OLTC Fault Diagnosis Method Based on Time Domain Analysis and Kernel Extreme Learning Machine

Yan Yan[1,3*], Hongzhong Ma[1], Dongdong Song[2], Yang Feng[4], Dawei Duan[1]

[1] College of Energy and Electrical Engineering, Hohai University, China,
[2] College of Mechanical Electrical and Engineering, Hebei Normal University of Science and Technology, China
[3] Electric Power Research Institute, State Grid Ningxia Power Co., Ltd., China
[4] Training Center of State Grid Ningxia Electric Power Co., Ltd., China
yanyan503986706@163.com

Abstract. Aiming at the problems of limited feature information and low diagnosis accuracy of traditional on-load tap changers (OLTCs), an OLTC fault diagnosis method based on time-domain analysis and kernel extreme learning machine (KELM) is proposed in this paper. Firstly, the time-frequency analysis method is used to analyze the collected OLTC vibration signal, extract the feature information and form the feature matrix; Then, the PCA algorithm is used to select effective features to build the initial optimal feature matrix; Finally, a kernel extreme learning machine optimized by improved grasshopper optimization algorithm (IGOA), is used to handle the optimal feature matrix for classifying fault patterns. Evaluation of algorithm performance in comparison with other existing methods indicates that the proposed method can improve the diagnostic accuracy by at least 7%.

Keywords: OLTC, IGOA, PCA, fault diagnosis, feature selection

1 Introduction

On-load tap changers (OLTC), the only movable parts in the transformer, are responsible for the tasks of power system reactive power regulation, voltage load stabilization, and power quality improvement. According to relevant statistics, transformer faults caused by OLTC account for more than 20% of its total faults, of which the main type of fault is mechanical fault. Therefore, improving the accuracy of OLTC mechanical fault diagnosis is of great significance for ensuring the safe operation of the power system [1-2].

Because the vibration signal of the OLTC in the process of shifting gears contains a large amount of mechanical state information, the analysis method based on vibration signal has gradually become one of the main methods for OLTC state monitoring. The basic principle of vibration analysis method is to firstly collect the vibration signal in the OLTC switching process through the acceleration sensor, and then extract and analyze its characteristic, so as to realize the detection of potential mechanical dangers. According to the time-frequency characteristics of the OLTC vibration signal, the literature [3-5] uses continuous wavelet transform to calculate the time-domain envelope, which is then regarded as the feature for classification. After establishing the OLTC feature-label library, ref. [3-5] employ Kohonen's self-organizing map network for fault diagnosis. Literature [6-7] employs wavelet method to analyzes abnormal states of some specific components that involving abnormal pulses and action sequences. The disadvantage of wavelet analysis methods is that its computation overhead is considerable. From the perspective of chaotic characteristics of vibration signals, literature [8] transformed one-dimensional vibration signals into high dimensions space through phase space reconstruction, and identified mechanical faults of vibration signals under different working conditions through weighted distributed information entropy. Similarly, according to the high-dimension spatial distribution of OLTC vibration signals, literature [9] proposed an OLTC mechanical fault identification method based on the K-means clustering algorithm. Unfortunately, due to the limitations of the clustering algorithm, the state information of reconstructed signal after clustering is not comprehensive. From the perspective of time-frequency analysis and inspired by ensemble empirical mode decomposition (EEMD), literature [10] proposed a narrow-band noise-assisted multivariate EMD (NA-MEMD) method, which added several narrow-band noise signals to the multi-channel acquisition signals and then carried out unified calculation, thus suppressing modal aliasing to a certain extent and having high diagnostic efficiency. However, the fault mechanism of OLTC is complex, and the fault types are diverse, and the different damage degrees of the key

---

components of the OLTC transfer switch will affect the diagnosis results, so there is an urgent need for efficient diagnosis methods to improve the fault accuracy of OLTC.

To further improve the mechanical fault diagnosis method of OLTC, based on the idea of reference [11], this paper extracts and analyzes the vibration signal by means of time-frequency analysis and proposes an OLTC fault diagnosis method based on time-domain analysis and IGOA-KELM:

(1) Aiming at the disadvantage of limited feature information in traditional OLTC fault diagnosis, which leads to a high misjudgment rate, a high-dimensional feature matrix based on time-domain is constructed from the perspective of multi-feature fusion;

(2) Since the combination of time-domain features leads to excessive redundancy and high correlation, principal component analysis (PCA) is used to select effective feature to build the optimal feature matrix;

(3) In view of the fact that KELM is sensitive to the settings of the nuclear parameter κ and the penalty parameter C, a method for optimizing the nuclear extreme learning machine based on the improved grasshopper algorithm (IGOA-KELM) is proposed. Finally, the effectiveness of the proposed method is verified by analyzing the vibration signals of the M-type under different working conditions.

2 Multi-feature Time-frequency Domain Analysis Method

When the OLTC performs contact switching, the vibration signal has strong nonlinearity and non-stationarity, and the duration is very short, which is usually only tens of milliseconds to about 160 milliseconds. Take the vibration signal under the normal working condition of OLTC as an example, its appearance is shown in Fig. 1.
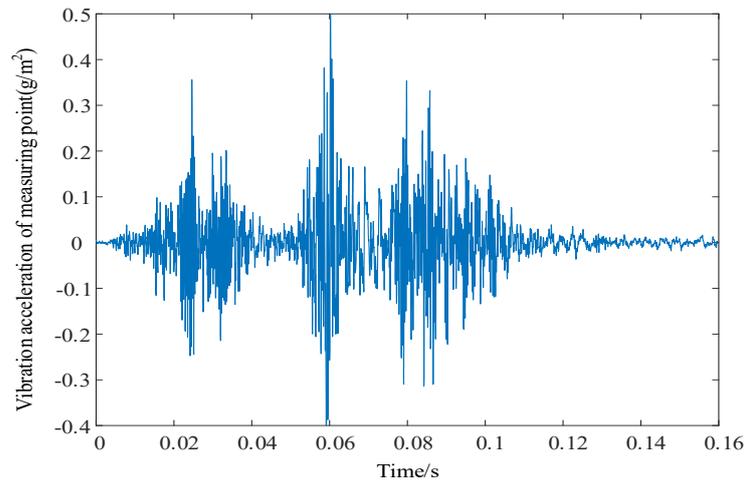

Fig. 1. Time domain diagram of OLTC vibration signal under normal working conditions

When OLTC switches contacts, the time-frequency parameters of the vibration signal will change greatly, so the time-frequency system time-domain analyzer can be used to characterize the change of vibration signal. The index can be divided into two categories: dimensional and dimensionless. Commonly used dimensional indexes are maximum, minimum, mean, root mean square, variance, and peak-to-peak of the vibration signal. While commonly used dimensionless indexes are square root amplitude value (SRA), Kurtosis, shape factor, impulse factor, clearance factor, peak factor, and Crest factor. In this paper, 13 kinds of features are extracted according to the time domain characteristics of OLTC vibration signals. The specific feature variables and mathematical expressions are shown in Table 1. In Table 1, x_i is the time-domain acceleration amplitude of the i-th sampling point of the vibration signal, and N is the number of sampling points [11-12].

The above features are fused to form a feature quantity matrix, fusion matrix $\boldsymbol{X'}$, which is specifically expressed as follows:

$$\boldsymbol{X'} = \begin{bmatrix} x'_{11} & x'_{12} & \cdots & x'_{1r} \\ x'_{21} & x'_{22} & \cdots & x'_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{m1} & x'_{m2} & \cdots & x'_{mr} \end{bmatrix}. \tag{1}$$

Where, m represents the population size, and r is the number of feature quantities. Since the combination of time-domain features will lead to excessive redundancy and increase the complexity of the system, this paper uses principal component analysis to reduce the redundancy of the system.

Table 1. Time domain feature list

NO.	Feature name	Calculation formula
1	Maximum	$C_1 = \max(x_i)$
2	Minimum	$C_2 = \min(x_i)$
3	Peak-to-peak	$C_3 = C_1 - C_2$
4	Average value	$C_4 = \dfrac{1}{N}\sum_{i=1}^{N} x_i$
5	Root mean square	$C_5 = \left(\dfrac{1}{N}\left(\sum_{i=1}^{N} x_i^2\right)\right)^{1/2}$
6	Average amplitude	$C_6 = \dfrac{1}{N}\sum_{i=1}^{N} |X_i|$
7	Amplitude variance	$C_7 = \left[\sum_{i=1}^{N}(x_i - C_5)^2\right] / N$
8	Margin factor	$C_8 = \dfrac{C_3}{C_7}$
9	Crest factor	$C_9 = \dfrac{C_3}{C_5}$
10	Impulse factor	$C_{10} = \dfrac{C_3}{C_6}$
11	Skewness factor	$C_{11} = \dfrac{C_3}{\sqrt{C_7}}$
12	Waveform factor	$C_{12} = \dfrac{C_3}{C_6}$
13	Kurtosis factor	$C_{13} = \left(\dfrac{C_5}{\sqrt{C_8}}\right)^4$

3 Feature Selection Based on Principle Component Analysis

Due to the combination of time-domain features, the redundant information among the parameters is too much and the correlation is too high. Therefore, the PCA method is used to reduce the dimensionality of the feature data of the OLTC vibration signal. The specific analysis steps of PCA are as follows [13-15]:

(1) Standardization of the original data: There are R indicators in the original data, the sample size is, and the sample matrix obtained is as shown in equation (1). Normalization of equation (1) can be obtained by the following standardization matrix:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1r} \\ y_{21} & y_{22} & \cdots & y_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mr} \end{bmatrix}. \tag{2}$$

Where Y_{ij} represents the sample elements in the i-th row and j-th column, $y_{ij} = \dfrac{x_{ij} - \overline{x}_{ij}}{s_r}$, $\overline{x}_j = \dfrac{1}{m} \sum\limits_{i=1}^{m} x_{ij}$ and the calculation formula of S_r can be expressed as equation (3):

$$s_r = \sqrt{\frac{1}{m\text{-}1} \sum_{i=1}^{m\text{-}1} (x_{ij} - \overline{x}_j)^2} \ . \tag{3}$$

(2) Calculate the covariance matrix Z as equation (4):

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1r} \\ z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{r1} & z_{r2} & \cdots & z_{rr} \end{bmatrix}. \tag{4}$$

Where $z_{ij}\,(i = 1, 2, \cdots, r,\ j = 1, 2, \cdots, r)$ is detailed as equation (5):

$$z_{ij} = \frac{\sum\limits_{i=1}^{m} (x_{ij} - \overline{x}_i)(x_{ij} - \overline{x}_j)}{m\text{-}1}. \tag{5}$$

(3) Calculate the eigenvalues and eigenvectors of equation $|\lambda E{-}Z| = 0$. After this operation, the feature value $\lambda_i\,(i = 1, 2, \cdots, r)$ and feature vector $u = (u_1, u_2, \cdots, u_r)$ are obtained.

(4) Unit orthogonalization feature vector. Rank the eigenvalues in descending order, that is, $\lambda_1' \geq \lambda_2' \geq \cdots \geq \lambda_n' \geq 0$, adjust the eigenvectors according to the corresponding eigenvalues to obtain $u' = (u_1', u_2', \cdots, u_r')$, and use Schmidt's orthogonal method to perform unit orthogonalization on u' to obtain a new eigenvector $Y' = (Y_1', Y_2', \cdots \ Y_n')$.

(5) Calculate the principal component projection. Take the eigenvector as the projection matrix, calculate the projection of the feature set X in the newly generated principal space according to equation (6).

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_k \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1r} \\ l_{21} & l_{22} & \cdots & l_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ l_{k1} & l_{k2} & \cdots & l_{kr} \end{bmatrix} \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_r' \end{bmatrix}. \tag{6}$$

Where, $f_1, f_2, \cdots, f_k$ represent the principal component element of $1 \square k$ and $l_1, l_2, \cdots, l_k$ are the eigenvalue.

(6) The main components in the original data are selected through the cumulative contribution rate, in which the contribution rate of the i-th principal component is:

$$P_i = \frac{\lambda_i}{\sum\limits_{i=1}^{r} \lambda_i}\,(i = 1, 2, \quad ,r) \ . \tag{7}$$

The formula for calculating the cumulative contribution rate of the first k principal components is as follows:

$$\xi = \frac{\sum\limits_{i=1}^{k} \lambda_i}{\sum\limits_{i=1}^{r} \lambda_i} (i = 1, 2, \cdots, r) . \tag{8}$$

Where $\xi(0 < \xi \leq 1)$ is the contribution rate corresponding to the main components of the original data, and when ξ is greater than 0.85, it indicates that the number of former principal components k is determined by the cumulative contribution rate, and the matrix composed of the corresponding sample elements of the principal component elements of the first k elements is shown in equation (9).

$$Y' = \begin{bmatrix} y'_{11} & y'_{12} & \cdots & y'_{1k} \\ y'_{21} & y'_{22} & \cdots & y'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y'_{m1} & y'_{m2} & \cdots & y'_{mk} \end{bmatrix} . \tag{9}$$

4 Feature Selection Based on Principle Component Analysis

4.1 KELM Method

KELM (Kernel extreme learning machine, KELM) [16-18] is a variant of the ELM (Extreme learning machine, ELM) [19-21], which is combined with a kernel function. It not only enhances the learning speed of single hidden layer feedforward neural network but also improves the generalization ability and model stability. Therefore, KELM is used to classify fault types of OLTC.

For the data set $\{u_i, v_i\}$ with any N samples, where, $u_i = [u_{i1}, u_{i2}, u_{i3}, \cdots, u_{in}]^{\mathrm{T}}$, $v_i = [v_{i1}, v_{i2}, v_{i3}, \cdots, v_{in}]^{\mathrm{T}} \in R^n$, n, l and s represent the number of nodes in the input layer, hidden layer and output layer respectively. If the output sample is approximated to zero error by using extreme learning, the process can be represented as follows:

$$v_i = \sum_{j=1}^{l} \beta_j g(w_j u_i + b_j) \qquad j = 1, 2, \cdots, l . \tag{10}$$

Where, β_i is the weight vector between the hidden layer and the input layer, $g(\square)$ is the activation function, w_j is the weight vector of the output layer and the hidden layer, and b_i is the threshold of the j-th neural node in the hidden layer. Convert equation (9) into matrix form as follows:

$$V = H\beta . \tag{11}$$

Where H is the output matrix of the hidden layer,

$$H = \begin{bmatrix} g(w_1 u_1 + b_1) & \cdots & g(w_l u_1 + b_l) \\ \vdots & \ddots & \vdots \\ g(w_1 u_N + b_1) & \cdots & g(w_l u_N + b_l) \end{bmatrix} . \tag{12}$$

Where $\beta = [\beta_1, \beta_2, ..., \beta_l]$ represents the weight matrix, $V = [v_1, v_2, \cdots, v_N]$ represents the output matrix, and the output layer weight vector is expressed as:

$$\beta^* = H^* Y . \tag{13}$$

Where, the calculation formula of H^* is

$$H^* = \mathrm{H}^{\mathrm{T}}(\frac{I}{C}+HH^{T})^{-1}.$$

(14)

In equation (14), C represents the penalty parameter.

In ELM training, the H matrix is generated by random assignment. Due to the uncertainty of random assignment, different H matrices will be generated and different output layer weight vectors are obtained when ELM is employed, resulting in weak generalization ability and stability. Considering the above problems, the kernel function is introduced into ELM, which not only makes the calculation easier but also solves the stability problem of the algorithm so that the algorithm has better generalized performance. The kernel matrix of the KELM algorithm is specifically expressed as follows:

$$\Omega_{ELM} = HH^{\mathrm{T}}\text{：}\ \Omega_{ELMij}=h(x_i)h(x_j) = \kappa(x_i,x_j).$$

(15)

Where, Ω is the symmetric matrix of $N \times N$ and $\kappa(x_i,x_j)$ is the kernel function. In this paper, the RBF (Radial basis function, RBF) [22] kernel function is selected for discussion.

The inspiration of grasshopper algorithm comes from modeling grasshopper foraging behavior in nature [23-27]. The grasshopper's life cycle can be divided into two stages: larvae and adults. Its population characteristics can be characterized mainly by slow movement in a small range during the larval stage and fast jumping motion during the adult stage. This corresponds to the exploration and development of the search process in the algorithm. In the exploration process, the leaping behavior of grasshopper encourages abrupt movements towards a global search, while in the development process, grasshopper leans towards more localized searches in small range.

The modeling process of grasshopper search range is as follows:

$$X = S_i+G_i + A_i.$$

(16)

Where, X_i is the position of the first grasshopper, S_i is the first grasshopper to receive the interaction from another grasshopper, G_i is the gravity influence of the first grasshopper, and A_i is the wind influence of the second grasshopper. If the influence of random environmental factors is considered, the equation (16) is changed to;

$$X_i = r_1S_i + r_2G_i + r_3A_i.$$

(17)

The values of r_1, r_2 and r_3 are random numbers between [0,1].

$$S_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} s(d_{ij})\hat{d}_{ij}.$$

(18)

Where, Eq. 18 conveys the interaction between grasshoppers, where $d_{ij} = |x_i - x_j|$ denotes the distance between the i-th grasshopper and the j-th grasshopper. The variable $\hat{d}_{ij}$ is equivalent to $\dfrac{x_i - x_j}{d_{ij}}$ which expresses the unit vector from the i-th to the j-th.

The S function is defined as a function of s affected by the interaction of other grasshoppers. It is given by the expression:

$$s(\tau) = fe^{\frac{-\tau}{l}} - e^{-\tau}.$$

(19)

Where, when $s(\tau)$ is greater than 0, the grasshoppers attract each other, and the range of τ is called the attraction domain. When $s(\tau)$ is less than 0, the grasshoppers are mutually exclusive, and the range of τ is the exclusion domain. However, when τ is very large (or approaches to infinity), the result is equal to zero [26]. At this point, τ is a non-comfortable distance, so the ranges of the τ and s values become critical. The two parameters of attraction strength f, and scale l determine the distribution of parameters τ. The values are usually $f = 1.5$ and $l = 0.5$.

Given that g is the gravity constant, N is the number of grasshoppers, Eq. (20) which provides the updated location can be replaced by:

$$X_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} s\left(\left|x_j - x_i\right|\right) \frac{x_j - x_i}{d_{ij}} - g\hat{e}_g + u e_w .$$

(20)

Where, $G_i = -g\hat{e}_g$ and $A = \mu\hat{e}_g$, $\hat{e}_g$ is the unit vector pointing to the center of the earth, $\hat{e}_w$ represents the unit vector in the same direction as the wind direction, and u is the drift constant [27].

However, the grasshopper updating position is usually determined by the grasshopper's current position, target value position, and the other grasshopper locations, without considering the influence of gravity and wind. Thus, Eq. (20) is transformed to

$$X_i^d = c_1 \left\{ \sum_{\substack{j=1 \\ j \neq i}}^{N} c_2 \frac{ub_d - lb_d}{2} s\left(\left|x_j^d - x_j^d\right|\right) \frac{x_j - x_i}{d_{ij}} \right\} + \hat{T}_d .$$

(21)

Where, ub_d and lb_d corresponds to the upper and lower bounds of grasshopper in the dimension (d), and $\hat{T}_d$ corresponds to the grasshopper swarm's target position. The coefficient c is calculated by the linear regressive equation

$$c = c_{max} - t \frac{c_{max} - c_{min}}{T_{max}} .$$

(22)

The coefficient c is used to balance the global search ability with the local search ability on the outside and to shrink the comfortable distance, the exclusion area, and the attraction area on the inside. Here, T_{max} is the maximum number of iterations, l is the current number of iterations, $c_{max} = 1$, and $c_{min} = 0.00001$.

4.2 Improved Grasshopper Algorithm Optimization

Like other heuristic algorithms, GOA (Grasshopper algorithm optimization, GOA) global search cannot always be well implemented, and it is easy to fall into local optimization and slow convergence. Therefore, an improved grasshopper algorithm is proposed in this paper. The improved GOA not only has all the advantages of the GOA but also has the advantages of fast convergence speed and high precision. The improved grasshopper algorithm can be divided into three stages:

(1) In this paper, chaotic opposition-learning strategy is used to initialize the population, so as to enhance the diversity of the population and improve the solution efficiency [28]. The chaotic variable is mapped into the solution space to obtain the positive population Y'^+ and the specific formula is as follows:

$$Y'^+ = \begin{bmatrix} y'^+_{1,1} & y'^+_{1\,2} & \cdots & y'^+_{r\,1} \\ y'^+_{2,1} & y'^+_{2\,2} & \cdots & y'^+_{r,2} \\ \vdots & \vdots & \ddots & \vdots \\ y'^+_{m,1} & y'^+_{m,2} & \cdots & y'^+_{m,k} \end{bmatrix} .$$

(23)

Where, $Y_{ij}^{+} = y_{ij}' + \mu(y_{\max,j}'^{+} - y_{\min,j}'^{+})$, $y_{\min,j}'^{+}$ and $y_{\max,j}'^{+}$ is shown in equation (24) and equation (25):

$$y_{\min,j}^{+} = \min_{i \in \{i=1,2,\cdots,k\}} \{y_{i,j}^{+}\}. \tag{24}$$

$$y_{\max,j}^{+} = \max_{i \in \{i=1,2,\cdots,k\}} \{y_{i,j}^{+}\}. \tag{25}$$

In the same way, the opposition population is obtained as follows:

$$Y'^{-} = \begin{bmatrix} y_{1,1}'^{-} & y_{1\,2}'^{-} & \cdots & y_{r\,1}'^{-} \\ y_{2,1}'^{-} & y_{2\,2}'^{-} & \cdots & y_{r\,2}'^{-} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m,1}'^{-} & y_{m,2}'^{-} & \cdots & y_{m,k}'^{-} \end{bmatrix}. \tag{26}$$

Where, $Y_{ij}'^{-} = y_{\min,j}' + y_{\max,j}' - y_{i,j}^{+}$.

The forward population and the reverse population are

$$Y'' = [Y'^{+}; Y'^{-}] = \begin{bmatrix} y_{1,1}'^{+} & y_{1\,1}'^{+} & \cdots & y_{1,k}'^{+} \\ y_{2,1}'^{+} & y_{2,2}'^{+} & \cdots & y_{2,k}'^{+} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m,1}'^{+} & y_{m,2}'^{+} & \cdots & y_{m,k}'^{+} \\ y_{1,1}'^{-} & y_{1,2}'^{-} & \cdots & y_{1,k}'^{-} \\ y_{2,1}'^{-} & y_{2,2}'^{-} & \cdots & y_{2,k}'^{-} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m,1}'^{-} & y_{m,2}'^{-} & \cdots & y_{m,k}'^{-} \end{bmatrix}. \tag{27}$$

The objective function values of the new population Y'' are calculated and sorted, and N individuals with the best fitness are selected to form a new initial population X, which is expressed in equation (28)

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,r} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,r} \end{bmatrix}. \tag{28}$$

The fitness function is expressed as follows [25]:

$$Fitness = 1 - \frac{y_t}{y_{sum}} \times 100\%. \tag{29}$$

Where, *Fitness* represents the classification accuracy, y_t is the correct number of classifications, and y_{sum} means the total number of samples.

(2) Considering that the inertia weight based on cloud model can better balance local search and global search, this paper introduces it into GOA for improvement. The specific steps are as follows [26].

① Firstly, calculate the fitness value of all grasshopper and calculate the average fitness value. Assumption the total number of grasshopper is N, and in the t-th iterations, the grasshopper individual X_i fitness value is f_i, and the

corresponding individuals value fitness is $\overline{f}_0^1 = \frac{1}{N} \sum_{i=1}^{N} f_i$.

② Secondly, the individuals can be divided into three sub-populations: when the fitness value $f_i > \overline{f}_1^0$, the average fitness is $\overline{f}_1^1$. When the fitness $f_i < \overline{f}_1^0$,the fitness value is $\overline{f}_1^2$, and the best fitness value is f_{min}. According to the above inertia weight strategy, grasshopper population can be divided into three different types of subgroups, and the expression is as follows:

$$E'_n = \text{normrnd}(E_n, H_e), \quad w = 0.9 - 0.2 \exp\left[-\frac{f_1 - E_x}{2E_n^2}\right]. \tag{30}$$

Where, the E_x represents the point of qualitative concept, Entropy E_n is a measure of the fuzziness of qualitative concepts. Super entropy he is a measure of the randomness of membership function. In this paper the $w = 0.9$ [26].

Incorporating the inertia weight into equation (21), then equation (21) is expressed as follows:

$$X_i^d = c_1 \left(\sum_{j=1\neq i}^{N} c_2 \left(\frac{|ub_d - lb_d|}{2}\right) s\left(|x_j - x_i|\right) \frac{x_j - x_i}{d_{ij}} \right) + \omega \cdot \hat{T}_d . \tag{31}$$

(3) In the later stage of iteration, the grasshoppers are prone to premature since their individuals in the population gather to the optimal individual position. Therefore, the individual chaotic search strategy is adopted to reduce the probability of GOA falling into local premature. The specific steps are as follows:

① Suppose the optimal individual position vector is $X_{best}^t = \{X_{best,i}^t, i = 1, 2, \cdots, k\}$, where $X_{best,i}^t$ represents the optimal individual position of the i-th dimension in the t-th iteration. Map each element of the optimal individual position vector into the interval (-1,1) and the specific formula is as follows:

$$y_{best,k} = \frac{2(X_{best,i}^t - y'_{min,j})}{(y'_{max,j} - y'_{min,j})} . \tag{32}$$

Where, $y_{best,k}$ represents the optimal individual variable corresponding to the dimension = k and this variable is need to be searched,

② Iterate $y_{best,k}$ through the self-mapping function to generate a chaotic sequence $y_{best,k}^t (t = 1, 2, \cdots, t_{max})$, where the self-mapping function expression is as follows:

$$y_{best,k}^t = 1 - 2(y_{best,k}^{t-1})^2 . \tag{33}$$

③ Use equation (33) to inversely transform $y_{best,k}$ into the original search neighborhood to generate a new individual vector $X''_{best,k}$:

$$X''_{best,k} = \frac{1}{2}(y'_{max,j} - y'_{min,j}) \times (y'_{best,k} + 1) . \tag{34}$$

④ Compare the fitness function value of X_{best} with that of $X''_{best,k}$ and update the current optimal individual; Judge whether the maximum number of chaotic iterations is reached [27]. If so, terminate the chaotic search, otherwise, return to step ② .

4.3 IGOA Optimizes KELM Parameters

In this paper, IGOA is used to evaluate the kernel parameter κ and penalty parameter C of KELM. Fig. 1 shows the process of optimizing KELM by IGOA. The specific optimization steps are as follows:

(1) Initialize the grasshopper population N and other related parameters (C_{min}, C_{max});

(2) Use the chaotic opposition learning strategy to initialize the population to obtain the forward population and the reverse population, and sort them according to formula (31) [27]. Among them, the N individuals with the best fitness reconstitute the new initial population X and record the best individual among them as T_d;

(3) Enter the iterative loop, update the position of each individual according to formula (32). Individuals beyond the upper and lower bounds are re-initialized, the positions of all grasshopper individuals are calculated, and in the later iterations, individual chaotic search strategies are used for grasshopper individuals to reduce the probability of falling into local precocity;

(4) Update the best individual T_d and iteration time [26], repeat Step (3) until the number of iterations reaches the maximum number of iterations T_{max}, then output the optimal individual;

(5) The IGOA-KELM fault diagnosis model is constructed according to the best two optimal parameters. The test data are input into the diagnosis model for classification and output results
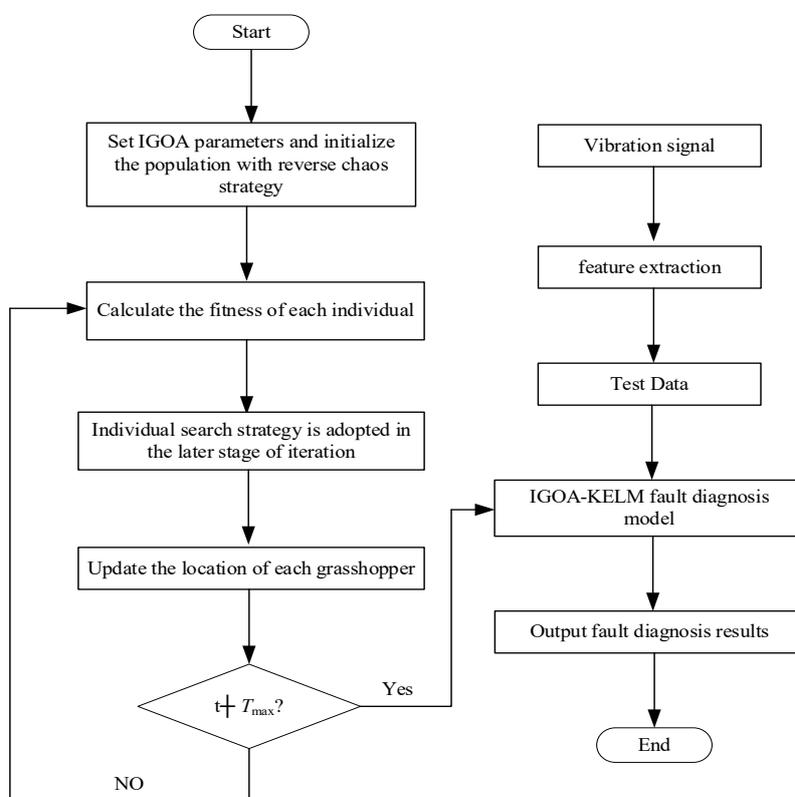

Fig. 1. IGOA-KELM fault identification flow chart

4.4 Fault Classification

The essence of OLTC fault diagnosis is a task of pattern recognition [26]. The operation of the tap changer in different states is regarded as different categories. Based on the time-frequency processing, PCA and IGOA-KELM, this paper proposes a new OLTC fault diagnosis method. First, use the time-frequency methods to extract the features of the vibration signal and feature fusion is constructed; then, the PCA is employed to select the features to obtain the optimal initial feature matrix; finally, the GOA algorithm is modified to improve the stability of the algorithm, and the optimal feature matrix is used as the input of the diagnosis model to obtain the optimal parameters of KELM. The established fault diagnosis method based on IGOA-KELM is shown in Fig. 2.
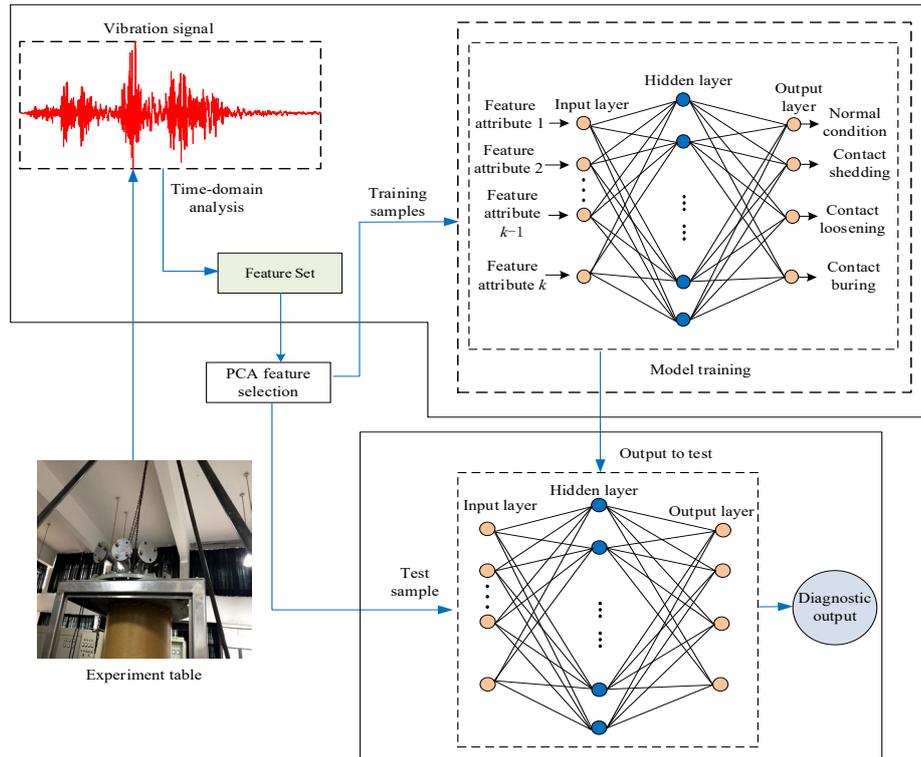
Fig. 2. KELM failure set

5 Experiment and Result Analysis

5.1 Experimental Setup

The OLTC type used in this simulation experiment is Huaming CM111-50-63B-10193W. The used sensor type is LC0151, which has the advantages of high resolution, low noise and strong anti-interference ability. The signal frequency is concentrated within 20kHz, so the sampling frequency is configured as 50kHz. the contact switching time is about 100~140ms, the sampling takes 6880 points, and the contact switching time is concentrated in 0~41ms. The installation position of the acceleration sensor is on the top of the tap changer, and the action mechanism is extracted from the inside of the tap changer by hoisting on a tripod. The installation scheme of the sensor is shown in Fig. 3(a), and the and the contact replacement position is shown in Fig. 3(b), the (b) contact burning is shown in Fig. 3(c).

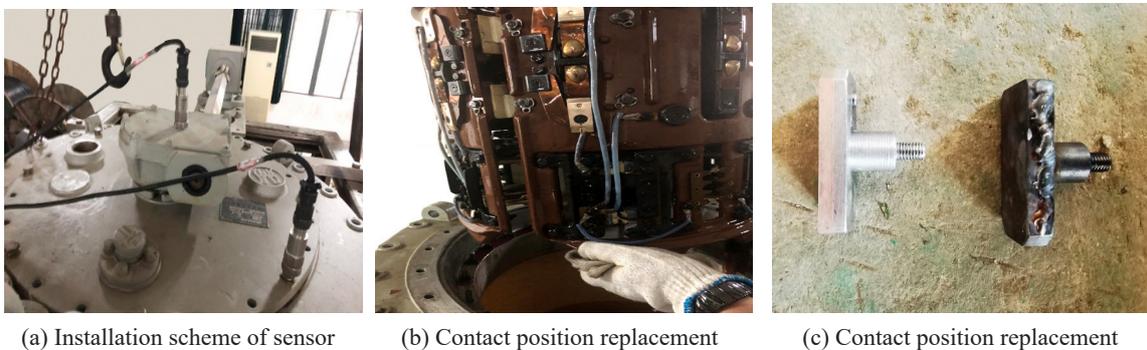(a) Installation scheme of sensor (b) Contact position replacement (c) Contact position replacement

Fig. 3. Experimental platform and contact fault setting

The four different types of vibration signals are shown in Fig. 4. The four typical working conditions are normal working conditions, contact burning, contact loosening, and contact shedding, respectively. There are 50 recordings for each type working condition, and the number of signal points in each recording is 6880. Each type is randomly selected according to the proportion of 50%, and 75% for training.
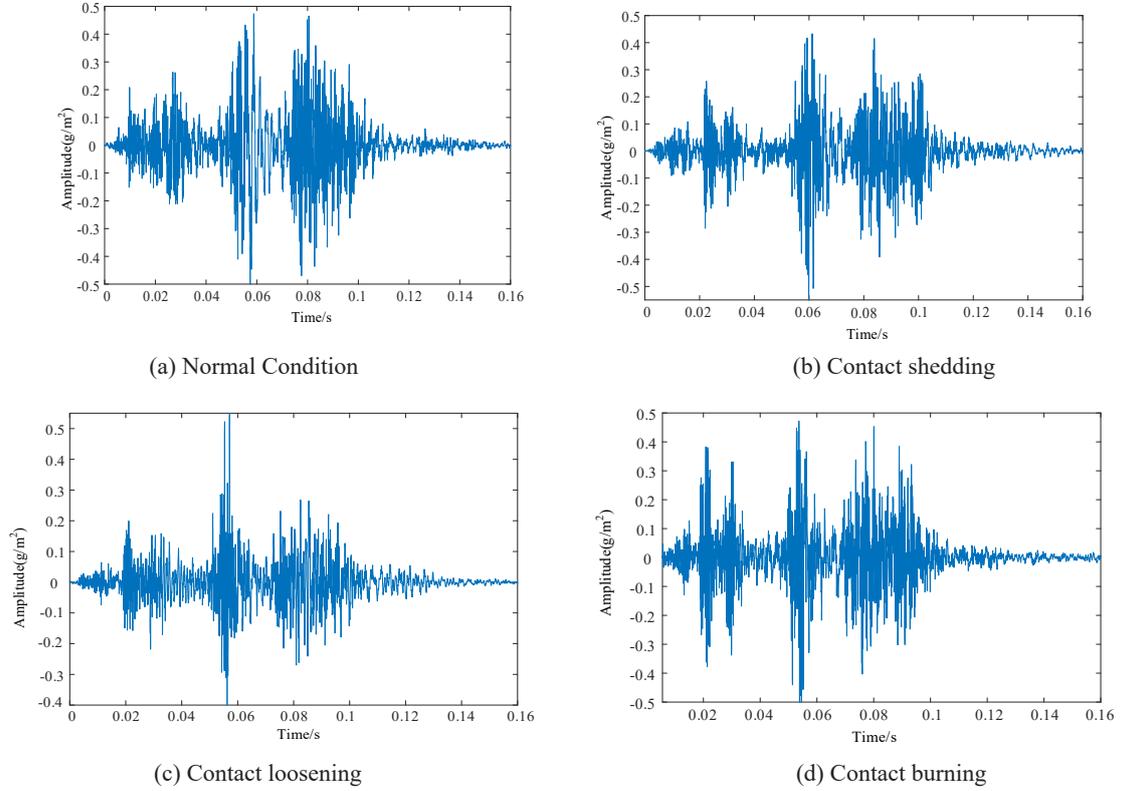

(a) Normal Condition

(b) Contact shedding

(c) Contact loosening

(d) Contact burning

Fig. 4. Vibration signals corresponding to four different operational states of OLTCs

5.2 Analysis of Experimental Results

Evaluation Index. In order to verify the proposed method, the average value, standard deviation, maximum value, and minimum value are used as the evaluation indicators of the algorithm [2]. The specific expressions are as follows:

(1) Average value:

$$Average = \frac{1}{N_{num}} \sum_{i=1}^{N_{num}} accuracy \,. \tag{35}$$

(2) Minimum value:

$$Minmum = \min_{1 \le j \le Num} accuracy \,. \tag{36}$$

(3) Maximum value:

$$Maximum = \max_{1 \le j \le Num} accuracy \,. \tag{37}$$

(4) Standard deviation:

$$STD = \sqrt{\frac{1}{N_{num} - 1} \sum_{i=1}^{N_{num}} (accuracy - Average)^2} . \tag{38}$$

5.3 Calculation Results

To verify the effectiveness of this method, ELM, SVM (support vector machine, SVM) [28], Particle Swarm Optimization-KELM (PSO-KELM) [29], Grey Wolf Optimizer-KELM (GWO-KELM) [30], Whale optimization algorithm-KELM (WOA-KELM) [31], and GOA-KELM [26] are employed for comparison. The parameters of PSO algorithm are $c_1 = 1.442$, $c_2 = 1.532$, population size $N = 20$, and the maximum number of iterations $T_{max} = 100$; The parameters of WOA, GOA and GWO are configured as population size $N = 20$, maximum number of iterations $T_{max} = 100$.
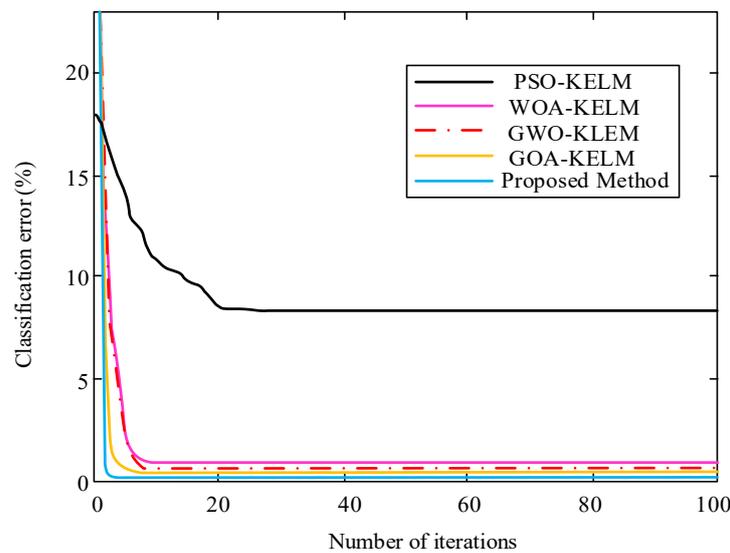

Fig. 5. Optimization iteration diagram

The 2 parameters of KELM are optimized as $[\kappa, C] = [86.73, 0.1]$, and the optimization iteration diagram is shown in Fig. 5. As can be observed from the figure, the classification error gradually decreases to 0 with the increase of the number of iterations. Among the five algorithms, the algorithm proposed in this paper has the best convergence speed, followed by GOA, GWO, and WOA, respectively. The algorithm with the worst effect is PSO.

Without considering the PCA, the eight methods are compared, and 50%, and 75% of the samples are selected for training, respectively. The calculation results are shown in Fig. 6, which are the results include 10 tests. It can be seen from the figure that with the continuous increase of sampling proportion, the calculation accuracy of diagnostic results continues to improve. When the sampling rate is 75%, the accuracy rate is the highest compared with the other two sampling results. It shows that under the diagnosis results of small samples, the more training samples, the better the test classification results. From the maximum, minimum, and average values in Fig. 6(a) to Fig. 6(b), it can be further found that the histogram distribution of IGOA is the highest compared with the other seven methods, indicating that the diagnostic accuracy of this paper is the highest and the diagnostic performance is the best.
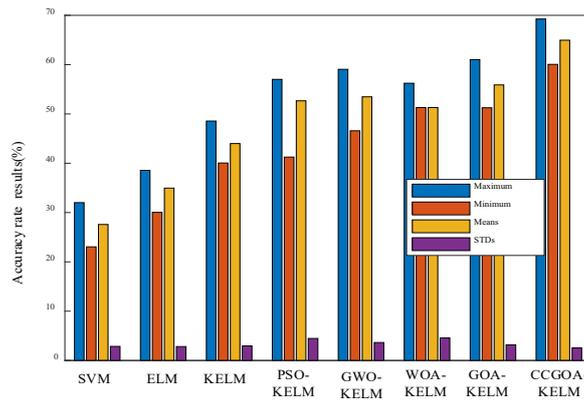
(a) Evaluation index results without PCA sampling of 50%
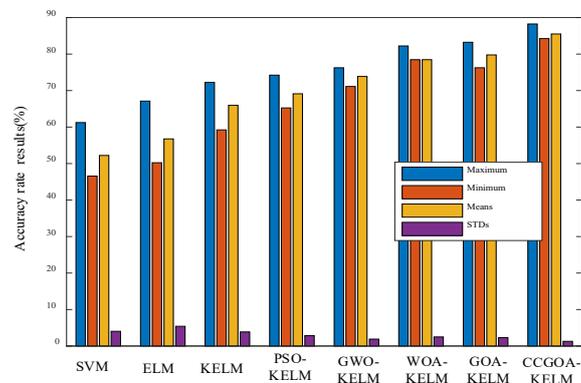
(b) Evaluation index results without PCA sampling of 75%

Fig. 6. Index results without PCA

Fig. 7 shows the results of 8 different algorithms obtained under the condition of PCA. The calculation results in Fig. 7 are greatly improved compared with those in Fig. 6(b), which shows that the PCA feature selection method is effective. In Fig. 7, the maximum, minimum, and average values of IGOA are the largest and the standard value is the smallest, indicating that the classification performance of this method is the best among the eight methods.
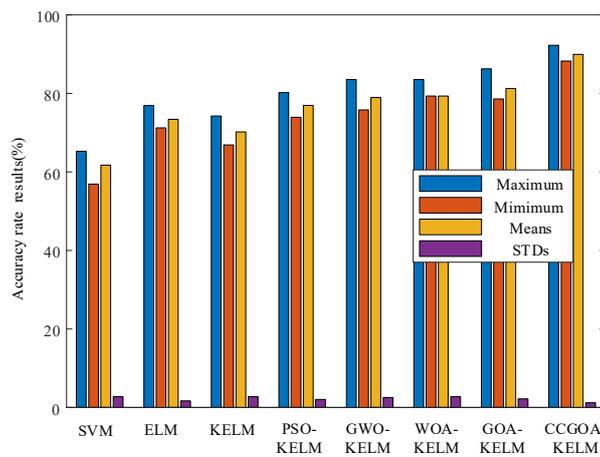

Fig. 7. Results under the condition of PCA

Fig. 8 The figure shows the calculation results of 50 times of different algorithms under the same fault, where the abscissa represents the fault type: 1, 2, 3, 4 are respectively represented as normal operation, contact loosening, contact burning, contact shedding. The ordinates 1, 2, 3, 4, 5, 6, 7, 8 respectively represent the algorithms SVM, ELM, PSO-KLEM, GWO-KLEM, WOA-KELM, GOA-KELM, CCGOA-KELM. CCGOA-KELM still has the highest fault recognition rate among all methods. The recognition rate of contact shedding is 91.25%, and the recognition rate of contact burning is relatively low, but it also reaches 89.45%. It further shows that this method has good classification effect under the same type of fault and different types of faults.
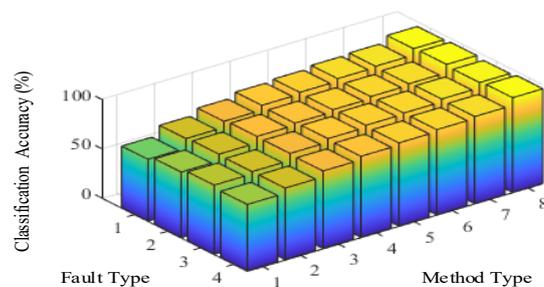

Fig. 8. Comparison of different algorithms under the same fault

6 Conclusion and Future Study

Aiming at the traditional OLTC diagnosis methods have the defects of limited feature information and low classification accuracy, this paper proposes an OLTC fault diagnosis method based on time-domain analysis and IGOA-KELM. The results are as follows:

1) The time-frequency analysis method is used to perform multi-dimension feature extraction for the OLTC vibration signal, which effectively alleviates the disadvantage of limited feature information in the traditional fault diagnosis method;

2) The employed PCA method can significantly reduce redundant information and correlation existing in the parameters, and thus obtain a compact feature;

3) The OLTC fault diagnosis model based on IGOA-KELM achieves a high recognition rate. Specifically, the two-parameter optimization of KELM conducted by IGOA effectively improves the fluctuation stability of KELM on the network structure caused by random assignment, and significantly improves the classification performance of the KELM algorithm.

Fundings

References

[1] Y. Xu, C. Zhou, J. Geng, S. Gao, P. Wang, A Method for Diagnosing Mechanical Faults of on-Load Tap Changer Based on Ensemble Empirical Mode Decomposition, Volterra Model and Decision Acyclic Graph Support Vector Machine, IEEE Access 7(2019) 84803-84816.

[2] Y. Yan, H. Ma, M. Wen, S. Dang, H. Xu, Multi-Feature Fusion-Based Mechanical Fault Diagnosis for On-Load Tap Changers in Smart Grid With Electric Vehicles, IEEE Sensors Journal 21(14)(2021) 15696-15708.

[3] E. Rivas, J.C. Burgos, J.C. Garcia-Prada, Condition Assessment of Power OLTC by Vibration Analysis Using Wavelet Transform, IEEE Transactions on Power Delivery 24(2)(2009) 687-694.

[4] P. Kang, D. Birtwhistle, Condition monitoring of power transformer on-load tap-changers. Part I: Automatic condition diagnostics, IEE Proceedings: Generation, Transmission and Distribution 148(4)(2001) 301-306.

[5] P. Kang, D. Birtwhistle, Condition monitoring of power transformer on-load tap-changers. Part II: detection of ageing from vibration signatures, IEE Proceedings: Generation, Transmission and Distribution 148(4)(2001) 307-311.

[6] E. Rivas, J.C. Burgos, J.C. Garcia-Prada, Vibration analysis as method diagnosis for power on load tap changers,

Ingenieria 13(1)(2008) 8-14.

[7] E. Rivas, J.C. Burgos, J.C. García-Prada, Vibration Analysis Using Envelope Wavelet for Detecting Faults in the OLTC Tap Selector, IEEE Transactions on Power Delivery 25(3)(2010) 1629-1636.

[8] P. Chen, X. Ma, F. Wang, F. Xie, D. Zhou, Mechanical Fault Diagnosis of On-load Tap Changer of Power Transformer by Clustering Algorithm, 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 2018.

[9] X. Zhou, F. Wang, J. Fu, J. Lin, Z. Jin, Mechanical Condition Monitoring of On-load Tap Changers Based on Chaos Theory and K-means Clustering Method, Proceedings of the CSEE 35(6)(2015) 1541-1548.

[10] R. Duan, F. Wang, Fault Diagnosis of On-Load Tap-Changer in Converter Transformer Based on Time–Frequency Vibration Analysis, IEEE Transactions on Industrial Electronics 63(6)(2016) 3815-3823.

[11] J.J. Saucedo-Dorantes, M. Delgado-Prieto, R.A. Osornio-Rios, R.J. Romero-Troncoso, Multifault diagnosis method applied to an electric machine based on high-dimensional feature reduction, IEEE Transactions on Industry Applications 53(3)(2017) 3086-3097.

[12] M. Van, H.-J. Kang, Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection, IET Science, Measurement & Technology 9(6)(2015) 671-680.

[13] N. Kwak, Principal Component Analysis Based on L1-Norm Maximization, IEEE Transactions on Pattern Analysis and Machine Intelligence 30(9)(2008) 1672-1680.

[14] Y. Kim, Detection of Eye Blinking Using Doppler Sensor With Principal Component Analysis, IEEE Antennas and Wireless Propagation Letters 14(2015) 123-126.

[15] P. Manfredi, S. Grivet-Talocia, Fast Stochastic Surrogate Modeling via Rational Polynomial Chaos Expansions and Principal Component Analysis, IEEE Access 9(2021) 102732-102745.

[16] L. Wang, M.J. Er, S. Zhang, A Kernel Extreme Learning Machines Algorithm for Node Localization in Wireless Sensor Networks, IEEE Communications Letters 24(7)(2020) 1433-1436.

[17] F. Luo, G. Liu, W. Guo, G. Chen, N. Xiong, ML-KELM: A Kernel Extreme Learning Machine Scheme for Multi-Label Classification of Real Time Data Stream in SIoT, IEEE Transactions on Network Science and Engineering 9(3)(2022) 1044-1055.

[18] Y. Zeng, Reply to "Comments on 'Traffic Sign Recognition Using Kernel Extreme Learning Machines With Deep Perceptual Features, IEEE Transactions on Intelligent Transportation Systems 20(10)(2019) 3762-3764.

[19] M. Min, X. Chen, Y. Xie, Constrained voting extreme learning machine and its application, Journal of Systems Engineering and Electronics 32(1)(2021) 209-219.

[20] R. Zheng, F. Pan, Soft sensor of glutamate concentration using extreme learning machine, in: Proc. of the 11th World Congress on Intelligent Control and Automation, 2014.

[21] J. Tang, C. Deng, G. Huang, Extreme Learning Machine for Multilayer Perceptron, IEEE Transactions on Neural Networks and Learning Systems 27(4)(2016) 809-821.

[22] F. Wang, Z.Q. Chao, L.B. Huang, H.-Y. Li, C.-Q. Zhang, Trajectory tracking control of robot manipulator based on RBF neural network and fuzzy sliding mode, Clustering Computing 22(3)(2019) 5799-5809.

[23] M.A. El-Shorbagy, A.M. El-Refaey, Hybridization of Grasshopper Optimization Algorithm With Genetic Algorithm for Solving System of Non-Linear Equations, IEEE Access 8(2020) 220944-220961.

[24] J. Liu, A. Wang, Y. Qu, W. Wang, Coordinated Operation of Multi-Integrated Energy System Based on Linear Weighted Sum and Grasshopper Optimization Algorithm, IEEE Access 6(2018) 42186-42195.

[25] P. Chen, X. Zhao, Q. Zhu, A Novel classification method based on ICGOA-KELM for fault diagnosis of rolling bearing, Applied Intelligence 50(9)(2020) 2833-2847.

[26] Y. Yan, H. Ma, Z. Li, An Improved Grasshopper Optimization Algorithm for Global Optimization, Chinese Journal of Electronics 30(3)(2021) 451-459.

[27] H.R. Tizhoosh, Opposition-based learning: A new scheme for machine intelligence, International Conference on Computational Intelligence for Modelling, in: Proc. Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05), 2005.

[28] Y. He, C.Y. Du, C.B. Li, A.G. Wu, Y. Xin, Sensor Fault Diagnosis of Superconducting Fault Current Limiter With Saturated Iron Core Based on SVM, IEEE Transactions on Applied Superconductivity 24(5)(2014) 1-5.

[29] S.K. Baliarsingh, S. Vipsita, K. Muhammad, B. Dash, S. Bakshi, Analysis of high-dimensional genomic data employing a novel bio-inspired algorithm, Applied Soft Computing 77(2019) 520-532.

[30] A.A. Heidari, R.A. Abbaspour, H. Chen, Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training, Applied Soft Computing 81(2019) 105521.

[31] X.-E. Wang, Y. Meng, J. Yang, X. Huang, F. Wang, H. Xu, Optimal kernel extreme learning machine model for predicting the fracture state and impact response of laminated glass panels, Thin-Walled Structures 162(2021) 107541.