ARCHIVES OF ELECTRICAL ENGINEERING VOL. 74(1), pp. 191 –208 (2025)

DOI 10.24425/aee.2025.153019

Power transformer fault warning combining support vector machine and improved grey wolf optimization algorithm

SHUZONG ZHAO ✉, **NORASAGE PATTANADECH**

School of Engineering, King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520, Thailand

e-mail: ✉ *lionogj@sina.com, norasage.pa@kmitl.ac.th*

(Received: 28.08.2024, revised: 10.02.2025)

Abstract: To optimize the parameter setting of the support vector machine and improve the classification performance and computational efficiency of power transformer fault diagnosis, this study proposes an improved grey wolf optimization algorithm. By optimizing the global search and local optimization capabilities of the grey wolf algorithm and combining them with stacked denoising autoencoders, a new power transformer fault warning model is constructed. Firstly, the grey wolf optimization algorithm is optimized through four strategies: elite reverse learning, nonlinear control parameters, Lévy flight, and particle swarm optimization, which improve its global search and local optimization capabilities. Secondly, the stacked denoising autoencoder is utilized to extract high-level features of fault data, and the improved GWO algorithm and SVM are combined to complete fault classification. The results indicated that the proposed diagnostic model achieved a diagnostic accuracy of 0.979, a recall rate of 0.986, and an F1 value of 0.983 in benchmark performance testing. In practical applications, the average fault diagnosis accuracy of this model could reach up to 99.21%, and the average diagnosis time was only 0.08 s. The developed power transformer fault warning model can provide an efficient and reliable technical solution for fault diagnosis in the power system.

Key words: fault diagnosis, HGWO, SDAE, SVM, transformer

1. Introduction

Power Transformer (PT) is one of the key equipment in the power system, and its operating status directly affects the safety and stability of the power system. Due to the susceptibility of transformers to various factors during long-term operation, faults or abnormal situations may occur, which may

lead to serious power accidents [1, 2]. Therefore, it is necessary to continuously update the existing fault warning technology for PTs to effectively ensure its safe and stable operation. In recent years, with the development of smart grid technology, the Fault Diagnosis and Warning (FDW) technology of PTs has received widespread attention and research. Traditional Fault Diagnosis Methods (FDMs) mainly rely on expert experience and regular manual maintenance, which have problems such as low efficiency, high cost, and low accuracy. To overcome these shortcomings, researchers have begun to introduce Machine Learning (ML) and data mining techniques to improve the accuracy and real-time performance of FDW [3, 4]. Among them, Support Vector Machine (SVM), as a classic ML method, has good classification performance and generalization ability and is widely used in fault diagnosis. However, SVM performance may be limited when dealing with high-dimensional and nonlinear data. To further improve the performance of FDW, Grey Wolf Optimizer (GWO), as an emerging intelligent optimization algorithm, has gradually been applied to optimize SVM parameters and improve classifier performance due to its ability to balance global and local search [5]. In addition, Stacked Denoising Autoencoder (SDAE), as a deep learning model, also has the ability to extract high-level features of data and can effectively handle complex nonlinear data. To optimize the parameter setting of SVM and improve its classification performance and computational efficiency in Power Transformer Fault Warning (PTFW) diagnosis, an Improved GWO (I-GWO) is proposed. By improving the global search and local optimization capabilities of GWO, the parameters of SVM can be adjusted more accurately to avoid model overfitting or underfitting problems. In addition, by combining the feature extraction capabilities of SDAE, the classification accuracy of complex fault data can be further improved, thereby achieving more efficient fault diagnosis and early warning. The innovation lies in its full utilization of the feature extraction capability of SDAE and the feature classification performance of SVM and GWO for complex data, enabling accurate identification and classification of different feature faults. This study contributes to improving the operational reliability of the power system and also has certain reference significance for FDW research in other similar fields.

2. Related work

The PT is one of the most essential components in the power system. Tang *et al*. proposed a new signal processing method, named the multivariate variational mode decomposition method, aimed at extracting features from motor rolling bearings using this method. This study used SVM for classification and recognition and employed GWO to optimize classifier parameters, thereby improving recognition accuracy. This scheme could achieve a diagnostic rate of 100% for normal working conditions, outer and inner ring faults, and rolling element faults under various load and speed conditions [6]. Liu *et al*. designed a fault diagnosis approach based on generalized multi-scale average permutation entropy and GWO Least Squares SVM (LSSVM). Firstly, it was based on multi-scale permutation entropy and replaced the coarsening process with multi-scale equalization, expanding the mean to variance and avoiding dynamic mutations in the original signal. Finally, by optimizing the model parameters through GWO, accurate identification of fault modes was achieved. This method could effectively achieve fault diagnosis of rolling bearings [7]. To lift the accuracy of Transformer Fault Diagnosis (TFD), Hui S. *et al*. suggested a grey wolf residual network FDM based on data augmentation. For imbalanced datasets, this

method used a Wasserstein Generative Adversarial Network (GAN) with a gradient penalty for data augmentation. Finally, GWO was taken to optimize the RNN in the early stages of training, and the model showed good performance in TFD [8].

To improve the accuracy of TFD, WANG H. *et al*. developed a TFD method based on a multi-scale Convolutional Neural Network (CNN) model. To address insufficient sample features, a feature extension method based on the proportional method was adopted, which expanded the sample features from 5 dimensions to 25 dimensions. To deal with small sample sizes and uneven distribution of fault samples, a sample size enhancement method based on GAN was also adopted to generate numerous simulated samples. The average fault diagnosis accuracy of this model was as high as 93.24% [9]. In the state assessment of PTs, insufficient feature parameters of raw data and uneven classification of existing data fault kinds could easily lead to training model distortion. To handle these obstacles, Haoran X. *et al*. proposed a PT state evaluation method built on GAN and CNN. This method extended the original parameters through GAN and generated an artificial dataset, and then jointly trained the data through CNN. In the end, the effectiveness of the method was verified by actual measurement data and experiments [10]. To improve the accuracy of oil-immersed TFD, Lu W. *et al*. put forward an FDM based on equidistant feature mapping and improved LSSVM. Firstly, this method used equidistant feature mapping to reduce the dimensionality of 14-dimensional fault feature data and eliminate redundancy. Secondly, the improved chimpanzee algorithm was utilized to optimize the LSSVM parameters and establish the optimal model. Ultimately, the diagnostic accuracy of the model could reach up to 90.83% [11].

In summary, significant progress has been made in the methods for power TFD, and the accuracy of fault diagnosis has been improved through various optimization algorithms and deep learning models. However, these methods still suffer from high computational complexity and time cost when dealing with high-dimensional features and large-scale data, and the diagnostic accuracy for rare fault types also needs to be enhanced. As a result, this study proposes a PTFW method that combines SVM, I-GWO, and SDAE, aiming to further improve the sensitivity and accuracy of model diagnosis by optimizing parameter settings and data augmentation techniques.

3. PTFW based on the improved hybrid GWO

To improve the PTFW effect and increase the stability of the power system, this study first introduces three strategies to optimize the traditional GWO, and then combines the optimized GWO, SVM, and SDAE technologies to build a new power TFD model.

3.1. Optimization design of GWO algorithm

GWO is proposed by Mirjalili *et al*. in 2014. This algorithm is inspired by the hunting behavior of grey wolves and simulates their social hierarchy and hunting strategies during the hunting process. In GWO, assuming N represents the population size of grey wolves, S_d is the search dimension for solving optimization problems using GWO, and the location of the i-th grey wolf individual is $X_i = \left\{X_i^1, X_i^2, \cdots, X_i^{S_d}\right\}$, $i = 1, 2, \cdots, N$. The process of getting surrounded by wolf

packs is shown in Eq. (1) [12, 13].

$$\begin{cases} D = |B \cdot X_L(t) - X(t)| \\ X(t+1) = X_L(t) - A \cdot D \end{cases}. \tag{1}$$

In Eq. (1), t means the number of iterations. $X_L(t)$ is the prey location and also the optimal solution location. $X(t)$ and $X(t+1)$ are the individual positions of grey wolves at the beginning and end of the wolf pack encirclement process. D is the distance between the grey wolf and its prey. A and B are two different variable coefficients, and their calculation formulas are shown in Eq. (2) [14].

$$\begin{cases} A = 2dr_1 - d \\ B = 2r_2 \end{cases}. \tag{2}$$

In Eq. (2), d is the distance control parameter. r_1 and r_2 are both random numbers between [0, 1]. When t satisfies the iteration condition, α, β, δ, and ω are the optima, the 2nd optima, the 3rd optima, and other solutions. Moreover, if the top three optimal solutions are relatively close, other solutions will adjust their output through position updates. When the termination condition is achieved, the position of the δ wolf no longer changes, and the fitness value of the α wolf is output as the optimum. The execution process of GWO is shown in Fig. 1 [15].

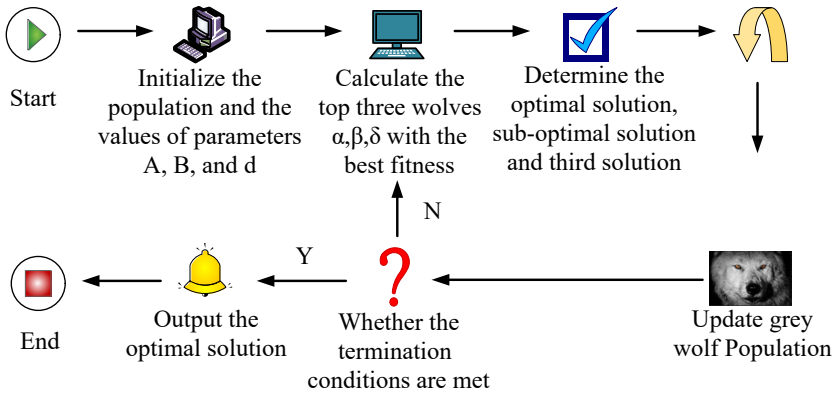


Fig. 1. The GWO flowchart

In Fig. 1, the operational process of GWO is mainly divided into four steps. Step 1 is to initialize the population and parameter values, randomly generate the initial grey wolf population, and initialize the position. Step 2 is to calculate the fitness value and evaluate the position of each grey wolf to determine the positions of α wolf, β wolf, and δ wolf. Step 3 is to update the position, updating the positions of other grey wolves based on the positions of α, β, and δ wolves, mimicking the process of surrounding, chasing, and attacking prey. Step 4 is to determine the termination conditions. If the predetermined number of iterations is reached or the accuracy requirements are met, the iteration stops. Otherwise, to get back to the Step 2 to recalculate the fitness value. Despite the GWO's apparent simplicity, its structure incorporates a limited number of adjustable parameters and a position update strategy that exhibits inherent deficiencies. Consequently, the optimal solution it generates is merely proximate to the original optimum, rather than the true

optimum of the problem [16]. Therefore, this study first introduces the Elite Reverse Learning Strategy (ERLS) to optimize GWO. ERLS is a strategy used to optimize algorithms by combining elite individuals and reverse individuals to enhance the global search ability. The core idea of ERLS is to evaluate not only the quality of the current solution but also the reverse solutions of these solutions, thereby increasing the diversity of the search space to escape from local optima. Firstly, letting the elite individuals in the grey wolf population be $X^e_{\sigma,\theta} = \left\{X^e_{\sigma,1}, X^e_{\sigma,2}, \cdots, X^e_{\sigma,S_d}\right\}$, $\sigma = 1, 2, \cdots s$, and $\theta = 1, 2, \cdots S_d$, and the solution for the reverse individuals is obtained as expressed by Eq. (3).

$$\overline{X}^e_{\sigma,\theta} = V\left(a_\theta + b_\theta\right) - X^e_{\sigma,\theta}. \tag{3}$$

In Eq. (3), V is a dynamic coefficient with a value range between 0 and 1. a_θ and b_θ are the dynamic boundary ranges of elite individual $X^e_{\sigma,\theta}$. a_θ and b_θ are the Minimum (Min) and Maximum (Max) dynamic boundary range values of $X^e_{\sigma,\theta}$. The existence of dynamic boundaries can enable elite individuals to obtain inverse solutions in a smaller space, thereby accelerating the iteration speed of the algorithm. When the solution of $\overline{X}^e_{\sigma,\theta}$ crosses the dynamic boundary or is a non-feasible solution, Eq. (4) is used for resetting.

$$\overline{X}^e_{\sigma,\theta} = \mathrm{rand}\left(a_\theta, b_\theta\right). \tag{4}$$

Secondly, considering that the value of variable coefficient A is determined by d, it cannot balance the global search capability of GWO well. Therefore, this study introduces a nonlinear control parameter strategy to control the range of values of d. The mathematical model of this strategy is shown in Eq. (5).

$$d = d_{\min} + \left(d_{\max} - d_{\min}\right) e^{\left(\frac{-1.5t}{T}\right)^4}. \tag{5}$$

In Eq. (5), T is the Max iteration times. $d_{\min}$ and $d_{\max}$ are the Min and Max values of d, where the Min value is 0 and the Max value is 2.

In addition to the two optimization methods mentioned above, GWO also has the drawback of the population easily falling into local optima. Therefore, this study further introduces the Lévy Flight Strategy (LFS) to expand GWO's search range for the population, aiming to improve the algorithm's adaptability. After introducing LFS, the GWO population position update formula is shown in Eq. (6).

$$X_G\left(t+1\right) = X\left(t\right) + L \oplus \lambda\left(X\left(t\right) - X_\alpha\left(t\right)\right). \tag{6}$$

In Eq. (6), $X_\alpha\left(t\right)$ is the optimal solution. $X_G\left(t+1\right)$ is the updated location after optimization. L is the Lévy flight path. λ is the adjustment coefficient. The control strategy is shown in Eq. (7).

$$\lambda = \lambda_{\min} + \left(\lambda_{\max} - \lambda_{\min}\right) e^{\left(\frac{-1.5t}{T}\right)^4}. \tag{7}$$

In Eq. (7), $\lambda_{\min}$ and λ_{maz} are the Min and Max values of λ, with a Min value of 0 and a Max value of 1. Considering the shortcomings of insufficient convergence speed and solution accuracy in GWO, the velocity change idea in Particle Swarm Optimization (PSO) is introduced to optimize GWO, thereby enhancing its parameter optimization ability in practical engineering [17, 18]. In PSO, the velocity and position update equations of particles are shown in Eq. (8).

$$\begin{cases} V_{L+1} = \omega_L \cdot V_L + c_{L1} \cdot r_{L1} \cdot \left(P_{Lb} - X_L\right) + c_{L2} \cdot r_{L2} \cdot \left(g_{Lb} - X_L\right) \\ X_{L+1} = V_{L+1} + X_L \end{cases}. \tag{8}$$

In Eq. (8), V_{L+1} and V_L are the updated and current particle velocities. X_{L+1} and X_L are the updated and current particle positions. r_{L1} and r_{L2} are two random numbers between 0 and 1, and they are not the same. P_{Lb} and g_{Lb} are the current individual optima and global optima. c_{L1} and c_{L2} are individual and social learning factors. ω_L is the Inertial Weight Factor (IWF). In PSO, dynamic IWF is introduced to optimize the range of ω_L values, as shown in Eq. (9).

$$\omega_L = \omega_{L\max} - (\omega_{L\max} - \omega_{L\min}) \cdot \left(\frac{t}{T}\right)^2. \tag{9}$$

In Eq. (9), $\omega_{L\min}$ and $\omega_{L\max}$ are the Min and Max values of ω_L. By sequentially incorporating the four optimization strategies mentioned above, a Hybrid Grey Wolf Optimizer (HGWO) is ultimately obtained, and its operational process is shown in Fig. 2.

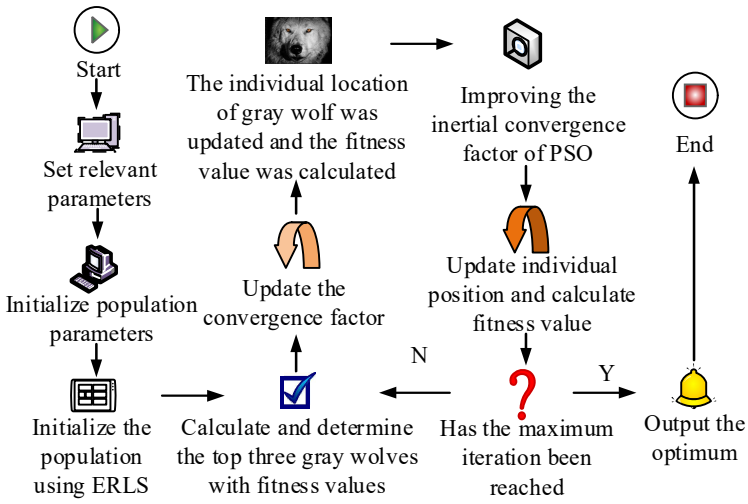


Fig. 2. The HGWO flowchart

In Fig. 2, ERLS is used to improve the population diversity, adjust d based on this, and optimize the optimization process of GWO algorithm using LFS. Finally, by combining PSO and GWO, Eq. (8) is used to calculate the updated position, velocity, and fitness values, thus completing the optimization of the entire HGWO algorithm.

3.2. Construction of TFD model combining SDAE and HGWS-SVM

PTs are critical equipment in the power system, and real-time diagnosis and warning of faulty PTs can ensure the secure operation. SVM is a supervised learning model that has been widely used in power TFD and early warning in recent years due to its excellent classification performance and powerful ability to process high-dimensional data. Due to the nonlinearity of transformer fault data, a nonlinear mapping algorithm is required to map the low-dimensional to the high-dimensional feature space to gain analyzable data. Therefore, this study adopts SVM as the nonlinear mapping algorithm in TFD tasks, which can find the hyperplane with the largest classification interval and accurately classify nonlinear samples. The optimal hyperplane structure of SVM is shown in Fig. 3 [19].

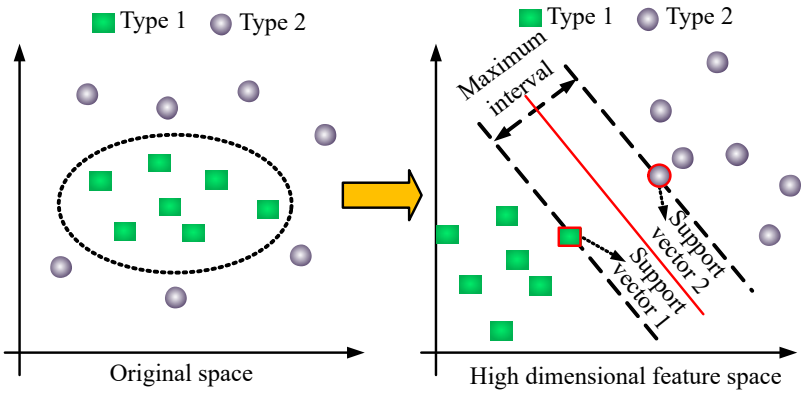


Fig. 3. Schematic diagram of the optimal SVM hyperplane

In Fig. 3, the data points in the original space are better separated in high-dimensional space. SVM determines the position and direction of the hyperplane in this process, thus achieving the best classification performance. Traditional SVM has high computational overhead, high training time and memory requirements when processing large-scale datasets, and also suffers from sensitivity to noisy data [20]. These issues are largely related to the parameters of SVM, especially the penalty parameters and kernel function parameters. These two parameters control the complexity of the model and balance the cost of misclassification, directly affecting the classifier's generalization ability. If the parameter selection is improper, the model may overfit or underfit [21]. Therefore, to overcome the limitations of traditional SVM, this study introduces HGWO to optimize the parameters of SVM, thereby improving its classification performance and computational efficiency. The SVM classification algorithm combined with HGWO will be referred to as HGWS-SVM, and its operational flowchart is shown in Fig. 4.

In Fig. 4, the process first initializes the relevant parameters and sets the penalty factor and the value range for the Kernel Parameters (KPs). Then, the grey wolf population is initialized, and based on the calculation of the diagnostic model, the top three optimal values are determined and the convergence factor is updated. LFS is used to update the location of the grey wolf population. Subsequently, PSO incorporating dynamic IWF is utilized to update the position of individual grey wolves. By continuously calculating the fitness values of individuals, the individual position with the highest fitness is selected as the optima. If the Max iterations have not been reached, the HGWO algorithm needs to be re-executed until the Max iteration is reached. Finally, the iteration of HGWO is completed and the optimal penalty factor and KPs are combined into SVM to construct a complete HGWS-SVM fault diagnosis model. Preprocessed transformer data are used as input, and then HGWS-SVM is used to implement TFD. In addition, in the real world, due to the unlabeled samples in the power TFD task, HGWO SVM cannot effectively recognize these data. Therefore, this study further integrates SDAE in deep learning to address this issue. To improve the recognition ability of SDAE for unlabeled samples, this study introduces Dropout mechanism and BN layer for improvement, as shown in Fig. 5.

Figures 5(a) and 5(b) show the Denoising Autoencoder (DAE) structure with dropout mechanism and the SDAE structure with BN layer. Among them, SDAE is composed of multiple stacked DAE structures. In Fig. 5(a), the introduction of Dropout mechanism in DAE can enhance

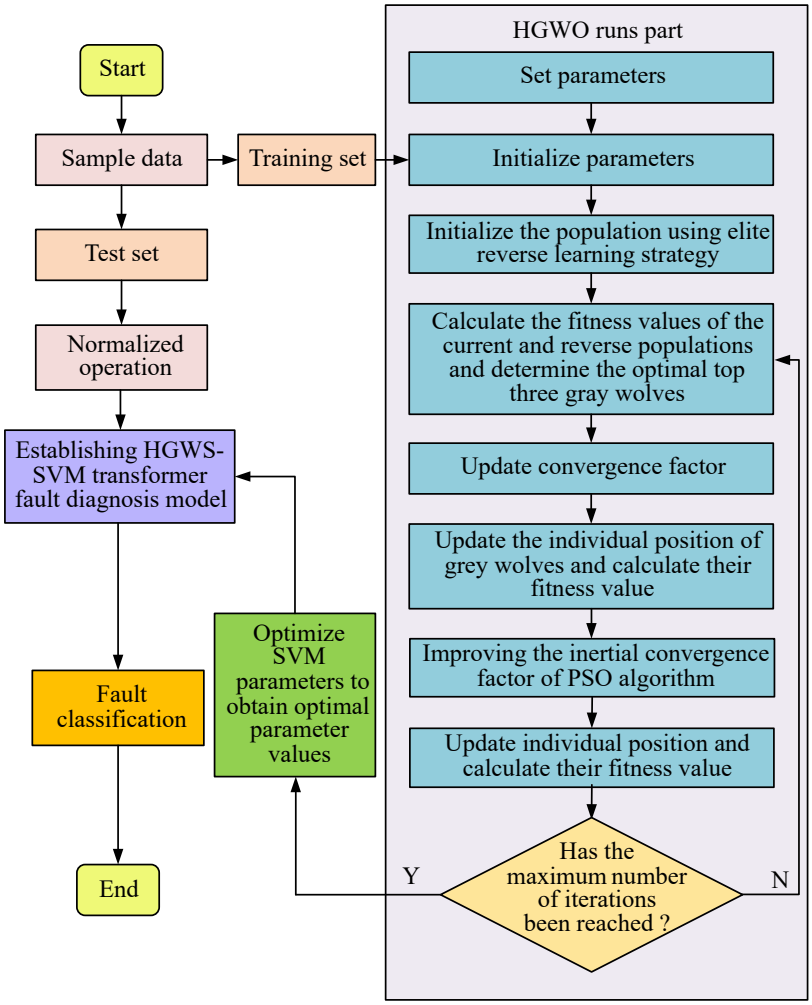
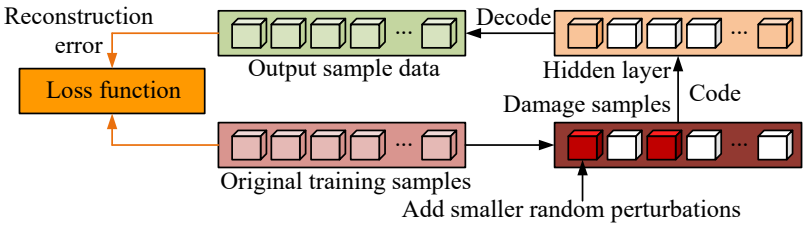


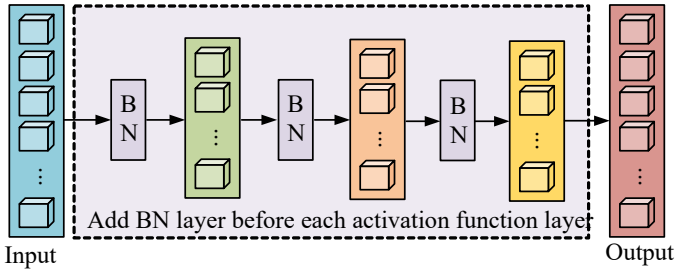
Fig. 4. Flowchart of HGWO-SVM

the model's ability to extract essential features, reduce the dependency relationship between nodes, improve classification accuracy, and effectively avoid overfitting by randomly masking the neurons in the hidden layer. In Fig. 5(b), the introduction of BN strategy can solve the problem of distribution shift and gradient disappearance in neural networks after multiple hidden layer transformations. By standardizing the input data, network training efficiency, convergence speed, and model generalization ability can be further improved. In SDAE, the formula for calculating BN on input data is shown in Eq. (10) [22, 23].

$$\mu_B = \frac{1}{n}\sum_{i=1}^{n} x_{Zi}. \tag{10}$$



(a) The DAE network architecture after introducing Dropout mechanism



(b) Topology structure of SDAE network after adding BN layer

Fig. 5. The two optimization methods of SDAE

In Eq. (10), μ_B is the mean of the input data, x_{Zi} is the input data, and n is the amount of data. The variance formula of the input data is shown in Eq. (11).

$$\sigma_B^2 = \frac{1}{n}\left(x_{Zi} - \mu_B\right)^2 . \tag{11}$$

In Eq. (11), σ_B^2 is the variance of the input data. The final TFD model constructed by combining SDAE, HGWO, and SVM is referred to as SDAE-HGWO-SVM, and its structure is shown in Fig. 6.

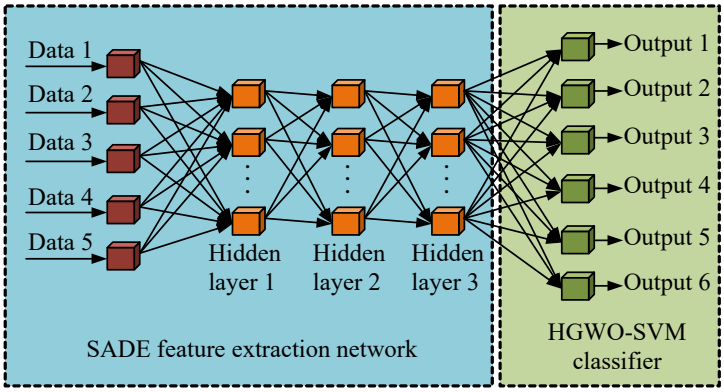


Fig. 6. Structural of the SDAE-HGWO-SVM diagnostic model

The fusion model in Fig. 6 mainly consists of two parts: SDAE and HGWS-SVM. The former is responsible for identifying data features, while the latter is responsible for fault classification of these features. The first is to obtain and classify actual fault data into two categories: unlabeled and labeled. Secondly, the SDAE is trained, with the Softmax classification layer removed but the feature extraction part retained. Then, the unlabeled data are input into the SDAE to extract features and segmented into training and testing sets. Next, the HGWS-SVM classifier is utilized to train and obtain unlabeled training set features. Finally, the test set features are input into the trained model to complete the PT fault classification task.

4. Performance testing of PTFW model based on SDAE-HGWO-SVM

To demonstrate the good performance of the SDAE-HGWO-SVM model, this study first introduced basic indicators like precision, recall, and F1 score to test the benchmark performance. Then, the actual fault warning application effect of the model was tested in four different types of PT faults.

4.1. Benchmark performance test results of SDAE-HGWO-SVM

To test the benchmark performance of SDAE-HGWO-SVM, the entire experiment is conducted on a computer equipped with an Intel Core i7-10750H processor, 16 GB of memory, and NVIDIA GeForce GTX 1650Ti graphics card with Windows 10 and Python 3.8. The main library used is TensorFlow. The dataset adopts the publicly available TFD dataset provided by IEEE Power & Energy Society. The dataset includes multiple features such as gas composition in oil, temperature, load current, etc. A total of 10 000 fault data are collected, of which 80% are from the training set and the remaining 20% are from the testing set.

In terms of parameter setting, the penalty factor range of SCM is determined by the I-GWO to be [0.1, 100] and the range of kernel function parameters is [0.01, 1]. The optimal parameters are optimized by the GWO algorithm within these ranges. The population size of GWO is set to 30 and the maximum number of iterations is set to 100. For SDAE, the number of hidden layer nodes is set to 500, the activation function is ReLU, the learning rate is initially set to 0.001, and the Adam optimizer is used for parameter update. To ensure the fairness of the comparative experiment, the parameters of other comparative models such as Differential Grey Wolf Optimizer (DGWO), I-GWO, and PSO-GWO also refer to the best practices in the literature. For example, the initial values of the inertia weight and acceleration factor in PSO are set to 0.9 and 2.0 respectively, and are dynamically adjusted during the iteration process. The parameters of DGWO and I-GWO are similar to those of the standard GWO, but they differ in differential evolution and adaptive mechanisms.

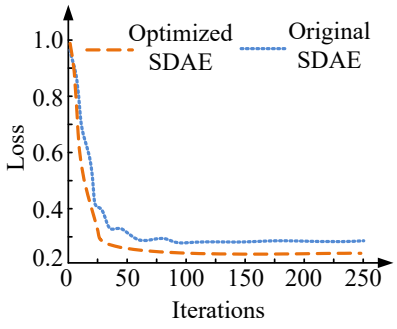
Due to the final SDAE-HGWO-SVM model is composed of multiple components, this study first uses ablation experiments to verify the benchmark performance of each component. Table 1 shows the ablation results.

Table 1 shows the combination of different fault diagnosis models. SVM+HGWO+SDAE is the SDAE-HGWO-SVM model proposed in this study. In Table 1, the research model has better benchmark performance compared to other models, with diagnostic precision, recall rate, and F1 values as high as 0.979, 0.986, and 0.983. In addition, SVM+HGWO also has good performance, with precision, recall, and F1 values of 0.954, 0.962, and 0.958. This indicates that adding SDAE

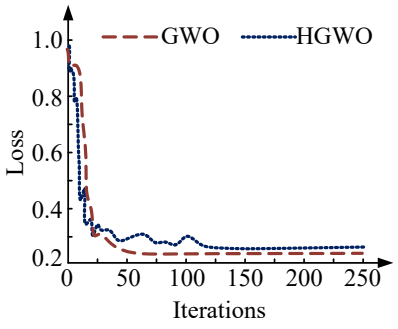
Table 1. Ablation results of the SDAE-HGWO-SVM model

Algorithm	Precision	Recall	F1
SVM	0.865	0.858	0.862
GWO	0.882	0.886	0.885
HGWO	0.902	0.914	0.911
SDAE	0.854	0.847	0.850
SVM+GWO	0.896	0.903	0.898
SVM+HGWO	0.954	0.962	0.958
SVM+SDAE	0.905	0.913	0.907
SVM+HGWO+SDAE	0.979	0.986	0.983

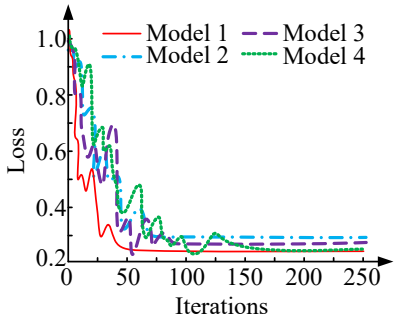
in deep learning to HGWS-SVM can lift the overall recognition performance for fault features, thereby enhancing the diagnostic accuracy. SDAE-HGWO-SVM, SVM+HGWO, SVM+SDAE, and SVM+GWO are respectively referred to as Models 1, 2, 3, and 4 for comparison, and the iteration situation of each model in the three cases is obtained, as displayed in Fig. 7.



(a) Loss curves of two SDAE



(b) Loss curves of GWO and HGWO



(c) Loss curves of four models

Fig. 7. Iteration of different models

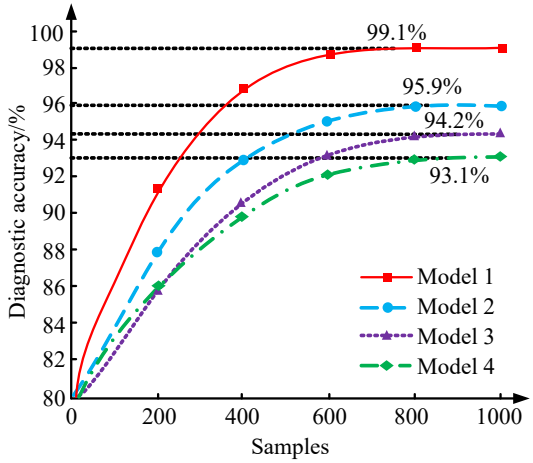
Figures 7(a) to 7(c) show the variation of loss functions with iteration times for optimizing SDAE, HGWO, and four comparative models. In Fig. 7(a), the optimized SDAE has a faster loss function drop rate in the early stage of iteration due to the introduction of the Dropout mechanism and the BN layer. It only takes 46 iterations to reach a stable state, showing good stability. The optimized SDAE has better stability than the traditional SDAE, and the number of iterations to the stable state is reduced by 46 times. The HGWO optimized with multiple strategies in Fig. 7(b) also has better stability than GWO, with a reduction of 78 iterations compared to GWO. This performance is due to the fact that HGWO adopts elite reverse learning, nonlinear control parameters, and LFS. These optimization strategies effectively improve the global search capability and reduce the risk of falling into local optima. In Fig. 7(c), Models 1−4 can achieve stability after 48, 77, 82, and 169 iterations, respectively, indicating that Model 1 has the best stability. This shows that the model that combines the SDAE feature extraction capability with the HGWO strategy can find the optimal solution faster and show smaller fluctuations during training. In contrast, the SVM+GWO model fluctuates greatly during the convergence process and requires 169 iterations to reach stability, indicating that its local optimization capability is insufficient and it is easy to fall into the local optimal. Figure 8 shows the results of testing the fault diagnosis accuracy of four models.

Figures 8(a) and 8(b) show the fault diagnosis accuracy of four models in two sets. The highest fault diagnosis accuracy of Models 1–4 can reach 99.1%, 96.0%, 94.7%, and 93.2%. Model 1 performs better in diagnosing fault data than the other three models. Finally, to verify the effect of the I-GWO proposed by the institute, three GWO variants, DGWO,I-GWO and PSO-GWO, are introduced as comparison models. The results of multi-index testing are shown in Table 2.

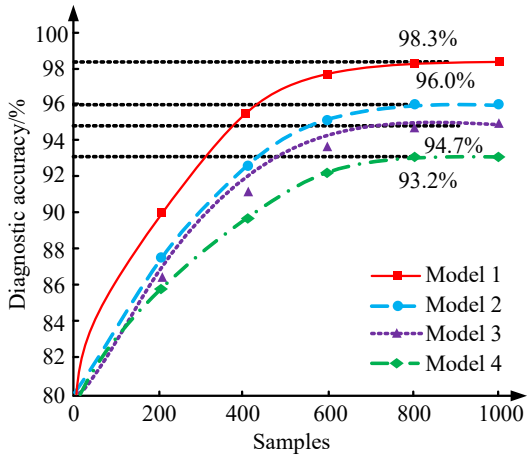
Table 2. Multi-index test results

Model	**Search efficiency (SE, solutions/time)**	**Model complexity (MC, MB)**	**Convergence stability (CS, %)**
GWO	137.45 solutions/sec	261.84 MB	8.37%
DGWO	149.57 solutions/sec	253.78 MB	7.19%
I-GWO	171.34 solutions/sec	279.62 MB	5.43%
PSO-GWO	192.85 solutions/sec	231.47 MB	4.58%
SDAE-HGWO-SVM	210.43 solutions/sec	241.52 MB	3.85%

In Table 2, the proposed SDAE-HGWO-SVM model performs better than other models in terms of search efficiency, model complexity, and convergence stability. Compared with the original GWO model, the SDAE-HGWO-SVM's search efficiency increases from 137.45 solutions/second to 210.43 solutions/second. This is due to the introduction of the combination of elite reverse learning, LFS, and PSO, which enables the model to perform global search. A better balance is achieved between local optimizations. In terms of model complexity, the complexity of the SDAE-HGWO-SVM model is 241.52 MB, which is about 10 MB more than the original GWO. For the overall benefit point, a 5% improvement in accuracy is valuable in fault diagnosis, especially in the application of key equipment such as PTs. Even a small improvement in accuracy may



(a) Training set



(b) Test set

Fig. 8. Fault diagnosis accuracy of different models in two datasets

significantly reduce the misjudgment rate, thereby significantly reducing the misjudgment rate and the risk of system failure. Although the complexity of the model has increased, compared to I-GWO's 279.62 MB, SDAE-HGWO-SVM has more efficient resource utilization and still maintains a high diagnostic accuracy. In addition, in terms of convergence stability, the fluctuation range of SDAE-HGWO-SVM is 3.85%, which is significantly better than the original GWO's 8.37%. This shows that the proposed model is more stable over multiple iterations and reaches convergence faster. Although PSO-GWO performs well in search efficiency and convergence stability, its model complexity is slightly lower than that of the SDAE-HGWO-SVM, resulting in slightly insufficient performance in some complex problems.

4.2. Analysis of the application effect of SDAE-HGWO-SVM in fault diagnosis

To further confirm the performance of the research model in practical TFD problems, this study collected transformer fault data from a certain power company over the past three years. After preprocessing, it was found that these data mainly contain four types of faults, namely short-circuit fault (Type 1), winding fault (Type 2), insulation fault (Type 3), and overheating fault (Type 4). Table 3 shows the diagnostic performance of Mean Diagnostic Time (MDT) and Average Diagnostic Accuracy Rate (ADAR) of four further tested models on four types of fault data.

Table 3. Diagnostic effects of different models on four fault types

Fault type	Index	Model 1	Model 2	Model 3	Model 4
Type 1	MDT/s	0.15 s	0.23 s	0.37 s	0.52 s
	ADAR/%	98.56%	95.14%	88.45%	85.16%
Type 2	MDT/s	0.08 s	0.21 s	0.29 s	0.44 s
	ADAR/%	99.21%	95.03%	89.01%	86.58%
Type 3	MDT/s	0.10 s	0.19 s	0.35 s	0.51 s
	ADAR/%	98.74%	94.51%	88.67%	85.03%
Type 4	MDT/s	0.11 s	0.26 s	0.32 s	0.48 s
	ADAR/%	98.92%	93.82%	88.47%	85.34%

In Table 3, the ADAR of Model 1 for all four types of faults is above 98%, with the highest reaching 99.21%, and the MDT is within 0.2 s, with the lowest requiring only 0.08 s. However, there are certain differences in the diagnostic performance of different fault types. The ADAR values of short-circuit faults and insulation faults are significantly higher than those of winding faults and overheating faults. This may be because the characteristics of short-circuit and insulation faults are more obvious and easy to be recognized by the model. The data distribution of winding and overheating faults is more complex, resulting in the model showing relatively low accuracy and longer diagnosis time under these fault conditions. Overall, the model performs better in fault types with obvious characteristics, while there is still room for improvement in performance under complex fault conditions. Figure 9 shows the error performance of four models during the diagnostic process.

Figures 9(a) to (9d) show the Average Diagnostic Error (ADE) values for four types of faults tested by Models 1 to 4. The ADE of Models 1–4 in actual diagnostic tasks can reach as low as 0.006%, 0.021%, 0.063%, and 0.125%. Similarly, for short circuit and insulation fault, two fault types with relatively regular data distribution, the model shows extremely low errors. In contrast, the models for winding faults and overheating faults exhibit greater complexity and larger errors. Figure 10 shows the CPU consumption of four models when completing fault diagnosis tasks.

In Fig. 10, as the amount of resources grows, the CPU execution time of each diagnostic model shows a continuous increasing trend. When the number of resources reaches 300, the CPU consumption of Models 1–4 is 13.7 s, 16.8 s, 18.2 s, and 23.5 s. The main reason for this phenomenon is the parallel processing capability of the system. When the number of resources reaches a certain level, the model can more effectively utilize parallel computing, thereby offsetting

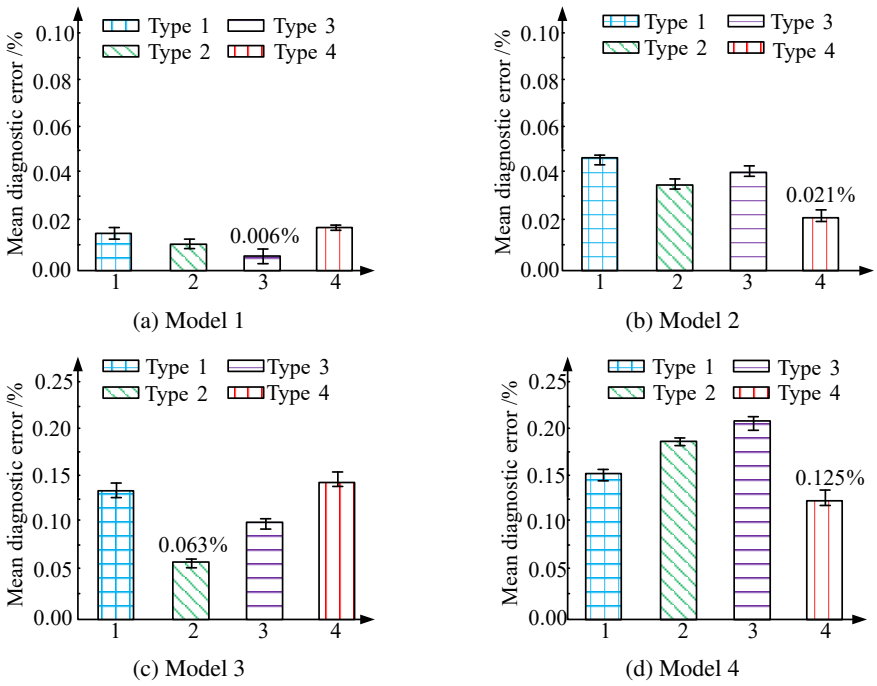


Fig. 9. Error representation of four types of faults diagnosed by different models

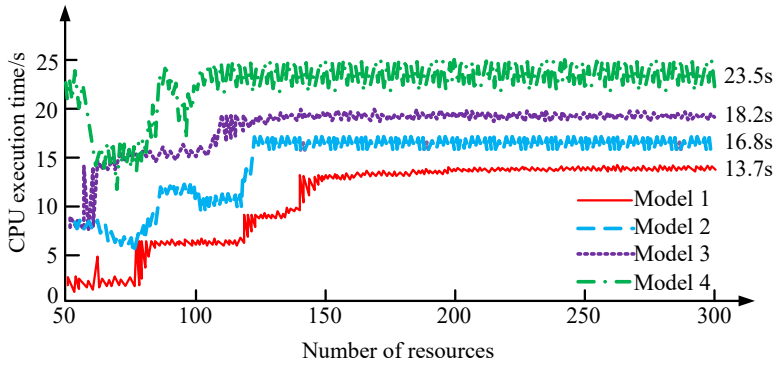


Fig. 10. CPU consumption of different models in the diagnosis process

the computational burden brought by additional resources. Therefore, for resources in excess of 150, the increase in time is no longer significant. While the CPU time no longer exhibits a notable increase, the incorporation of additional resources continues to enhance the analytical outcomes. However, this impact tends to diminish gradually. This is consistent with the diagnostic accuracy trend in Figure 8. After the resources increase to a certain number, the performance improvement tends to be saturated and the model reaches its optimal state.

5. Conclusion and future works

Aiming at the problem of slow diagnosis speed and low accuracy of traditional PT fault diagnosis model, a new SDAE-HGWO-SVM model was proposed. The enhanced GWO algorithm was mainly used to optimize the parameter tuning of SVM. The enhanced GWO algorithm aimed to improve classification performance and reduce calculation time, thereby achieving faster and more accurate fault diagnosis. The ablation test of this research model showed that its precision, recall, and F1 value in benchmark performance testing were as high as 0.979, 0.986, and 0.983, significantly higher than other combination types. During the iterative training process, the research model only needed to iterate 48 times to achieve stability. The diagnosing precision in the training and testing sets was 99.1% and 98.3%, respectively. In practical applications, the ADAR of the research model for diagnosing Type 2 was the highest, reaching 99.21%, while its MDT was also the lowest, completing the diagnostic task in only 0.08 s. In terms of error performance, the ADE of the research model was the lowest at 0.006%. In summary, the SDAE-HGWO-SVM model demonstrates significant superiority and practicality in the field of power TFD. Although the proposed SDAE-HGWO-SVM model shows excellent performance in PT fault diagnosis, it still has some limitations. First, the computational complexity of the model is high, which may prolong the training time when dealing with large-scale datasets or real-time fault detection. Second, although it performs well in the experimental dataset, the model effect needs to be further verified in the case of unbalanced datasets or lack of labeled samples. Finally, the parameters of the model may need to be readjusted in different scenarios. Future research can optimize the computational efficiency and complexity of the model, especially in large-scale data sets and real-time detection. To solve the problem of unbalanced data sets and sample scarcity, data enhancement or transfer learning techniques can be combined to improve the adaptability of the model. In addition, the model can also be applied to other equipment fault detection in smart grids or abnormal diagnosis of industrial systems to verify its generalization ability.

References

[1] Deng Y., Ruan J., Dong X., Huang D., Zhang C., *Inversion detection method of oil-immersed transformer abnormal heating state*, IET Electric Power Applications, vol. 17, no. 1, pp. 134–148 (2023), DOI: 10.1049/elp2.12249.

[2] Hebbi C., Mamatha H., *Comprehensive dataset building and recognition of isolated handwritten kannada characters using machine learning models*, Artificial Intelligence and Applications, vol. 1, no. 3, pp. 179–190 (2023), DOI: 10.47852/bonviewAIA3202624.

[3] Wang C., Wu Z., Huang T., Wu X., Huang H., *Research on automatic detection of gradual fault of high voltage electric energy metering transformer based on fuzzy rough set and whale optimization algorithm*, Journal of Vibroengineering, vol. 26, no. 3, pp. 551–566 (2024), DOI: 10.21595/jve.2023.23596.

[4] Yan P., Chen F., Kan X., Zhang H., Wang J., Li G., *Research on transformer fault diagnosis based on an IWHO optimized MS1DCNN algorithm and LIF spectrum*, Analytical Methods, vol. 15, no. 29, pp. 3562–3576 (2023), DOI: 10.1039/d3ay00713h.

[5] Yan X., Lin Z., Lin Z., Vucetic B., *A novel exploitative and explorative GWO-SVM algorithm for smart emotion recognition*, IEEE Internet of Things Journal, vol. 10, no. 11, pp. 9999–10011 (2023), DOI: 10.1109/JIOT.2023.3235356.

[6] Tang J., Zhao Q., *Motor rolling bearing fault diagnosis based on MVMD energy entropy and GWO-SVM*, Journal of Vibroengineering, vol. 25, no. 6, pp. 1096–1107 (2023), DOI: 10.21595/jve.2023.23046.

[7] Liu L., Liu Z., Qian X., *Rolling bearing fault diagnosis based on generalized multiscale mean permutation entropy and GWO-LSSVM*, IET Science, Measurement & Technology, vol. 17, no. 6, pp. 243–256 (2023), DOI: 10.1049/smt2.12149.

[8] Hui S., Longxiang Y., Shuangquan G.U.O., *GWO-ResNet power transformer fault diagnosis method based on data augmentation and feature attention mechanism*, Modern Electric Power, vol. 41, no. 2, pp. 392–400 (2024), DOI: 10.19725/j.cnki.1007-2322.2022.0163.

[9] Wang H., Yao H., Guo Q., Yu X., Zhang X., Cong L., *A transformer fault diagnosis method based on multiscale 1DCNN*, Journal of Electric Power Science and Technology, vol. 38, no. 4, pp. 104–112 (2023), DOI: 10.19781/j.issn.1673-9140.2023.04.011.

[10] Haoran X., Ziyi W., *Condition evaluation and fault diagnosis of power transformer based on gan-cnn*, Journal of Electrotechnology, Electrical Engineering and Management, vol. 6, no. 3, pp. 8–16 (2023), DOI: 10.23977/jeeem.2023.060302.

[11] Lu W., Shi C., Fu H., Xu Y., *Research on transformer fault diagnosis based on ISOMAP and IChOA-LSSVM*, IET Electric Power Applications, vol. 17, no. 6, pp. 773–787 (2023), DOI: 10.1049/elp2.12302.

[12] Men Z., Hu C., Li Y.H., Bai X., *A hybrid intelligent gearbox fault diagnosis method based on EWCEEMD and whale optimization algorithm-optimized SVM*, International Journal of Structural Integrity, vol. 14, no. 2, pp. 322–336 (2023), DOI: 10.1108/IJSI-12-2022-0145.

[13] Tian H., Wu H., *Fault Distance measurement method based on wavelet energy spectrum and BWO algorithm optimized CNN-GRU hybrid neural network*, Academic Journal of Science and Technology, vol. 11, no. 1, pp. 247−256 (2024), DOI: 10.54097/s30txd46.

[14] Seyyedabbasi A., Kiani F., *I-GWO and Ex-GWO: improved algorithms of the Grey Wolf Optimizer to solve global optimization problems*, Engineering with Computers, vol. 37, no. 1, pp. 509–532 (2021), DOI: 10.1007/s00366-019-00837-7.

[15] Kalita K., Pal S., Haldar S., Chakraborty S., *A hybrid TOPSIS-PR-GWO approach for multi-objective process parameter optimization*, Process Integration and Optimization for Sustainability, vol. 6, no. 4, pp. 1011–1126 (2022), DOI: 10.1007/s41660-022-00256-0.

[16] Ma R., Karimzadeh M., Ghabussi A., Zandi Y., Baharom S., Selmi A., Maureira-Carsalade N., *Assessment of composite beam performance using GWO–ELM metaheuristic algorithm*, Engineering with Computers, vol. 38, no. 3, pp. 2083–2099 (2022), DOI: 10.1007/s00366-021-01363-1.

[17] Deng Z., Zhao C., Leng J., Zhai G., Wang X., *Evaluation method of transformer insulation aging state based on IWOABP algorithm*, Journal of Electric Power Science and Technology, vol. 38, no. 5, pp. 253–261 (2024), DOI: 10.19781/j.issn.1673-9140.2023.05.026.

[18] Li H., Dou L., Li S., Kang Y., Yang X., Dong H., *Abnormal state detection of OLTC based on improved fuzzy C-means clustering*, Chinese Journal of Electrical Engineering, vol. 9, no. 1, pp. 129–141 (2023), DOI: 10.23919/CJEE.2023.000002.

[19] Jin Z., He D., Lao Z., Wei Z., Yin X., Yang W., *Early intelligent fault diagnosis of rotating machinery based on IWOA-VMD and DMKELM*, Nonlinear Dynamics, vol. 111, no. 6, pp. 5287−5306 (2023), DOI: 10.1007/s11071-022-08109-8.

[20] Xue R., Cai Y., *Optimization of parallel SVM algorithm for big data*, Journal of Computational Methods in Sciences and Engineering, vol. 24, no. 2, pp. 1253–1266 (2024), DOI: 10.3233/JCM-247335.

[21] Liu S., Liu Y., Shan L., Wang Q., Sun Y., He L., *Hybrid Conditional Kernel SVM for Wire Rope Defect Recognition*, IEEE Transactions on Industrial Informatics, vol. 20, no. 4, pp. 6234–6244 (2024), DOI: 10.1109/TII.2023.3344135.

[22] Saufi S.R., Isham M.F., Ahmad Z.A., Hasan M.D.A., *Machinery fault diagnosis based on a modified hybrid deep sparse autoencoder using a raw vibration time-series signal*, Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 4, pp. 3827–3838 (2023), DOI: 10.1007/s12652-022-04436-1.

[23] Jang J.G., Noh C.M., Kim S.S., Shin S.C., Lee S.S., Lee J.C., *Vibration data feature extraction and deep learning-based preprocessing method for highly accurate motor fault diagnosis*, Journal of Computational Design and Engineering, vol. 10, no. 1. pp. 204–220 (2023), DOI: 10.1093/jcde/qwac128.