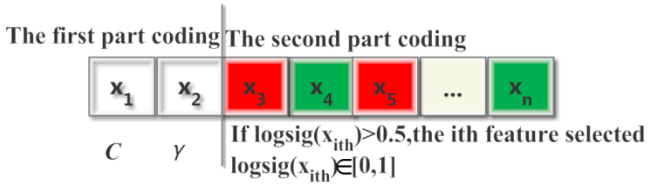


Fig. 3. CMFO coding for parameters and feature selection

The classification accuracy is set as the fitness function:

$$fitness = averageACC = \frac{\sum_{i=1}^{k} testACC_i}{k} \tag{23}$$

where *averageACC* in the above function refers to the average test classification accuracy obtained by the *KELM* classifier via the 5-fold CV. The process of selecting the parameters and features is shown below:

The steps of parameters optimization and feature selection

Begin

Step 1: Parameter Initialization. *Initialize the generation counter G, the population size P, the search boundary and the dimensionality of the space.*

Step 2: Population Initialization. *Set the initialization parameters to initialize the positions with the chaos sequence and then code the key parameters C, γ, and the features as shown in figure 5. Calculate the fitness of each agent and keep the P numbers for the best moth as the initialization population.*

*k=0; /*Control the disturbance mechanism as a counter.*/*

*Best_Moth_fitness=0; /*Set the current best fitness of moth as 0.*/*

Step 3: **While** *G<MaxGeneration* **do**

Code the key parameters C, γ, and the features **.**

Calculation Fitness.

for *i*=1:*P*

*/*Calculate the fitness of each moth with the C, γ and the selected sub-features.*/*

Moth_fitness(*i*)=*averageACC*$_i$

End

if *G* = 1

*[fitness_sorted, I] = sort (Moth_fitness , 'descend'); /*Sort the first population of moths;*/*

sorted_population = Moth_pos (I, :);

*best_flames = sorted_population; /*Update the flames;*/*

best_fitness_fitness = fitness_sorted;

else

Sort the moths and update the flames and save the Moth_fitness.

if *Best_Moth_fitness< fitness_sorted(1);*

Best_Moth_fitness= fitness_sorted(1);

k=0;

else

k=k+1;

end

if *k=5 /*Adopt the chaotic disturbance mechanism to update the corresponding position;*/*

update the position of the current best agent using Eq. (21)

end

for *i = 1 : size* (*positions*, 1)

Update the position of MFO according the corresponding mechanism;

end

Control the position in the search boundary.

G = G +1;

Step 4: End while

Step 5: *Save the position of the final best moth and then the first two elements are the BestC, Best ɣ; Decode the remaining moths to obtain the best feature subset.*

Return *BestC, Best ɣ, and Best feature subset;*

End

The computational complexity of the proposed method *MFO-KELM* depends on the number of samples (*L*), the number of generation (*g*), the population number (*n*), and the parameters dimensions (*d*). Therefore the overall computational complexity is *O* (*KELM*, *MFO*) = *O* (Initialization) + *g** (*O* (*KELM*) + (*MFO* position updating)) + *O* (*MFO* sort mechanism of flames) + O (Output). As we all know, KELM's computational complexity on *L* samples is *O* (L^3). The computational complexity of initialization is *O* (*n***d**O (L^3)) + *O* (*nlogn*), which is composed of the fitness calculation and sorting mechanism of flames. Therefore, the final computational complexity of the proposed method is *O* (*KELM*, *MFO*) = *O* (*n** *d* **O* (L^3)) + *O* (*nlogn*) + *g**(*O* (*n***d* **O* (L^3))) + *O* (*n***d*) + *O* (*nlogn*) + *O* (1)) ≈*O* (*n***d***O* (L^3)) + *g**(*O* (*n***d* **O* (L^3))).

4 Experimental studies

4.1 Data description

In order to evaluate the effectiveness of the proposed *CMFOFS-KELM* methodology, we take two typical disease diagnosis problems, namely, Parkinson's disease and breast cancer, as examples. Parkinson's disease is a degenerative central nervous system disorder which can cause partial or complete loss of speech, behavior and mental control, and other serious physiological and psychological functions. Breast cancer is one of the most common and deadly cancers in women. The early detection and accurate diagnosis of these diseases is vital so that effective intervention measures can be implemented in the early disease stages. The medical datasets from the UCI machine learning repository were adopted in this study, including the Oxford Parkinson dataset and the Wisconsin breast cancer dataset. The Oxford Parkinson dataset is composed of a range of biomedical voice measurements from 31 people, 23 of whom have Parkinson's disease (*PD*). The features of the Wisconsin breast cancer dataset are computed from a digitized image of a fine needle aspiration of a breast mass. They describe the characteristics of the cell nuclei that are present in the image. The Wisconsin dataset has some missing values.

For further verifying the effective and efficient of the proposed method, it also has been applied to diagnose the paraquat-poisoned patients and predict overweight statuses. These two datasets are both collected from the First Affiliated Hospital of Wenzhou Medical University. For diagnosis of paraquat-poisoned patients, there are 31 participants taken in this study and the blood of each sample is analyzed by gas chromatography coupled with mass spectrometry (*GC-MS*). For diagnosis of overweight statuses, 476 participants were introduced in terms of 18 blood indexes,

16 biochemical indexes, and other three conventional indexes.

4.2 Experimental setup

To avoid difficulties with numerical calculation and the domination of greater numerical ranges over smaller numerical ranges, the Oxford Parkinson's dataset and Wisconsin breast cancer were both scaled to the range [-1,1] before constructing the *KELM* classifiers. In order to make the results valid, the classic 10-flod CV strategy was adopted to appraise the property of the classifiers. However, only one single repetition of the 10-fold CV cannot guarantee the unbiased classification results, because of the influence of randomness. The reported final results were calculated by averaging the results of 10 runs of the 10-fold CV.

To evaluate the performance of the proposed methodology, the following methods were implemented, and a rigorously compared: the original *MFO* based *KELM* with feature selection (*MFOFS-KELM*), *PSO* based *KELM* with feature selection (*PSOFS-KELM*), *GA* based *KELM* with feature selection (*GAFS-KELM*), and the original *MFO* based *KELM* without feature selection (*MFO-KELM*), *PSO* based *KELM* without feature selection (*PSO-KELM*), and *GA* based *KELM* without feature selection (*GA-KELM*). All the methods were implemented in the MATLAB platform on a Windows 7 system with Intel (R) Core (TM) i7-4790 CPU @3.60GHz with 8GB of RAM. The same number of generations and the same population swarm size were set for all of the methods. The number of generations and swarm size were set as 100 and 50 respectively. Detailed descriptions of the parameter settings can be found in Table 1. This table shows that the two key values C and γ fell within the ranges of $C \in \{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$ and $\gamma \in \{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$, respectively. For *GA*, the crossover and mutation rates were set as 0.95 and 0.05, respectively. For *PSO*, the maximum velocity of the *PSO* was approximately 65% of the dynamic ranges of the variable on each dimension, with two acceleration coefficients of approximately 2.05 and an inertial weight of 1. For the original *MFO* and *CMFO*, the parameter *b* was set 1 for construing the spiral function.

Table 1 Parameter setting used in GA, PSO, MFO, and CMFO

Parameters	**GA**	**PSO**	**MFO**	**CMFO**
Population size	50	50	50	50
Max iteration	100	100	100	100
Search space	$\{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$	$\{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$	$\{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$	$\{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$
Crossover rate	0.95	-	-	-
Mutation rate	0.05	-	-	-
c_1 and c_2	-	2.05	-	-
Inertia weight	-	1	-	-
Maximum velocity	-	65%	-	-
μ				4
b (spiral function)	-	-	1	1

For further verifying the proposed method, other three common methods including KELM, original ELM, and Logit were also introduced in this study. For KELM, a grid-search technique was adopted to search the optimal parameters values of the RBF kernel. The range of the two key parameters also varied in the ranges of $C \in \{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$ and $\gamma \in \{2^{-8}, 2^{-6}, 2^{-4}...,2^{8}\}$. For original *ELM*, the activation functions such as sigmoid function, sine function, hard-limit function,

triangular basis function and radial basis function were used in turn to choose the suitable function. The hidden neuron was varied from 1-250 to obtain the best optimal hidden neuron. According to repeated attempts, the sigmoid activation function with 50 was the best choose to construct the *ELM*. For Logit model, the program from WEAK tool was called by the main program in MATLAB platform.

4.3 Performance measures

The properties of the proposed method were evaluated using classification accuracy (*ACC*), the area under the receiver operating characteristic curve (*AUC*), sensitivity, and specificity. The formulations of these metrics are defined as follows:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{24}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{25}$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \tag{26}$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives, and FP is the number of false positives. *AUC* is the area under the ROC cure, which is one of the most commonly used methods for evaluating the binary classifiers. The value of *AUC* is between 0 and 1. The classifier has a higher value of *AUC*, which can be treated as having greater properties.

5 Experimental results

5.1 Parkinson's disease diagnosis

The Parkinson's disease dataset has 195 biomedical sound measurements and 22 features. This dataset contains the mean, maximum and minimum sound fundamental frequency, amplitude irregularity measurements, measurements of the harmonics and the noise ratio, nonlinear fundamental frequency change measurements, nonlinear dynamic complexity measurements, and fractional exponent signal. Detailed descriptions of this dataset can be found in Table 2 in terms of the statistical minimum, maximum, and average value of each feature.

Table 2 Descriptions of the Parkinson dataset in terms of feature definitions and statistical values

Label	Attribute	Description	Minimum Value	Maximum value	Average value
1	MDVP:Fo (Hz)	Average vocal fundamental frequency	88.33	260.105	154.22
2	MDVP:Fhi (Hz)	High vocal fundamental frequency	102.14	592.03	197.10
3	MDVP:Flo (Hz)	Low vocal fundamental frequency	65.476	239.17	116.32
4	MDVP:Jitter (%)	Jitter percent	0.00168	0.03316	0.00622

5	MDVP:Jitter (Abs)	Absolute jitter	7 9 *10-6	0.00026	4.39 *10-5
6	MDVP:RAP	Relative average perturbation	0.00068	0.02144	0.0033
7	MDVP:PPQ	Period perturbation quotient	0.00092	0.01958	0.0034
8	Jitter:DDP	Difference of differences of periods	0.00204	0.06433	0.0099
9	MDVP:Shimmer	Shimmer percent	0.00954	0.11908	0.0297
10	MDVP:Shimmer (dB)	Shimmer in dB	0.085	1.302	0.2822
11	Shimmer:APQ3	Amplitude perturbation quotient	0.0045	0.0564	0.0156
12	Shimmer:APQ5	Quotient of amplitude perturbation in 3-point	0.0057	0.0794	0.0178
13	MDVP:APQ	Quotient of amplitude perturbation in 5-point	0.00719	0.1377	0.0240
14	Shimmer:DDA	Mean absolute difference between consecutive amplitude differences of consecutive periods	0.01364	0.1694	0.0469
15	NHR	Noise-to-harmonics ratio	0.00065	0.3148	0.0248
16	HNR	Harmonics-to-noise ratio	8.441	33.047	21.885
17	RPDE	Recurrence period density entropy	0.2565	0.6851	0.49853
18	D2	Correlation dimension	0.57428	0.825	0.7180
19	DFA	De-trended fluctuation analysis	-7.9649	-2.434	-5.684
20	Spread1	Quantifications the fundamental	0.00627	0.4504	0.2265
21	Spread2	Frequency in variation	1.423	3.6711	2.3818
22	PPE	Pitch period entropy	0.04453	0.5273	0.2065

Table 3 shows the average *ACC*, *AUC*, sensitivity, and specificity results obtained by *CMFO-KELM*, *MFO-KELM*, *PSO-KELM*, and *GA-KELM* on the Parkinson dataset without feature selection with the two key parameters being continuously adjusted. *CMFO-KELM* had the highest property among the four methods with the smallest standard deviation and with the average results of: 96.78% *ACC*, 96.47% *AUC*, 97.51% sensitivity, and 95.43% specificity. *MFO-KELM* followed, producing slightly lower results, with an *ACC* of 96.42%, *AUC* of 91.56%, sensitivity of 94.80%, and specificity of 97.52%. Next was *PSO-KELM*, which yielded an *ACC* of 95.04%, *AUC* of 92.27%, sensitivity of 96.71%, and specificity of 88.49%. *GA-KELM* performed the worst among the four methods, producing an *ACC* of 94.09%, *AUC* of 91.22%, sensitivity of 96.50%, and specificity of 85.91%. The results indicate that the *CMFO-KELM* offers consistently higher classification accuracy and more stable results than the other three methods. A more detailed comparison of the 10-fold CV results of the 10 runs of the four models, from the *ACC* perspective, is exhibited in Fig. 4. This figure shows that across all 10 runs, *CMFO-KELM* has significant advantages over the other three methods. The reason that *CMFO-KELM* has such good results may be because the chaos theory has improved the global search ability of *MFO* by the use of the *KELM* model in this dataset.

Table 3 Detailed results of all the models on the Parkinson dataset without feature selection

Method	**Metrics**			
	ACC	**AUC**	**sensitivity**	**specificity**
CMFO-KELM	0.9687±0.0017	0.9647±0.0076	0.9751±0.0056	0.9543±0.0192
MFO-KELM	0.9642±0.0024	0.9480±0.0100	0.9752±0.0056	0.9196±0.0192
PSO-KELM	0.9504±0.0087	0.9247±0.0216	0.9671±0.0083	0.8842±0.0404

GA-KELM	0.9409±0.0110	0.9122±0.0271	0.9650±0.0116	0.8591±0.0489

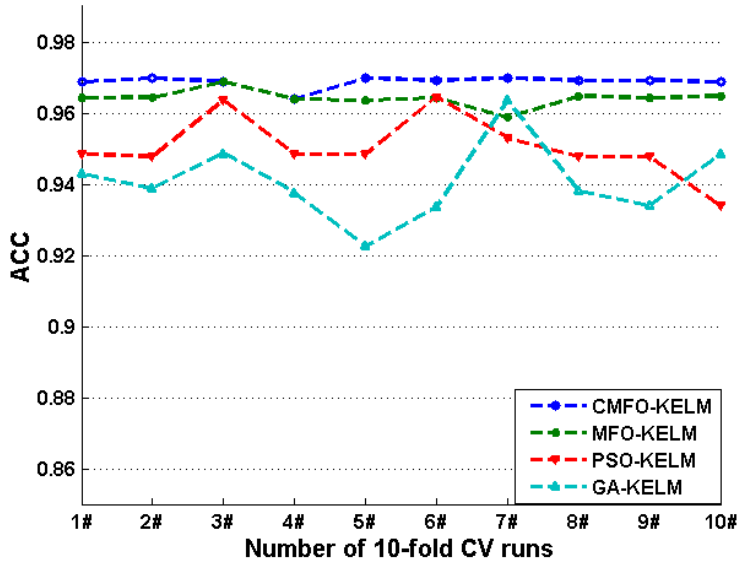


Fig. 4. The ACC for each run on the Parkinson dataset without feature selection

Table 4 illustrates the detailed results of the four methods, *CMFOFS-KELM*, *MFOFS-KELM*, *PSOFS-KELM*, and *GAFS-KELM* in terms of *ACC*, *AUC*, sensitivity, and specificity on the Parkinson dataset with feature selection. In this setting, the two key parameters of optimization and feature selection were conducted simultaneously. The table shows that *CMFOFS-KELM* has better experimental results than the other three methods in terms of the four performance metrics. In addition, *CMFOFS-KELM* had the smallest standard deviation among the four methods with an *ACC* of 97.45%, *AUC* of 97.37%, sensitivity of 97.63%, and specificity of 97.12%. *MFOFS-KELM* was next and yielded an *ACC* of 96.87%, *AUC* of 96.60%, sensitivity of 97.52%, and specificity of 95.67%. *PSOFS-KELM* followed, yielding an *ACC* of 95.04%, *AUC* of 92.27%, sensitivity of 96.71%, and specificity of 88.49%. *GAFS-KELM* produced the worst results with an *ACC* of 94.09%, *AUC* of 91.22%, sensitivity of 96.50%, and specificity of 85.91%. Fig. 5 shows a detailed comparison of the results from10 runs of 10-fold CV in terms of *ACC*. It can be seen from the figure that *CMFOFS-KELM* clearly performs better than the other three methods over all of the runs. When these results are combined with the results shown in Tables 3 and 4, it can be concluded that taking the optimization parameters together with the feature selection produces better results than optimization on a single parameter. This is true with regard to *ACC*, *AUC*, sensitivity, specificity, and also stability.

Table 4 Results of all models on the Parkinson dataset with feature selection

Method	**Metrics**			
	ACC	**AUC**	**Sensitivity**	**Specificity**
CMFOFS-KELM	0.9745±0.0002	0.9737±0.0045	0.9763±0.0043	0.9712±0.0126
MFOFS-KELM	0.9687±0.0014	0.9660±0.0048	0.9752±0.0066	0.9567±0.0161
PSOFS-KELM	0.9542±0.0067	0.9373±0.0162	0.9653±0.0086	0.9094±0.0340
GAFS-KELM	0.9455±0.0114	0.9321±0.0303	0.9569±0.0096	0.9072±0.0527

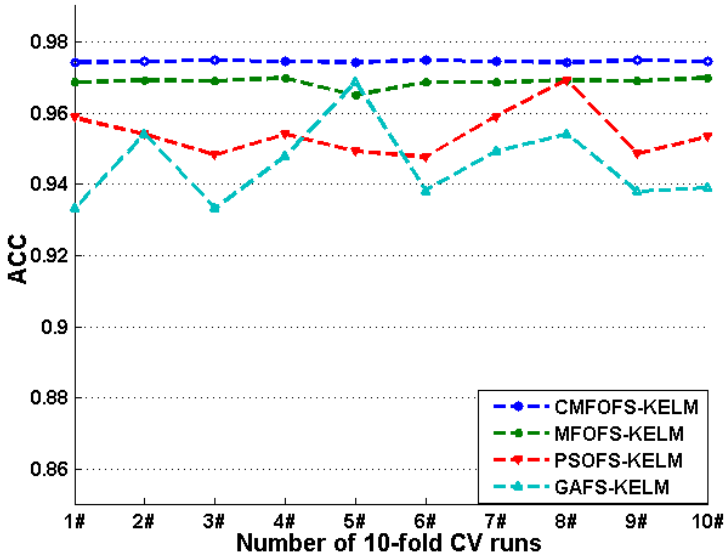


Fig.5. The ACC obtained for each run on the Parkinson dataset with feature selection

Fig. 6 shows the number of times that each feature was selected through 10 runs of 10-fold CV. This figure shows that the bars can be divided into two categories. The first category encompasses the features that are selected in large amounts. The second category contains the features which are selected in small amounts. Further, the more selected times was obtained by the *KELM* based on *CMFO* compared to others methods in the first category. However, *KELM* constructed by *CMFO* selected fewer times than others in the second category. Therefore, compared to the other methods, *KELM* based on *CMFO* has the best ability to distinguish between the pivotal features and the rest of the features. Moreover, a reasonable assumption can be derived that features selected over 80 times during the entire 100 folds are pivotal features. Therefore, the features MDVP:Fo, MDVP:Jitter, MDVP:RAP, Shimmer:APQ3, Shimmer:APQ5, RPDE, D2, DFA, and PPE are critical features in identifying patients with Parkinson disease. Doctors should give these features adequate attention when making diagnoses.

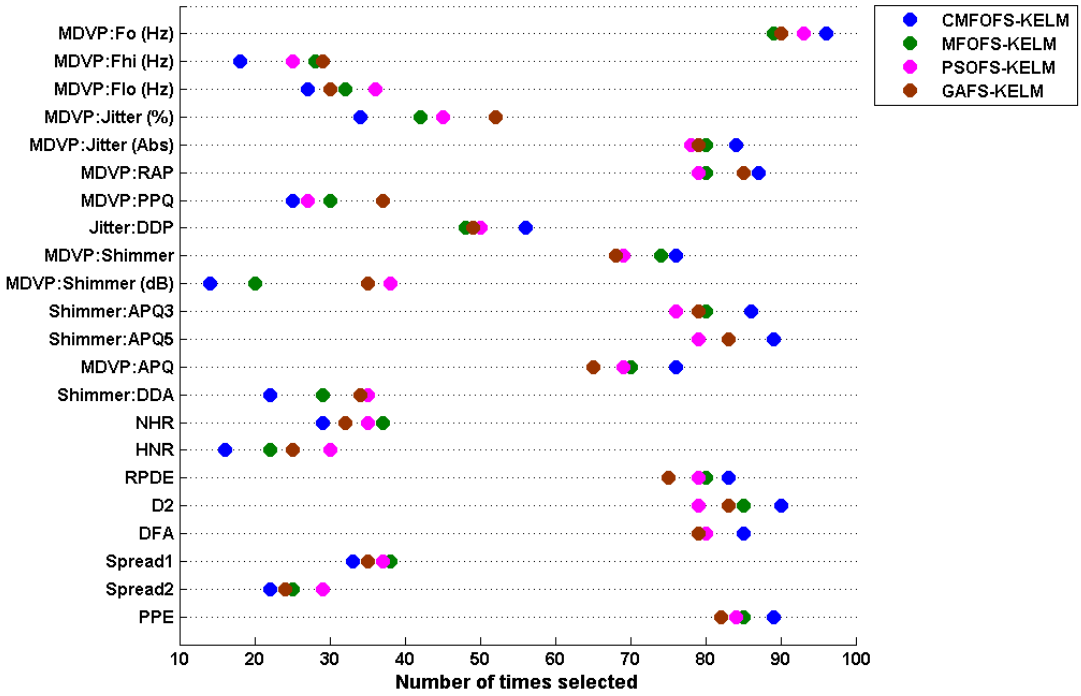


Fig.6. Numbers of selected features via 10 runs of 10-fold CV

Fig. 7 shows the evolutionary process of each method optimized by different strategies on the training dataset, in order to further compare the properties of the methods. As shown in this figure, the four fitness curves with feature selection gradually improved from iteration 1 to iteration 100. No obvious improvements were noted for *CMFOFS-KELM*, *MFOFS-KELM*, *PSOFS-KELM*, and *GAFS-KELM* after iterations 18, 23, 29, and 78, respectively. In contrast, *CMFO-KELM*, *MFO-KELM*, *PSO-KELM*, and *GA-KELM* do not improve at all after iterations 31, 37, 48, and 84, respectively. It can be observed that the convergence rate of each method is improved noticeably with the feature selection strategy. *CMFOFS-KELM* obtained the best convergence rate of all of the methods. In the enlarged portion of the graph it can be observed that the final training fitness plot of the *CMFO-KELM* is smoother than *MFOFS-KELM*, even though the convergence rate of *MFOFS-KELM* is faster than *CMFO-KELM*. This indicates that the chaotic strategy could greatly improve the convergence rate. A reasonable conclusion can be made that the property of the *KELM* based on the optimization parameters, in conjunction with feature selection using *MFO*, can be improve extensively when the chaotic strategy is introduced.

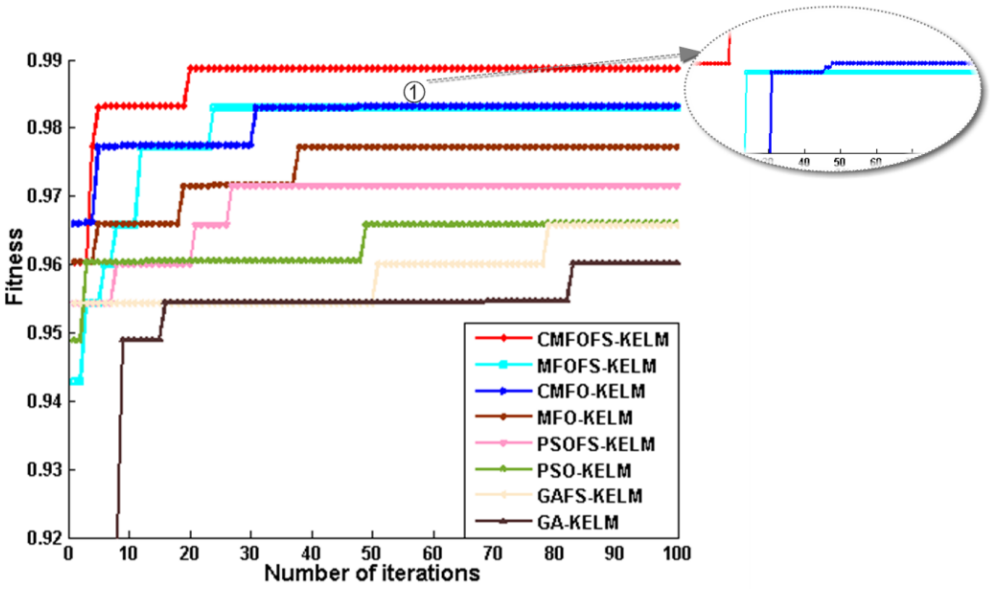


Fig. 7. The evolutionary processes of the methods on the Parkinson dataset

In order to reveal subtle differences in the classification performances of the methods, a paired t-test was carried out. A p-value of less than 0.05 indicates statistical significance in the experiment. Table 5 shows that *CMFOFS-KELM* has nearly-significantly better results than the other seven methods in terms of *ACC*, *AUC*, sensitivity, and specificity. In Table 5, significant values that are greater than 0.05 are shown in bold. Positive values mean that the *i*th classifier fulfills better properties than the *j*th one.

Table 5 Paired t-test results of *CMFOFS-KELM* and the comparison methods on the Parkinson dataset.

Method	**t-value (significance)**			
	ACC	**AUC**	**sensitivity**	**specificity**
MFOFS-KELM	14.201(0.001)	3.27(0.010)	5.717(0.002)	3.664(0.003)
PSOFS-KELM	9.403(0.000)	6.142(0.000)	3.261(0.010)	4.785(0.001)
GAFS-KELM	7.965(0.000)	4.491(0.002)	6.122(0.000)	3.909(0.004)
CMFO-KELM	10.481(0.000)	9.643(0.000)	7.881(0.000)	2.685(0.025)
MFO-KELM	13.905(0.000)	2.685(0.025)	2.731(0.0024)	**1.909(0.089)**
PSO-KELM	8.833(0.000)	6.931(0.000)	1.430(0.730)	6.248(0.000)
GA-KELM	9.643(0.000)	6.720(0.000)	2.271(0.022)	5.911(0.000)

5.2 Breast cancer disease diagnosis

The breast cancer dataset is composed of 699 samples, 65% of which are benign cases, and 35% of which are malignant cases. The nine features have values that vary between 1 and 10. Table 6 displays detailed descriptions of the dataset. The dataset has some missing values. In this study, the missing categorical attributes were replaced with the mode of the attributes, and the missing continuous values were replaced by the mean of the attributes.

Table 6 Descriptions of breast cancer dataset in terms of features definition and statistical value

Attribute	Description	Domain	Mean Value
1	Clump thickness	1–10	4.44
2	Uniformity of cell size	1–10	3.15
3	Uniformity of cell shape	1–10	3.22
4	Marginal adhesion	1–10	2.83
5	Single epithelial cell size	1–10	2.23
6	Bare nuclei	1–10	3.54
7	Bland chromatin	1–10	3.45
8	Normal nucleoli	1–10	2.87
9	Mitoses	1–10	1.60

Table 7 contains the experimental results of the various methods in terms of the average *ACC*, *AUC*, sensitivity, and specificity on the Breast cancer dataset without feature selection. The *CMFO-KELM* had the highest *ACC* among the four methods and also achieved the smallest standard deviation. *CMFO-KELM* yielded average results of 96.93% *ACC*, 96.87% *AUC*, 97.26% sensitivity, and 96.58% specificity. *MFO-KELM* yielded an *ACC* of 96.45%, *AUC* of 96.21%, sensitivity of 97.13%, and specificity of 94.39%. The *PSO-KELM* yielded an *ACC* of 95.38%, *AUC* of 94.88%, sensitivity of 96.26%, and specificity of 94.88.49%. The *GA-KELM* yielded an *ACC* of 95.09%, *AUC* of 94.70%, sensitivity of 95.89%, and specificity of 93.16%. *CMFO-KELM* consistently had a higher classification accuracy and also more stable results than the other three methods. A more detailed comparison of the 10 runs of the 10-fold CV results for the four methods through the perspective of the *ACC* is exhibited in Fig. 8. From this figure, the same phenomenon can be viewed as in the Oxford Parkinson's dataset; *CMFO-KELM* out-performed the other three *KELM* methods in most runs.

Table 7 Detailed results of all methods on the Breast cancer dataset without feature selection

Method	Metrics			
	ACC	AUC	sensitivity	specificity
CMFO-KELM	0.9693±0.0060	0.9687±0.0078	0.9726±0.0047	0.9658±0.0130
MFO-KELM	0.9645±0.0088	0.9621±0.0100	0.9713±0.0136	0.9439±0.0257
PSO-KELM	0.9538±0.0181	0.9488±0.0220	0.9626±0.0118	0.9488±0.0284
GA-KELM	0.9509±0.0159	0.9470±0.0256	0.9589±0.0249	0.9316±0.0631

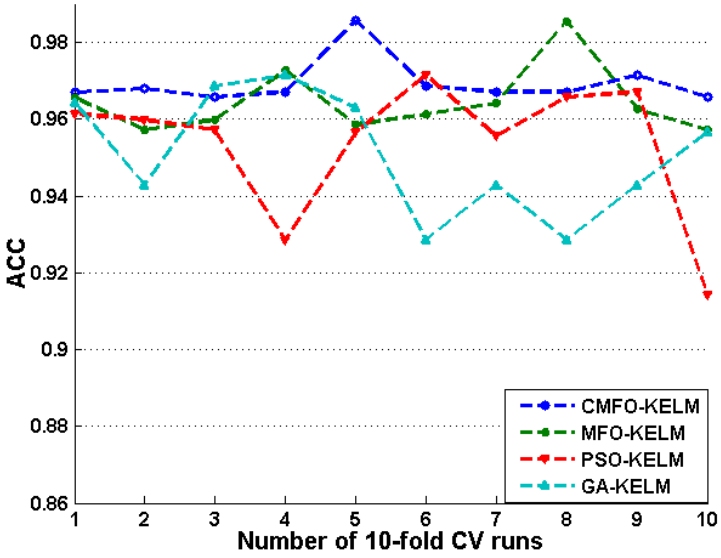


Fig. 8. The *ACC* obtained from each run on the Breast cancer dataset without feature selection

Table 8 illustrates the classification results, on the Breast cancer dataset, of the four methods constructed through parameter optimization and feature selection synchronously in terms of *ACC*, *AUC*, sensitivity, and specificity. This table shows that the performance of *CMFOFS-KELM* is advantageous to the other three competitors in terms of the *ACC*, *AUC*, sensitivity, and specificity. *CMFOFS-KELM* achieved the best performance with an *ACC* of 98.13%, *AUC* of 97.31%, sensitivity of 97.70%, and specificity of 96.92%. It also had the smallest standard deviation. The *MFOFS-KELM* yielded an *ACC* of 96.94%, *AUC* of 96.82%, sensitivity of 97.35%, and specificity of 96.16%. The *PSOFS-KELM* yielded an *ACC* of 96.22%, *AUC* of 96.02%, sensitivity of 96.93%, and specificity of 95.22%. The *GAFS-KELM* yielded the worst results with an *ACC* of 95.99%, *AUC* of 95.92%, sensitivity of 96.49%, and specificity of 94.99%. Fig. 9 shows the comparison results of 10 runs of the 10-fold CV, in terms of *ACC*, obtained by these four methods. The figure shows that *CMFOFS-KELM's* performance is again superior to that of the other three meta-heuristic-based methods. Using a strategy that synchronously optimizes the parameters and selects features can improve the performance of *KELM*.

Table 8 Detailed results of all the models on the Breast cancer dataset with feature selection

Method	**Metrics**			
	ACC	**AUC**	**Sensitivity**	**Specificity**
CMFOFS-KELM	0.9813±0.0038	0.9731±0.0092	0.9770±0.0031	0.9692±0.0129
MFOFS-KELM	0.9694±0.0089	0.9682±0.0104	0.9735±0.0098	0.9616±0.0150
PSOFS-KELM	0.9622±0.0108	0.9602±0.0137	0.9693±0.0060	0.9522±0.0281
GAFS-KELM	0.9599±0.0155	0.9592±0.0156	0.9649±0.0140	0.9499±0.0312

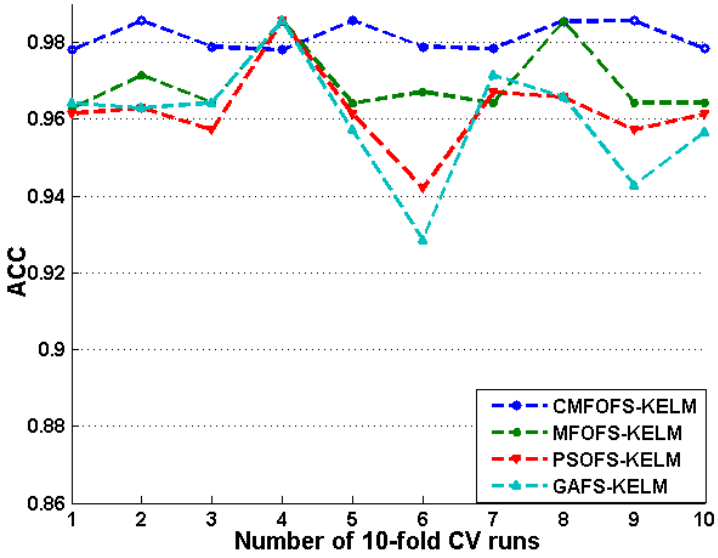


Fig. 9. The *ACC* obtained on each run on the Breast cancer dataset with feature selection

Fig. 10 shows the number of times, from a statistical point of view, that each feature was selected through 10 runs of 10-fold CV on the Breast cancer dataset. This figure shows that the features 1, 3, 6, 7, and 9 were selected numerous times. Among these features, *CMFOFS-KELM* always obtained the most selected times. The features 2, 4, 5, 8 were selected a limited number of times, and used in this dataset, the *KELM* based on the *CMFO* strategy has the strongest ability to distinguish the pivotal features from the whole features. Therefore, the features of Clump thickness, Uniformity of cell shape, Bare nuclei, Bland chromatin, and Mitoses, can be treated as vital features for distinguishing malignant tumors from benign ones. This also implies that medical personnel should pay more attention to the valuable information given by these 5 features.

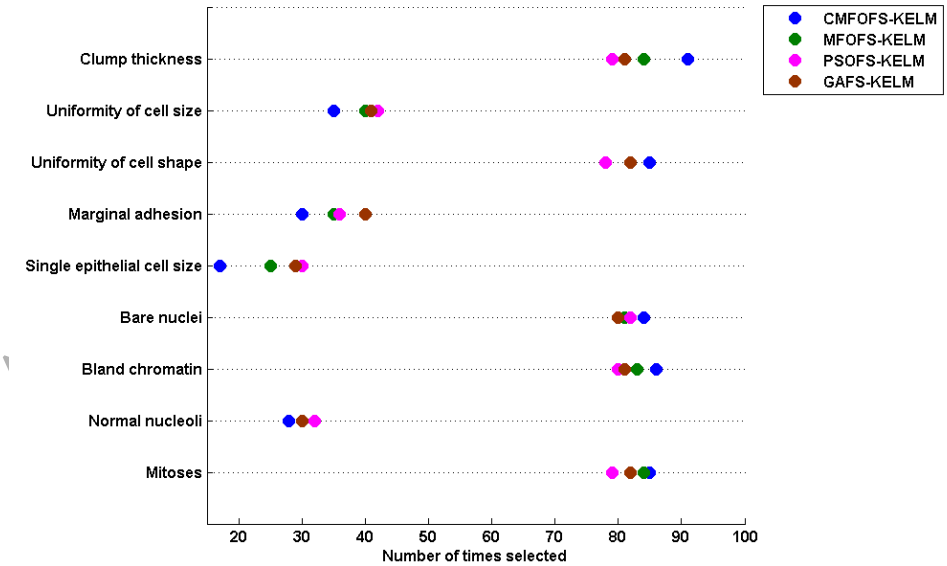


Fig. 10. The numbers of selected features via 10 runs 10-fold CV

For further comparison of the properties of these methods, Fig. 11 shows the evolutionary process

of each *KELM* method optimized by different meta-heuristics on the training dataset extracted from the whole Breast cancer dataset. As shown in this figure, the four fitness curves of *CMFOFS-KELM*, *MFOFS-KELM*, *PSOFS-KELM*, and *GAFS-KELM* gradually reached the optimal fitness after iterations 30, 32, 40, and 65, respectively. In contrast, *CMFO-KELM*, *MFO-KELM*, *PSO-KELM*, and *GA-KELM* do not improve at all after iterations 47, 44, 52, and 69, respectively. The convergence rate of each *KELM* is improved significantly with the feature selection strategy. Among all of these methods, *CMFOFS-KELM* obtains the best convergence rate. The convergence rate of the *CMFOFS-KELM* was better than that of *MFOFS-KELM*; the convergence rate of *CMFO-KELM* was also better than that of *MFO-KELM*, which indicates that the chaos strategy can greatly improve the convergence rate. It can be concluded that the property of the *KELM* based on synchronously optimizing the parameters and selecting the features using *MFO* can be extensively improved, especially when the chaos operation is creatively introduced.

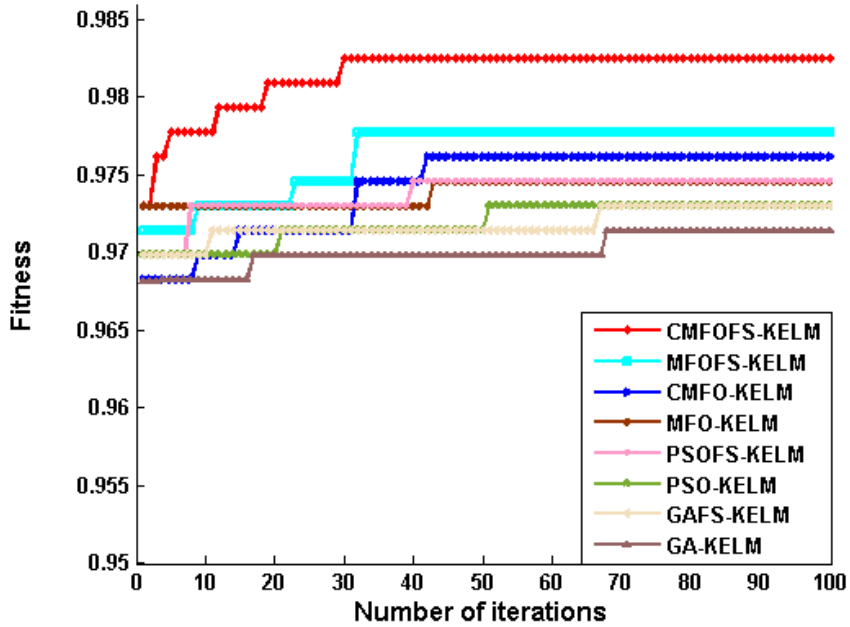


Fig. 11. The evolutionary processes of the these methods on the Breast cancer dataset

A paired t-test was also conducted on the eights methods, using the Breast cancer dataset, for validating the significance of the proposed method. In this experiment, a p-value less than 0.05 was considered statistically significant. Table 9 shows that *CMFOFS-KELM* has nearly significantly better results than the other seven competitors in terms of the *ACC*, *AUC*, and sensitivity. In the table, positive values represent that the *i*th classifier has fulfilled a better property than the *j*th one.

Table 9 Paired t-test results of *CMFOFS-KELM* and the other seven methods in terms of the four metrics on the Breast cancer dataset

Method	**t-value (significance)**			
	ACC	**AUC**	**sensitivity**	**specificity**
MFOFS-KELM	4.153 (0.002)	1.500(0.003)	1.259(0.002)	3.664(0.003)
PSOFS-KELM	5.159(0.001)	2.347(0.000)	3.861(0.004)	1.543(0.040)
GAFS-KELM	4.054(0.003)	2.346(0.002)	3.015(0.000)	1.684(0.004)
CMFO-KELM	7.479(0.000)	1.130(0.000)	2.106(0.004)	**0.504(0.016)**

MFO-KELM	5.822(0.000)	2.467(0.001)	1.349(0.00)	1.984(0.001)
PSO-KELM	5.172(0.001)	3.885(0.004)	3.572(0.001)	**2.700(0.024)**
GA-KELM	5.434(0.000)	2.796(0.000)	2.219(0.002)	1.845(0.008)

5.3 Influence of parameters on model performance

The performance of the most meta-heuristic is largely affected by the initialization parameters. As described above, MFO has three key parameters that can make a prominent effect on the property of the algorithm. The first one is the number of the moths (agents), and the second parameter is the *a* which is can be treated as the most important parameter due to the reason that this parameter can guarantee the trade-off between exploration and exploitation in the MFO. The third one is the maximum number of iterations in the whole process of optimization. In this part, analysis of three parameters is conducted in order to explore influence of these three parameters on the performance of the algorithm. Fig. 12 illustrates the average fitness of the proposed method on the Parkinson's disease dataset and Breast cancer dataset respectively with different number of moths. It can be seen from this figure that the efficiency of the proposed method is obviously influenced by the number of moths. The average fitness increases with the number of moths, 10, 30, and 50; however, it gets some decreased when the number of moths reaches 70. So the number of moth is set to 50 in solving the problems in this study.

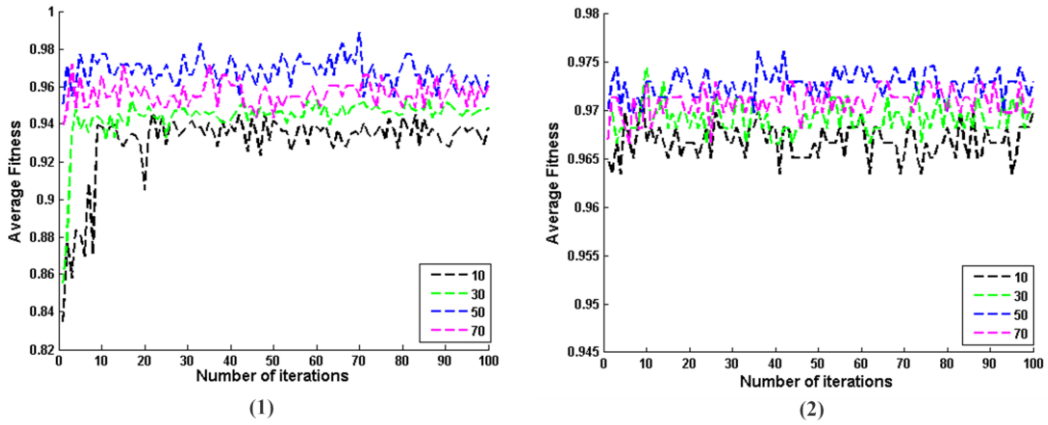


Fig.12. Effect of change in the number of search agents on the performance of the proposed method

The second parameter is the *a* that can get a better balance between the exploration and exploitation of the search space. The effect of the change in this parameter on the two experimental datasets is exhibited in Fig 13. From this figure, the proposed method can achieve an obvious superiority over other conditions on this problem when parameter *a* is set as -1. So, the -1 was chosen as the value of *a* in this study.

The third parameter, maximum number of iterations plays an important role of the proposed method. The detailed descriptions of this parameter can be vividly shown in Fig 14. As shown, the performance of the proposed method can be improved with the increase of the maximum number of iterations from 20 to 100; a fact cannot be ignored is that the performance of the proposed method has no obvious improvement after iteration 120. In short, what is given above explains why the number of iterations was set as 100 in this study.

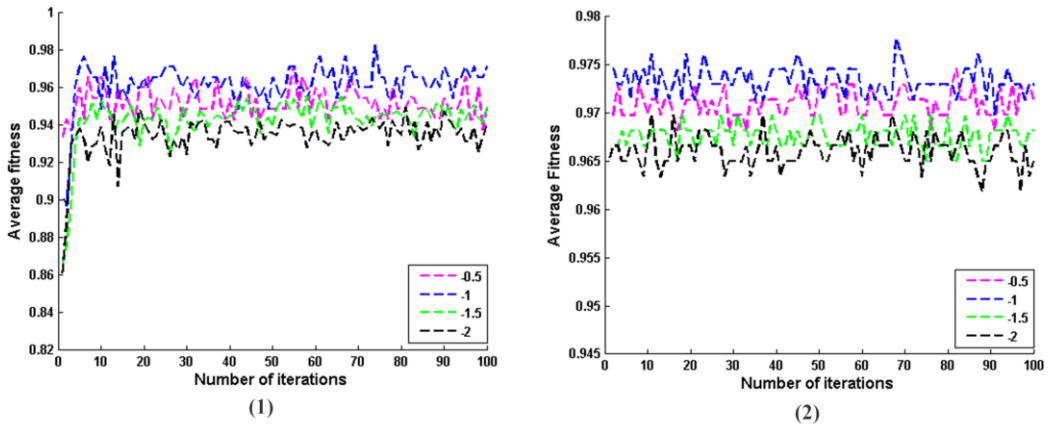


Fig.13. Effect of change in the constant on the performance of the proposed method

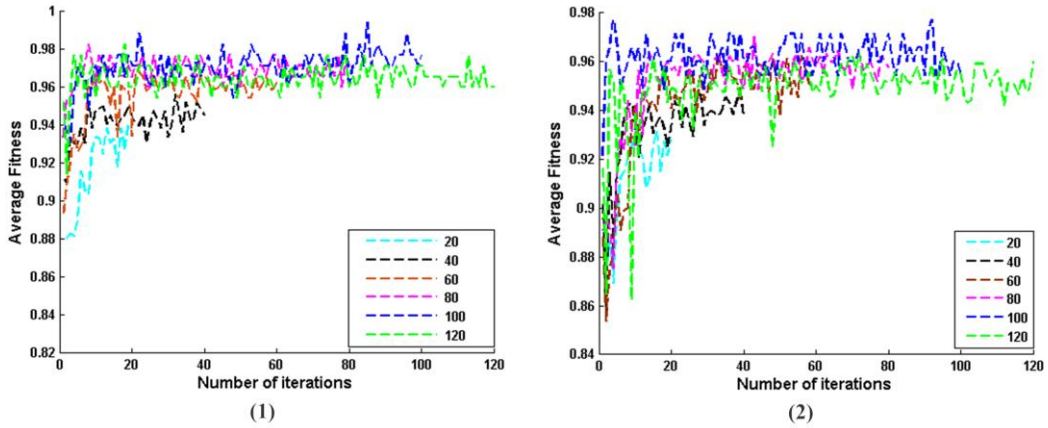


Fig.14. Effects of change in the maximum of iterations on the performance of the proposed method

6 Application in the practical problems

6.1 Diagnosis of paraquat-poisoned patients

Paraquat (1.1′-dimethyl-4, 4′-bipyridum dichloride, PQ) has long been deemed as the toxic pesticide for human where this herbicide is widely used in the world. Acute ingesting can cause serious organ failure which may directly leads to death if the patients do not get treatment timely. As we all know, the PQ concentration in blood can reveal the actual situation of poisoning. However, PQ is absorbed poorly from stomach and distributed into blood within 5h and the suspected patients cannot clearly express the history of PQ poisoning with the disturbance of language or consciousness. A fact has to be acknowledged that the early diagnosis of the PQ poisoning is quite difficult. In this study, the proposed method *CMFOFS-KELM* is applied to diagnosis the PQ poisoning by gas chromatography coupled with mass spectrometry (*GC-MS*) based on plasma metabolomics. The experimental dataset is collected from Medical Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University. This dataset is

composed of 15 patients whose timeframe of PQ poisoning was 5h to 12h and 16 healthy volunteers were chosen as control. All these samples were analyzed by the *GC-MS* method in metabolomics. Based on this technology, many peaks can be obtained from blood and identified by using retention time (RT) and mass spectra data (more detailed information can be found in [3]). So, 119 peaks in blood samples were detected, and 20 peaks were identified. Some of their match qualities are beyond 50% in Table 10. In this study, each sample is composed of 119 features. For comparison purpose, the other methods have also been conducted including MFOFS-KELM, MFO-KELM, KELM, original ELM, Logit model, and the reported final results were also calculated by averaging the results of 10 runs of the 10-fold CV. The detailed experimental results of this dataset are exhibited in Table 10. RT means the time it takes for an ingredient to go through the column of gas chromatograph. MD is the matched-degree of detected peaks with reference peaks.

Table.10.The main plasma metabolites detected by GC-MS in blood samples

No	RT (min)	Metabolites	MD (%)
1	6.044	Butanoic acid	66
2	7.007	1-Alanine	74
3	7.449	Piperazine	50
4	9.356	L-Valine	74
5	10.375	L-Leucine	78
6	10.461	glycerol	80
7	10.564	Phosphate	80
8	12.22	L-threonine	90
9	13.98	L-Proline	83
10	15.206	Glutamine	90
11	15.278	L-phenylalanine	78
12	16.036	Xylitol	64
13	16.515	Arabitol	64
14	17.262	9H-Purine	72
15	18.679	L-Tyrosine	50
16	18.787	glucitol	50
17	19.01	Inositol	64
18	19.205	D-Altrose	50
19	20.314	Uric acid	90
20	21.11	9,12-Octadecadienoic acid	90

Table 11 vividly exhibits the experimental results when the proposed method *CMFOFS-KELM* was applied to diagnosis the PQ poisoning patients. It can be seen that, *CMFOFS-KELM* has not only obtained a high performance with ACC of 0.9333, AUC of 0.9500, sensitivity of 0.9521, and specificity of 0.9533 respectively than other competitors, but also the smallest standard deviation. *MFOFS-KELM* without feature selection yielded ACC of 0.9293, AUC of 0.9321, sensitivity of 0.9432, and specificity of 0.9351. The *MFO-KELM* with ACC of 0.9245, AUC of 0.9163, sensitivity of 0.9204, and specificity of 0.9221, which is obviously inferior to *MFOFS-KELM. KELM* with grid-search technique yielded ACC of 0.9162, AUC of 0.9152, sensitivity of 0.9135, and specificity of 0.9172. Further, the original *ELM* yielded ACC of 0.9000, AUC of 0.9250, sensitivity of 0.9000, and specificity of 0.9100. The linear Logit model just obtained ACC of 0.8667, AUC of 0.8583, sensitivity of 0.8667, and specificity of 0.8500. In short, the property of

MFO has been significantly improved due to the chaos strategy and then it was successfully applied to construct an optimal KELM model for early diagnosis of paraquat-poisoned patients. This study will be a significant supplement in diagnosis of PQ poisoning.

Table 11 Detailed experimental results on the PQ dataset

Method	Metrics			
	ACC	**AUC**	**Sensitivity**	**Specificity**
CMFOFS-KELM	0.9333±0.0092	0.9500±0.0082	0.9521±0.0090	0.9533±0.089
MFOFS-KELM	0.9293±0.0153	0.9321±0.0143	0.9432±0.0120	0.9351±0.0126
MFO-KELM	0.9245±0.0165	0.9163±0.0131	0.9204±0.0143	0.9221±0.0189
KELM	0.9162±0.0187	0.9152±0.0163	0.9135±0.0143	0.9172±0.0201
ELM	0.9000±0.0221	0.9250±0.0231	0.9000±0.0201	0.9100±0.0211
Logit	0.8667±0.0238	0.8583±0.0259	0.8667±0.0212	0.8500±0.0238

For further comparison of the properties of these methods, Fig. 15 shows average fitness of the method including *CMFOFS-KELM*, *MFOFS-KELM*, *MFO-KELM*, *KELM*, original *ELM*, and *Logit* on the PQ dataset. As shown in this figure, the proposed method *CMFOFS-KELM* also has the highest average fitness value than other five competitors apparently. In sum, a conclusion can be made that the proposed method may be a meaningful tool with an excellent performance for diagnosis of paraquat-poisoned patients in the practical application.

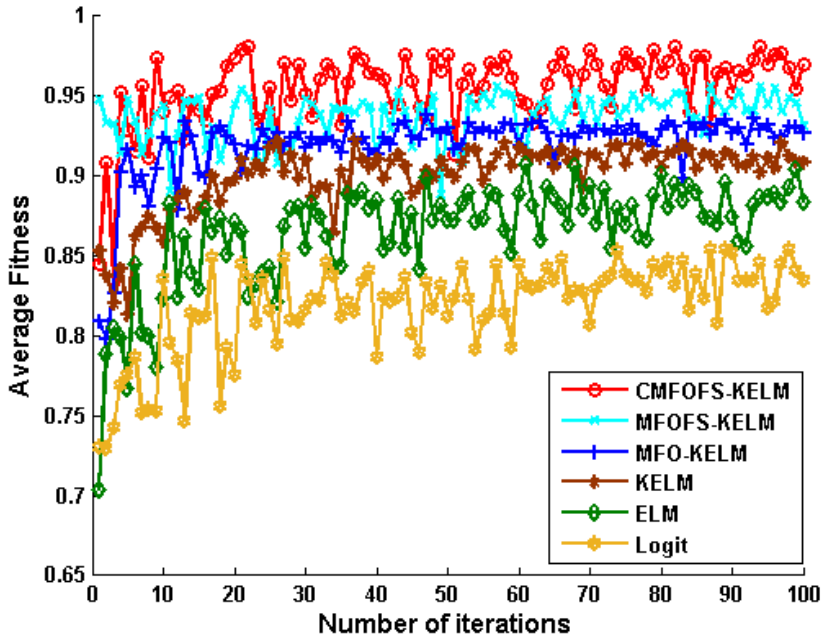


Fig.15. The average fitness of the methods on the PQ dataset in the whole iterations process

6.2 Diagnosis of overweight statuses

Recently, the human's overweight or obesity problem is getting more and more attention all over the world. Some studies have validated that being overweight can increase the possibility to suffer from disease such as diabetes, coronary heart disease, blood glucose, sleep-breathing disorder and some forms of cancer. Therefore, it is critical to early identify the overweight statuses

for decreasing health risks. In this study, the proposed method *CMFOFS-KELM* is used to discriminate overweight from healthy subjects based on the information from blood samples in the early stages of obesity. All samples were also strictly collected from the First Affiliated Hospital of Wenzhou Medical University. 476 participants were introduced in this experiment. Based on the definition of the World Health Organization, some samples mean overweight people if their body mass index (BMI) are $26.17\pm0.84 kg/m^2$, and other samples means healthy people if their BMI are 21.79 ± 0.92 kg/m^2. In this study, 37 indicators were used including 18 blood indexes, 16 biochemical indexes, age, red blood cell count, and white blood cell in urine. All of these features in this study are listed in Table 12. The more detailed information can refer to [4].

Table 12 The brief description of each feature used in this study

Feature	Description	Feature	Description
1	Age	20	Absolute value of monocyte (AVM)
2	triglyceride	21	red blood cell (RBC)
3	glucose	22	red blood cell in urine (RBCU)
4	Low density	23	hematocrit (HCT)
5	High density lipoprotein (LDL)	24	Percentage of leukomonocyte (PLC)
6	total cholesterol (CHO)	25	absolute value of leukomonocyte (AVLC)
7	alanine transaminase (ALT)	26	mean corpuscular volume (MCV)
8	Aspartate aminotransferase (AST)	27	mean corpuscular hemoglobin concentration (MCH)
9	γ-glutamyl transpeptise (γ-GT)	28	Mean corpuscular hemoglobin concentration (MCHC)
10	Total protein (TP)	29	Mean platelet volume (MPL)
11	Albumin (ALB)	30	Absolute value of eosinophils (PE)
12	Creatinine (CR)	31	percentage of eosinophils
13	Urea nitrogen (BUN)	32	hemoglobin (HB)
14	Alkaline phosphatase (AKP)	33	Blood platelet (PLT)
15	total bilirubin (TBIL)	34	thrombocytocrit (THR)
16	direct bilirubin (DBIL)	35	percentage of neutrophils (PN)
17	uric acid (UA)	36	absolute value of neutrophils (AVN)
18	White blood cell in urine (WBCU)	37	red cell volume distribution width (RBCVD)
19	percentage of monocyte (PMC)		

The detailed experimental result of each method to predict overweight statuses is exhibited in Table 13. From this table, *CMFOFS-KELM* obviously fulfills the best performance in terms of four metrics and the smallest standard deviation. MFOFS-KE*LM* with ACC of 0.9361, AUC of 0.9178, sensitivity of 0.9354, and specificity of 0.9312 is second in the rank followed by MFO-KELM, and KELM. The original ELM has only ACC of 0.9032, AUC of 0.8998, sensitivity of 0.8395, and specificity of 0.9602 and the linear method Logit performs worst among all methods with ACC of 0.8510, AUC of 0.8706, sensitivity of 0.8000, and specificity of 0.9412. Based on the observation above, the improved *MFO* based on chaos can construct an optimal

KELM method for diagnosis of overweight statuses.

Table 13 Detailed experimental results of prediction of overweight statuses

Method	Metrics			
	ACC	**AUC**	**Sensitivity**	**Specificity**
CMFOFS-KELM	0.9379±0.98	0.9388±1.07	0.9546±0.97	0.9243±1.21
MFOFS-KELM	0.9361±1.32	0.9378±1.90	0.9524±1.45	0.9231±1.56
MFO-KELM	0.9232±1.89	0.9178±2.12	0.9354±1.78	0.9312±1.83
KELM	0.9148±2.21	0.9194±2.09	0.8889±2.34	0.9500±2.32
ELM	0.9032±2.09	0.8998±2.23	0.8395±4.97	0.9602±2.67
Logit	0.8510±2.03	0.8706±2.10	0.8000±3.54	0.9412±3.21

Fig. 16 clearly shows average fitness of each method for prediction of overweight statuses during the process of constructing model. It can be seen that the average fitness of the proposed method *CMFOFS-KELM* express the obvious superiority over other five competitors significantly, which means that the proposed method has a positive significance for early identifying the overweight statuses and potential practical value for decreasing health risks.

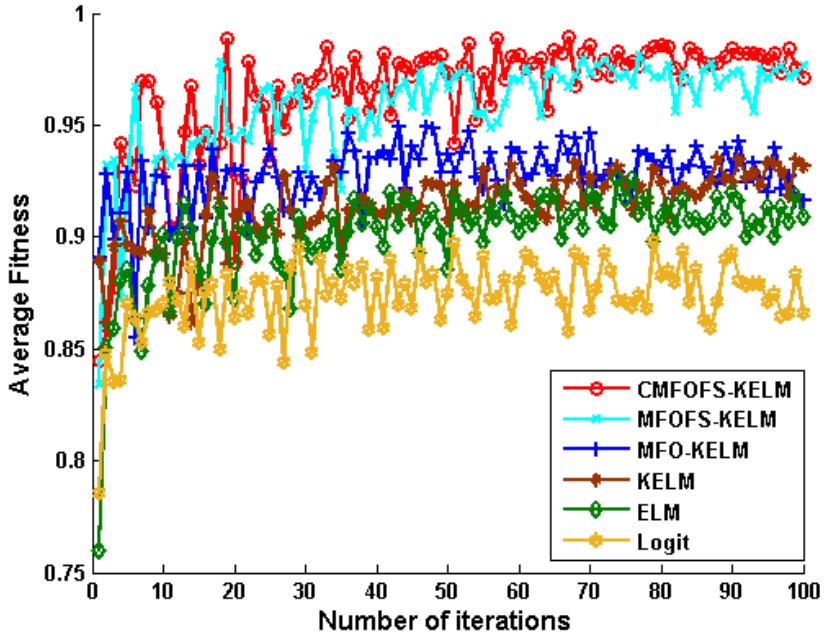


Fig.16. The average fitness of the methods to diagnosis of overweight statuses in the whole iterations process

7 Discussion

We have proposed a novel learning scheme for *KELM* with applications in diagnosing medical conditions. In the proposed scheme, the chaos enhanced *MFO* strategy was adopted to dynamically and simultaneously identify the two key parameters and the optimal feature subset for *KELM*. To the best of our knowledge, this is the first time that the improved *MFO* strategy has

been employed for training the *KELM* model. In the proposed improved *MFO* strategy, we introduce two vital mechanisms based on chaos theory: population initiation and chaotic disturbance. The chaotic population initiation mechanism can encourage *KELM* to learn from a good initial solution, which dramatically improves the convergence rate. The best evidence for the efficacy of this method is found in Figs. 7 and 11. The chaotic disturbance mechanism can efficiently prevent *MFO* from falling into the local optimum. In order to adaptively control the degree of disturbance, we have further introduced the coefficient λ during the disturbance process. This ensures an improved balance between exploitation and exploration. In order to validate the proposed methodology, we have carried out experiments on two common disease diagnosis problems, and then it was been successfully used to diagnose of paraquat-poisoned patients and predict overweight statues in the practical application. The experimental results show that the proposed *CMFO* enhanced *KELM* method has significant superiority over other meta-heuristic-based *KELM* methods. This implies that integrating chaos *MFO* and *KELM* can generate robust and stable results for the two disease diagnosis problems that were investigated.

A few studies have been conducted related to the issue of parameter setting in *KELM*. Yang et al. [40] adopted original *PSO* to optimize key parameters of *KELM* for pressure prediction of coal slurry transportation pipelines. Lu et al. [41] proposed quantum-behaved *PSO* with binary encoding to select the appropriate feature subset and parameters for *KELM*. Liu et al. [20] used a quantum genetic algorithm to optimize the two key parameters of KELM for the reconstruction of defective profiles. Few of these works have investigated the simultaneous parameter setting and feature selection for *KELM*, and none have applied the proposed methods to medical diagnosis problems. Our study is novel because of both the joint investigation of the parameter setting and feature selection issues by introducing a new set of mechanisms for *MFO*, and also because it is the first time that the developed methods have been applied to common disease diagnosis problems.

This study has offered the medical field a new, effective decision support tool. Doctors often rely on their own clinical experience when making medical diagnoses. Therefore, the diagnostic results may be subjective and it is possible that inaccurate diagnosis is made. In recent years, significant progress has been made in the field of medical diagnosis through artificial intelligence (AI) techniques. Compared with doctors, AI methods are able to make more objective decisions, with fewer mistakes. There have been many studies of AI in the literature. Khan et al. [42] proposed a method for the diagnostic prediction of cancers based on gene expression profiling and artificial neural networks. The experimental results show that the proposed method was capable of predicting cancers. Olanow et al. [43] constructed Parkinson's disease treatment guidelines for clinicians, based on a decision tree. Sarkar et al. [44] used k-nearest neighbors to diagnose breast cancer and improved the classification results by 1.17%. Staiano et al. [45] investigated the use of random forests for diagnosing familial combined hyperlipidemia (FCH). This work reported on further gene association with FCH. Shen et al. [46] proposed a novel support vector machine, based on fruit fly optimization, for several medical diagnosis problems. They found that the proposed method could perform well on diseases like diabetes and thyroid disease. Chen et al. [4] investigated *ELM* to successfully predict patient overweight status using blood indexes. In this study, the proposed novel *KELM* based approach can assist doctors in making accurate disease diagnoses and in making timely treatment decisions. This can be of particular benefit to doctors working in remote areas. It can address issues that result from doctor shortages and mixed quality

problems.

This research is not without limitations. The proposed *CMFO* learning scheme has no special advantage over other meta-heuristic-based schemes in terms of computational cost. During the training procedure, our proposed method had a slightly longer computational time than the other methods. This may be because of the frequent position updates of the flames in the iterations. Future work can investigate the parallel deployment of the proposed methodology as done in [47, 48] in order to reduce the heavy computational burden. The proposed method can be further explored on big data, with the aid of high performance tools. It also should be noted that, due to space limitations, this study only examined two kinds of disease diagnosis problems. Before wide application, more diseases should be studied using our method.

There are a number of aspects that may be beneficial to study in future work. In *MFO*, the multi swarm strategy can be introduced as a new element and then the population can be divided into sub-swarms for collaborative optimization. A hybrid algorithm can also be considered in *MFO* for further enhancing the capability of model selection as conducted in [49, 50]. Further, the fitness function can be converted into a multi-objective function in terms of classification accuracy, length of feature subset, and the computation time based on the Pareto condition.

8 Conclusion

In this study, a new effective learning scheme, *CMFOFS-KELM*, has been proposed for *KELM*. This new scheme has been successfully applied to the diagnosis of Parkinson's disease and breast cancer. This paper proposes a new *CMFO* approach, in order to maximize the generalization capability of the *KELM* classifier by simultaneously addressing the parameter setting and feature selection. The experimental results of the two common disease diagnosis problems have demonstrated that *KELM* based on the *CMFO* learning strategy has significant advantages over other competitive *KELM* methods constructed by the original *MFO*, *PSO* and *GA*, in terms of a set of performance metrics. Therefore, it can be concluded that the proposed *CMFOFS-KELM* may be treated as a valuable decision support tool.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. of 61303113, 61572226 and 61373053. This research is also funded by the Science and Technology Plan Project of Wenzhou of China under Grant Nos. of G20140048, H20110003, Y20160070. Jilin Province Natural Science Foundation under grant No. of 20150101052JC, Zhejiang Provincial Natural Science Foundation of China under grant Nos. of LY17F020012, LY16H300005, LY14F020035 and LQ13G010007, Guangdong Natural Science Foundation under grant No. of 2016A030310072, the open project program of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University under Grant No. 93K172013K01, Graduate Innovation Fund of Wenzhou University under Grant No. 3162016028.

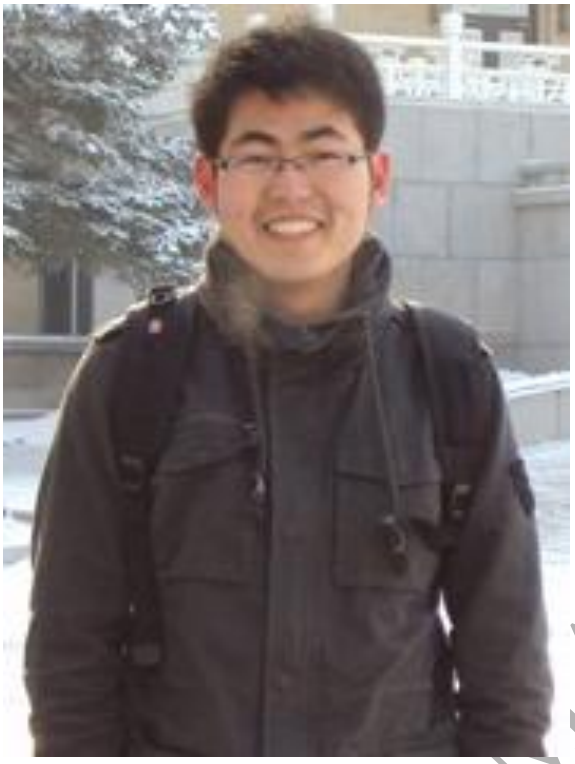
References

1. Huang, G.-B., Q.-Y. Zhu, and C.-K. Siew, *Extreme learning machine: Theory and applications.* Neurocomputing, 2006. **70**(1–3): p. 489-501.
2. Zhang, R., et al., *Multicategory classification using an extreme learning machine for microarray gene expression cancer diagnosis.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2007. **4**(3): p. 485-494.
3. Hu, L., et al., *An efficient machine learning approach for diagnosis of paraquat-poisoned patients.* Computers in Biology & Medicine, 2015. **59C**: p. 116-124.
4. Chen, H., et al., *Using Blood Indexes to Predict Overweight Statuses: An Extreme Learning Machine-Based Approach.* PLoS ONE, 2015. **10**(11): p. e0143003.
5. Mohammed, A.A., et al., *Human face recognition based on multidimensional PCA and extreme learning machine.* Pattern Recognition, 2011. **44**(10-11): p. 2588-2597.
6. Suresh, S., R. Venkatesh Babu, and H.J. Kim, *No-reference image quality assessment using modified extreme learning machine classifier.* Applied Soft Computing, 2009. **9**(2): p. 541-552.
7. Cao, J., et al., *Extreme learning machine and adaptive sparse representation for image classification.* Neural Networks, 2016. **81**(C): p. 91.
8. Cao, J., et al., *Excavation Equipment Recognition Based on Novel Acoustic Statistical Features.* IEEE Transactions on Cybernetics, 2016. **PP**(99): p. 1-13.
9. Pal, M., *Extreme-learning-machine-based land cover classification.* International Journal of Remote Sensing, 2009. **30**(14): p. 3835-3841.
10. Moreno, R., et al., *Extreme learning machines for soybean classification in remote sensing hyperspectral images.* Neurocomputing, 2014. **128**: p. 207-216.
11. Huang, G.-B., et al., *Extreme learning machine for regression and multiclass classification.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2012. **42**(2): p. 513-529.
12. Pal, M., A.E. Maxwell, and T.A. Warner, *Kernel-based extreme learning machine for remote-sensing image classification.* Remote Sensing Letters, 2013. **4**(9): p. 853-862.
13. Chen, C., et al., *Spectral-Spatial Classification of Hyperspectral Image Based on Kernel Extreme Learning Machine.* Remote Sensing, 2014. **6**(6): p. 5795-5814.
14. Liu, T., et al., *A fast approach for detection of erythemato-squamous diseases based on extreme learning machine with maximum relevance minimum redundancy feature selection.* International Journal of Systems Science, 2015. **46**(5): p. 919-931.
15. Ma, C., et al., *An Efficient Diagnosis System for Parkinson's Disease Using Kernel-Based Extreme Learning Machine with Subtractive Clustering Features Weighting Approach.* Computational and mathematical methods in medicine, 2014. **2014**.
16. Li, Q., et al., *An Enhanced Grey Wolf Optimization Based Feature Selection Wrapped Kernel Extreme Learning Machine for Medical Diagnosis.* Computational and Mathematical Methods in Medicine, 2017. **2017**: p. 15.
17. Chen, H.-L., et al., *An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease.* Neurocomputing, 2016. **184**: p. 131-144.
18. Zhao, D., et al., *An Effective Computational Model for Bankruptcy Prediction Using Kernel Extreme Learning Machine Approach.* Computational Economics, 2017. **49**(2): p. 325-341.

19. Deng, W.-Y., Q.-H. Zheng, and Z.-M. Wang, *Cross-person activity recognition using reduced kernel extreme learning machine.* Neural Networks, 2014. **53**: p. 1-7.
20. Liu, B., et al., *2-D defect profile reconstruction from ultrasonic guided wave signals based on QGA-kernelized ELM.* Neurocomputing, 2014. **128**: p. 217-223.
21. Zhao, X., et al., *An Efficient and Effective Automatic Recognition System for Online Recognition of Foreign Fibers in Cotton.* IEEE Access, 2016. **4**: p. 8465-8475.
22. Jiang, Y., J. Wu, and C. Zong, *An effective diagnosis method for single and multiple defects detection in gearbox based on nonlinear feature selection and kernel-based extreme learning machine.* Journal Of Vibroengineering, 2014. **16**(1): p. 499-512.
23. Safarzadeh, O., et al., *Loading pattern optimization of PWR reactors using Artificial Bee Colony.* Annals of Nuclear Energy, 2011. **38**(10): p. 2218-2226.
24. Hsieh, T.J. and W.C. Yeh, *Knowledge Discovery Employing Grid Scheme Least Squares Support Vector Machines Based on Orthogonal Design Bee Colony Algorithm.* Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics, 2011. **41**(5): p. 1198-1212.
25. Song, G., F. Xue, and C. Zhang, *A model using texture features to differentiate the nature of thyroid nodules on sonography.* Journal of Ultrasound in Medicine, 2015. **34**(10): p. 1753-1760.
26. Xiao, C. and W.A. Chaovalitwongse, *Optimization Models for Feature Selection of Decomposed Nearest Neighbor.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016. **46**(2): p. 177-184.
27. Mirjalili, S., *Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm.* Knowledge-Based Systems, 2015. **89**: p. 228-249.
28. Allam, D., D.A. Yousri, and M.B. Eteiba, *Parameters extraction of the three diode model for the multi-crystalline solar cell/module using Moth-Flame Optimization Algorithm.* Energy Conversion and Management, 2016. **123**: p. 535-548.
29. Emary, E. and H.M. Zawbaa, *Impact of chaos functions on modern swarm optimizers.* PLoS ONE, 2016. **11**(7).
30. Li, Z., et al., *Lévy-Flight Moth-Flame Algorithm for Function Optimization and Engineering Design Problems.* Mathematical Problems in Engineering, 2016. **2016**.
31. Li, C., S. Li, and Y. Liu, *A least squares support vector machine model optimized by moth-flame optimization algorithm for annual power load forecasting.* Applied Intelligence, 2016. **45**(4): p. 1166-1178.
32. Sayed, G.I., M. Soliman, and A.E. Hassanien, *Bio-inspired swarm techniques for thermogram breast cancer detection*, in *Studies in Computational Intelligence*. 2016. p. 487-506.
33. Zhang, L., et al., *Intelligent facial emotion recognition using moth-firefly optimization.* Knowledge-Based Systems, 2016. **111**: p. 248-267.
34. Saremi, S., S. Mirjalili, and A. Lewis, *Biogeography-based optimisation with chaos.* Neural Computing and Applications, 2014. **25**(5): p. 1077-1097.
35. Coelho, L.d.S. and V.C. Mariani, *Use of chaotic sequences in a biologically inspired algorithm for engineering design optimization.* Expert Systems with Applications, 2008. **34**(3): p. 1905-1913.
36. Huang, Y., Q. Huang, and Q. Wang, *Chaos and Bifurcation Control of Torque-Stiffness-Controlled Dynamic Bipedal Walking.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016. **PP**(99): p. 1-12.

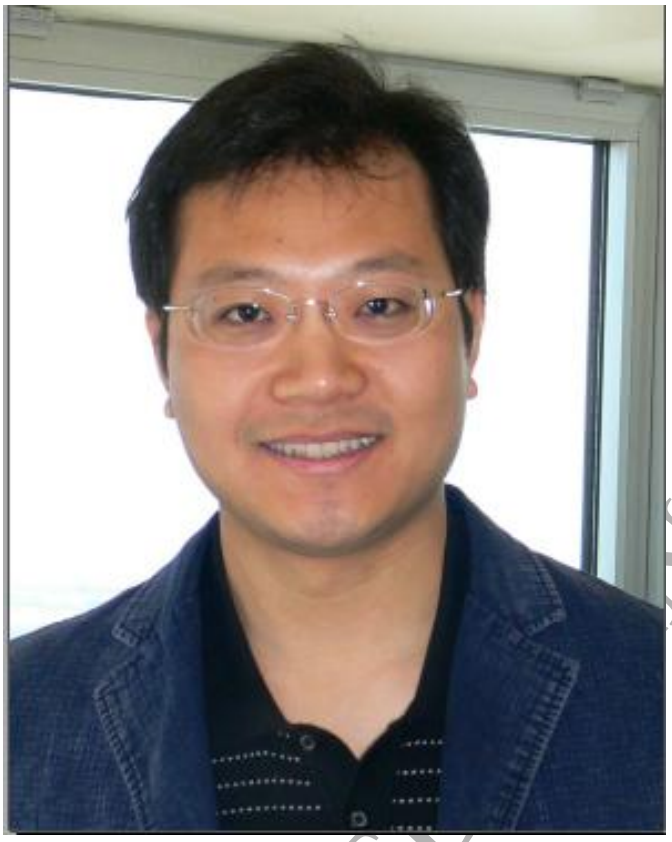
37. Chenglin, Z., et al., *Fault diagnosis of sensor by chaos particle swarm optimization algorithm and support vector machine.* Expert Systems with Applications, 2011. **38**(8): p. 9908-9912.
38. Rahnamayan, S., H.R. Tizhoosh, and M.M.A. Salama, *Opposition-Based Differential Evolution.* IEEE Transactions on Evolutionary Computation, 2008. **12**(1): p. 64-79.
39. Rahnamayan, S., H.R. Tizhoosh, and M.M.A. Salama, *A novel population initialization method for accelerating evolutionary algorithms.* Computers & Mathematics with Applications, 2007. **53**(10): p. 1605-1614.
40. Yang, X.C., X.R. Yan, and C.F. Song, *Pressure Prediction of Coal Slurry Transportation Pipeline Based on Particle Swarm Optimization Kernel Function Extreme Learning Machine.* Mathematical Problems in Engineering, 2015. **2015**.
41. Lu, L., et al., *Video analysis using spatiotemporal descriptor and kernel extreme learning machine for lip reading.* Journal of Electronic Imaging, 2015. **24**(5).
42. Khan, J., et al., *Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks.* Nature medicine, 2001. **7**(6): p. 673-679.
43. Olanow, C.W. and W.C. Koller, *An algorithm (decision tree) for the management of Parkinson's disease Treatment guidelines.* Neurology, 1998. **50**(3 Suppl 3): p. S1-S1.
44. Sarkar, M. and T.Y. Leong, *Application of K-nearest neighbors algorithm on breast cancer diagnosis problem.* Proceedings / AMIA ... Annual Symposium. AMIA Symposium, 2000: p. 759-763.
45. Staiano, A., et al., *Investigation of Single Nucleotide Polymorphisms Associated to Familial Combined Hyperlipidemia with Random Forests*, in *Neural Nets and Surroundings: 22nd Italian Workshop on Neural Nets, WIRN 2012, May 17-19, Vietri sul Mare, Salerno, Italy*, B. Apolloni, et al., Editors. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 169-178.
46. Shen, L., et al., *Evolving support vector machines using fruit fly optimization for medical data classification.* Knowledge-Based Systems, 2016. **96**: p. 61-75.
47. Chen, H.L., et al., *Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy.* Applied Mathematics and Computation, 2014. **239**: p. 180-197.
48. Chen, H.L., et al., *A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method.* Knowledge-Based Systems, 2011. **24**(8): p. 1348-1359.
49. Feng-Tse, L., K. Cheng-Yan, and H. Ching-Chi, *Applying the genetic approach to simulated annealing in solving some NP-hard problems.* IEEE Transactions on Systems, Man, and Cybernetics, 1993. **23**(6): p. 1752-1767.
50. Chia-Feng, J., *A hybrid of genetic algorithm and particle swarm optimization for recurrent network design.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2004. **34**(2): p. 997-1006.

Mingjing Wang is currently a graduate student majoring in computer software and theory at Wenzhou University, China. He received his bachelor degree in department of computer science and technology at Jilin Agricultural University, China. His research interests are data mining, pattern recognition, evolutionary computation, as well as their applications such as medical diagnosis, bankruptcy prediction, among others.



Huiling Chen is currently a lecturer in the department of computer science and technology at Wenzhou University, China. He received his Ph.D. degree in department of computer science and technology at Jilin University, China. His present research interest centers on machine learning and data mining, as well as their applications such as medical diagnosis, bankruptcy prediction, face recognition, among others. He is currently a Reviewer for IEEE Transactions on Systems, Man, and Cybernetics, Part B. He has published more than 80 papers in international journals and conference proceedings, including Expert Systems with Applications, Knowledge-Based Systems, Neurocomputing, Soft Computing, PloS One, PAKDD, and among others.

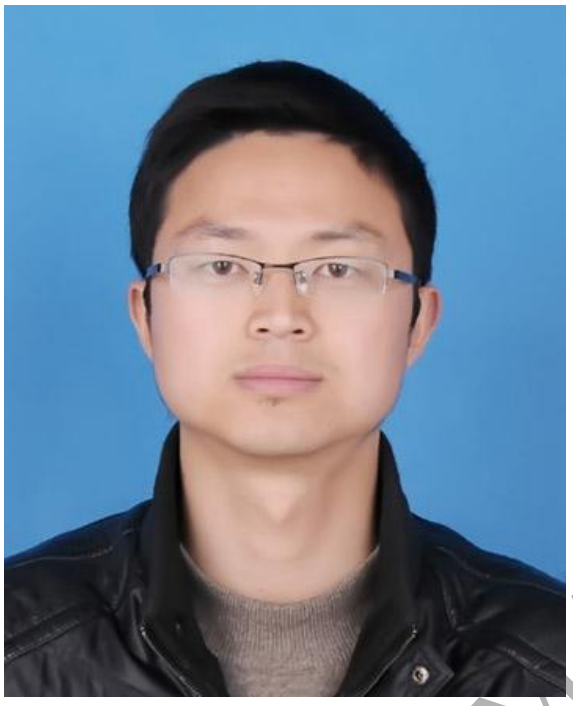
Bo Yang is currently a professor in the College of Computer Science and Technology, Jilin University. He is also the director of the Key Laboratory of Symbolic Computation and Knowledge Engineering, Ministry of Education, China. His current research interests are in the areas of data mining, complex network analysis, self-organized and self-adaptive multi-agent systems, with applications to knowledge engineering and intelligent health informatics.



Xuehua Zhao received the Ph.D. degree in College of Computer Science and Technology, Jilin University in 2014. He is currently a lecturer in the School of Digital Media, Shenzhen Institute of Information Technology. His main research interests are related to machine learning and data mining.

Lufeng Hu, pharmacology Master, pharmacist-in-charge, majored in clinical pharmacology, working at department of pharmacy, the first affiliated hospital of Wenzhou medical University. When I graduated from Wenzhou medical college at 2008.7, I began to carry out the blood drug concentration monitoring of many kinds of antibiotic, antitumor drugs, antiepileptic drugs and pesticide detection. Now, I have developed the determination method of paraquat, vancomycin, voriconazole and fluconazole, rina thiazole amine, piperacillin, beauty e.faecalis, MTX, imatinib, oqa xiping, lamotrigine and so on.

Zhennao Cai is pursuing the Ph.D. degree in the school of computer at Northwestern Polytechnical University, China. He is now also a researcher at Wenzhou University, China. He received his Master's degree from school of software Engineering of Huazhong University of science and technology. His main research interests include machine learning, pattern recognition, and data mining.



Hui Huang is a lecturer at Intelligent Information Systems Institute, Wenzhou University, Wenzhou, China. He is now working for his Ph.D. in Northwestern Polytechnical University. His research interests include image processing, parallel computing and machine learning.

Changfei Tong is currently an associate professor in the department of computer science and technology at Wenzhou University, China. He received his PhD degree in control science and engineering at Zhejiang University, Hangzhou, China, in 2008. His research interests include sparse data processing, urban computing with transportation data and pattern recognition. He has published more than 20 papers in international journals and conference proceedings, including Sensors, Abstract and Applied Analysis, Journal of Pharmacological and Toxicological Methods, and among others.