ARCHIVES OF ELECTRICAL ENGINEERING VOL. X(Y), pp. 1-23 (2027)

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Transformer fault diagnosis based on improved pelican algorithm optimized light gradient boosting machine combined with synthetic minority over-sampling technique

CHENG LIU[1*], MINGFEI ZHU[2], WEIMIN YANG[1]

[1]*Jiangsu Vocational Institute of Commerce*
Nanjing, China

[2]*State Grid Zhejiang Electric Power Co., Ltd.*
Jiaxing, China
e-mail: 200022@jvic.edu.cn

Early Access

Abstract: Accurate and timely diagnosis of incipient faults in power transformers is critical to power grid reliability, yet it remains challenging owing to severe class imbalance, insufficient discriminative features, and inadequate model tuning in dissolved gas analysis data. This paper presents an integrated diagnosis framework combining feature engineering, the synthetic minority over-sampling technique (SMOTE), and an improved pelican optimization algorithm (IPOA) to optimize the base classifier. The IPOA strengthens the original algorithm by introducing individual historical memory, global optimal guidance, and dynamic inertia weight decay, enhancing convergence stability and search efficiency. The SMOTE alleviates class imbalance by generating synthetic minority fault samples. Experiments on a real transformer dataset show that the proposed model yields outstanding performance with an accuracy of 97.00%. Ablation studies confirm that domain-informed SMOTE balancing, and IPOA-based hyperparameter optimization each significantly contribute to performance, and their integration is essential for robust and balanced fault recognition.

Key words: fault diagnosis, hyperparameter optimization, improved pelican optimization algorithm, light gradient boosting machine

1. Introduction

Electricity serves as the cornerstone of modern societal operations and economic development. As a complex network for power transmission and distribution, the stability of the power system is paramount for ensuring an uninterrupted supply to industrial, commercial, and residential sectors. Transformers, as critical components of this infrastructure, facilitate efficient power transmission through voltage conversion. They play an irreplaceable role in both step-up transmission at power plants and step-down distribution at user endpoints; consequently, their operational stability directly dictates the reliability and safety of the entire power system [1].

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Given their critical role, transformer failures are among the most severe incidents in power systems, often precipitating widespread outages and inflicting substantial economic losses on utilities and society [2].

In operation, transformers are subjected to multiple electrical, thermal, mechanical, and environmental stresses, such as overvoltages, load variations, thermal cycling, moisture, and contamination. These stresses accelerate equipment aging and insulation degradation, thereby increasing the probability of catastrophic failures. Beyond causing localized outages that impede residential and industrial activities, such malfunctions can precipitate cascading grid collapses, thereby jeopardizing the stability of the entire power network [3]. In oil-immersed transformers, incipient faults are primarily categorized into partial discharge (PD), low-energy discharge (LD), high-energy discharge (HED), medium-low temperature overheating (MLT), and high-temperature overheating (HT). Crucially, each fault mechanism induces distinct thermal or electrical decomposition of the insulating oil and paper, generating unique dissolved gas signatures that serve as the fundamental basis for diagnostic analysis [4].

Dissolved gas analysis (DGA) has emerged as one of the most widely adopted techniques for transformer fault diagnosis due to its non-invasive nature and ability to detect incipient faults before they escalate into catastrophic failures [5]. The technique is based on the principle that various fault conditions generate characteristic gases through thermal and electrical decomposition of transformer oil and solid insulation. The key gases monitored include hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), and ethane (C_2H_6), whose concentrations and ratios provide valuable insights into the nature and severity of developing faults [6]. Traditional DGA interpretation methods, such as the Duval Triangle, Rogers' Ratio, and IEC three-ratio methods, rely on fixed thresholds and empirical rules to classify fault types [7]. While these methods have been widely used in industry practice, they suffer from several limitations. First, the boundary conditions in these methods often lead to ambiguous or contradictory diagnoses when gas ratios fall near classification boundaries [8]. Second, these methods struggle to handle complex fault patterns where multiple fault types coexist or transition between states [9]. Third, the fixed thresholds may not adapt well to variations in transformer design, operating conditions, and oil types [10]. These interpretation principles are codified in the fundamental literature of the field, including the IEEE and IEC guides on the interpretation of dissolved gases [39, 40], the classic Rogers ratio code [41], and the fault taxonomy proposed by Duval [42], on which the ratio-based methods mentioned below are built.

With the rapid advancement of artificial intelligence and machine learning technologies, data-driven approaches have garnered significant attention in transformer fault diagnosis [11]. Machine learning (ML) algorithms, including support vector machines (SVMs), artificial neural networks (ANNs), random forests (RFs), and gradient boosting machines, have demonstrated superior performance compared to traditional rule-based methods by automatically learning complex nonlinear relationships between gas concentrations and fault types from historical data [12]. Among these, ensemble methods based on gradient boosting decision trees, particularly the light gradient boosting machine (LGBM), have exhibited exceptional performance due to their high efficiency, robustness to noise, and ability to handle high-dimensional features [13].

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Despite these promising results, three critical challenges persist in practical applications. First, class imbalance is prevalent in transformer fault datasets [14]. Standard machine learning algorithms trained on such imbalanced data tend to prioritize majority classes, resulting in poor recognition rates for minority fault types. These minority classes are often critical for preventing catastrophic failures [15]. Second, feature representation requires careful consideration. While raw gas concentrations provide valuable information, they may insufficiently capture the complex relationships between gases and fault characteristics [16]. Previous studies have shown that gas ratios and derived features can significantly enhance diagnostic accuracy by normalizing concentration values and highlighting relative gas production patterns [17]. However, the optimal feature set remains an open question, with existing studies proposing varying combinations without systematic evaluation [18]. Third, hyperparameter optimization is crucial for maximizing model performance. Ensemble methods involve numerous hyperparameters that profoundly impact diagnostic accuracy [19]. Traditional approaches, such as grid search and random search, are computationally expensive and often fail to identify optimal parameter combinations within feasible timeframes [20]. Efficient automated hyperparameter optimization strategies are essential to unlock the full potential of machine learning models in transformer fault diagnosis.

Recently, metaheuristic optimization algorithms inspired by natural phenomena have shown great promise for solving complex optimization problems, including hyperparameter tuning for machine learning models [21]. These algorithms, such as particle swarm optimization (PSO), a genetic algorithm (GA), grey wolf optimizer (GWO), and sparrow search algorithm (SSA), emulate the collective behavior or evolutionary processes observed in nature to efficiently explore high-dimensional search spaces [22]. The pelican optimization algorithm (POA), proposed by Trojovský and Dehghani in 2022, is a recently developed metaheuristic that mimics the cooperative hunting behavior of pelicans [23]. While the POA has demonstrated competitive performance on various benchmark functions, it suffers from premature convergence and a lack of memory mechanism, which limit its effectiveness for complex optimization tasks [24].

To address transformer fault diagnosis challenges, this paper proposes a unified framework integrating three key components: domain-knowledge-based feature engineering for constructing informative gas ratio features, the SMOTE for mitigating class imbalance through synthetic minority sample generation, and an improved pelican optimization algorithm (IPOA) for automated light gradient boosting machine (LGBM) hyperparameter tuning. The framework achieves three core contributions:

1) an IPOA incorporates historical best memory, adaptive inertia weight for exploration-exploitation balance, and refined position updates using personal and global best information, enabling faster convergence and superior solution quality.

2) The SMOTE-IPOA-LGBM pipeline synergistically addresses data scarcity and hyperparameter sensitivity, ensuring robust recognition of all fault types including rare minority classes.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

3) Comprehensive validation on 220 kV+ transformer DGA datasets confirms 97.00% accuracy through state-of-the-art comparisons, ablation studies, sensitivity analysis, and statistical testing, demonstrating significant performance improvement over existing methods.

The remainder of this paper is organized as follows. Section 2 introduces the fault mechanisms of oil-immersed transformers and the fundamentals of DGA, including the gas signature of each fault type and the interpretation of the Duval triangle. Section 3 reviews related work, focusing on machine learning-based fault diagnosis and metaheuristic algorithms for hyperparameter tuning. Section 4 presents the detailed methodology, including the proposed IPOA, the LGBM classifier, the SMOTE, and the overall framework. Section 5 presents the experimental results, including the dataset, benchmark function tests, feature selection analysis, model comparison, robustness and ablation studies. Finally, Section 6 concludes the paper.

2. Fault mechanisms, DGA fundamentals, and the Duval Triangle

The incipient faults that develop inside an oil-immersed transformer are conventionally grouped, following IEC 60599 and IEEE C57.104, according to the energy dissipated and the temperature reached at the fault site. The dataset used in this work distinguishes six conditions, and the same six are used throughout the paper: PD, LD, HED, MLT, HT, and the normal state (NS). Table 1 summarizes the dominant dissolved gas signatures for each fault category adopted in this paper.

Table 1. Dominant dissolved gas signatures for each fault category

Fault class	Dominant gases
PD	H_2 (dominant), CH_4; negligible C_2H_2
LD	C_2H_2, H_2
HED	C_2H_2, C_2H_4, H_2
MLT	CH_4, C_2H_6 (< 300°C); C_2H_4 (300–700°C)
HT	C_2H_4 (dominant), H_2; trace C_2H_2 for concurrent arcing
NS	Low and stable concentrations of all hydrocarbon gases

PD refers to low-energy corona discharge occurring in insulation voids, gas bubbles, or floating-potential regions. Such degradation progresses slowly without carbonizing insulating paper, producing hydrogen as the primary gas product and a small amount of methane, while acetylene is almost absent. PD is widely regarded as an early precursor of severe insulation failure and is closely associated with oil gas saturation and insulation dampness. LD represents low-energy sparking and surface tracking under local electric stress, which leaves discrete

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

carbonized traces on insulation surfaces and generates detectable acetylene and hydrogen. In contrast, HED corresponds to violent continuous arcing, which commonly occurs between winding turns, between leads and the transformer tank, or inside defective tap-changers. Intense arcing yields substantial hydrogen, acetylene, and ethylene, accompanied by extensive paper carbonization and even metal fusion, generally requiring immediate equipment outage and maintenance.

Thermal faults are divided into MLT and HT in this study. MLT covers low-and medium-temperature overheating below 700°C, mainly caused by overload, insufficient oil circulation, local winding hotspots, poor electrical connections, or circulating currents. Temperatures below 300°C mainly generate methane and ethane with only slight insulation discoloration, whereas temperatures ranging from 300°C to 700°C gradually take ethylene as the dominant gas and initiate mild paper carbonization. HT represents severe overheating above 700°C, originating from poor contact, core/tank circulating currents, or shorted core laminations. The gas signature is dominated by ethylene with abundant hydrogen, and trace acetylene indicates simultaneous arcing during overheating. Finally, NS denotes the normal operating state, where all hydrocarbon gases maintain low and temporally stable concentrations. Reliable diagnosis depends on long-term trending variations rather than individual sampling values.

The above fault gas correlations constitute the physical basis for mainstream DGA diagnostic methods, including Rogers ratios, IEC three-ratio criteria, and the Duval Triangle. As one of the most authoritative and widely applied graphical diagnostic tools, the Duval Triangle interprets transformer faults based on the normalized proportional distribution of three key hydrocarbon gases: CH_4, C_2H_4, and C_2H_2. The three gas components are normalized such that their summed proportions equal 100 % and projected into an equilateral triangular coordinate system, where each vertex corresponds to one of the three target hydrocarbon gases. The triangular plane is partitioned into standard diagnostic zones corresponding to PD, low-energy discharge, high-energy discharge, and three thermal zones, as well as a mixed thermal–electrical zone. A measured DGA sample is projected as a coordinate point within the triangle, and the residing zone determines the most probable fault type. Nevertheless, samples near zone boundaries often lead to ambiguous or conflicting diagnoses, limiting the reliability of standalone Duval interpretation. In this work, the physical prior knowledge embedded in Duval zoning and traditional ratio-based criteria is adopted to construct interpretable gas-ratio features for fault classification.

Early Access

3. Literature review

3.1. Intelligent diagnosis based on machine learning

With the evolution of artificial intelligence, data-driven ML methods have gradually superseded traditional ratio methods, becoming the dominant paradigm in DGA for fault diagnosis. These methods are capable of automatically excavating complex nonlinear mapping

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

relationships between characteristic gas concentrations, their ratios, and fault types, thereby significantly enhancing both the automation level and accuracy of diagnosis.

Early research primarily focused on shallow ML models, represented by ANNs and SVMs. Although traditional back propagation (BP) neural networks possess nonlinear fitting capabilities, their application to complex fault scenarios is constrained by defects such as gradient vanishing and susceptibility to local optima [25]. To overcome this bottleneck, Jin *et al*. (2023) proposed a diagnostic model based on a residual BP neural network. By stacking multiple residual modules to deepen the network hierarchy, this approach effectively mitigates the gradient vanishing problem, achieves deep fusion and expansion of gas feature information, and significantly improves the robustness of feature extraction [26]. Regarding small-sample datasets, the SVM, grounded in the principle of structural risk minimization, constructs an optimal hyperplane in high-dimensional space via kernel function mapping, demonstrating superior generalization performance. However, SVM performance is highly dependent on the selection of the penalty factor and kernel parameters; manual trial-and-error methods often fail to identify the global optimum. Consequently, the introduction of meta-heuristic algorithms for adaptive optimization has become a critical pathway. Bermani Ali *et al*. (2021) developed a bat algorithm-optimized SVM model, leveraging the global search capability of the bat algorithm to automatically optimize key parameters [27]. Similarly, Hu *et al*. (2022) combined SVM with a logarithmic arctangent transform of DGA features, which alleviates the sensitivity of SVM to parameter selection and improves the precision of classification boundary delineation [28]. Zhao and Pattanadech (2025) further combined the SVM with an improved grey wolf optimization algorithm for power transformer fault warning, enhancing parameter optimization and diagnostic reliability [43].

Despite the excellent performance of these improved shallow models, their capacity for automatic feature extraction remains insufficient as data scales expand. In recent years, deep learning techniques represented by convolutional neural networks and long short-term memory networks have been applied to the DGA diagnosis, enabling end-to-end extraction of high-order spatiotemporal features [29, 30]. Nevertheless, these approaches often face challenges such as overfitting and poor interpretability. Conversely, ensemble learning algorithms based on Boosting strategies have demonstrated exceptional performance in DGA diagnosis, emerging as a current research hotspot. Specifically, the LGBM incorporates a histogram-based decision tree algorithm and a leaf-wise growth strategy with depth constraints. This architecture substantially reduces memory consumption and accelerates training speed, making it particularly well-suited for processing the high-dimensional, sparse, and noisy DGA data prevalent in power systems [31]. In addition, Li *et al*. (2026) proposed a transformer fault diagnosis method based on multilevel acoustic information, offering a complementary perspective for multi-source intelligent diagnosis [44].

Early Access

3.2. Metaheuristic optimization for hyperparameter tuning

Hyperparameter optimization represents a critical yet computationally demanding phase in machine learning model development. Traditional approaches, such as grid search and random

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

search, suffer from exponential complexity as parameter dimensions increase, often failing to identify optimal configurations within feasible timeframes [32]. Consequently, metaheuristic algorithms, inspired by natural phenomena and evolutionary processes, have emerged as efficient alternatives for navigating high-dimensional hyperparameter spaces.

Particle swarm optimization (PSO) is a nature-inspired computational technique grounded in the social behavior of bird flocks or fish schools. It has been extensively applied to hyperparameter tuning [33]. For instance, Li *et al*. demonstrated the effectiveness of PSO in optimizing hyperparameters for recurrent neural networks in electric load forecasting. Their study confirmed that PSO significantly reduces computational complexity while yielding superior hyperparameter configurations compared to traditional methods under equivalent computational budgets [34]. Similarly, genetic algorithms (GAs), drawing inspiration from mechanisms of natural selection, have shown immense potential in complex optimization landscapes. Yokoyama *et al*. proposed a GA-based framework for machine learning hyperparameter tuning, validating its efficacy for multi-objective optimization across various classification and regression tasks [35].

However, classical swarm intelligence algorithms still face challenges such as premature convergence and entrapment in local optima when addressing high-dimensional, multi-modal, and strongly coupled hyperparameter spaces. To overcome these limitations, researchers have developed various improvement strategies to better balance global exploration and local exploitation. Yang *et al*. utilized the ergodicity and stochasticity of chaotic sequences to generate initial populations, replacing traditional random initialization. This approach enhances population diversity and prevents the algorithm from stagnating in local regions during early iterations [36]. Jiao *et al*. integrated Lévy flight mechanisms into the Sparrow Search Algorithm (SSA). This strategy enables individuals to perform large-scale jumps, effectively escaping local optima and strengthening global search capabilities [37]. Ma *et al*. improved the hunting behavior of the grey wolf optimizer by incorporating dual adaptive weights and dimensional learning. This modification allows the algorithm to prioritize global exploration in early stages while shifting focus to fine-grained local exploitation in later stages, thereby achieving a dynamic balance in search strategies [38].

In recent years, a new generation of metaheuristic algorithms inspired by novel biological behaviors has emerged, aiming to address the structural deficiencies of traditional methods at a mechanistic level. Among these, the POA has garnered significant attention due to its unique foraging mechanism [23]. Although the POA demonstrates substantial potential, standard implementations may still exhibit insufficient convergence precision or limited adaptability to specific optimization landscapes when tackling the high-dimensional, non-linear, and multi-modal hyperparameter optimization problems inherent in transformer fault diagnosis. Therefore, drawing upon the aforementioned improvement strategies, this study proposes an IPOA. By tailoring POA to address these specific challenges and applying it to optimize LGBM hyperparameters, we aim to establish a critical pathway for enhancing the accuracy and robustness of fault diagnosis.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

4. Proposed method

4.1. Overview of the proposed framework

The proposed transformer fault diagnosis framework, designated as the SMOTE-IPOA-LGBM, integrates three key components: feature engineering, data balancing through the SMOTE, and hyperparameter optimization through an IPOA. The overall workflow consists of five sequential stages, as illustrated in Fig. 1.

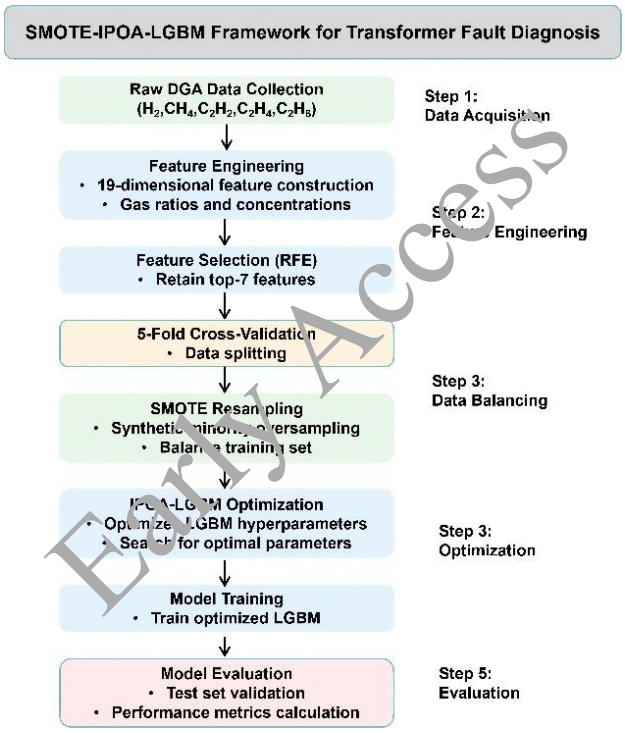


Fig. 1. Overall framework of the SMOTE-IPOA-LGBM method

Stage 1: Feature engineering

Raw DGA data containing concentrations of five key gases (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6) is collected from transformer oil samples. A 23-dimensional feature set is constructed by combining raw gas concentrations with various gas ratios that reflect the relationships between different gases and fault characteristics. Recursive feature elimination (RFE) is then applied to select the most informative features, resulting in a compact yet effective feature subset.

Stage 2: Data partitioning

The dataset is randomly divided into a training set (80%) and an independent hold-out test set (20%). Within the training set only, 5-fold cross-validation is employed during recursive

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

feature elimination to determine the optimal number of features; the test set is never involved in feature selection or hyperparameter optimization and is used exactly once for final evaluation. The SMOTE is applied exclusively to the training folds to prevent any information leakage from synthetic samples into the validation or test data.

Stage 3: Data balancing

The SMOTE is applied to the training set to address class imbalance by generating synthetic minority class samples. This step is performed independently for each fold to prevent data leakage between training and testing sets.

Stage 4: Hyperparameter optimization

The IPOA is employed to optimize the hyperparameters of the LGBM classifier. The IPOA searches for the optimal combination of parameters that maximizes diagnostic performance on the balanced training set.

Stage 5: Model training and evaluation

The optimized LGBM model is trained on the balanced training set and evaluated on the independent testing set. Multiple performance metrics, including accuracy, f1 score, miou, and kappa coefficient, are computed to comprehensively assess model performance.

Early Access

4.2. Improved pelican optimization algorithm

The pelican optimization algorithm (POA) is a nature-inspired metaheuristic that mimics the hunting behavior of pelicans, consisting of two phases: exploration and exploitation. However, the original POA suffers from premature convergence and lack of memory mechanism due to its reliance on random individuals and absence of historical information.

To address these limitations, we propose an improved pelican optimization algorithm (IPOA) that enhances search efficiency through three key strategies:

1) incorporation of individual and global best memory,
2) adaptive inertia weight for dynamic balance between exploration and exploitation, and
3) refined update rules grounded in the original POA framework.

Let x_i^t denote the position of the i-th pelican at iteration t, and $x_i^{p\text{best}}$ is the best position ever found by individual i, and $x_i^{g\text{best}}$ is the global best position in the population.

An adaptive inertia $w(t)$ is introduced to modulate search intensity, in the early stage of algorithm iteration, a larger inertia weight maintains strong velocity inertia of individuals, supporting large-scale global search in the solution space and avoiding premature convergence to local optima. In the later stage of iteration, a smaller inertia weight reduces velocity, enhances local fine-grained search capability, and helps the algorithm converge near the optimal solution. The calculation formula is shown in Eq. (1).

$$w(t) = 0.9 - \frac{t}{2 \cdot T}, \tag{1}$$

where T is the maximum number of iterations, as t increases from 1 to T, the weight $w(t)$ decreases gradually and eventually reaches 0.4. In the exploration phase, a random individual k is selected, and k is not the same as i, the update follows a guided strategy as shown in Eq. (2).

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

$$x_i^{\text{new}} = \begin{cases} w(t) \cdot x_i^t + c_1 r_1 (x_i^{p\text{best}} - x_i^t) + c_2 r_2 (x^{g\text{best}} - x_i^t) & \text{if } f(x_k^t) < f(x_i^t) \\ w(t) \cdot x_i^t + c_1 r_1 (x_i^{p\text{best}} - x_i^t) + c_2 r_2 (x_i^t - x_k^t) & \text{otherwise} \end{cases}, \tag{2}$$

where c_1 and c_2 are both positive learning factors (default 1.5), and r_1 and r_2 follow a uniform distribution over the interval from 0 to 1. Subsequently, in the exploitation phase, a local refinement is performed as shown in Eq. (3).

$$x_i^{\text{new}} = x_i^t + w(t) \cdot (1 - \frac{t}{T}) \cdot (2r - 1) \odot x_i^t , \tag{3}$$

where r is in [0,1], and $\odot$ denotes element-wise multiplication. Finally, the personal best is updated if a better solution is found, the calculation formula is shown in Eq. (4).

$$x_i^{p\text{best}} = \begin{cases} x_i^{\text{new}} & \text{if } f(x_i^{\text{new}}) < f(x_i^{p\text{best}}) \\ x_i^{p\text{best}t} & \text{otherwise} \end{cases}. \tag{4}$$

The computational complexity of the IPOA is analyzed as follows. Let N be the population size, D be the problem dimension, and T be the maximum number of iterations. The complexity of each component is:

1) Population initialization: $O(ND)$;
2) Fitness evaluation: $O(NTC_f)$, where C_f is the cost of a single fitness evaluation;
3) Position updates: $O(NTD)$;
4) Memory updates: $O(NT)$.

Therefore, the overall computational complexity of the IPOA is $O(NT(D + C_f))$, which is comparable to other swarm intelligence algorithms such as PSO and GWO. The additional operations for memory maintenance and adaptive parameter calculation do not significantly increase the computational overhead.

Early Access

4.3. The LGBM model and SMOTE method

The LGBM is an efficient gradient boosting framework based on decision trees. It achieves fast training and low memory usage for large-scale high-dimensional data through gradient based one side sampling and exclusive feature bundling techniques. Its core lies in constructing an ensemble of multiple regression trees and iteratively optimizing a differentiable loss function with a regularization term to improve prediction accuracy and avoid overfitting. For a given input sample x_i, its final predicted value $\hat{y}_i$ is the weighted sum of the predictions from all trees, and the formula as below:

$$\hat{y}_i = \sum_{m=1}^{M} w_m \cdot h_m(x_i), \tag{5}$$

in which, M is the total number of decision trees in the ensemble model, $h_m(x_i)$ is the prediction output of the m-th decision tree for the i-th sample, and w_m is the weight of the m-th decision tree, which adjusts the contribution of this tree to the final prediction, in standard gradient boosting, this weight is typically a constant or optimized via line search.

To address the class imbalance issue in transformer fault diagnosis datasets, the SMOTE method is adopted. The SMOTE generates synthetic minority class samples by linear

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

interpolation between minority samples and their nearest neighbors, rather than simply duplicating existing samples. The synthetic sample generation formula is shown in Eq. (6).

$$x_{\text{new}} = x_i + \delta \cdot (x_{\text{nn}} - x_i), \tag{6}$$

here x_i is a minority class sample, x_{nn} is a randomly selected nearest neighbor of x_i within the same class, and δ is a random scalar uniformly distributed between 0 and 1, that controls the interpolation distance.

4.4. Evaluation metrics

The specific calculation formulas of the metrics are as follows, the accuracy refers to the proportion of correctly predicted samples to the total number of samples, reflecting the overall diagnostic correctness of the model, the calculation formula is shown in Eq. (7).

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

here TP is the count of correctly predicted positive samples, TN is the count of correctly predicted negative samples, FP is the count of negative samples incorrectly classified as positive, and FN is the count of positive samples incorrectly classified as negative.

The mean intersection over union (miou) is a key metric for evaluating the overlap between predicted and true labels, widely used in multi-classification tasks. The calculation formula is shown in Eq. (8).

$$\text{miou} = \frac{1}{C}\sum_{c=1}^{C} \frac{TP_c}{TP_c + FP_c + FN_c}, \tag{8}$$

in which C denotes the total number of fault classes in the dataset, TP_c denotes the number of correctly predicted samples of class c, FP_c denotes the number of samples incorrectly predicted as class c and FN_c denotes the number of samples of class c incorrectly predicted as other classes.

The kappa measures the consistency between the model's predicted results and the true labels, eliminating the impact of random guessing. The calculation formula is shown in Eq. (9).

$$\begin{cases} \text{kappa} = \dfrac{p_o - p_e}{1 - p_e} \\ p_e = \sum_{c=1}^{C} \dfrac{N_c \cdot P_c}{N \cdot N} \end{cases}, \tag{9}$$

in which p_o is the actual accuracy of the model, and p_e is the expected consistency probability under random guessing, and N_c denotes the number of true samples of class c, P_c denotes the number of predicted samples of class c, and N denotes the total number of samples in the dataset.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

5. Experiments and results

5.1. Experimental setup and dataset

All experiments were conducted on a workstation equipped with an Intel Core i9-10900X CPU, 64 GB RAM, and an NVIDIA RTX 3090 GPU. The software environment was based on Python 3.10, with key libraries including scikit-learn 1.6.1, lightgbm 4.6.0, imblearn 0.11.0, numpy 1.26.0, and pandas 2.3.2, running on Windows 11.

Each sample corresponds to one DGA record, i.e., one set of data consisting of the concentrations of the five characteristic gases (H_2, CH_4, C_2H_6, C_2H_4, and C_2H_2) measured in μL/L (ppm) from a single oil sample by chromatographic analysis, together with its verified diagnostic label. The diagnostic labels were established by cross-checking each DGA record against the maintenance and commissioning test records of the corresponding transformer, including electrical test results and, where available, internal inspection reports. The dataset contains 1000 such records, collected from power transformers operating at voltage levels of 220 kV and above in Nanjing, China, between January 2019 and December 2023, and distributed over the six diagnostic classes: HED (202), HT (219), MLT (248), PD (253), LD (52) and NS (26).

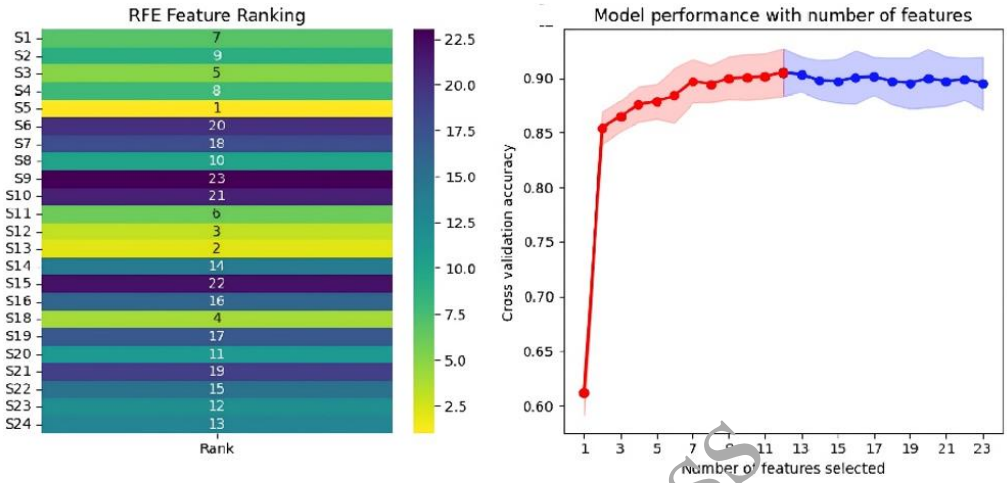
5.2. Experimental results and comparative analysis

The dataset is randomly divided into a training set (800 records, 80%) and an independent hold-out test set (200 records, 20%). The resulting test-set class counts are HT (49), MLT (35), PD (47), HED (52), LD (11), and NS (6). Within the training set, RFE with internal 5-fold cross-validation is used to select features, and the IPOA optimizes hyperparameters on the SMOTE-balanced training data.

RFE iteratively removes less informative features based on the importance scores of the LGBM classifier, yielding a ranked list of input variables. As shown in Fig. 2, features S5 (rank 1), S12 (rank 2), and S1 (rank 3) are identified as the most discriminative. Further analysis of model performance across varying numbers of selected features reveals that accuracy peaks at approximately 0.91 when seven features are retained. Beyond this point, additional features do not significantly improve performance and may introduce noise, as evidenced by the increased variance. Therefore, the top-7 features were selected to construct a compact yet effective diagnostic model. The retained features are listed in Table 2.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592



(a) (b)

Fig. 2. (a) RFE-based feature ranking; (b) cross-validation accuracy with number of selected features

To comprehensively evaluate the performance of the proposed IPOA, six benchmark functions (F7–F12) from the CEC2022 suite were employed. As shown in Fig. 3, the IPOA consistently outperforms the comparison algorithms: bald eagle search (BES), the bird swarm algorithm (BSA), sparrow search algorithm (SSA), and POA on functions F7 to F11, demonstrating superior convergence speed and solution quality, particularly in the early stages of optimization. However, on function F12, a high-dimensional composite function, the POA achieves slightly better results, indicating that the IPOA may face challenges in handling highly nonlinear and strongly coupled variable landscapes. This is consistent with the No Free Lunch theorem, which implies that no optimizer dominates universally across all problems. Overall, the IPOA achieves the best or near-best performance across most test functions, confirming its effectiveness and robustness in solving complex optimization problems.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Table 2. The seven features retained by recursive feature elimination, with their RFE rank

Feature	Definition	RFE rank
S5	C_2H_2	1
S12	$CH_4 + C_2H_6 + C_2H_4 + C_2H_2$ (total combustible hydrocarbon)	2
S1	H_2	3
S3	C_2H_6	4
S2	CH_4	5
S23	$(H_2 + C_2H_2)/(CH_4 + C_2H_6 + C_2H_4 + 1)$	6
S11	$C_2H_2/(C_2H_4 + C_2H_6 + 1)$	7

Early Access

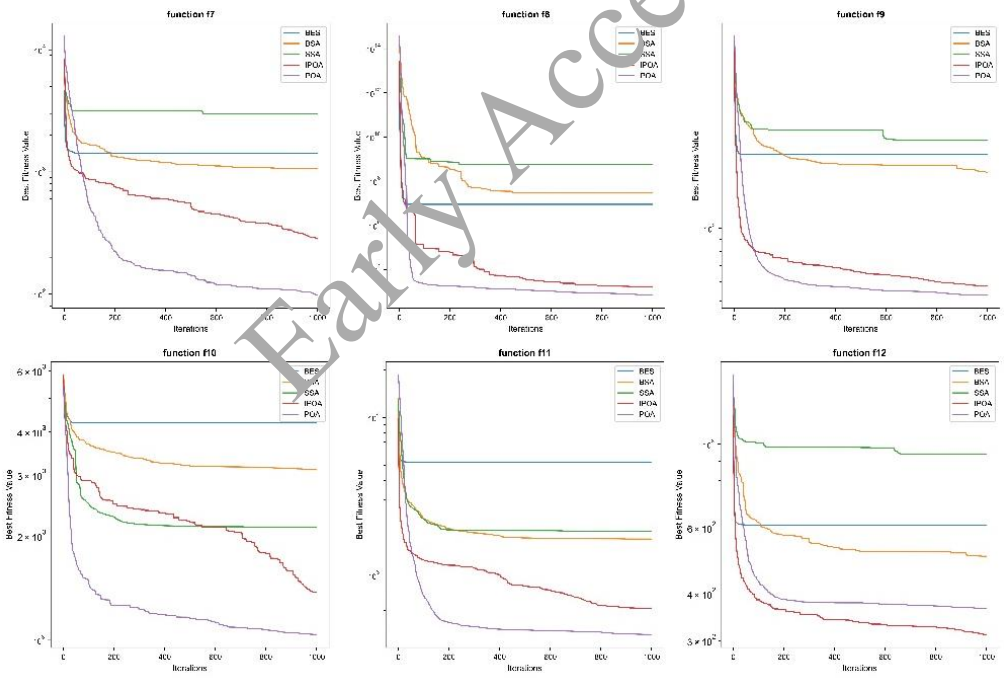


Fig. 3. Convergence comparison on CEC2022 functions (F7–F12)

Following feature engineering, the constructed input features were fed into four representative ensemble learning models, the XGB, LGBM, RF, and the gradient boosting machine (GBM). To address class imbalance in the dataset, the SMOTE was applied to the training set prior to model optimization. The hyperparameters of these models were

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

automatically optimized using the IPOA. The search ranges and the best-found configurations are summarized in Table 3.

Table 3. Search space and optimal hyperparameters

Algorithm	Parameters	Search space	Optimal values
XGB	*n*_estimators	[10, 50]	38
	learning_rate	[0.01, 0.2]	0.20
	max_depth	[10, 50]	37
	min_child_weight	[5, 20]	5
	subsample	[0.5, 1]	0.87
	gamma	[0.01, 0.2]	0.11
LGBM	*n*_estimators	[10, 50]	48
	learning_rate	[0.01, 0.2]	0.158
	max_depth	[10, 50]	33
	num_leaves	[10, 50]	39
RF	*n*_estimators	[20, 100]	23
	max_depth	[10, 50]	10
	min_samples_split	[5, 50]	5
	min_samples_leaf	[2, 10]	2
GBM	*n*_estimators	[10, 100]	50
	learning_rate	[0.01, 1]	0.02
	max_depth	[3, 20]	10

Early Access

As shown in Fig. 4, the IPOA-based optimization strategy consistently enhances the performance of all baseline models across multiple evaluation metrics. Notably, the SMOTE-IPOA-LGBM achieves the best overall results, with an accuracy of 97.00%, while also attaining the highest scores in both miou and Cohen's kappa coefficient. This demonstrates IPOA's superior global exploration capability and its ability to adaptively configure hyperparameters in complex search spaces.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

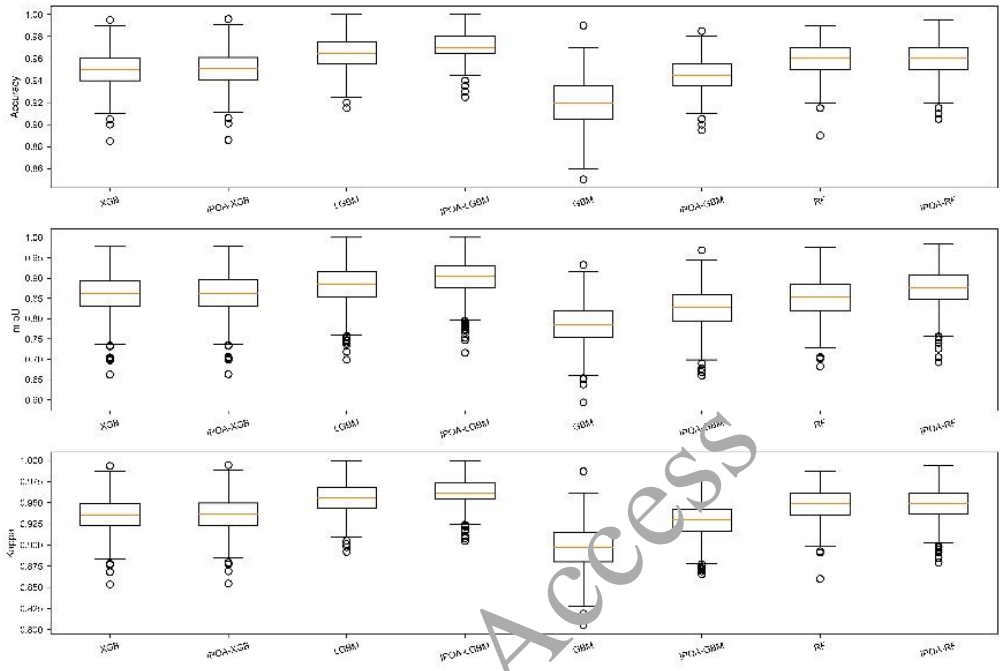


Fig. 4. Performance of ensemble models with and without IPOA across different metrics

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Table 4. Class-wise classification accuracy (%) of models for different types of faults

Models	HT	MLT	LD	PD	NS	HED
SVM	89.80	88.57	54.55	89.36	0.00	100.00
NB	95.92	85.71	72.73	93.62	83.33	94.23
KNN	87.76	80.00	72.73	95.74	83.33	96.15
DT	95.92	88.57	63.64	100.00	66.67	100.00
RF	95.92	97.14	81.82	97.87	83.33	98.08
SMOTE-IPOA-LGBM	95.92	97.14	90.91	100.00	83.33	98.08
XGB	95.92	91.43	81.82	97.87	83.33	98.08
DNN	93.88	88.57	72.73	87.23	0.00	100.00
LSTM	79.59	77.14	72.73	95.74	66.67	98.08
CNN	91.84	77.14	72.73	89.36	0.00	100.00
Three-Ratio	63.27	71.43	27.27	78.72	16.67	84.62
Rogers Ratio	55.10	65.71	18.18	72.34	16.67	78.85
Duval Triangle	69.39	74.29	36.36	82.98	16.67	90.38

Early Access

Table 4 presents the per-class classification accuracy of various models across six transformer states. Among all methods, the SMOTE-IPOA-LGBM achieves the most balanced and robust performance, delivering 97.14% accuracy on MLT and 90.91% on LD, the highest among all models, and perfect recognition (100%) for PD. Notably, LD proves to be the most challenging fault type for most classifiers, with many models such as the SVM and decision tree (DT) scoring below 65%. In contrast, the IPOA-LGBM significantly outperforms its counterparts on this critical class, demonstrating its enhanced discriminative power.

In stark contrast, the three conventional diagnostic methods, the Three-Ratio, Rogers Ratio, and Duval Triangle, exhibit substantially inferior performance across all fault categories. Samples that cannot be assigned to any defined ratio zone or triangle region by these rule-based methods are counted as misclassifications, so the denominators are identical to those of the learning-based models. These rule-based approaches achieve accuracies ranging from only 16.67% to 90.38%, with particularly poor recognition of the NS, where all three methods score merely 16.67%. The Rogers Ratio method performs the worst overall, with accuracies below 66% for four out of six classes. Even the best-performing conventional method, the Duval Triangle, fails to exceed 90.38% on any class and struggles notably with the NS (16.67%) and LD (36.36%). These results underscore the inherent limitations of fixed-ratio thresholds and geometric partitioning rules when applied to complex, real-world DGA data, where gas concentration patterns often deviate from idealized boundaries. The proposed SMOTE-IPOA-

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

LGBM framework, by learning data-driven decision boundaries, overcomes these shortcomings and delivers consistently superior accuracy across all diagnostic classes.

Furthermore, several models, including the SVM, deep neural network (DNN), and convolutional neural network (CNN), completely fail to identify the NS type, with 0% accuracy, which indicates a strong bias toward fault classes, likely due to data imbalance. Conversely, tree-based ensemble methods consistently recognize the NS with 83.33% accuracy, highlighting their reliability in distinguishing normal operation from faulty conditions. Overall, the SMOTE-IPOA-LGBM not only matches or exceeds the best results in every category but also exhibits superior stability across all fault types, confirming the effectiveness of the proposed IPOA-based hyperparameter optimization framework. The confusion matrix in Fig. 5 illustrates the classification behavior of the model across all six fault types, clearly showing the distribution of true and misclassified samples. Although the proposed model achieves 100% class-wise accuracy on the PD class, this result should be interpreted with caution because the test set is small and the confidence interval is wide. It does not imply that the model is error-free; the overall accuracy is 97.00%, and the mean class-wise accuracy is 94.23%.

Early Access

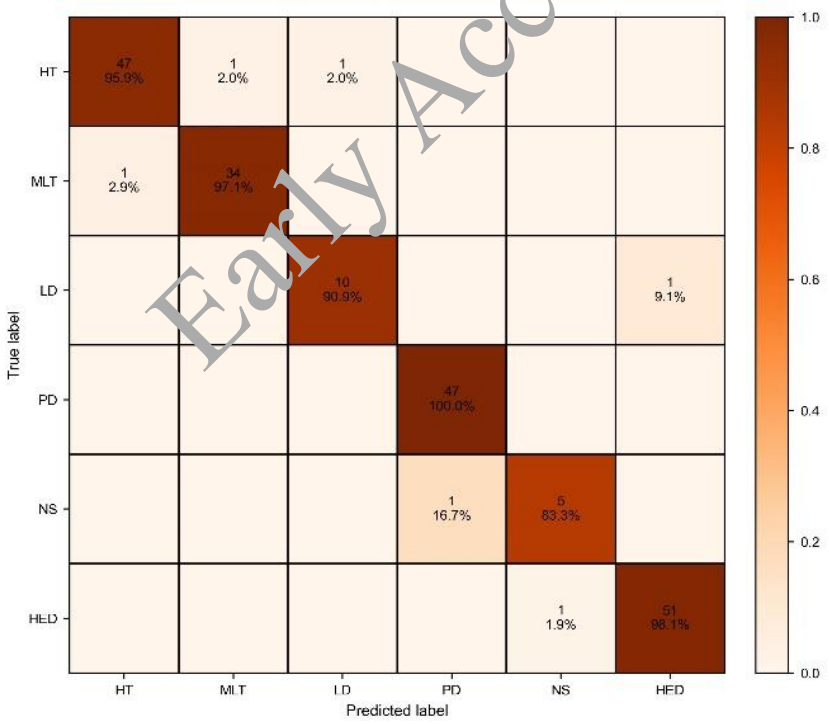


Fig. 5. Confusion matrix of the SMOTE-IPOA-LGBM model on the test set

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

5.3. Robustness and ablation study

Figure 6 illustrates the trend of model accuracy across varying test set ratios. Overall, a gradual decline in accuracy is observed as the test set ratio increases from 0.10 to 0.80. Specifically, the model maintains high performance with an average accuracy of approximately 96% when the test set ratio is relatively small (≤ 0.4). However, as the ratio expands beyond 0.70, the accuracy exhibits significant fluctuations and drops to its minimum value of roughly 91.86%. The shaded region represents the standard deviation (± SD), indicating the stability of the model under different random splits.

The ablation results shown in Table 5 validate the effectiveness of each module and their synergistic effect. Comparing the LGBM with the SMOTE-LGBM, SMOTE increases accuracy by 1.00% and miou by 0.028, which confirms it effectively addresses class imbalance caused by insufficient rare fault samples, reduces the model's bias toward majority classes, and enhances the recognition ability for minority faults. By comparing the LGBM with the IPOA-LGBM, the IPOA-based hyperparameter optimization boosts accuracy by 1.50%, demonstrating IPOA's efficiency in searching for optimal hyperparameter combinations to improve the model's fitting and generalization performance while avoiding overfitting. The SMOTE-IPOA-LGBM achieves the best performance across all metrics. To assess the statistical significance of this improvement, we apply McNemar's test to the paired predictions of the SMOTE-IPOA-LGBM and the baseline LGBM on the same hold-out test set. The resulting p-value is < 0.05, indicating that the performance gain is statistically significant. This confirms that data balancing and hyperparameter optimization act synergistically to enhance diagnostic accuracy and stability, verifying the rationality and superiority of the proposed SMOTE-IPOA-LGBM method.

Early Access

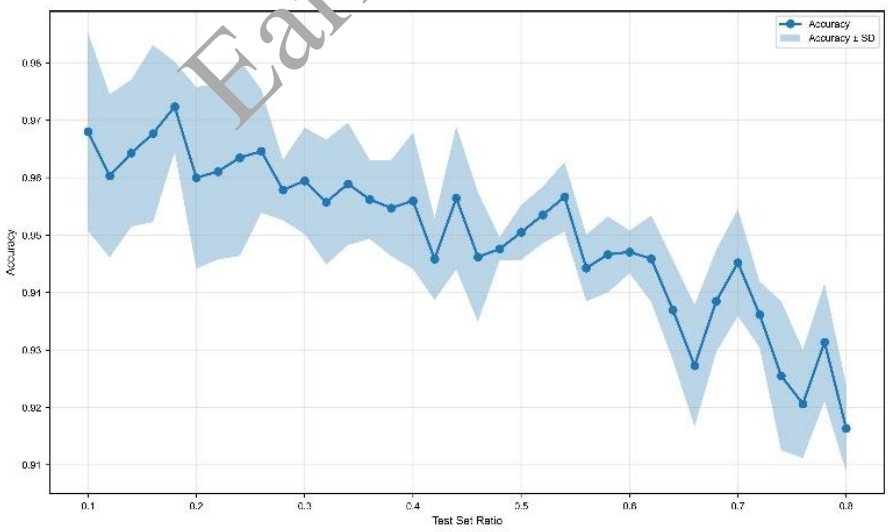


Fig. 6. Performance stability of the model under varying test set proportions

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

Table 5. Ablation experiment results

Models	Accuracy(%)	miou	kappa
LGBM	92.50	0.861	0.907
SMOTE-LGBM	93.50	0.889	0.914
IPOA-LGBM	94.00	0.896	0.922
SMOTE-IPOA-LGBM	97.00	0.903	0.961

Early Access

6. Conclusion

This paper presents an SMOTE-IPOA-LGBM framework for diagnosing incipient faults in oil-immersed power transformers from DGA. The improved pelican optimization algorithm enhances the original algorithm by incorporating individual historical memory, global-best guidance, and a dynamically decaying inertia weight, which stabilizes the search process and makes LGBM hyperparameter tuning more reproducible. The SMOTE is used to mitigate the bias toward majority classes that otherwise dominates the training of the classifiers evaluated in this study.

On the 1000-record dataset, the proposed framework achieves an accuracy of 97.00% on the independent hold-out test set, with a mean intersection over union of 0.903 and a Cohen's kappa of 0.961. Ablation experiments quantify the contribution of each module: The SMOTE alone improves accuracy by 1.00% and mean intersection over union by 0.028 over a plain LGBM, and IPOA-based hyperparameter optimization provides a further gain of 1.50%. Recursive feature elimination shows that seven of the 23 candidate features, acetylene concentration, total combustible hydrocarbon content, H_2, C_2H_6, CH_4, and two physically motivated ratios, are sufficient, and adding further features degrades performance. Repeated random-split experiments indicate that the model maintains an accuracy of about 92% or higher as the test-set ratio increases from 0.10 to 0.80; the accuracy remains stable at about 96% when the test fraction is no larger than 0.4 and decreases to 91.86% at 0.8, indicating reasonable robustness to the amount of labeled data.

Several limitations remain. The two smallest classes, low energy discharge and normal state, still account for most misclassifications, and the framework has so far been evaluated only offline on historical records. Future work will focus on few-shot augmentation for rare fault types, online deployment integrated with online dissolved gas analysis monitoring devices, and a tighter fusion of Duval triangle knowledge with data-driven features.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

References

[1] Liu C., Yu S., *Power transformer fault diagnosis based on MIC feature extraction and INGO-SVM*, Engineering Research Express, vol. 7, no. 4, 045347 (2025), DOI: 10.1088/2631-8695/ae180b.

[2] Song Y., Zhao Z., Yuan J., Cui N., Yang S., Yu H., Li G., *Transformer short-circuit fault diagnosis using convolutional neural networks*, Energy Reports, vol. 15, 109115 (2026), DOI: 10.1016/j.egyr.2026.109115.

[3] Liu C., Ji L., *Transformer Fault Diagnosis Based on Parallel AdaBoost-NB Algorithm on Spark Cloud Platform*, International Journal of Grid and High Performance Computing, vol. 17, no. 1, pp. 1–23 (2025), DOI: 10.4018/IJGHPC.381295.

[4] Cong H., Quan S., Hu X., Zhang X., Li Q., Xu M., *Microscopic Mechanism Effect of Oxygen and Moisture on Pyrolysis of Transformer Insulating Paper*, IEEE Transactions on Dielectrics and Electrical Insulation, vol. 32, no. 2, pp. 1008–1016 (2025), DOI: 10.1109/TDEI.2024.3456771.

[5] Ma Y., Ma X., Wang R., Li J., Guan K., Chen X., *A field-deployable calibration platform for on-line dissolved gas monitoring systems in power transformer oil*, Electric Power Systems Research, vol. 255, 112782 (2026), DOI: 10.1016/j.epsr.2026.112782.

[6] Yan C., Li M., Liu W., *Transformer Fault Diagnosis Based on BP-Adaboost and PNN Series Connection*, Mathematical Problems in Engineering, vol. 2019, no. 1, 1019845 (2019), DOI: 10.1155/2019/1019845.

[7] Saleh A.G., Azzam A.E., Attiya G., Ibrahim E., *Detecting Incipient Faults in Power Transformers through Hybrid Model of DGA and Machine Learning*, Electric Power Systems Research, vol. 256, 112877 (2026), DOI: 10.1016/j.epsr.2026.112877.

[8] Lakehal A., Tachi F., *Bayesian Duval Triangle Method for Fault Prediction and Assessment of Oil Immersed Transformers*, Measurement and Control, vol. 50, no. 4, pp. 103–109 (2017), DOI: 10.1177/0020294017707461.

[9] Kunjal Jane, *Dissolved Gas Analysis in Transformer using Three Gas Ratio Method and Fuzzy Logic based on IEC Standard*, International Journal of Electrical and Electronics Engineering, vol. 2, no. 4, pp. 1–4 (2015), DOI: 10.14445/23488379/IJEEE-V2I4P101.

[10] Taha I.B.M., Hoballah A., Ghoneim S.S.M., *Optimal ratio limits of rogers' four-ratios and IEC 60599 code methods using particle swarm optimization fuzzy-logic approach*, IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 1, pp. 222–230 (2020), DOI: 10.1109/TDEI.2019.008395.

[11] Mashifane L.D., Mendu B., Monchusi B.B., *State-of-the-Art Fault Detection and Diagnosis in Power Transformers: A Review of Machine Learning and Hybrid Methods*, IEEE Access, vol. 13, pp. 48156–48172 (2025), DOI: 10.1109/ACCESS.2025.3546861.

[12] Bjelić M., Brković B., Žarković M., Miljković T., *Machine learning for power transformer SFRA based fault detection*, International Journal of Electrical Power & Energy Systems, vol. 156, pp. 109779 (2024), DOI: 10.1016/j.ijepes.2023.109779.

[13] Yan P., Chen F., Zhao T., Zhang H., Kan X., Liu Y., *Transformer fault diagnosis research based on LIF technology and IAO optimization of LightGBM*, Analytical Methods, vol. 15, no. 3, pp. 261–274 (2023), DOI: 10.1039/D2AY01745H.

[14] Rajesh K.N.V.P.S., Rao U.M., Fofana I., Rozga P., Paramane A., *Influence of Data Balancing on Transformer DGA Fault Classification With Machine Learning Algorithms*, IEEE Transactions on Dielectrics and Electrical Insulation, vol. 30, no. 1, pp. 385–392 (2023), DOI: 10.1109/TDEI.2022.3230377.

[15] He H., Garcia E.A., *Learning from Imbalanced Data*, IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284 (2009), DOI:10.1109/TKDE.2008.239.

[16] Xu G., Zhang M., Chen W., Wang Z., *Transformer Fault Diagnosis Utilizing Feature Extraction and Ensemble Learning Model*, Information, vol. 15, no. 9, 561 (2024), DOI: 10.3390/info15090561.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

[17] Zhu Y., Guo Z., Zhan X., Huang X., *Research on transformer fault diagnosis models with feature extraction*, Review of Scientific Instruments, vol. 95, no. 11, 115109 (2024), DOI: 10.1063/5.0225204.
[18] Ram V.S.S., Kayastha N., Sha K., *OFES: Optimal feature evaluation and selection for multi-class classification*, Data & Knowledge Engineering, vol. 139, 102007 (2022), DOI: 10.1016/j.datak.2022.102007.
[19] Jiang J., Liu Z., Wang P., Yang F., *Improved Crow Search Algorithm and XGBoost for Transformer Fault Diagnosis*, Journal of Physics: Conference Series, vol. 2666, no. 1, 012040 (2023), DOI: 10.1088/1742-6596/2666/1/012040.
[20] Rimal Y., Sharma N., Alsadoon A., *The accuracy of machine learning models relies on hyperparameter tuning: Student result classification using random forest, randomized search, grid search, bayesian, genetic, and optuna algorithms*, Multimedia Tools and Applications, vol. 83, no. 30, pp. 74349–74364 (2024), DOI: 10.1007/s11042-024-18426-2.
[21] Zito F., Talbi E.-G., Cavallaro C., Cutello V., Pavone M., *Metaheuristics in automated machine learning: Strategies for optimization*, Intelligent Systems with Applications, vol. 26, 200532 (2025), DOI: 10.1016/j.iswa.2025.200532.
[22] Hou J., Wang X., Su Y., Yang Y., Gao T., *Parameter Identification of Lithium Battery Model Based on Chaotic Quantum Sparrow Search Algorithm*, Applied Sciences, vol. 12, no. 14, 7332 (2022), DOI: 10.3390/app12147332.
[23] Trojovský P., Dehghani M., *Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications*, Sensors, vol. 22, no. 3, 855 (2022), DOI: 10.3390/s22030855.
[24] SeyedGarmroudi S., Kayakutlu G., Kayalica M.O., Çolak Ü., *Improved Pelican optimization algorithm for solving load dispatch problems*, Energy, vol. 289, 129811 (2024), DOI: 10.1016/j.energy.2023.129811.
[25] Chen X., *Fault diagnosis of high power grid wind turbine based on particle swarm optimization BP neural network during COVID-19 epidemic period*, Journal of Intelligent & Fuzzy Systems, vol. 39, no. 6, pp. 9027–9035 (2020), DOI: 10.3233/JIFS-189301.
[26] Jin Y., Wu H., Zheng J., Zhang J., Liu Z., *Power Transformer Fault Diagnosis Based on Improved BP Neural Network*, Electronics, vol. 12, no. 16, 3526 (2023), DOI: 10.3390/electronics12163526.
[27] Benmahamed Y., Kherif O., Teguar M., Boubakeur A., Ghoneim S.S.M., *Accuracy Improvement of Transformer Faults Diagnostic Based on DGA Data Using SVM-BA Classifier*, Energies, vol. 14, no. 10, 2970 (2021), DOI: 10.3390/en14102970.
[28] Hu Q., Mo J., Ruan S., Zhang X., *Fault Diagnosis of Oil-Immersed Power Transformers Using SVM and Logarithmic Arctangent Transform*, IEEJ Transactions on Electrical and Electronic Engineering, vol. 17, no. 11, pp. 1562–1569 (2022), DOI: 10.1002/tee.23678.
[29] Taha I.B.M., Ibrahim S., Mansour D.-E.A., *Power Transformer Fault Diagnosis Based on DGA Using a Convolutional Neural Network with Noise in Measurements*, IEEE Access, vol. 9, pp. 111162–111170 (2021), DOI: 10.1109/ACCESS.2021.3102415.
[30] Luo Z., Zhang Z., Yan X., Qin J., Zhu Z., Wang H., Gao Z., *Dissolved Gas Analysis of Insulating Oil in Electric Power Transformers: A Case Study Using SDAE-LSTM*, Mathematical Problems in Engineering, vol. 2020, pp. 1–10 (2020), DOI: 10.1155/2020/2420456.
[31] Ghoneim S.S.M., Baz M., Alzaed A., Zewdie Y.T., *Predicting the insulating paper state of the power transformer based on XGBoost/LightGBM models*, Scientific Reports, vol. 15, no. 1, 17836 (2025), DOI: 10.1038/s41598-025-03033-4.
[32] Bergstra J., Bengio Y., *Random search for hyper-parameter optimization*, Journal of Machine Learning Research, vol. 13, pp. 281–305 (2012).
[33] Valdez F., Vazquez J.C., Melin P., Castillo O., *Comparative study of the use of fuzzy logic in improving particle swarm optimization variants for mathematical functions using co-evolution*, Applied Soft Computing, vol. 52, pp. 1070–1083 (2017), DOI: 10.1016/j.asoc.2016.09.024.
[34] Li Y., Zhang Y., Cai Y., *A New Hyper-Parameter Optimization Method for Power Load Forecast Based on Recurrent Neural Networks*, Algorithms, vol. 14, no. 6, pp. 163 (2021), DOI: 10.3390/a14060163.

Early Access

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication.
Citation information: DOI 10.24425/aee.2027.17592

[35] Yokoyama A.M., Ferro M., Schulze B., *Multi-objective hyperparameter optimization approach with genetic algorithms towards efficient and environmentally friendly machine learning*, AI Communications, vol. 37, no. 3, pp. 429–442 (2024), DOI: 10.3233/AIC-230063.
[36] Yang T., Fang J., Jia C., Liu Z., Liu Y., *An improved Harris hawks optimization algorithm based on chaotic sequence and opposite elite learning mechanism*, PLOS ONE, vol. 18, no. 2, e0281636 (2023), DOI: 10.1371/journal.pone.0281636.
[37] Jiao C., Zhou Q., Zhang W., Zhang C., *The Modified Sparrow Search Algorithm with Brown Motion and Levy Flight Strategy for the Class Integration Test Order Generation Problem*, Biomimetics, vol. 10, no. 4, 195 (2025), DOI: 10.3390/biomimetics10040195.
[38] Ma S., Fang Y., Zhao X., Liu Z., *Multi-swarm improved Grey Wolf Optimizer with double adaptive weights and dimension learning for global optimization problems*, Mathematics and Computers in Simulation, vol. 205, pp. 619–641 (2023), DOI: 10.1016/j.matcom.2022.10.007.
[39] International Electrotechnical Commission, *Mineral oil-filled electrical equipment in service — Guidance on the interpretation of dissolved and free gases analysis (IEC 60599:2015, Edition 3.0)*, International Electrotechnical Commission (2015), available at: https://webstore.iec.ch/en/publication/23323.
[40] Institute of Electrical and Electronics Engineers, *IEEE Std C57.104-2019: IEEE guide for the interpretation of gases generated in mineral oil-immersed transformers* (2019), DOI: 10.1109/IEEESTD.2019.8890040.
[41] Rogers R.R., *IEEE and IEC Codes to Interpret Incipient Faults in Transformers, Using Gas in Oil Analysis*, IEEE Transactions on Electrical Insulation, vol. EI-13, no. 5, pp. 349–354 (1978), DOI: 10.1109/TEI.1978.298141.
[42] Duval M., *A review of faults detectable by gas-in-oil analysis in transformers*, IEEE Electrical Insulation Magazine, vol. 18, no. 3, pp. 8–17 (2002), DOI: 10.1109/MEI.2002.1014963.
[43] Zhao S., Pattanadech N., *Power transformer fault warning combining support vector machine and improved grey wolf optimization algorithm*, Archives of Electrical Engineering, vol. 74, no. 1, pp. 191–208 (2025), DOI: 10.24425/aee.2025.153019.
[44] Li X., Wu P., Shao J., Liang J., Li Q., Lu Y., Wang T., *Transformer fault diagnosis method based on multilevel acoustic information*, Archives of Electrical Engineering, vol. 75, no. 1, pp. 23–40 (2026), DOI: 10.24425/aee.2026.156800.

Early Access