Mathematical Biosciences and Engineering

MBE, 20(3): 4877–4895.
DOI: 10.3934/mbe.2023226
Received: 21 October 2022
Revised: 11 December 2022
Accepted: 16 December 2022
Published: 04 January 2023

http://www.aimspress.com/journal/MBE

Research article

Research on imbalanced data fault diagnosis of on-load tap changers based on IGWO-WELM

Yan Yan[1,2,*], Yong Qian[1], Hongzhong Ma[2] and Changwu Hu[3]

[1] State Grid Ningxia Electric Power Research Institute, Yinchuan 750002, China
[2] College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China
[3] Zhongwei Power Supply Company of State Grid Ningxia Electric Power Co. Ltd, Zhongwei 755000, China

* **Correspondence:** Email: yanyan503986706@163.com.

Abstract: Aiming at the problem of on-load tap changer (OLTC) fault diagnosis under imbalanced data conditions (the number of fault states is far less than that of normal data), this paper proposes an OLTC fault diagnosis method based on an Improved Grey Wolf algorithm (IGWO) and Weighted Extreme Learning Machine (WELM) optimization. Firstly, the proposed method assigns different weights to each sample ac-cording to WELM, and measures the classification ability of WELM based on G-mean, so as to realize the modeling of imbalanced data. Secondly, the method uses IGWO to optimize the input weight and hidden layer offset of WELM, avoiding the problems of low search speed and local optimization, and achieving high search efficiency. The results show that IGWO-WLEM can effectively diagnose OLTC faults under imbalanced data conditions, with an improvement of at least 5% compared with existing methods.

Keywords: imbalanced data; OLTC; IGWO; WELM; IGWO-WELM

1. Introduction

An on-load tap changer (OLTC) is the core component in a load-ratio voltage transformer and the only movable component in a transformer. As the mechanical structure of an on-load tap-changer is complicated and the voltage is frequently regulated, it experiences frequent faults. According to international transformer fault data, the faults caused by OLTCs account for more than 20% of total transformer faults, and the fault rate is on the rise [1]. In view of the characteristics of OLTCs, such as high failure rate, frequent actions, and wide influence range, it is necessary to further improve the online monitoring of OLTCs to ensure the high-quality, safe and stable operation of power systems [2].

As one of the most effective fault diagnosis methods at present, the vibration analysis method is widely used in mechanical fault diagnosis because of its advantages of simple operation, accurate analysis and nondestructive testing [3,4]. This method has already achieved good results in large power equipment such as transformers. The application of the vibration analysis method to OLTC mechanical condition detection started late, and was first proposed by Bengtsson and others of ABB Company in the 1990s, and has been paid increasing attention by Chinese and foreign scholars since then [5]. How to extract the effective characteristic quantity from OLTC vibration and build a fault diagnosis model for OLTCs is the key to this detection technology. For example, P. Kang et al. proposed to judge the mechanical state of an OLTC by using the characteristics of envelope [6] by setting typical OLTC mechanical faults, collecting vibration signals to extract envelope, and distinguishing different states. Continuous Wavelet Transform (CWT) has been used to analyze the vibration signal of OLTCs [7,8]. The state database of the OLTC is established by a two-dimensional wavelet coefficient "ridge distribution map", and the different working conditions of the OLTC are judged. In ref [9], the phase space reconstruction method is used to reconstruct the vibration signal from low dimension to high dimension, and the phase trajectory diagram of the vibration signal is clustered by K-means, and different working conditions of the OLTC are judged. Ref [10] and Ref [11] use the empirical mode decomposition method and variational mode decomposition method to decompose the OLTC vibration signal to obtain natural frequency and then applies an optimized correlation vector machine and support vector machine to classify and diagnose the OLTC under different working conditions. The OLTC fault diagnosis model based on the vibration signal power spectrum and hidden Markov model is proposed in ref [12]. The experimental results show that the method has a good classification effect. In ref [13], an OLTC mechanical fault diagnosis model based on homologous and heterogeneous data fusion is proposed. Through the fusion of four features and image features, the fusion data is used to train the support vector machine for diagnosis. The results show that this method has good accuracy. However, in an actual fault diagnosis environment, the data collected on the spot generally have the problem of imbalanced distribution of categories (more samples in the normal state and fewer samples in the abnormal state), which cannot meet the requirements of model training.

However, in the actual fault diagnosis environment, the data collected on site usually has the problem of imbalanced distribution of categories (more samples under normal conditions and fewer samples under abnormal conditions), which cannot meet the requirements of model training. Therefore, it is necessary to further adopt appropriate mathematical models to solve the problem of OLTC imbalanced data failure. Many scholars have studied imbalanced data using AdaBoost, CNN, LSTM, Smote and WELM. Among them, refs [14–16] deals with data imbalance through the combination of AdaBoost and SVM, GUS-LSTM and an improved AdaBoost algorithm, respectively. Refs [17,18] uses layered CNN and adaptive cost sensitive CNN to train samples as data, so as to solve the problem of imbalanced data. Ref [19] uses the advantages of AdaBoost adaptive weight assignment, and combines with CNN to solve the problem of imbalanced data. Ref [20] uses the advantage of time perception to improve the LSTM model, so as to solve the problem of imbalanced data. Refs [21–23] uses K-medoids-Smote, ACC Smote and Smote ASVM to deal with the problem of imbalanced data, and the results show that they have high classification accuracy. Ref [24] and ref [25] respectively use the PSO algorithm and DA to optimize WELM correlation against the defects of WELM. The results showed that the classification accuracy of the optimized WELM was significantly improved compared with the unmodified WELM.

The WELM method is used to solve the problem of OLTC imbalanced data based on the idea of ref [21]. Based on the advantages of GWO in parameter optimization, this paper uses the GWO algorithm to optimize WELM input weight and hidden layer offset parameters. However, similar to other methods, in the late iteration period of GWO algorithm, the grey wolf individual search speed gradually decreases, and the overall convergence is premature, which increases the probability of falling into local optimum. Considering that the PSO algorithm has better search ability and higher execution strategy, this paper introduces the PSO algorithm into the update equation of the grey wolf algorithm, and proposes the improved grey wolf optimization (IGWO) and OLTC fault diagnosis method of WELM (IGWO-OLTC for short). As an improvement strategy, IGWO-OLTC can improve the search ability and development ability of the whole algorithm, and reduce the probability of falling into local optimum.

This paper presents an OLTC fault diagnosis method based on improved Grey Wolf Optimization (IGWO) algorithm and WELM. The research consists of three parts: 1) Aiming at the problem that the classification results of traditional machine learning algorithms are not accurate when dealing with imbalanced data, a WELM based OLTC fault diagnosis model is proposed; 2) Because WELM is easily affected by input weight and hidden layer deviation, GWO is used to optimize WELM; 3) Considering that the GWO can easily fall into local optimum and the convergence speed is slow, the particle swarm optimization algorithm is used to optimize it, and the IGWO-WELM fault diagnosis model is proposed. Through the analysis of simulation data and experimental data, the proposed fault diagnosis model has high accuracy.

2. IGWO method

2.1. GWO Method

Marjiali et al. proposed a new swarm intelligence algorithm based on the tightly organized system and hunting behavior of grey wolves, which includes three parts: tracking prey, surrounding prey, attacking prey, and other optimization processes, summarized as follows [26–28]:

1) Rank stratification of wolf pack:

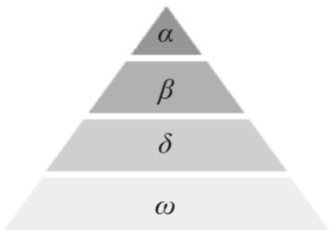

Figure 1. Hierarchy of grey wolf individuals.

Grey wolves mainly live in groups, and the group follows the social hierarchy, as shown in Figure 1. It can be seen from the figure that the α Wolf is the leader of the social group and is mainly responsible for making decisions about activities such as predation, while the rest of the wolves obey the command of the α Wolf. Level 2: β Wolf, obeying and assisting α Wolf, can dominate all the wolves except for *α* Wolf. Level 3: *δ* Wolf, obeying the arrangement of *α* and *β* Wolf at the same time, can dominate the rest of the remaining wolf pack, and rank *ω* is the lowest level. The overall predation behavior of grey wolves is led by *α* wolves, and the task of other wolves is to besiege the prey.

2) Surrounding prey:

Grey wolves surround their prey as they hunt. The mathematical model of encircling prey is as follows:

$$D = \left| C \cdot X_p(t) - X(t) \right| \tag{1}$$

where $X(t)$ represents the position of grey wolves, and X_p represents the position vector of prey:

$$X(t+1) = X_p - A \cdot D \tag{2}$$

where A and C represent coefficient vectors, and the calculation formula is as follows:

$$A = 2a \cdot (r_1 - 1) \tag{3}$$

$$C = 2r_2 \cdot t \tag{4}$$

where t represents the current number of iterations, and $a = 2\ (1\text{-}t/T_{\max})$ represents that the variable decreases linearly from 2 to 0, r_1, r_2 $\in [0,1]$ during the iteration process.

3) Hunting prey:

Grey wolves can identify prey and surround it. The search process is *α* Wolf commands and leads, *β* and *δ* sometimes, they will take part in hunting. Hypothesis *α*, *β* and *δ* The wolf can have a deeper understanding of the potential location of prey, and accordingly, during the algorithm iteration process, save the best location of the three wolves in the current population, and mark them as *α*, *β* and *δ*. Then, according to the position of the three parameters ϖ Wolf individuals are updated, and the mathematical model is as follows:

$$\begin{aligned} X_1(t+1) &= X_\alpha(t) - A_1 \cdot \left| C_1 \cdot X_\alpha(t) - A_1 \cdot X(t) \right| \\ X_2(t+1) &= X_\beta(t) - A_2 \cdot \left| C_2 \cdot X_\beta(t) - A_2 \cdot X(t) \right| \\ X_3(t+1) &= X_\delta(t) - A_3 \cdot \left| C_3 \cdot X_\sigma(t) - A_3 \cdot X(t) \right| \end{aligned} \tag{5}$$

$$X(t+1) = \frac{X_1(t+1) + X_2(t+1) + X_3(t+1)}{3} \tag{6}$$

where X represents the position of the grey wolves. When $|A| > 1$, the grey wolves will try to disperse in each area to search for prey. When $|A| < 1$, the wolves will search for prey in a predetermined area.

2.2. IGWO method

The GWO algorithm has been successfully applied in the fields of job shop scheduling, power system analysis, economic forecasting, etc. However, like other algorithms, the GWO is prone to fall into the local optimum and has a slow convergence speed [28]. Therefore, in order to improve the global convergence and convergence speed, this paper uses the Particle Swarm Optimization (PSO) algorithm to improve the grey wolf algorithm, namely IGWO [27]. The main reason for choosing the PSO algorithm is that the search process is simple and easy to implement, and the convergence speed and search speed are fast. The specific formula is as follows:

$$v_i(t+1) = \varpi X_i(t) + b_1 rand \cdot (P_{gbest,\ i}(\mathrm{t}) - X_i(t)) + \mathrm{b}_2 rand \cdot (P_{gbest,_i}(\mathrm{t}) - X_i(t)) \tag{7}$$

$$X(t+1) = X_i(t) + v_i(t+1) \tag{8}$$

where, b_1 and b_2 are learning factors, $P_{\mathrm{gbest,t}}$ and t are the best positions experienced by the i-th grey wolf individual, ω is the inertial weight, and the inertial weight formula is as follows:

$$\varpi = -(\varpi_{\max} - \varpi_{\min})t / T_{\max} \tag{9}$$

Where, $\varpi_{\max}$ is the maximum weight value, $\varpi_{\min}$ is the minimum weight value, and $T_{\max}$ is the maximum number of iterations.

2.3. IGWO algorithm

In order to verify the effectiveness of the algorithm, eight common standard test functions are used in this paper to verify the IGWO, GOA, PSO, MFO, GWO and SCA algorithms [26,29]. The test function expressions are shown in Table 1. In order to verify the effectiveness of the proposed algorithm, the average value, the lowest value, the best fitness value, the standard deviation, the precision rate and the optimization success rate are used as evaluation indexes to calculate.

Table 1. Test functions.

Functions	Range	Dim	$f_{\min}$
$F_1(x) = \sum_{i=1}^{n} x_i^2$	[−100,100]	30	0
$F_2(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	[−10,10]	30	0
$F_3(x) = \sum_{i=1}^{n} (\sum_{j=1}^{i} x_j)^2$	[−100,100]	30	0
$F_4(x) = \max_i \{\lvert x_i \rvert, 0 \le x_i \le n\}$	[−100,100]	30	0
$F_5(x) = \sum_{i=1}^{n} i x_i^2 + \mathrm{random}[0.1]$	[−1.28,1.28]	30	0
$F_6(x) = -20\exp[-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) - \exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)] + 20 + \mathrm{e}$	[−32,32]	30	0
$F_7(x) = \frac{1}{4000}\sum_{i=1}^{D} x_i^2 \cdot \prod_{i=1}^{D}\cos(\frac{x_i}{\sqrt{x_i}}) + 1$	[−600,600]	30	0
$F_8 = \frac{\pi}{n}\{10\sin^2(\pi y_1) + \sum_{j=1}^{n-1}(y_i - 1)^2[1 + 10\sin^2(\pi \mathrm{y}_i + 1)] + (y_n - 1)^2\} + \sum_{i=}^{n} u(x_i, 10, 100, 4)$, $y_i = \frac{x_i + 1}{4}, u(x_i a, k, m) = \begin{cases} k(x_i - a)^m, x_i > a \\ 0, -a < x_i < a \\ k(-x_i - a)^m, x_i < -a \end{cases}$	[−50,50]	30	0

This paper tests each function in the table 20 times, including 30 algorithm populations and 500 iterations. The final calculation results for IGWO and GOA, PSO, MFO, GWO and SCA, are shown in Table 2.

Table 2. Test functions F1−F8.

Functions	Index	IGWO	GWO	PSO	MFO	GOA	SCA
F1	Best	7.28×10^{-11}	5.06×10^{4}	3.25×10^{-5}	1.850	7.1251	0.8100
	Worst	3.24×10^{-7}	6.77×10^{4}	3.71×10^{-4}	99.453	135.78	10.110
	Ave.	1.14×10^{-7}	6.15×10^{4}	1.42×10^{-4}	23.23	41.65	6.1320
	STD	1.43×10^{-7}	6.47×10^{3}	1.44×10^{-4}	42.650	53.00	3.8269
	SR%	100	0	100	0	0	0
F2	Best	6.14×10^{-10}	3.140×10^{6}	0.0093	20.08	7.464	0.0053
	Worst	5.93×10^{-8}	5.960×10^{12}	0.0300	70.00	26.02	0.0300
	Ave.	1.98×10^{-8}	2.950×10^{12}	0.0156	48.03	15.24	0.0156
	STD	2.75×10^{-8}	2.756×10^{12}	0.0093	21.64	7.354	0.0093
	SR%	100	0	0	0	0	0
F3	Best	2.61×10^{-5}	8.16×10^{4}	39.61	8.67×10^{3}	1.94×10^{3}	2.38×10^{3}
	Worst	5.21×10^{-4}	1.21×10^{5}	133.52	2.62×10^{4}	4.78×10^{3}	1.98×10^{4}
	Ave.	1.92×10^{-4}	1.00×10^{5}	79.503	1.80×10^{4}	2.86×10^{3}	8.10×10^{3}
	STD	2.85×10^{-4}	1.90×10^{4}	35.098	7.14×10^{3}	1.15×10^{3}	7.44×10^{3}
	SR%	100	0	0	0	0	0
F4	Best	2.44×10^{-15}	489.57	3.39×10^{-6}	0.64	1.004	0.970
	Worst	1.15×10^{-11}	660.11	0.0074	90.98	1.121	1.141
	Ave.	2.61×10^{-12}	598.50	0.0045	18.88	1.080	1.077
	STD	5.05×10^{-12}	66.991	0.0041	40.30	0.047	0.065
	SR%	100	0	100	0	0	0
F5	Best	0.0081	76.313	0.103	0.170	0.024	0.030
	Worst	0.0752	133.39	0.236	27.00	0.099	0.790
	Ave.	0.0310	116.04	0.144	6.091	0.046	0.234
	STD	0.0309	23.214	0.053	11.746	0.030	0.316
	SR%	0	0	0	0	0	0
F6	Best	3.22×10^{-10}	19.959	0.0062	3.503	3.890	9.463
	Worst	5.88×10^{-8}	19.962	1.1564	19.96	5.597	20.27
	Ave.	2.17×10^{-8}	19.960	0.2511	15.90	4.585	16.64
	STD	3.22×10^{-8}	9.83×10^{-4}	0.5063	7.028	0.757	4.974
	SR%	100	0	0	0	0	0
F7	Best	2.44×10^{-15}	489.57	3.39×10^{-6}	0.64	1.004	0.970
	Worst	1.15×10^{-11}	660.11	0.0074	90.98	1.121	1.141
	Ave.	2.61×10^{-12}	598.50	0.0045	18.88	1.080	1.077
	STD	5.05×10^{-12}	66.991	0.0041	40.30	0.047	0.065
	SR%	100	0	100	0	0	0
F8	Best	2.14×10^{-5}	3.56×10^{8}	4.35×10^{-4}	1.228	6.561	1.575
	Worst	0.0069	6.69×10^{8}	0.103	12.18	16.54	1.36×10^{4}
	Ave.	0.0031	5.27×10^{8}	0.041	7.024	10.89	2.74×10^{3}
	STD	0.0034	1.20×10^{8}	0.056	4.062	3.989	6.12×10^{3}
	SR%	100	0	100	0	0	0
Ave.	SR%	75%	0	37.5%	0	0	0
Fiderman Average		1.0000	6.0000	2.1875	4.5000	3.6250	3.6875
Fiderman STD		1.125	5.500	2.3125	4.7500	3.2500	4.0625
Wilcoxon		—	7.93×10^{-7}	7.93×10^{-7}	7.94×10^{-7}	8.73×10^{-7}	7.94×10^{-7}

It can be seen from the total statistical values of SRs in Table 2 that the number of IGWOs is 6, the average value is 75%, the number of PSOs is 3, the average value of PSOs is 37.5%, and the value of the other 4 SRs is 0, indicating that the optimization ability of the method presented in this paper is the best of the six methods. According to the table, the standard deviation and mean value of IGWO are the smallest of the eight algorithms. Friedman test was performed on the means and standard deviations of all the algorithms, and the results are ranked as follows: IGWO < PSO < GOA < SCA < MFO < GWO. The Wilcoxon test shows that the progressive significance of IWOA and the five optimization algorithms in the same dimension is less than 0.05, which proves that there is a significant difference between IWOA and the five different algorithms mentioned above. The above verification shows that IGWO has excellent optimization accuracy and stability.

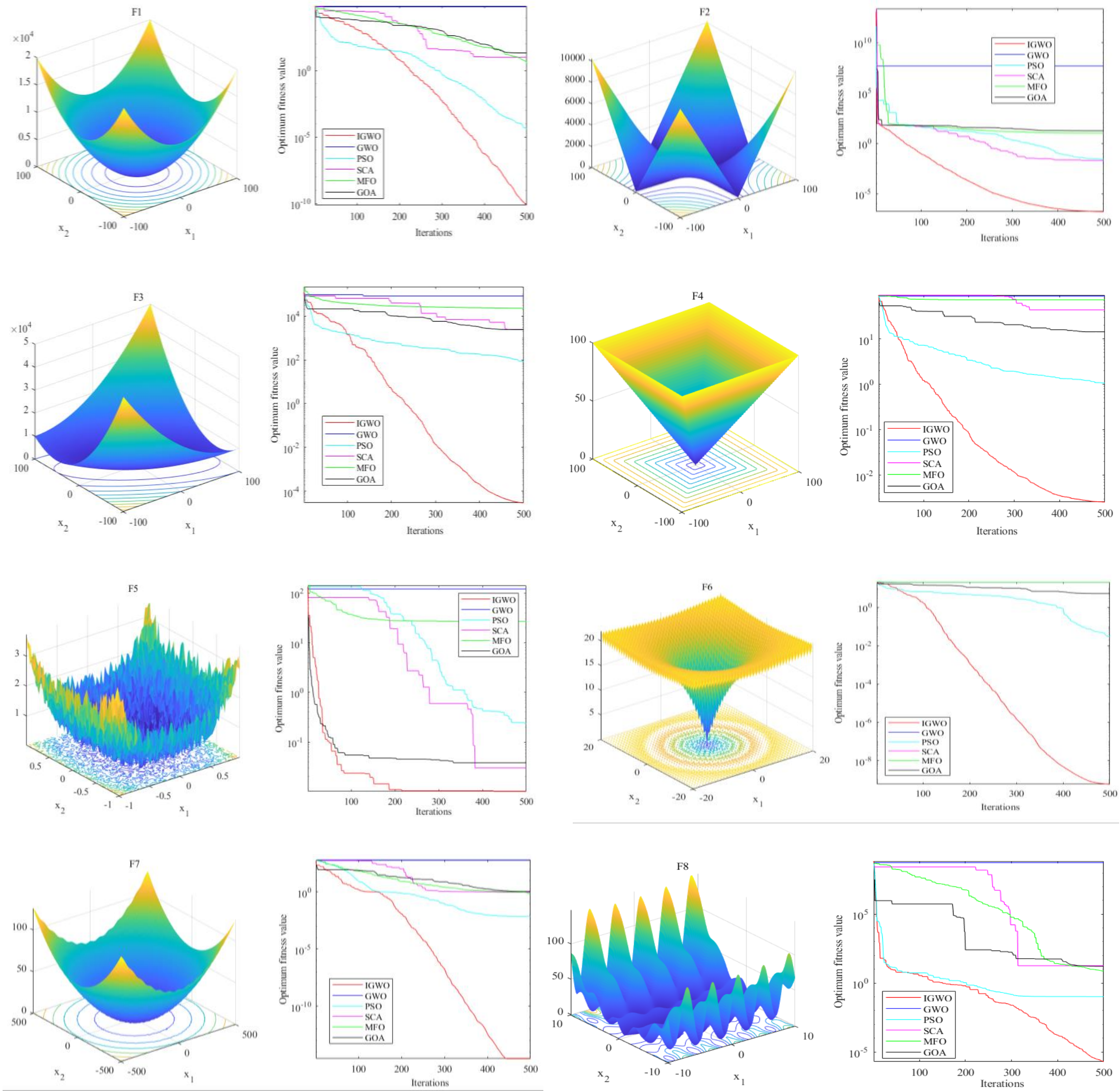

Figure 2. Test function simulation results.

Figure 2 records the convergence results of IGWO, GWO, PSO, MFO, GOA and SCA in each test function. It can be seen from the eight figures that the method in this paper has a faster convergence speed compared with the other seven methods. As can be seen from the F1–F3 and F6–

F8 iteration graphs, the IGWO algorithm significantly outperforms the other algorithms in convergence speed and reaches the optimal value at the end of the iteration. With the increase of the number of iterations, the convergence rate of IGWO is the fastest in all algorithms except F4 and F5, which further indicates that the improvement of GWO algorithm by PSO is effective.

3. Imbalanced data fault diagnosis model based on IGWO-WELM

3.1. WELM algorithm

Weighted Extreme Learning Machine (WELM) was proposed by Zong et al. [30] in 2013. This method retains the advantages of ELM, such as easy implementation and wide classification of mapping functions, and can be directly used to deal with data imbalance problems.

The WELM correlation principle is implemented based on the cost-sensitive idea. Each sample xi is weighted by introducing a weighting matrix, and the diagonal matrix W of $N_s \times N_s$ is formed by weighting, and the elements on the diagonal are the weight values of corresponding samples. If x_i belongs to the majority class, a smaller weight is assigned; conversely, if x_i is a minority class, a larger weight is assigned. After the weight ω is introduced, the optimization problem of WELM can be obtained according to the solution idea of extreme learning machine in the last section, and the mathematical problem can be modeled. The expression of WELM is as follows [31]:

$$\text{Minimize:}\frac{1}{2}\|\phi\|^2+\frac{1}{2}CW\sum_{i=1}^{N_s}\|\xi_i\|^2 \tag{10}$$

The constraint expression is as follows:

$$\phi h(a_i x_j+b_i)=q_i+\xi_i, i=1,2,\cdots,L,\ j=1,2,\cdots,N_s \tag{11}$$

The corresponding Lagrangian form is:

$$M_{\text{WELM}}(\phi,\lambda,\xi)=\frac{1}{2}\|\phi\|^2+\frac{1}{2}CW\sum_{i=1}^{N_s}\|\xi_i\|^2-\sum_{i=1}^{N_s}\lambda(H_i\phi-q_i-\xi_i) \tag{12}$$

According to KKT theory, the Lagrange penalty factor λ is assumed to be constant. Let the partial derivative of WELM with respect to Φ, λ, and ξ be 0, and the specific equation is as follows [30]:

$$\begin{cases}\dfrac{\partial M(\phi,\lambda,\xi)}{\partial\phi}=0\rightarrow & \phi=H^{\mathrm{T}}\lambda\\ \dfrac{\partial M(\phi,\lambda,\xi)}{\partial\xi_i}=0\rightarrow & \lambda=CW\xi_i\\ \dfrac{\partial M(\phi,\lambda,\xi)}{\partial\lambda_i}=0\rightarrow & H\phi-q_i-\xi_i=0\end{cases} \tag{13}$$

The corresponding Φ expression is shown in the following equation:

$$\phi=\begin{cases}(\frac{\mathrm{I}}{C}+H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WT & L\ll N_s\\ H^{\mathrm{T}}(\frac{\mathrm{I}}{C}+WHH^{\mathrm{T}})WT & L\gg N_s\end{cases} \tag{14}$$

where *I* represents the identity matrix and L represents the number of hidden layers in the network. For binary classification problems, the decision function of WELM classifier is $f(x)=\operatorname{sign} h(x)\phi$, and the specific expression is as follows:

$$f(x)=\operatorname{sign} h(x)\phi=\begin{cases}\operatorname{sign} h(x)(\frac{\mathrm{I}}{C}+H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WH & L\ll N_s\\ \operatorname{sign} h(x)H^{\mathrm{T}}(\frac{\mathrm{I}}{C}+WHH^{\mathrm{T}})WT & L\gg N_s\end{cases} \tag{15}$$

3.2. IGWO-WELM imbalance diagnostic model

Although the WELM algorithm is widely used for data imbalance, WELM, as a variant of ELM derived from the weighting idea, has similar problems to ELM. The randomly selected hidden layer bias and input weight may lead to model ill-conditioning problems, resulting in an unsatisfactory diagnosis. To solve the above problems and further improve the fault diagnosis accuracy of WELM, this paper uses IGWO to optimize the input weight and implicit bias of WELM and establishes an OLTC data imbalance fault diagnosis model based on IGWO-WELM (Weighted Extreme Learning Machine Based on Improved Grey Wolf Algorithm).

(a) Design of fitness function

To evaluate and select the next generation of grey wolf individuals, appropriate evaluation criteria must be selected as the fitness function of IGWO. The commonly used performance evaluation index of conventional machine learning algorithms is Accuracy (ACC). However, when ACC is used as an evaluation index to evaluate the performance of imbalanced data classification algorithms, the algorithm results will be biased toward most categories, resulting in high classification accuracy and the possibility of a high false-negative rate. Therefore, ACC is not suitable as a classification index for imbalanced data, and an evaluation index that can take into account both majority and minority classification results is needed.

Table 3. The confusion matrix table.

Predicted labels	The actual label	
	Positive category	Negative category
Positive category	*TP*	*FP*
Negative category	*FN*	*TN*

For binary classification problems, the minority class is usually defined as a positive category, and the majority class is defined as a negative category. In order to evaluate the classification results, the sample set is assumed to be composed of *P* anode and *N* cathode samples, and *TP*, *FN*, *TN* and

FP are defined respectively, where *TP* represents the number of correctly classified samples in the positive category, *FN* represents the number of incorrectly classified samples in the positive category, *TN* represents the number of correctly classified samples in the negative category, and *FP* represents the number of misclassified samples in the negative category, according to this confusion matrix, as shown in Table 3 [32,33].

Two indexes are obtained according to Table 3, namely Recall and G-mean, which evaluate the classification results of positive categories. A larger Recall value means that most positive category samples are detected, and G-mean is a good index for overall evaluation. The calculation formulas for Recall and G-mean are as follows [32,33]:

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$G\text{-}mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \tag{17}$$

According to Eq (16), the fitness function expression of IGWO-WELM is as follows:

$$fitness = \frac{1}{c}\sum_{i=1}^{c}\sqrt{G\text{-}mean} \tag{18}$$

Where, *c* represents the number of categories and fitness represents the fitness function.

3.3. IGWO-WELM imbalance diagnostic model

The main steps of the IGWO-WELM algorithm are as follows:

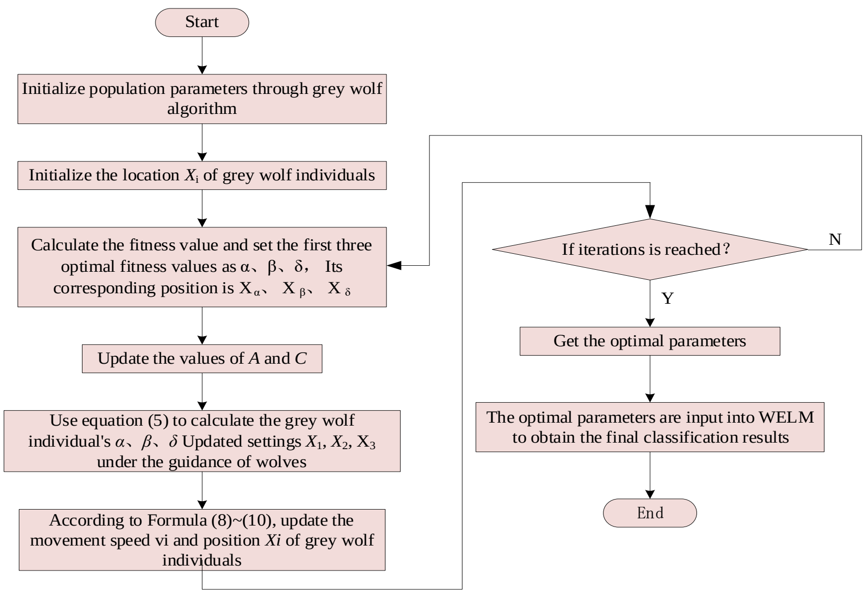

Figure 3. Imbalanced data fault diagnosis model based on IGWO-WELM.

1) Set the initial parameters A, C and a of the algorithm, the maximum number of iterations T_{max}, and select the appropriate number of wolves N;

2) According to the order from large to small, the fitness is calculated by Eq (18). The individuals corresponding to the first three fitness values are α, β, δ, and the corresponding positions of each grey wolf are X_α, X_β, and X_δ, respectively.

3) Calculate A and C according to Eqs (3) and (4);

4) According to Eq (6), the update positions of individual gray wolves under the guidance of α, β and δ wolves are calculated as X_1, X_2 and X_3, respectively;

5) According to Eqs (7)−(9), the moving speed v_i and moving position X_i of grey wolf individuals are updated with the idea of particle swarm optimization;

6) Judge whether t reaches T_{max}, and if so, obtain the optimal input weight and hidden layer offset; Otherwise, return to step 2;

The WELM model is tested on the test set with the optimal weight and hidden layer bias, and the final classification results are obtained. The specific process of the diagnosis model is shown in Figure 3.

4. Analysis of fault diagnosis results

4.1. Analysis of experimental results of imbalanced datasets based on KEEL

To observe the generalization ability of IGWO-WELM, eight datasets in the KEEL database are used to verify the proposed method, which contain binary and multi-classification datasets, and all of them have data imbalance problems. The specific parameters of the data are shown in Table 4. In order to illustrate the generalization ability, the proportion of data imbalance in the table increases gradually from top to bottom.

In order to comprehensively analyze the IGWO-WELM method, GWO-WELM [34], GOA-WELM [34], GA-WELM [34], WOA-WELM [34], PSO-WELM [24], WELM, and Support Vector Machine (SVM) are used as the over sampling algorithms of the base classifier (SMOTE-SVM, SSVM), and Kernel Extreme Learning Machine (KELM) [32] are used as the over sampling algorithms of the base classifier (SMOTE-KELM, SKELM) [35]. The improved oversampling algorithm - Borderline SMOTE for random forest (RF) (Borderline SMOTE- Random F the sampling size of each algorithm N = 10, the maximum number of iterations T_{max} = 30, the kernel parameter g of SVM is 1, and the penalty factor c is 2. The specific parameters of each algorithm are shown in Table 5.

Table 4. The characteristics of KEEL datasets.

Dataset	Abbreviation	Scale	Dim	Imbalance ratio
wine	win	178	13	1.5
contraceptive	con	1473	9	1.89
newth-yroid2	ny2	215	5	5.14
dermatology	der	366	34	5.55
segment0	seg	2308	19	6.02
zoo3	zo3	101	16	19.2
lymphography	ly	148	18	40.5
shuttle	shu	2175	9	853

Table 5. Parameter comparison table.

Algorithm	Parameters
PSO	$w_{\max} = 0.9$, $w_{\min} = 0.2$, $C_1 = 2$, $C_2 = 2$
GOA	$C_{\max} = 1$, $C_{\min} = 0.00004$
MFO	$t \in [-1,1]$. b = 1
SVM	Nuclear parameter $g_1 = 1$, Penalty factor $c_3 = 2$
KELM	Nuclear parameter $g_2 = 1$, Penalty factor $c_4 = 2$
RF	The number of decision trees $s_1 = 10$, the maximum number of features $c_5 = 42$

80% of each category in the eight data sets is randomly selected as the training set and 20% as the test set. In order to avoid the randomness brought by the algorithm, each algorithm repeats the calculation 30 times to obtain the G-mean value and average it. It should be noted that SKELM, SSVM, and BSRF are used for training. First, the training set is oversampled, and different types of data samples in the training set are balanced. Then KELM, SVM and RF are trained to establish a classification model for the balanced training set. Finally, the test samples are input into the established classification model to verify the performance of the oversampling algorithm.

To verify the optimization performance of IGWO, the IGWO algorithm, GWO, GOA, WELM, MFO and GA algorithms are used to verify the optimization performance of WELM. The iteration diagram is shown in Figure 4. It can be seen from the figure that with the increase of iteration times, the advantages of IGWO are gradually highlighted, and it is optimal in all eight data sets.

Table 6. Classification results of the different algorithms.

Method	G-mean (%)							
	win	con	ny2	der	seg	zo3	ly	Shu
IGWO-WELM	100	68.43	99.19	99.56	85.28	86.70	74.37	98.78
GWO-WELM	100	52.47	83.65	87.34	46.01	75.57	67.43	78.69
WOA-WELM	94.66	52.51	77.98	66.34	48.57	66.53	62.50	65.67
GOA-WELM	73.48	34.75	47.07	47.17	43.08	50.16	57.12	34.06
GA-WELM	97.72	51.65	76.74	79.60	49.01	72.51	62.61	91.02
PSO-WELM	100	58.87	98.46	98.18	66.81	81.73	72.22	97.01
WELM	89.88	53.25	64.76	53.67	27.93	55.88	56.71	78.59
SSVM	86.57	53.48	91.52	90.25	75.54	71.85	56.95	89.26
SKELM	87.35	63.54	92.59	89.65	77.62	78.56	72.45	91.56
BSRF	92.51	70.26	95.23	93.56	81.56	82.35	73.52	92.15

Table 6 shows the imbalanced classification results the different algorithms. It can be seen from the results that the evaluation indicators of the remaining 8 KEEL datasets of IGWO-WELM have the best results, except that the evaluation indicators of the contractual dataset are less than those of BSRF. To sum up, IGWO-WELM is better than the other nine methods in imbalanced data classification.

In order to evaluate the impact of imbalanced data on the proposed model, models with an imbalanced ratio of training data of 2:1, 3:1, 4:1, 5:1, 6:1, and 7:1 were used to verify IGWO-WELM. When designing the training data of imbalanced data, 48 samples are randomly selected from the normal state feature set composed used as in ref [2], and 24 samples are randomly selected from other fault sample feature sets to form the 2:1 imbalanced monitoring data and conduct training. Similarly, 24 samples are randomly selected from the feature set, 24 samples are randomly selected from other fault sample feature sets, and all the remaining samples form a test set. Calculate 30 times under each proportion, and calculate the average value as the final test result.

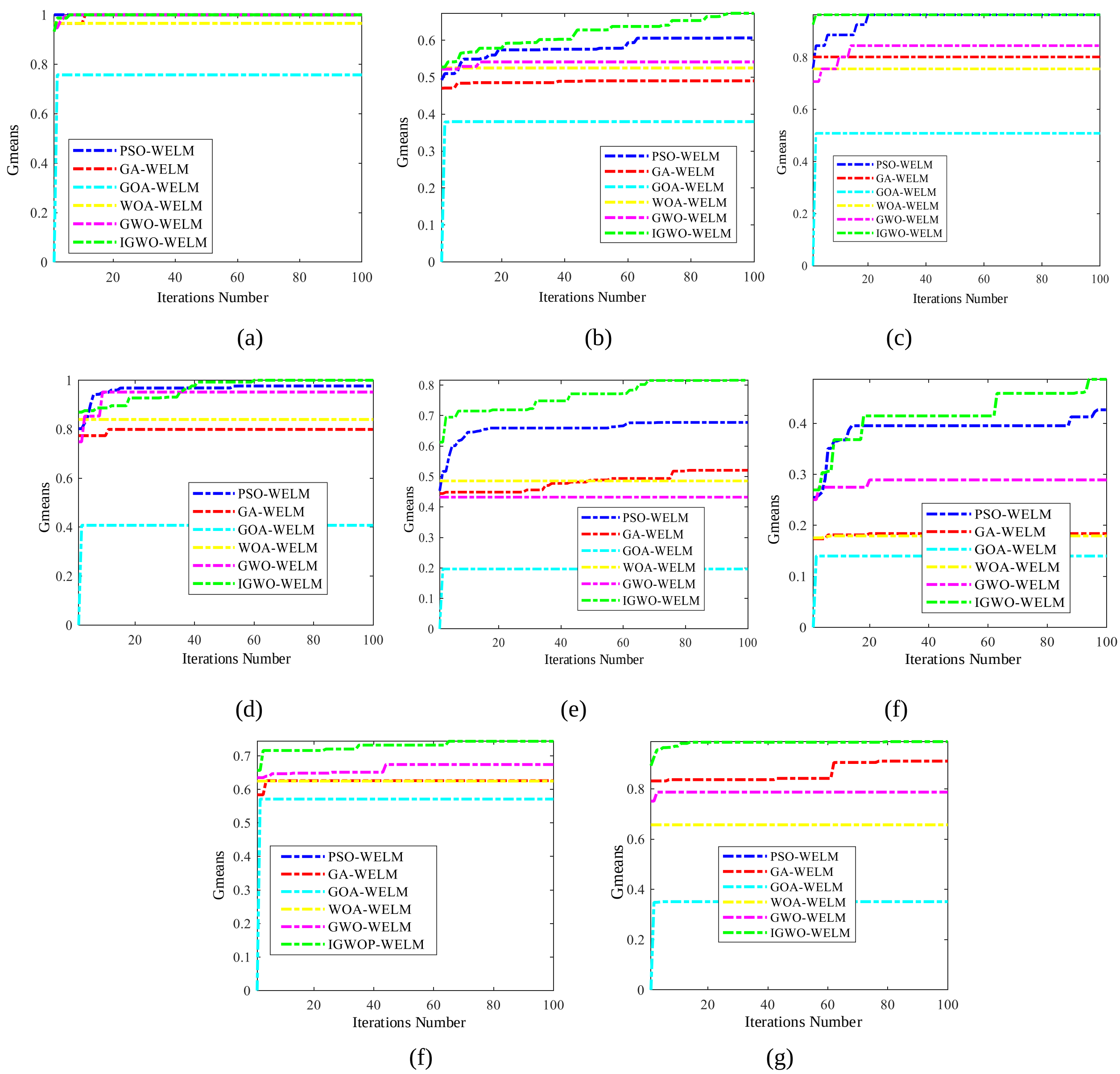

Figure 4. Imbalanced data fault diagnosis model based on IGWO-WELM. (a) Iteration of different algorithms for wine data. (b) Iteration of different algorithms for contraceptive data. (c) Iteration of different algorithms for newth-yroid2 data. (d) Iteration of different algorithms for dermatology data. (e) Iteration of different algorithms for segment data. (f) Iteration of different algorithms for zoo data. (g) Iteration of different algorithms for lymphography data. (h) Iteration of different algorithms for shuttle data.

Table 7. Classification results of the different algorithms.

Method	G-mean (%)						
	2:1	3:1	4:1	5:1	6:1	7:1	Mean
IGWO-WELM	97.57	94.56	93.23	92.01	91.25	89.59	93.05
GWO-WELM	91.80	89.56	87.12	86.25	85.23	84.37	87.38
WOA-WELM	73.82	72.15	70.51	68.15	65.12	57.61	67.75
GOA-WELM	87.52	85.23	83.56	82.12	75.65	72.89	81.16
GA-WELM	92.43	90.12	88.23	85.12	82.65	79.25	86.30
PSO-WELM	93.15	91.25	89.32	85.64	83.82	82.74	87.65
WELM	60.08	58.23	56.45	52.37	53.42	46.29	55.47
SSVM	87.52	84.23	82.15	78.61	74.23	72.56	79.88
SKELM	89.76	87.56	85.65	81.72	79.52	76.89	83.52
BSRF	92.61	89.23	88.56	84.75	82.32	80.56	86.34

Furthermore, the proportion of fault samples mistaken as normal samples in the total number of fault samples is calculated for statistics, which is defined as the false alarm rate, as shown in Table 8.

Table 8. False alarm rate of the different algorithms.

Method	False alarm rate under different imbalanced proportions (%)						
	2:1	3:1	4:1	5:1	6:1	7:1	Mean
IGWO-WELM	9.292	9.580	10.23	11.21	12.42	13.13	12.43
GWO-WELM	17.25	17.86	18.54	19.23	20.42	21.23	19.01
WOA-WELM	28.50	31.64	32.45	33.56	34.52	35.62	32.72
GOA-WELM	18.64	20.15	21.28	22.62	23.45	25.35	21.92
GA-WELM	24.63	25.45	26.68	27.52	28.23	29.52	27.17
IPSO-WELM	12.16	13.05	13.45	15.42	16.58	19.21	14.98
WELM	18.70	19.50	19.86	20.21	21.23	22.34	20.31
SSVM	17.56	19.25	20.68	21.56	23.26	24.85	21.19
SKELM	13.25	15.51	17.62	18.65	19.52	21.68	17.70
BSRF	9.895	11.62	12.75	15.53	16.65	18.54	14.16

As shown in Table 7, the overall performance of the IGWO-WELM algorithm in OLTC imbalance data diagnosis is better than that of the other nine methods. As the imbalance of the imbalance data gradually deepens, the advantages of the IGWO-WELM algorithm become increasingly obvious. The G-mean values of IGWO-WELM under different proportions are higher than those of the other nine methods, and the average values are higher than PSO-WELM, GA-WELM, GOA-WELM, WOA-WELM, GWO-WELM, WELM, SSVM, SKELM and BSRF at 11.89, 25.3, 5.67, 37.58, 13.17, 9.53 and 9.41% respectively. Secondly, the BSRF is better, which shows that the over-sampling algorithm is feasible to change the training set method, but it is still not as effective as IGWO-WELM. The worst method is WELM. This is due to the influence of hidden layer bias and input weight on the model, which makes the diagnosis accuracy low. This shows the importance of WELM parameter optimization. Table 7 further shows that the optimization effect of WOA-WELM is the worst among the six optimization methods, followed by GOA-WELM, which is caused by the performance defect of the algorithm itself.

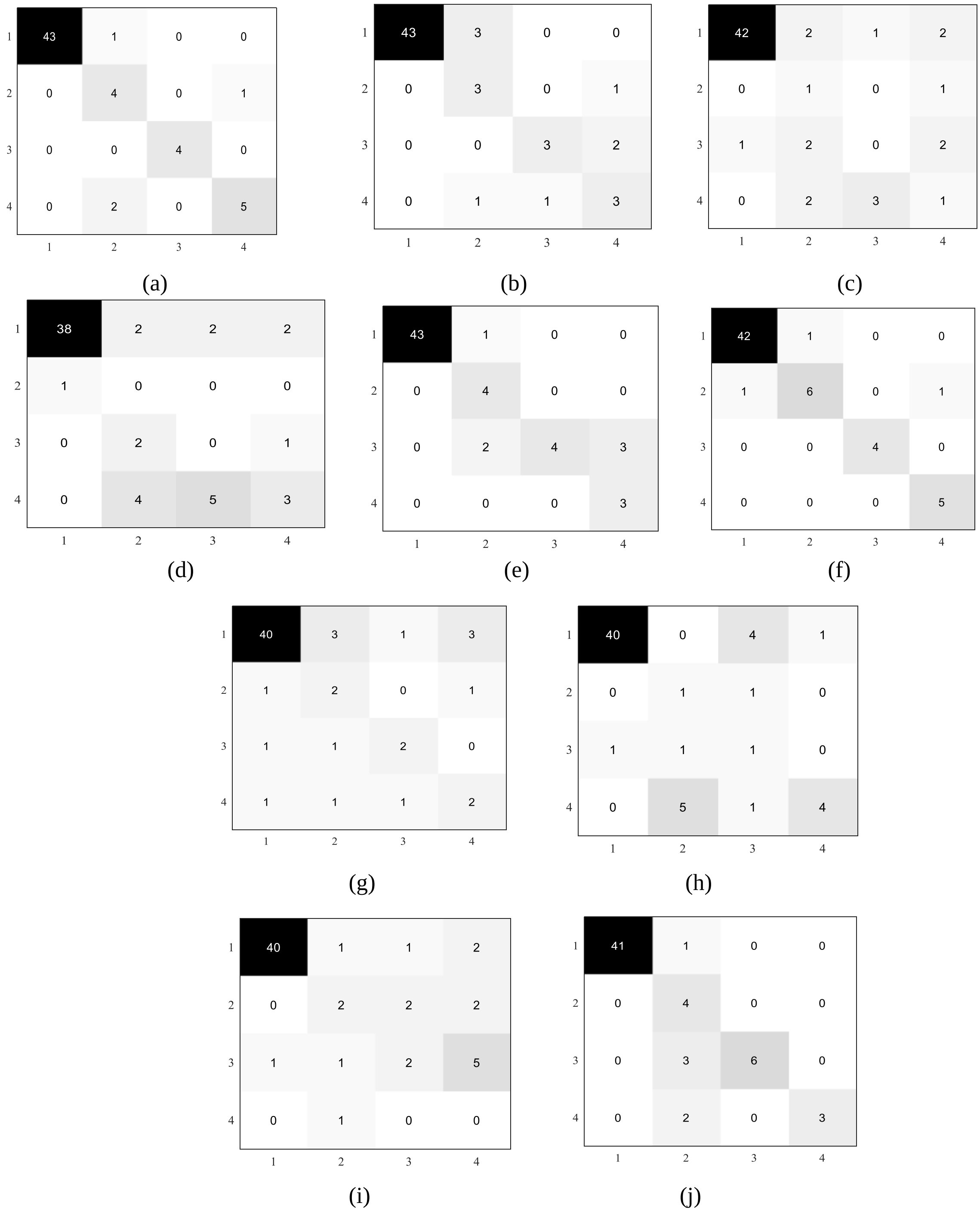

Figure 5. Classification results of different algorithms. (a) PSO algorithm classification results. (b) GA algorithm classification results. (c) GOA algorithm classification results. (d) WOA algorithm classification results. (e) GWO algorithm classification results. (f) IGWO algorithm classification results. (g) WELM algorithm classification results. (h) SSVM algorithm classification results. (i) SKELM algorithm classification results. (j) BSRF algorithm classification results.

From Table 8, we can see that IGWO-WELM has the lowest false alarm rate among all the methods and can maintain the false alarm rate at a low level, significantly lower than the other nine algorithms. It can be further seen from the table that WOA-WELM has the highest false alarm rate among the 10 methods, followed by GA-WELM and GOA-WELM. For further explanation, it can be seen from Figure 5 that IGWO-WELM has the best effect among all classification methods, followed by PSO-WELM, GA-WELM, GGO-WELM, SSVM, SKELM and BSRF, and the worst are WOA-WELM and WELM. WELM is not optimized, resulting in poor results. To sum up, it can be further shown that the WOA WELM algorithm is not suitable for OLTC unbalanced data fault diagnosis.

5. Conclusions

Aiming at the problems of classification bias and model invalidation when traditional machine learning algorithms deal with OLTC imbalanced data distribution, this paper proposes a fault diagnosis method for OLTC imbalanced distribution based on IGWO-WELM. The main conclusions are as follows:

1) The particle algorithm is used to improve GWO, and the IGWO algorithm is proposed. This algorithm can overcome the problem that the GWO algorithm can easily fall into the local optimum and has slow convergence.

2) IGWO-WELM algorithm is proposed by using IGWO's good global search and fast convergence ability to optimize the input weight and implicit offset of WELM, and G-mean is used as the fitness function of IGWO-WELM.

3) By comparing other classical methods of imbalanced data fault diagnosis with the method in this paper through the KEEL datasets and OLTC dataset, the method in this paper shows improvement of least 5%, which has certain theoretical research and practical engineering significance.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. 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 Sens. J.*, **21** (2021), 15696−15708. https://doi.org/10.1109/JSEN.2020.3021920
2. K. Wongon, K. Sunuwe, J. Jingyo, H. Kim, H. Lee, B. D. Youn, Digital twin approach for on-load tap changers using data-driven dynamic model updating and optimization-based operating condition estimation, *Mech. Syst. Signal PR.*, **181** (2022), 1−17. https://doi.org/10.1016/j.ymssp. 2022.109471
3. Q. Yang, J. Ruan, Z. Zhuang, Fault diagnosis of circuit breakers based on time–frequency and chaotic vibration analysis, *IET Gener. Transm. Dis.*, **14** (2020), 1214–1221. https://doi.org/10.1049/iet-gtd.2019.0985
4. R. Y. Shang, C. Q. Peng, P. F. Shao, R. Fang, FFT-based equal-integral-bandwidth feature extraction of vibration signal of OLTC, *Math. Biosci. Eng.*, **18** (2021), 1996–1980. https://doi.org/10.3934/mbe.2021102

5. C．Bengtsson, Status and trends in transformer monitoring, *IEEE Trans. Power deliver.*, **11** (1996), 1379–1384. https://doi.org/10.1109/61.517495
6. P. Kang，D. Birtwhistle．Condition monitoring of power transformer on-load tap-changers．Part I: Automatic condition diagnostics, *IEE P-Gener. Transm. D.*, **148** (2001), 301–306. https://doi.org/10.1049/ip-gtd:20010389
7. P. Kang, D. Birtwhistle．Condition monitoring of power transformer on-load tap-changers．Part II：Detection of ageing from vibration signatures, *IEE P-Gener. Transm. D.*, **148** (2001), 307–311. https://doi.org/10.1049/ip-gtd:20010388
8. P. Kang, D. Birtwhistle. Condition assessment of power transformer on-load tap changers using wavelet analysis and self-organizing map: field evaluation, *IEEE Trans. Power deliver.*, **18** (2003), 78–84. https://doi.org/10.1109/TPWRD.2002.803692
9. X. Zhou, F. H. Wang, J. Fu, J. Lin, Mechanical condition monitoring of on-load tap changers based on chaos theory and k-means clustering method, *Proc. CSEE*, **35** (2015), 1541–1548. https://doi.org/10.13334/j.0258-8013.pcsee.2015.06.031
10. J. X. Liu, G. Wang, T. Zhao, L. Zhang, Fault diagnosis of on-load tap-changer based on variational mode decomposition and relevance vector machine, *Energies*, **10** (2017), 946–959. https://doi.org/10.3390/en10070946
11. X. Duan, T. Zhao, T. Li, J. Liu, L. Zou, L. Zhang, Method for diagnosis of on-load tap-changer based on wavelet theory and support vector machine, *J. Eng. Ny.*, **13** (2017), 2193–2197. https://doi.org/10.1049/joe.2017.0719
12. Q. Li, T. Zhao, Z. Li, J. Lou, Mechanical fault diagnosis of on load tap changer within power transformers based on hidden Markov model, *IEEE Trans. Power deliver.*, **27** (2012), 596–601. https://doi.org/10.1109/TPWRD.2011.2175454
13. X. Liang, Y. Wang, H. Gu, A mechanical fault diagnosis model of on-load tap changer based on same-source heterogeneous data fusion, *IEEE Trans. Instrum. Meas.*, **71** (2022). https://doi.org/10.1109/TIM.2021.3064808
14. L. Zheng, G. Liu, C. Yan, C. Jiang, et al. Improved TradaBoost and its application to transaction fraud detection, *IEEE Trans. Comput. Social Syst.*, **7** (2020), 1304–1316. https://doi.org/10.1109/TCSS.2020.3017013
15. S. Dhote., C. Vichoray, R. Pais, et al., Hybrid geometric sampling and AdaBoost based deep learning approach for data imbalance in E-commerce, *Electron. Commer. Res.*, **20** (2020), 259–274. https://doi.org/10.1007/s10660-019-09383-2
16. W. Lee, C. H. Jun, J. Lee, Instance categorization by support vector machines to adjust weights in AdaBoost for imbalanced data classification, *Inform. Sci.*, **381** (2017), 92–103. https://doi.org/10.1016/j.ins.2016.11.014
17. Y. Gao, L. Gao, X. Li, S. Cao, A hierarchical training-convolutional neural network for imbalanced fault diagnosis in complex equipment, *IEEE Trans. Ind. Inform.*, **18** (2022), 8138–8145. https://doi.org/10.1109/TII.2022.3177662
18. Y. Geng, X. Y. Luo, Cost-sensitive convolutional neural networks for imbalanced time series classification, *Intell. Data Anal.*, **23** (2019), 357–370. https://doi.org/10.3233/IDA-183831
19. A. Taherhnai, G. Cosma, T. T. McGinnity, AdaBoost-CNN: An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning, *Neurocomputing*, **404** (2020), 351–366. https://doi.org/10.1016/j.neucom.2020.03.064

20. Y. Xie, G. Liu, C. Yan, C. Jiang, M. Zhou, M. Li, Learning transactional behavioral representations for credit card fraud detection, *IEEE Trans. Neural Networks Learn. Syst.*, 2022. https://doi.org/10.1109/TNNLS.2022.3208967
21. Q. X. Zhu, X. W. Wang, N. Zhang, Novel K-Medoids based SMOTE integrated with locality preserving projections for fault diagnosis, *IEEE Trans. Instrum. Meas.*, **71** (2022). https://doi.org/10.1109/TIM.2022.3218551
22. Y. Gao, Q. C. Liu, An over sampling method of unbalanced data based on ant colony clustering, *IEEE Access*, **9** (2021), 130990–130996. https://doi.org/10.1109/ACCESS.2021.3114443
23. J. Shen, J.C. Wu, M. Xu, D. Gan, B. An, F. Liu, A hybrid method to predict postoperative survival of lung cancer using improved SMOTE and adaptive SVM, *Comput. Math. Method Med.*, **2021** (2021). https://doi.org/10.1155/2021/2213194
24. K. Hu, Z. Zhou, L. Weng, J. Liu, L. Wang, Y. Su, et al., An optimization strategy for weighted extreme learning machine based on PSO, *Int. J. Pattern. Recogn.*, **31** (2017). https://doi.org/10.1142/S0218001417510016
25. Z. Z. Li, M. Huang, G. J. G. Liu, Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection, *Expert Syst. Appl.*, **175** (2021). https://doi.org/10.1016/j.eswa.2021.114750
26. S. Saremi, S. Z. Mirjalili, S. M. Mirjalili. Evolutionary population dynamics and grey Wolf optimizer, *Neural Comput*. *Appl*., **26** (2015), 1257–1253. https://doi.org/10.1007/s00521-014-1806-7
27. G. M. Komaki, V. Kayvanfar. Grey Wolf optimizer for the two-stage assembly flow shop scheduling problems with release time, *J. Comput*. *Sci*., **8** (2015), 109–120. https://doi.org/10.1016/j.jocs.2015.03.011
28. A. K. Mishra, S. R. Das, P. K. Ray, R. K. Mallick, A. Mohanty, D. K. Mishra, PSO-GWO optimized fractional order PID based hybrid shunt active power filter for power quality improvements, *IEEE Access*, **8** (2020), 74497–74512. https://doi.org/10.1109/ACCESS.2020.2988611
29. H. H. Zhu, G. J. Liu, M. C., Zhou, Y. Xie, Q. Kang, Dandelion algorithm with probability-based mutation, *IEEE Access*, **7** (2019), 97974–97985. https://doi.org/10.1109/ACCESS.2019.2927846
30. A. K. Dutta, B. Qureshi, Y. Albagory, Ma. Alsanea, M. Al Faraj, A. R. W. Sait, Optimal weighted extreme learning machine for cybersecurity fake news classification, *Comput. Syst. Sci. Eng.*, **44** (2023), 2395–2409. https://doi.org/10.32604/csse.2023.027502
31. G. B. Allende, G. Still, Solving bilevel programs with the KKT-approach, *Math. Program.*, **138** (2013), 309–332. https://doi.org/10.1007/s10107-012-0535-x
32. C. Lei, S. Wan, Intelligent fault diagnosis of high-voltage circuit breakers using triangular global alignment kernel extreme learning machine, *ISA Trans.*, **109** (2020), 368–379. https://doi.org/10.1016/j.isatra.2020.10.018
33. Z. C. Li, M. Huang, G. J. Liu, C. Jiang, A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection, *Expert Syst. Appl.*, **175** (2021), 114750. https://doi.org/10.1016/j.eswa.2021.114750
34. C. B. Liu, H. F. Ke, G. Zhang, Y. Mei, H. Xu, An improved weighted extremely learning machine for imbalanced data classification, *Memetic Comput.*, **20** (2019), 27–34. https://doi.org/10.1007/s12293-017-0236-3

35. L. Chen, S. Wan, L. Dou, Improving diagnostic performance of High-Voltage circuit breakers on imbalanced data using an oversampling method, *IEEE Trans. Power deliver.*, **37** (2022), 2704–2716. https://doi.org/10.1109/TPWRD.2021.3114547