International Journal of Advanced Science

https://doi.org/10.70731/48tvx723

Power Transformer Fault Diagnosis Based on Multi Class SVM and IPSO

Zihan Bai [a,*], **Ziyu Zhao** [b], **Mingshen Xu** [c]

[a] *School of Automation and Electrical Engineering, Lanzhou Jiaotong University.Lanzhou,Gansu 730070, China*
[b] *International Business School, Hainan University, Haikou 570228, China*
[c] *Department of Mathematics And Physics, North China Electric Power University, Baoding 071003, China*

KEYWORDS

Power Transformer,
Fault Diagnosis,
PSO,
SVM

ABSTRACT

This article focuses on classifying a variety of faults in power transformers with high precision, using an Improved Particle Swarm Optimization-Support Vector Machine (IPSO-SVM) system designed for fault diagnosis. The process begins with the identification of five distinct gases dissolved in oil, serving as diagnostic features. Minimal output encoding is then used to construct multiple binary support vector machine (SVM), facilitating a multiclass classification of transformer faults. While other studies often combine traditional Particle Swarm Optimization (PSO) algorithms with SVMs, our approach employs an enhanced PSO algorithm. This improved algorithm allows for the optimization of inertia and learning factors, values of which adapt based on iteration counts. The PSO is then leveraged on the optimization of the penalty factor and the radial basis function of SVM, thereby improving its classification performance. Simulation results indicate that our IPSO-SVM methodology achieves 90% and 92% accuracy in training and testing sets, respectively. This method significantly enhances the accurateness of transformer malfunction diagnosis, exhibiting superior diagnostic precision compared to traditional power transformer malfunction diagnosis methods.

1. Introduction

Power transformers, a vital portion of the safe performance of the power grid, serve as essential energy conversion and transmission equipment [1]- [2]. Gas chromatography analysis of dissolved gases in oil has become a sensitive and effective method for early diagnosis of internal latent faults and their development degree. Similar to the "blood test" of the human body, the analysis of dissolved gases in oil has become an important tool for status diagnosis and fault classification of power transformers. When transformers work normally, because of aging and deterioration of solid insulation and oil, a little gas, such as H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, CO_2, etc, will be decomposed. When overheating faults, discharge

* Corresponding author. E-mail address: 15002544850@163.com

Received 20 December 2024; Received in revised from 31 January 2025; Accepted 12 February 2025; Available online 28 February 2025.

faults, or internal insulation dampness occur in the interior of transformer, the content of these gases will gradually rise. Therefore, different concentrations of gases represent different faults, and this method is used as the basis for this article [9].

Malfunction diagnosis methodology for transformers on the base of dissolved gas processing in transformer oil include the characteristic gas method, the IEC three - ratio method, the Rogers ratio method, and the Duval triangle method. However, most of these methods rely heavily on empirical knowledge, leading to some degree of subjectivity and uncertainty. Therefore, methods such as artificial neural networks (ANN) [10], random forest (RF) [12], extreme learning machine (ELM) [13], and SVM have been widely researched and employed in transformer fault malfunction. ANN demands plenty of data samples during the training phase, resulting in slower convergence speed and increased learning costs. RF tends to fall into local optima during the search phase, compromising diagnostic accuracy. Despite ELM's high learning rapidity and strong generalization ability, the weights and thresholds of its import and hidden layers are stochasticly set and fixed, rendering it susceptible to overfitting and compromising stability. Conversely, SVM demonstrates simplicity, high training efficiency, and strong generalization ability. To fall into local optima is slightly possible and it is more apt for small sample learning—making it particularly useful for addressing the challenge of limited transformer fault samples and diverse, complex faults.

This article combines multi classification SVM and dissolved gas analysis technology to diagnose transformer faults. In the second section, the classification principle of SVM and its application core skills are introduced, and multi classification SVM using combined coding method is introduced. In the third section, the principle of PSO and the improved PSO method are explained to establish the best diagnosis model of SVM classifier. The adoption of cross validation method improved the overall generalization performance of the model. In the fourth section, it was verified through a transformer fault diagnosis example and the methodology this paper presents has higher diagnostic accuracy than GWO algorithm.

2. Multi Classification SVM

SVM is a commonly used machine learning algorithm that can be used for fault diagnosis. Mapping data points into high - dimensional space and discovery a hyperplane to segment diverse types of data points is the main contain of SVM. Mapping data points into high - dimensional space and searching for a hyperplane to segment diverse types of data points are the main contains of SVM. In fault diagnosis, SVM can be applied on the classification of diverse sorts faults and predict unknown fault situations based on known fault data.

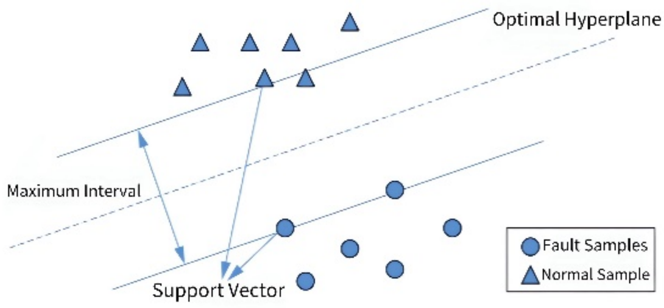

Figure 1 I SVM Classification

2.1. Classification Model

SVM optimization uniquely considers both the minimization of empirical risk and structural risk. This dual focus ensures more stable classification and enhances generalization ability. SVM shines especially when the sample data size is limited, as it can still identify the optimal solution. To discovery a plane—or an optimal hyperplane—in the sample or feature space is the basic idea of SVM, that maintains the largest possible distance from each category of samples.

The SVM principle is underpinned by three core concepts: margin, duality, and kernel techniques. At its essence, SVM aims to discover a hyperplane in the feature space that affords the maximum margin classifier. The nearest sample points to the hyperplane are the support vectors, and its working principle is shown in the Figure.1.

The fundamental principle of SVM is to assume a given sample set $\left\{ (x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n) \right\}$ among $x_i \in r^p, i \in 1,2,\ldots,n$, n is the number of p-dimensional vectors.

$$f(x) = wx + b \tag{1}$$

In the equation, $f(x)$ is the predicted value refunded by the regression function; w, b are parameters in the SVM model.

Transforming regression problems into convex optimization problems

$$\min(\|w\|^2/2) \tag{2}$$

The constraint conditions are

$$\begin{cases} y_i - wx_i - b\tilde{N}\epsilon \\ wx_i + b - y_i\tilde{N}\epsilon \end{cases} \tag{3}$$

In the inequation, y_i is the corresponding true value; ϵ is the error of insensitive loss function.

There is a portion of correctly classified training sample data in the training model, introducing relaxation variables ξ, ξ_i^*.

The above problem is transformed into

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* \tag{4}$$

The constraint conditions are

$$\begin{cases} y_i - wx_i - b\tilde{N}\epsilon + \xi_i \\ wx_i + b - y_i\tilde{N}\epsilon + \xi_i \\ \xi_i > 0 \\ \xi_i^* > 0 \end{cases} \tag{5}$$

Using Lagrange functions to solve convex optimization problems

$$L = \min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i + \xi_i^* - \sum_{i=1}^{n} \alpha_i^* \left(\xi_i + \epsilon - y_i + wx_i + b\right) - \sum_{i=1}^{n} \left(n_i\xi_i + n_i\xi_i^*\right) \tag{6}$$

According to the KKT conditions, it can be concluded that

$$\begin{cases} \frac{\partial L}{\partial b} = \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right) = 0 \quad 0\tilde{N}\alpha_i, \alpha_i^*\tilde{N}C \\ \frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i = 0 \rightarrow w = \\ \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)x_i = 0 \\ w\left(\alpha_i - \alpha_i^*\right) = -\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\cdot \\ \left(x_i x_j\right) + \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)y_i + \sum_{i=1}^{n}\left(\alpha_i - \alpha_i^*\right)\varepsilon \end{cases} \tag{7}$$

In the equation, α_i, α_i^* are the Lagrange multiplier.

Calculate from equation (4) α_i, α_i^* substitute into equation (2) and $f(x)$ to obtain the regression function.

$$f(x) = \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)K(x_i x_j) + b \tag{8}$$

Where $x_i x_j$ is the new input situation, $K(x_i x_j)$ is the SVM kernel function. There are several possibilities for selecting kernel functions, such as linear ker-

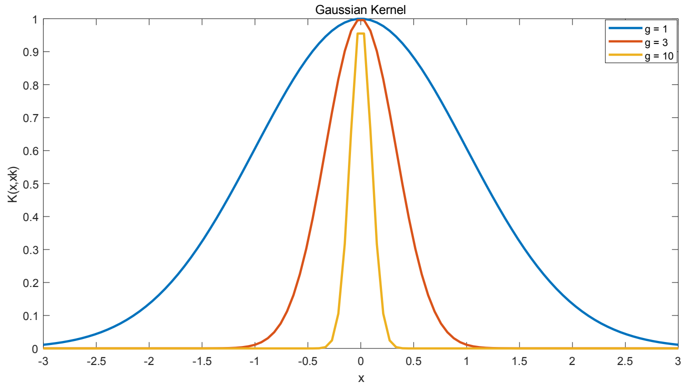

Figure 2 l Diagnostic Process

nels, polynomial kernels, and Gaussian radial basis function (RBF) kernels. Gaussian RBF has simpler model, lighter computational burden, and higher computational efficiency than other kernel functions. Therefore, in this article, RBF[5]-[6] is selected as the kernel function of SVM, as follows.

$$\sum_{i,j=1}^{n} K(x_i x_j) = \exp\left(\frac{-\|x_k - x\|_2^2}{\gamma_l^2}\right) \tag{9}$$

γ_l^2 is the width of RBF kernel function, which determines the complexity of sample data distribution, and further affects the generalization performance of the optimal classification hyperplane obtained by SVM in the feature space.

Figures 2 shows the function image of the Gaussian kernel function at gamma, g=1, 3, and 10, reflecting the relationship between vector similarity and the distance between two points. The larger the value of g, the narrower the image, and vice versa. Flatness or sharpness determine the criteria for judging similarity by the Gaussian kernel function. The image becomes flatter as the g value becomes smaller. Even if the distance is far, there is still obvious similarity. The resemblance between data points is amplified, which can make data points easier to be divided by hyperplane. In the process of calculating the separation hyperplane, the spatial characteristics of these points need to be considered. So how to choose a suitable g value is also very important.

2.2. Multiple Classification Methods

The standard SVM is a 2-class classifier, so this section needs to extend the 2-class SVM to the K_0 class classification. There are 7 categories to be distinguished in this paper, namely, high-energy discharge, low-energy discharge, low-temperature overheating, overheating high-temperature, medium temperature overheating, partial discharge and normal.

This article chooses the minimum output encoding method. At present, the enhancement of multi-classification in support vector machines primarily relies on two methods [3-4], multi-objective optimization and combination coding. The multi-objective optimization approach attempts to tackle all classification problems simultaneously, but this increases the number of variables to be solved and complicates the solving process. Thus, this study opts for a combination coding method, which using the method of building multi-class classification to facilitates multiple 2-class SVMs. This method integrates the combination encoding technique, which includes one-to-many encoding, one-to-one encoding, error-correcting output encoding, and minimal output encoding. This article chooses the minimum output encoding method.

Notably, the minimal output encoding method necessitates the least number of 2 - class SVM classifiers[7] for classifying K_0 class problems. It operates on the principle of treating each category as a binary classification problem, training multiple binary classifiers, and then combining them to achieve multi-class classification.

3. Fault Diagnosis Based on Improved Particle Swarm Optimization SVM

The improved PSO has better search ability and global optimization ability than the traditional PSO , and can find the optimal solution faster; Higher search accuracy can more accurately locate the global optimal solution in the search space, improving search accuracy; Better robustness can cope with different optimization problems, and also has better robustness against noise and interference; Better convergence can be achieved by introducing adaptive weights and dynamically adjusting parameters to make the algorithm more convergent; Algorithm complexity is reduced by simplifying the algorithm process, reducing computational complexity, and improving operational efficiency.

3.1. Traditional Particle Swarm Optimization Algorithm

Position and velocity vectors label each particle in the PSO algorithm, where the probable solution of the problem is signified by the position vector, and the direction and magnitude of position changes are signified by the velocity vector [10]. At the assumption of the entire amounts of particles is M, the position and velocity of the nth particle in dimension d are as follows.

$$\begin{cases} X'_n = \left(x'_{n1}, x'_{n2}, \cdots, x'_{nd}\right)^T, \quad n = 1,2,\cdots,M \\ V'_n = \left(v'_{n1}, v'_{n2}, \cdots, v'_{nd}\right)^T, \quad n = 1,2,\cdots,M \end{cases} \tag{10}$$

Through tracking their individual optimal position and group optimal position before each particle, adjust the speed and position of each particle. The optimal location can be seen as a measure of the "importance" of different aspects of problem-solving, adjusting algorithm behavior by reducing their weight to achieve better optimization. The optimal location is.

$$\begin{cases} P'_n = \left(P'_{n1}, P'_{n2}, \cdots, P'_{n,d}\right) \\ P'_g = \left(P'_{g1}, P'_{g2}, \cdots, P'_{g,d}\right) \end{cases} \tag{11}$$

In the equation, P'_n is the optimal individual position for the nth particle; P'_g is the group optimal position obtained from all particles in the previous iteration.

The speed and position update formula of the PSO algorithm is as follow:

$$\begin{cases} X_{nd}^{k+1} = X_{nd}^{k} + V_{nd}^{k+1} \\ V_{nd}^{k+1} = w'V_{nd}^{k} + c'_1 r'_1 (P_{nd}^{k} - X_{nd}^{k}) + c'_2 r'_2 (P_{gd}^{k} - X_{nd}^{k}) \end{cases} \tag{12}$$

In the formula, w is the inertia weight factor. Increasing the inertia weight factor should not be trapped in local minima, making it easier for global search. Reducing the inertia weight factor is advantageous to local search and algorithm convergence; 1c′ and 2c′ , learning factors, respectively express the self-learning capability and social learning capability of particles; 1r′ and 2r′ are stochastic numbers uniformly dispersed in [0, 1].

3.2. Improved Particle Swarm Optimization Algorithm

The performance of PSO algorithm is affected by the selection of parameters. The invariance of fixed inertia weight factor and fixed learning factor of traditional PSO algorithms makes it prone to fall into local optima. In response to this drawback, particle swarm optimization algorithm has been enhanced from two sides: inertia weight factor and learning factor. The modified approach is expressed as follows.

$$w = w_e + \frac{(w_s - w_e)(MI - IT)}{MI} \tag{13}$$

$$\begin{cases} c_1 = c_{1s} + (c_{1e} - c_{1s})\dfrac{IT^2}{MI^2} \\ c_2 = c_{2s} + (c_{2e} - c_{2s})\dfrac{IT^2}{MI^2} \end{cases} \tag{14}$$

In the equation, IT is the present number of iterations,

MI is the entire number of iterations; w_s and w_e are the original and ending values of the inertia weight factor. At the start of the iteration, a large w makes the algorithm unsuitable for getting trapped into local minima and benefit global search. In the later stage of iteration, a lesser w is beneficial for local search and algorithm convergence; C_{1s} and C_{1e} are the initial and stopping values of C_1, with C_{1s} greater than C_{1e}; C_{2s} and C_{2e} are the initial and stopping values of, and C_{2s} is less than C_{2e}. In the early stages of iteration, the large C_1 and small C_2 enable particles to have excellent self-learning capability and bad social learning capability, benefiting global search. In the later period of iteration, small C_1 and large C_2 enable particles to attain excellent social learning capability and bad self-learning capability, making the algorithm more convergent.

3.3. PSO-SVM Model Principle

The experiment was conducted on the open-source software LiBSVM and Matlab platforms. LiBSVM is a brief, convenient to use, rapid software package which can effectively recognize and regress at SVM pattern. This software can either offers compiled executable files utilized on Windows series systems, or provides source code that is easy to improve, modify, and apply on additional operating systems. This software has the following characteristics.

1) It supports different SVM algorithm.
2) Efficient multi classification.
3) Model selection for cross validation.
4) Probability estimation.
5) Support weighted SVM for imbalanced data.
6) Provide both C + + and Java source code.

When using, it is necessary to determine the appropriate penalty factor C and kernel function parameter g. In SVM, the penalty factor is controlled by the regularization parameter C. The larger the C, the stronger the penalty factor, and SVM will pay more attention to the correctness of classification rather than the maximum interval. On the contrary, if C is smaller, SVM will pay more attention to the maximum interval and be less sensitive to misclassified samples. Choosing the appropriate C value requires adjusting according to the features of the specific puzzle and the size of the dataset. High or low C values can lead to poor SVM classification performance. Generally speaking, the selection of parameter C needs determining according to the specific problem and the characteristics of the dataset, and both unduly large and unduly small C values may result in a decline in model performance.

The penalty factor C and the kernel parameter g have an important influence on the classification accurateness. In view of the excellent global search capability of particle swarm optimization, C and g are optimized by utilizing the PSO algorithm, and the

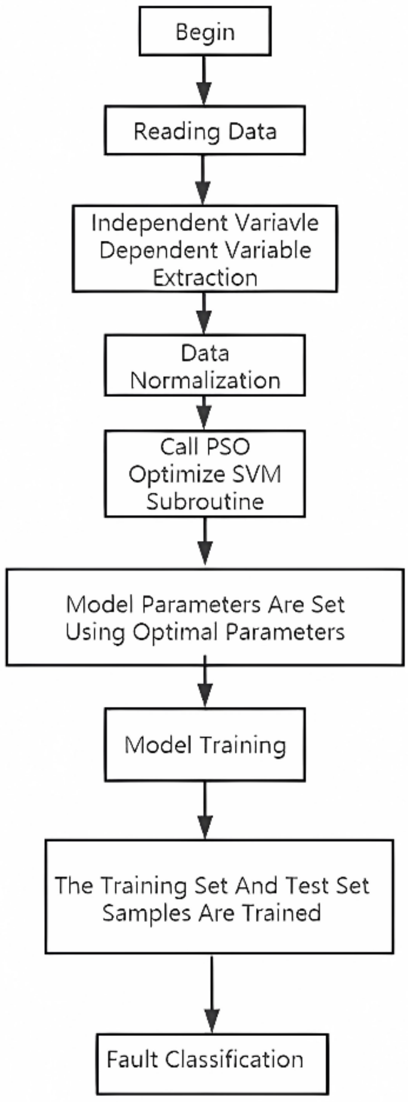

Figure 3 I Diagnostic Process

SVM recognition accurateness is taken as its fitness function.

The procedures of the fault diagnosis of oil immersed transformers are as follow:

1) Organize the training and testing sets in advance for easy reading before starting.
2) Gas parameters are independent variables and fault types are dependent variables, and the data is normalized.
3) The particle swarm optimization is utilized on the optimization of the parameters and train them in the SVM model.
4) Determine the type of transformer fault and output diagnostic results.

The PSO-SVM diagnostic process is shown in the Figure3, and the specific steps are as follows.

4. Fault Diagnosis Results and Comparison

Comparing the original SVM with PSO-SVM can clearly demonstrate the optimization improvement of the algorithm by PSO. The PSO algorithm can find the globally optimal parameter solution, while the SVM method may only find the local optimal solution. The globally optimal parameter can construct a model with better performance, avoiding the highly redundant calculation of grid search and finding the optimal solution more efficiently. PSO-SVM can optimize both C and g parameters simultaneously. Avoid overfitting caused by adjusting a parameter separately. Joint tuning of two parameters can find a more balanced solution[11].

4.1. Data and Parameter Preparation

In this paper, 337 groups of data are used to divide the faults of oil immersed transformers into the following seven types: high energy discharge, low energy discharge, low temperature overheating, high temperature overheating, medium temperature overheating, partial discharge and normal.

fault type		Corresponding code
Low energy discharge		1
Low temperature overheating		2
High energy discharge		3
High temperature overheating		4
Partial discharge		5
Normal		6
Medium temperature overheating		7

Table 1 I Types of power transformer faults and their corresponding codes

The PSO parameters are set as follows: the number of particles is 20; The number of iterations is 80.

The local learning factor, global learning factor, and inertia factor are determined by the current number of iterations, total number of iterations, and initial parameter values, its initial parameters are w_s=1.5, w_e =0.8, C_{1s}=2, C_{1e}=1, C_{2s}=2, C_{2e}=1.

Use their respective codes as the output values for transformer fault classification diagnosis. The fault types and their corresponding codes are shown in Table 1.

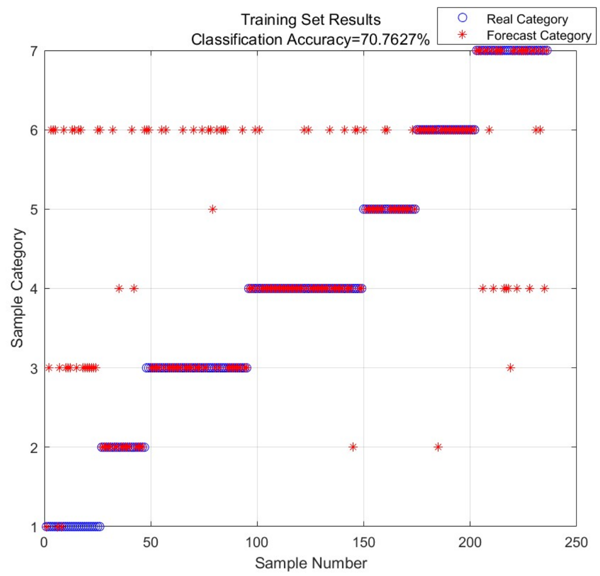

Figure 4 I Training Set Data of SVM

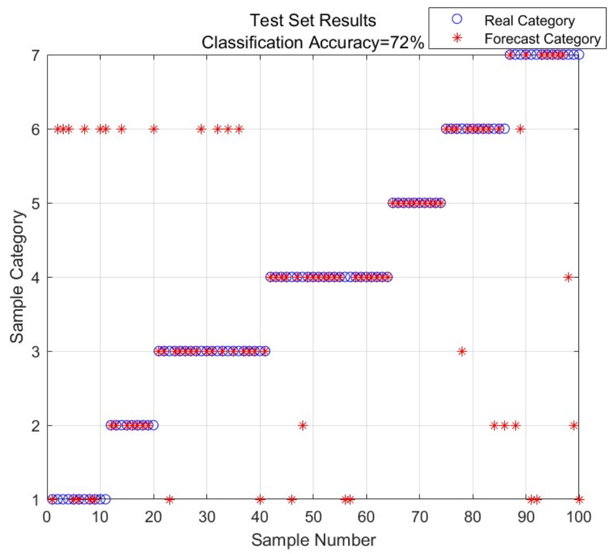

Figure 5 I Test Data of SVM

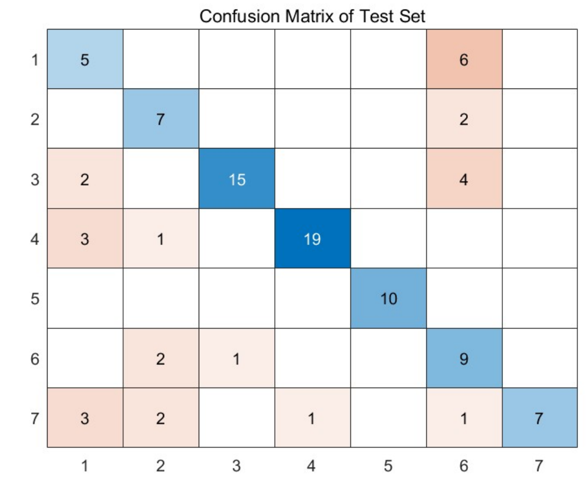

Figure 6 l Confusion of Test Set

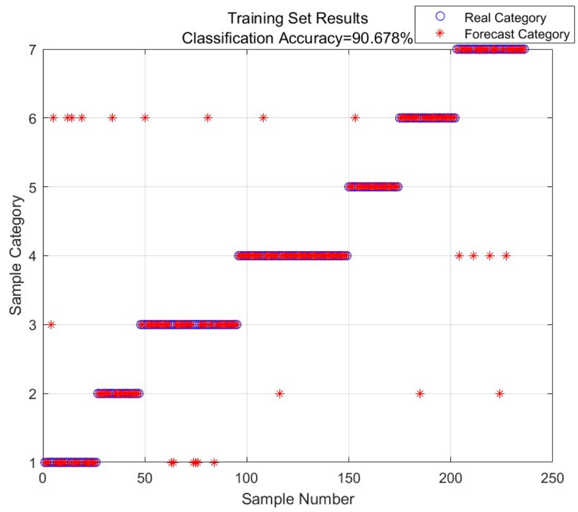

Figure 7 l Training Set Data of PSO-SVM

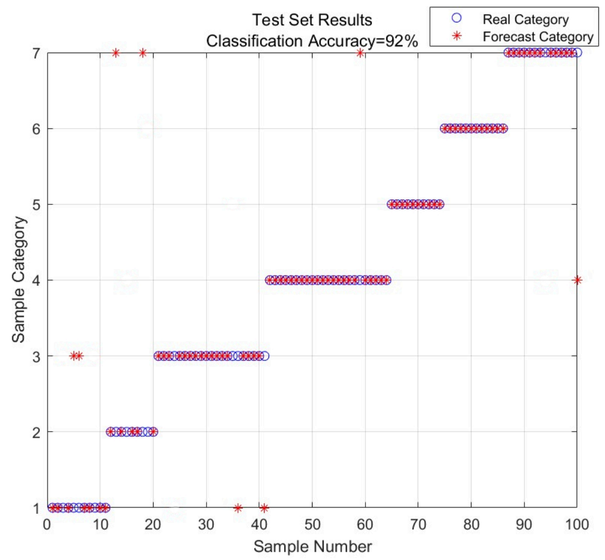

Figure 8 l Test Data of PSO-SVM

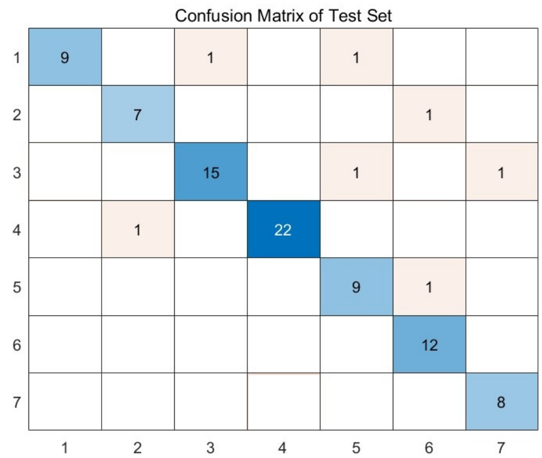

Figure 9 l Confusion of Test Set

4.2. Example Analysis

For comparison purposes, PSO-SVM and ordinary SVM were used for transformer fault diagnosis based on the same training and testing samples.

The diagnostic accuracy of using SVM and PSO-SVM for the test set and training set is shown in Figure 4, Figure 5, and Figure 7, Figure 8. The final classification accuracy of each model is summarized in the Table 2

From Figures 4 and 5, it can be seen that the diagnostic accuracy of traditional SVM is 71% and 72%.

Figure 6 shows the confusion matrix of the SVM model on the test set. The confusion matrix is a tool used to evaluate the performance of classification models. It shows the comparison between the categories predicted by the model and the real categories. Each type of label has errors, so it is neces-

Table 2 l Comparison of classification accuracy of various methods

Item	SVM	PSO-SVM
Low energy discharge	76.60%	81.80%
Low temperature overheating	90%	100%
High energy discharge	68.30%	90.00%
High temperature overheating	91.70%	95.70%
Partial discharge	100%	100%
Medium temperature overheating	91.10%	100%
Normal	51.70%	78.00%

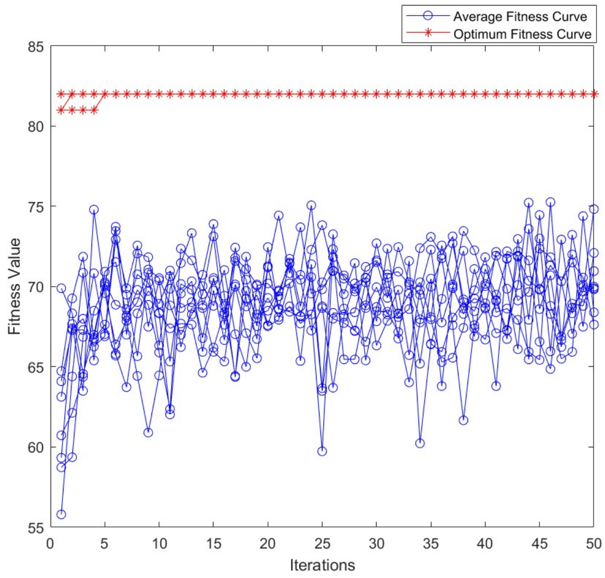

Figure 10 l Fitness convergence curve

sary to optimize the parameters of SVM to improve the overall prediction accuracy of the model.

The diagnostic accuracy of PSO-SVM is 90% and 92%. It can be seen that the improved particle swarm optimization algorithm coupled with support vector machine improves the fault recognition accuracy by 21% and 20%, respectively, indicating that using particle swarm optimization can effectively improve the fault recognition accuracy.

From Table 2, it can be seen that using IPSO-SVM to predict 100 sets of test samples resulted in a total of 8 erroneous prediction samples. The accuracy of six types of faults and their normal operation has been improved, with the accuracy of Low temperature overeating, Partial discharge, and Medium temperature overeating reaching 100%. The overall prediction accuracy of IPSO-SVM has significantly improved, indicating that IPSO has found more accurate C and g through search.

After parameter optimization by PSO optimization algorithm, the fitness convergence curve of SVM fault diagnosis model is shown in Figure 6. The horizontal axis represents evolutionary algebra, and the vertical axis represents fitness. The optimal parameters of the model are C=662.3047, g=10. These two curves represent the change rule of the average fitness value and the maximum fitness value of all particles in the population in each generation. From Figure 10, it can be analyzed that in the first 5 generations, the convergence speed of the curve was relatively fast, and then with the increase of evolutionary algebra, the convergence speed of the curve significantly slowed down, and the curve gradually stabilized and approached consistency. Therefore, this method achieves the goal of optimizing support vector machine parameters.

4.3. Grey Wolf Optimization Algorithms

The Grey Wolf Optimizer (GWO) is a heuristic optimization algorithm, based on the behavior of the natural gray wolf population. The design inspiration for this optimization algorithm comes from the social structure and cooperative hunting strategy of the grey wolf, and these behavioral strategies are cleverly introduced in the algorithm design to help find the optimal solution to the optimization problem.

The GWO algorithm simulates three key behaviors of the grey wolf population, tracking, besieging, and searching. In tracking behavior, gray wolves track their prey and attack at the best opportunity. In the siege behavior, the gray wolves will try their best to surround their prey and find the best attack path. In search behavior, if the prey escapes, the gray wolves will conduct a large-scale search. These natural behaviors are assigned to the optimization process to solve complex optimization problems.

Compared with another widely used PSO, GWO algorithm has similar characteristics, such as being able to deal with high-dimensional optimization problems, and both are swarm intelligence optimization methods. However, GWO exhibits more significant advantages in dealing with high-dimensional problems, as it can effectively avoid falling into local optima and has better global search capabilities.

Due to the excellent performance of GWO, this algorithm has gradually gained widespread recognition in some optimization problems. The optimization effect of Grey Wolf optimization algorithm can even surpass that of particle swarm optimization. This is also thanks to GWO's unique design philosophy and excellent search strategy, making it outstanding in handling complex optimization problems.

For ease of comparison, GWO-SVM is used for transformer fault diagnosis based on the same training and testing samples. The diagnostic accuracy of the training and testing sets using GWO-SVM is shown in Figures 11 and 12.

From Figures 11 and 12, it can be seen that the accuracy of the training and testing sets of GWO-SVM is 84.3% and 82.1%, respectively, which are lower than IPSO-SVM. The accuracy of IPSO-SVM exceeds GWO-SVM, indicating that the IPSO algorithm has achieved results.

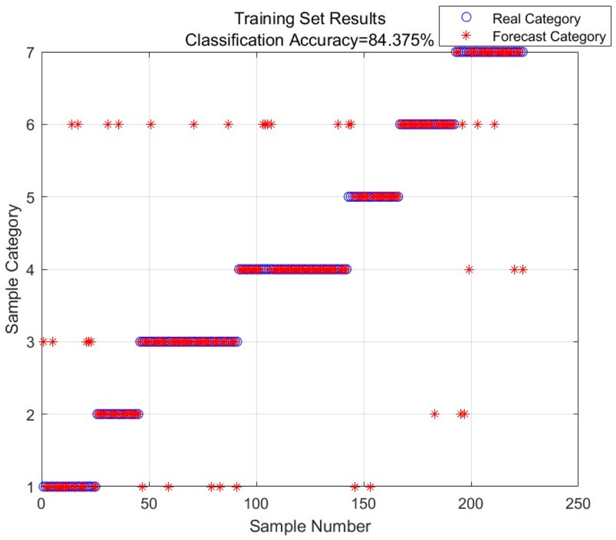

Figure 11 | Train Data of GWO-SVM

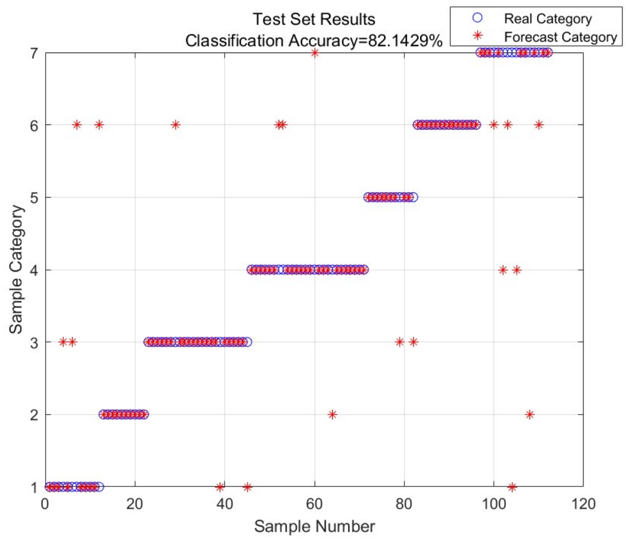

Figure 12 | Train Data of GWO-SVM

5. Conclution

In this study, we apply a multi-class SVM approach to the fault diagnosis of power transformers, constructing several 2-class SVMs to facilitate a multi-class classification for fault diagnosis. Utilizing the particle swarm optimization algorithm, we optimize the kernel function g and the penalty factor C, thereby establishing a fault diagnosis model grounded in PSO-SVM. Analysis of transformer fault diagnosis data reveals that the proposed improved PSO-SVM multi-classification method can accurately and effectively diagnose faults in power transformers, offering a higher fault diagnosis accuracy compared to traditional SVM methods. In the future, we expect to improve and expand the model, enhance its diagnostic capabilities, and explore its applications in other forms of fault detection and analysis. Explore other optimization algorithms to incorporate into SVM, or use PSO to optimize other models.

References

1. Sun, C. X. (2005). Present situation and development of condition on-line monitoring and diagnosis technology for power transformation equipment. *Electric Power, 38*(2), 1-7.
2. Chen, X. Q., Liu, J. M., & Huang, Y. W. (2012). Transformer fault diagnosis using improved artificial fish swarm with rough set algorithm. *High Voltage Engineering, 38*(6), 1403-1409.
3. Wang, Y. Q., Lv, F. C., & Li, H. M. (2006). Synthetic fault diagnosis method of power transformer based on rough set theory and Bayesian network. *Proceedings of the CSEE, 26*(8), 137-141.
4. Meng, K., Dong, Z. Y., & Wang, D. H. (2010). A self-adaptive RBF neural network classifier for transformer fault analysis. *IEEE Transactions on Power Systems, 25*(3), 1350-1360.
5. Keerthi, S. S., & Lin, C. J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation, 15*(7), 1667-1689.
6. Fei, S. W., & Sun, Y. (2008). Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm. *Electrical Power Systems Research, 78*(3), 507-514.
7. Chris, H. Q., & Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics, 17*(4), 349-358.
8. Hu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 13*(2), 415-425.
9. Hossain, M. A., Pota, H. R., & Squartini, S. (2019). Energy scheduling of community microgrid with battery cost using particle swarm optimization. *Applied Energy, 254*(15), 1-14.
10. Zheng, H. B., Wang, W., & Li, X. G. (2014). Power transformer fault diagnosis method based on multi-class least squares SVM and improved particle swarm optimization algorithm. *High Voltage Engineering, 40*(11), 3424-3429.
11. Hsu, C. W. (2009). *A practical guide to support vector classification*. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf
12. Wang, X., & Han, T. (2021). Transformer fault diagnosis based on Bayesian optimized random forest. *Electrical Measurement & Instrumentation, 58*(6), 167-173.
13. Wu, J. K., Qin, W. M., & Liang, H. H. (2019). Transformer fault identification method based on self-adaptive extreme learning machine. *Journal of Electrical Engineering, 39*(10), 181-186.