Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 25-35.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

A Novel Fault Prediction Technique for Oil-Immersed Transformers Based on Advanced Gradient Boosting and Particle Swarm Optimization (PSO)

Elahe Moradi*

Department of Electrical Engineering, YI.C., Islamic Azad University, Tehran, Iran.

Article Info

Article History:
Received 10 May 2025
Revised 17 September
Accepted 23 September

DOI:10.22044/jadm.2025.16214.2745

Keywords:
Power Transformer, Machine Learning, Gradient Boosting; Particle Swarm Optimization (PSO), Fault Prediction, DGA.

**Corresponding author: Elahe.Moradi@iau.ac.ir (E. Moradi).*

Abstract

Fault prediction in power transformers is pivotal for safeguarding operational reliability and reducing system disruptions. Leveraging dissolved gas analysis (DGA) data, AI-driven techniques have recently been employed to enhance predictive performance. This paper introduces a novel machine-learning framework that integrates Histogram Gradient Boosting (HGB) with a metaheuristic Particle Swarm Optimization (PSO) algorithm for hyperparameter tuning, thereby ensuring classifier robustness. The proposed method underwent a two-stage evaluation: first, Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), and HGB were benchmarked, revealing HGB as the most effective method; second, PSO was applied to optimize HGB's hyperparameters, yielding further performance improvements. Experimental results demonstrate that the hybrid HGB-PSO model achieves an accuracy of 97.85%, precision of 98.90%, recall of 97.33%, and an F1-score of 98.99%. All simulations and comparative analyses against state-of-the-art methods were implemented in Python, and confusion-matrix analysis was employed to assess predictive performance comprehensively. These findings demonstrate that the hybrid HGB-PSO method achieves superior accuracy and robustness in transformer fault prediction.

1. Introduction

Power transformers are indispensable assets in power generation plants and high-voltage substations. Failures or performance degradations can precipitate interruptions in energy production and compromise transmission reliability, imposing substantial costs for system reinstatement, unserved- energy penalties, and transformer repair or replacement. Accordingly, the adoption of comprehensive protection schemes and condition-based maintenance strategies is essential for optimizing asset lifecycle management and ensuring resilient grid operation [1–3].

One of the most effective and widely used techniques for diagnosing and predicting faults caused by electrical, thermal, and mechanical stresses in transformers is the analysis of dissolved gases in oil through gas chromatography. These stresses can degrade the oil and paper insulation, leading to the release of various gases. When the concentration of these gases exceeds certain thresholds, they may even result in catastrophic transformer failures, including explosions. The formation of these gases is typically associated with abnormal energy losses inside the transformer, such as overheating, partial discharges, and arcing. Among various diagnostic tools, gas chromatography is recognized as the most accurate and reliable method for detecting internal faults in high-voltage equipment, particularly oil-filled transformers [4–6].

Despite significant advancements in the measurement of dissolved gases in transformer oil in recent years, the interpretation of the resulting data remains a challenging task. Traditional diagnostic techniques, such as the Doernenburg ratio method [7, 8], IEC ratio method [9, 10], Duval triangle method [11, 12], and Roger's ratio method [13, 14], exhibit limitations in terms of both accuracy and reliability. Unlike these ratio-based approaches, the Duval Pentagon method [15, 16] employs the central positioning of fault points within a pentagon diagram, offering enhanced diagnostic precision, particularly for thermal faults. Although conventional methods are computationally simple and require no complex programming, their application becomes increasingly inefficient when dealing with large-scale DGA datasets. To overcome these challenges, recent studies have increasingly adopted artificial intelligence (AI) algorithms, particularly machine learning (ML) and deep learning (DL), to enable automated fault detection and prediction in oil-immersed power transformers [17].

The authors of [18] devised a fault classification model for transformers by integrating the C-set method with the fuzzy C-means clustering algorithm. Their method specifically addressed challenges such as data imbalance, outliers, and boundary class overlap. By generating labeled expert training data through unsupervised clustering and training a one-vs-one multiclass SVM, they achieved an accuracy of 88.9%. The authors of [19] proposed a transformer fault diagnosis based on DGA, employing an enhanced LightGBM ensemble model with a dual-branch structure. Furthermore, an improved grey wolf optimizer was utilized to fine-tune hyperparameters, while Jacobian regularization was applied to mitigate noise sensitivity and improve model robustness. The authors of [20] developed an intelligent diagnostic system to increase the accuracy of fault detection in transformers with DGA. Recognizing the limitations of traditional DGA methods such as the IEC Code, Rogers Ratio, and Duval Triangle, the study incorporated optimization procedures to advance the decision-making capabilities of these models. By comparing the outputs of multiple DGA techniques, the proposed system achieved an accuracy of 89.12%, outperforming conventional approaches. The authors of [21] utilized the Common Vector Approach to classify incipient transformer faults based on DGA data. By incorporating both raw and extracted features, their method demonstrated superior accuracy and faster computation compared to conventional and intelligent classifiers, particularly under limited data conditions.

The authors of [22] examined a deep belief network-based DGA approach for transformer fault diagnosis, aiming to overcome the rigidity of traditional methods by enabling flexible input combinations. Their model improved diagnostic accuracy by customizing input features beyond fixed gas ratios. The authors of [23] developed a CNN-based model utilizing DGA data to classify transformer fault types under varying noise conditions. By leveraging conventional, novel, and hybrid gas ratio inputs, the model was trained on 589 samples, demonstrating robust performance even with noise levels reaching ±20%. The authors of [24] proposed an intelligent fault classification technique using key DGA attributes and an ANFIS model, optimized via the Black Widow Optimization Algorithm. By integrating feature selection through association rule learning, the model achieved increased accuracy and robustness in transformer fault detection. The study in [25] introduced a dynamic fault prediction framework employing a long short-term memory (LSTM) model to forecast future DGA trends and assess transformer conditions. Among several AI classifiers tested, the LSTM-KNN model yielded the highest predictive accuracy for identifying potential transformer faults.

The study of [26] addressed a DGA-based fault diagnosis scheme combining the dual pentagon method with several tree-based classifiers and data scaling techniques. Among the evaluated models, the Light-GBM classifier demonstrated superior performance, achieving 96.08% accuracy and outperforming conventional methods. The authors of [27] explored the impact of data-level balancing methods on transformer fault classification using DGA data. They compared three ML algorithms, including Support Vector Machine (SVM), Decision Tree, and Random Forest, with ENN combined with SVM achieving the highest classification performance, with 88% accuracy. The study of [28] utilized a fusion of machine learning and sensor-level integration to improve transformer fault diagnosis via DGA. Applying the Sequential Kalman Filter alongside Majority Voting and Dempster–Shafer methods, the approach achieved over 90% estimation accuracy. The authors of [29] introduced a Seasonal Autoregressive Integrated Moving Average (SARIMA)-based model for forecasting dissolved gas concentrations in transformer oil. By analyzing periodicity and temperature correlation, SARIMA showed superior accuracy and stability over the autoregressive and LSTM models, especially when external factors were included in the prediction.

The research of [30] assessed fuzzy logic and neural networks to detect and predict transformer failures, aiding timely maintenance decisions. The proposed model achieved up to 95% accuracy, emphasizing the role of predictive maintenance and offering a practical tool to support maintenance teams in preventing unexpected faults. The authors of [31] proposed a hybrid Genetic algorithm (GA)–SVM method for improving transformer fault diagnosis using DGA data. It employs adaptive sampling, arctangent transformation, and five filter methods for feature ranking. Optimal features are selected via GA with SVM, achieving improved accuracy through five-fold cross-validation on the IEC TC10 dataset. The authors of [32] employed a DNN-based diagnostic model for transformer faults, enhanced by SMOTE to handle class imbalance. Hyperparameters are optimized using grid search, random search, and Bayesian optimization methods. Experiments on real datasets show superior performance, with 94.6% testing accuracy, outperforming traditional classifiers on imbalanced data. The study of [33] provided a transformer fault diagnosis method with DGA combined with data transformation techniques and six optimized machine learning (OML) algorithms, including decision tree, SVM, discriminant analysis, Naïve Bayes, KNN, and ensemble classifiers.

To the best of our knowledge, no prior study has employed the hybrid integration of Histogram-based Gradient Boosting (HGB) with Particle Swarm Optimization (PSO) for transformer fault prediction on this dataset. While advanced ensemble learning methods have rarely been explored in this field, most recent research has concentrated on conventional machine learning and deep learning approaches, typically relying on hyperparameter tuning via grid search, random search, or Bayesian optimization. In contrast, the present work introduces a novel framework that combines an advanced ensemble learning technique with a powerful metaheuristic optimization algorithm. The key contributions of this study can be summarized as follows: (i) applying HGB to enhance the accuracy of transformer fault prediction, (ii) utilizing PSO for efficient hyperparameter optimization, and (iii) achieving superior predictive accuracy and robustness compared with state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 describes the DGA dataset and preprocessing steps, including outlier removal via the IQR method, z-score normalization, and train-validation-test splitting. Section 3 details the proposed methodology, introducing and comparing Gradient Boosting (GB), XGBoost, and Histogram-based Gradient Boosting (HGB) classifiers, and then presenting the PSO-based hyperparameter optimization for HGB. Section 4 reports experimental results and provides an in-depth discussion of model performance across standard metrics. Finally, Section 5 concludes the study.

2. DGA Dataset

2.1. Dissolved Gas Concentrations

Electrical and thermal stresses cause an oil-immersed transformer's insulation system to decompose, and the dissolved gases in the transformer insulating oil usually contain the following: the product of paper decomposition-carbon monoxide (CO), carbon dioxide (CO_2), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), methane (CH_4), and hydrogen (H_2) [25, 34]. By analyzing the composition or ratio of these gasses, the type of transformer failure may be recognized. DGA identifies transformer faults by analyzing the concentration or ratios of various gases, employing distinct calculation methods that utilize different gas ratios, such as Rogers' method (CH_4/, C_2H_2/C_2H_4, C_2H_4/C_2H_6, C_2H_6/CH_4), Dornenburg's method (CH_4/H_2, C_2H_2/C_2H_4, C_2H_2/CH_4, C_2H_6/C_2H_2), and the IEC 60599 method (CH_4/H_2, C_2H_2/C_2H_4, C_2H_4/C_2H_6) [22, 26].

2.2. Duval Pentagon Method

In this study, the Duval Pentagon method is employed for the preliminary identification of faults in oil-immersed transformers. This diagnostic technique utilizes five key dissolved gases: hydrogen (H2), acetylene (C2H2), methane (CH4), ethane (C2H6), and ethylene (C2H4). The Duval Pentagon method is capable of identifying seven distinct fault types, including partial discharge (PD), low-energy and high-energy electrical discharges, normal aging, as well as thermal faults occurring at various temperature ranges [26]. The seven aforementioned fault types are presented and described in Table 1.

2.3. Dataset

The dataset utilized in this research was derived from the datasets presented in [15] and [26]. This dataset, based on DGA, includes five features as inputs: hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), and ethane (C_2H_6). It is labeled with seven fault types: T1, T2, T3, PD, D1, high-energy discharge D2, and S, as detailed in Table 1. The number of samples associated with each fault type is as follows: PD – 241, S – 227, D1 – 233, D2 – 237, T1 – 239, T2 – 241, and T3 – 240.

Table 1. Categorization of faults based on the Duval Pentagon method.

Type of fault	Acronyms	Description of faults
Type 1	PD	Corona Partial Discharges
Type 2	D1	Discharge of low-energy
Type 3	D2	Discharge of high-energy
Type 4	T1	Low Thermal fault (Temperature range < 300 °C)
Type 5	T2	Medium Thermal fault (300 °C <Temperature range < 700 °C)
Type 6	T3	High thermal fault (Temperature range > 700 °C)
Type 7	S	Stray gassing at low temperatures

2.4. Data Pre-Processing

Data pre-processing is a significant stage in the field of AI, involving the cleaning and structuring of raw data to facilitate the development and training of AI models. Real-world data are often incomplete, inconsistent, and/or lack identifiable patterns or trends, and they may contain various errors. In simple terms, data pre-processing is a data mining method that transforms raw data into a readable and interpretable format suitable for intelligent systems [26, 35].

Initially, missing value handling and outlier detection were performed on the DGA dataset. This dataset contains no missing values. To detect outliers, the Interquartile Rang (IQR) method was applied [36], and its mathematical formulations are presented in Equations (1) to (3).

$$IQR = Q_3 - Q_1 \tag{1}$$

$$LB = Q_1 - 1.5 \times IQR \tag{2}$$

$$UB = Q_3 + 1.5 \times IQR \tag{3}$$

here, the third quartile (Q_3) and the first quartile (Q_1), along with the lower bound (LB) and upper bound (UB), are used to identify extreme values.

Subsequently, data encoding for categorical labels was carried out using the Label Encoder method, which converted the seven identified fault types into numerical values. After that, feature scaling was performed using the z-score normalization technique [37], as described in Equation (4).

$$x_{zs} = \frac{x - x_{mean}}{x_{std}} \tag{4}$$

In this equation, x, and x_{zs} represent the values of samples the before and after scaling, respectively. Additionally, x_{mean} denotes the mean and x_{std} the standard deviation of across all samples [38].

To facilitate a more rigorous evaluation of the dataset, it was partitioned into three distinct subsets: a training set (70%), a validation set (15%), and a testing set (15%). The splitting process was conducted using the “train_test_split” function from the Scikit-learn library in Python. All preprocessing procedures and simulations in this study were implemented using Python, with the support of several well-established libraries, including NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, PySwarms, and XGBoost.

3. Proposed Method

This section provides a comprehensive description of the methodology proposed for fault diagnosis in oil-immersed transformers. The approach integrates systematic data preprocessing techniques with advanced ensemble classifiers to accurately predict and classify various fault types based on the DGA dataset.

3.1. Classifiers

– *Gradient Boosting (GB)*

Gradient Boosting (GB) is a powerful ensemble learning technique that iteratively constructs a strong classifier by combining multiple weak learners, typically shallow decision trees, into a single predictive model. At each iteration, GB fits a new tree to the negative gradient (pseudo-residuals) of the loss function with respect to the current ensemble's predictions, effectively performing gradient descent in function space to minimize training error [39]. This methodology was first formalized by Friedman [40] and has since become a cornerstone of advanced machine learning applications in classification tasks.

– *Extreme Gradient Boosting (XGB)*

Extreme Gradient Boosting (XGB) is a highly efficient and scalable machine learning algorithm widely recognized for its superior performance in classification and regression tasks. XGBoost enhances the traditional gradient boosting framework by implementing an additive model where decision trees are sequentially constructed to minimize a composite objective function, comprising a loss term and a regularization component to prevent overfitting [41, 42]. For transformer fault prediction, XGBoost's capability to process complex dissolved gas analysis (DGA) data, coupled with its computational efficiency and adaptability to high-dimensional datasets, positions it as a powerful tool for enhancing diagnostic accuracy and reliability in power systems.

– *Histogram-based Gradient Boosting (HGB)*

Histogram-based Gradient Boosting (HGB) is an advanced ensemble learning technique that improves the efficiency of traditional gradient boosting. It achieves this by using histogram-based discretization of continuous input features, which significantly reduces computational complexity while preserving high predictive accuracy. HGB groups feature values into discrete intervals, allowing for faster training through optimized memory usage and parallel processing. This makes it especially well-suited for large-scale datasets, such as dissolved gas analysis (DGA) data used in transformer fault prediction. Modern machine learning libraries, such as Scikit-learn, implement HGB with these optimizations, leading to robust performance in classification tasks. Recent studies have shown its effectiveness in high-dimensional settings [43, 44]. Consequently, HGB's capability to handle complex DGA features with lower computational demands makes it a powerful tool for improving the reliability and precision of transformer fault diagnosis.

3.2. Criteria for the Proposed Classifiers Evaluation

For the evaluation of the proposed classifiers in this study, standard performance metrics, including accuracy, precision, recall, F_1-score, and the area under the receiver operating characteristic curve (AUC), were employed.

Accuracy is defined as the proportion of correctly classified instances, both positive and negative, relative to the total number of samples, as given by Eq. (5).

$$Accuracy = \frac{TP + TN}{FN + TN + FP + TP} \tag{5}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. True negatives (TN) occur when negative instances are correctly classified, whereas true positives (TP) refer to accurately identified positive cases. A false negative (FN) arises when a positive instance is mistakenly labeled as negative, while a false positive (FP) occurs when a negative instance is erroneously classified as positive.

Precision represents the proportion of correctly predicted positive instances to the total number of instances predicted as positive, as expressed in Eq. (6).

$$\Pr ecision = \frac{TP}{FP+TP} \tag{6}$$

Recall, also known as the true positive rate, quantifies the proportion of actual positive cases that are correctly identified by the classifier. It is formally defined in Eq. (7).

$$\operatorname{Re} call = \frac{TP}{FN + TP} \tag{7}$$

The F-Score is a harmonic mean that balances both recall and precision, providing a single measure of performance, as formulated in Eq. (8).

$$F1 - score = \frac{2 \times \left(\Pr ecision \times \operatorname{Re} call \right)}{\Pr ecision + \operatorname{Re} call} \tag{8}$$

The AUC is a widely used assessment metric derived from the Receiver Operating Characteristic (ROC) curve. It measures the classifier's ability to distinguish between classes, with higher values indicating better classification performance Furthermore, confusion matrix analysis was incorporated to support a more detailed interpretation of the classifiers' predictive capabilities [26, 45, 46].

3.3. Optimized HGB Classifier Using PSO Algorithm

HGB is a powerful tree-based ensemble learning technique capable of handling large-scale data with high predictive efficiency. However, its performance is highly influenced by the selection of hyperparameters, including "learning_rate", "max_iter", and "max_depth". To enhance the accuracy and robustness of the HGB classifier, this study employs the Particle Swarm Optimization (PSO) algorithm for hyperparameter tuning. PSO, initially proposed by Kennedy and Eberhart in 1995 [47], is a population-based stochastic optimization technique inspired by the social behavior of birds flocking or fish schooling. It has been extensively used in various engineering optimization problems and is particularly effective for tuning complex models. The velocity and position of each particle in the swarm are updated using the following equations [48-51]:

$$v_i(t+1) = \omega \times v_i(t) + c_1 r_1 \left(pbest_i - x_i(t) \right) + c_2 r_2 \left(gbest - x_i(t) \right) \tag{9}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{10}$$

where:

- $v_i(t)$ and $x_i(t)$ represent the velocity and position of particle i at iteration t, respectively.
- ω is the initial weight.
- c_1 and c_2 are the cognitive and social coefficients.
- $pbest_i$ is the personal best position of particle i.
- $gbest$ is the global best position discovered so far.
- r_1 and r_2 are uniformly distributed random variables in [0, 1].

The implementation of the PSO-HGB framework was based on optimizing the accuracy of the HGB classifier. The search space and parameter bound used in this optimization are shown in Table 2.

Table 2. Hyperparameter search space for HGB optimized by PSO.

Hyperparameter	Description	LB	UB
learning_rate	Learning rate for boosting	0.01	0.3
max_iter	Number of boosting iterations	50	500
max_depth	Maximum depth of trees	3	10

The algorithmic parameters employed in the PSO algorithm for hyperparameter tuning are listed in Table 3. The proposed PSO-HGB optimization procedure is outlined in Algorithm 1.

Table 3. PSO parameter settings.

Parameter	Description	Value
ω	Inertia weight	0.5
c_1	Cognitive coefficient (individual learning)	2.0
c_2	Social coefficient (swarm learning)	2.0
N	Number of particles in the swarm	20
T	Maximum number of PSO iterations	30

The PSO parameters employed in this study, as summarized in Table 3, were determined empirically through a trial-and-error procedure. The inertia weight ω is inherently a positive parameter, and prior studies [50,52] commonly recommend values within the range of 0.4–0.9. In this work, $\omega = 0.5$ was identified as the most appropriate setting, offering a balanced trade-off between exploration and exploitation.

4. Results and Discussion

In this study, the Dissolved Gas Analysis (DGA) dataset, comprising five input features (H_2, C_2H_2, CH_4, C_2H_6, and C_2H_4) and seven fault labels defined by the Duval Pentagon method, was employed. Data preprocessing included outlier removal via the Interquartile Range (IQR) method and feature standardization using z-score normalization. The dataset was then split into training (70%), validation (15%), and testing (15%) subsets. Advanced classifiers, including Gradient Boosting (GB), Extreme Gradient Boosting (XGBoost), and Histogram-based Gradient Boosting (HGB), were applied to the DGA data. Table 4 presents their performance according to standard metrics: accuracy, precision, F1-score, recall, and AUC.

Table 4. A comparative performance analysis of the advanced classifiers

Classifier	Criteria (%)				
	Accuray	Precision	F1-Score	Recall	AUC
GB	91.13	91.55	91.17	91.13	99.48
XGB	93.95	94.50	94.00	93.95	99.75
HGB	96.37	96.43	96.37	96.37	99.83

For visual comparison, bar plots of each classifier's performance are shown in Figures 1, 2, and 3. Among these, HGB achieved the highest results: accuracy of 96.37%, precision of 96.43%, F1-score of 96.37%, recall of 96.37%, and AUC of 99.83%.

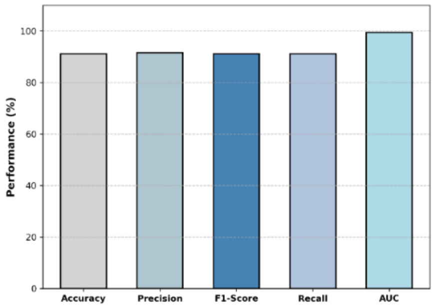

Figure 1. Performance evaluation of the Gradient Boosting (GB) classifier.

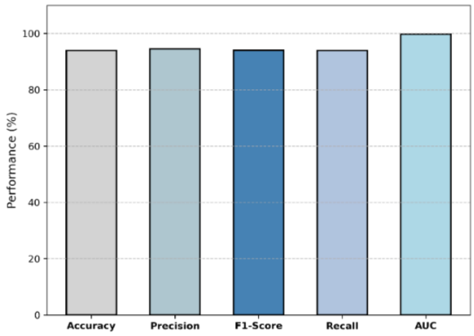

Figure 2. Performance evaluation of the Extreme Gradient Boosting (XGB) classifier.

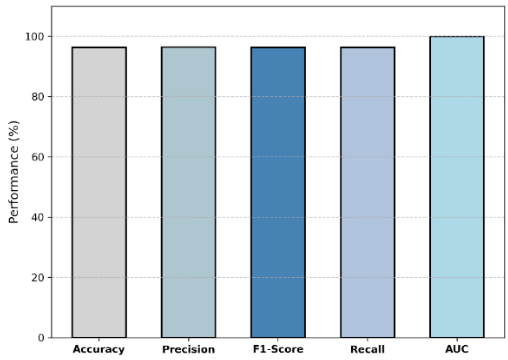

Figure 3. Performance evaluation of the Hist Gradient Boosting (HGB) classifier.

For a comprehensive comparison of the three advanced classifiers GB, XGBoost, and HGB, their performance line plots based on the standard metrics of accuracy, precision, recall, F1-score, and AUC are demonstrated in Figure 4.

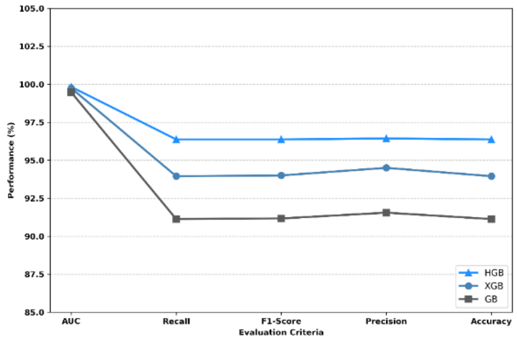

Figure 4. Comparative performance of GB, XGB and HGB classifiers.

Also, Table 5 compares the HGB classifier with a state-of-the-art LightGBM ensemble model proposed in [26]. The results indicate that HGB outperforms the reference classifier across all evaluated metrics. To further enhance the predictive accuracy of transformer fault classification, the Particle Swarm Optimization (PSO) algorithm was employed to fine-tune the hyperparameters of the HGB model. This metaheuristic strategy was used to optimize the learning rate, number of boosting iterations, and maximum tree depth by maximizing the validation accuracy within a predefined hyperparameter search space. The optimization process followed the steps outlined in Algorithm 1.

The optimal configuration obtained through PSO included a learning rate of 0.2679, 292 boosting iterations, and a maximum depth of 9, as summarized in Table 6. This configuration yielded an accuracy of 97.98%, demonstrating a significant enhancement in the predictive capability of the HGB classifier for fault classification in oil-immersed transformers. These improvements are notably superior to both the unoptimized HGB classifier and the LightGBM-based method referenced in [26]. For a more comprehensive comparison, we additionally employed the Genetic Algorithm (GA) to optimize the HGB model, and the corresponding results are presented in Table 5 and Figure 6.

Table 5. Comparative evaluation of HGB, HGB-GA, HGB-PSO, and the LightGBM model in [26].

Criteria	Classifier			
	HGB	[26]	HGB-GA	HGB-PSO
Accuracy (%)	96.37	96.08	97.58	97.98
Precision (%)	96.43	96.09	97.75	98.07
F1-Score (%)	96.37	96.06	97.58	97.99
Recall (%)	96.37	96.08	97.60	97.98
AUC (%)	99.83	99.64	99.89	99.96

As evidenced by these results, the HGB-PSO consistently outperforms all other methods, including the HGB-GA, thereby underscoring its superior optimization effectiveness. Furthermore, the runtime analysis indicates that the HGB-PSO required 3806 seconds, whereas the HGB-GA required 4846 seconds, highlighting the greater computational efficiency of the HGB-PSO.

Table 6. Optimized hyperparameters of the HGB classifier via PSO.

Hyperparameter	Description	Optimized Value
Learning rate	Learning rate used in boosting	0.2679
Max iterations	Number of boosting iterations	292
Max depth	Maximum depth of individual trees	9

The two-sample paired t-test was conducted to statistically compare the performance of the HGB model before and after PSO-based optimization. The analysis yielded a p-value of 0.00863, which is below the 0.05 significance threshold. Consequently, the null hypothesis of equal means is rejected, indicating that the improvement in classification accuracy achieved by the HGB-PSO model is statistically significant.

Figure 5 presents a bar plot of HGB-PSO performance, and Figure 6 illustrates a comparative line chart among HGB, the model from [26], HGB-GA, and HGB-PSO.

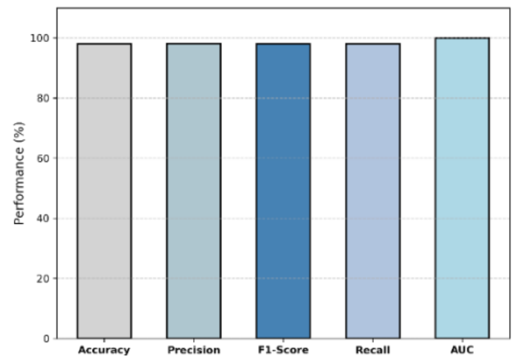

Figure 5. Performance evaluation of the HGB classifier with PSO algorithm.

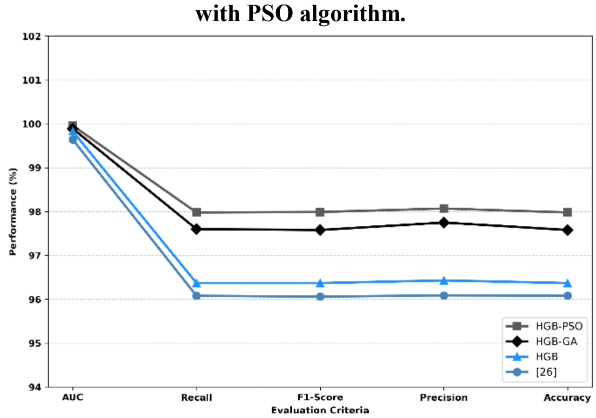

Figure 6. Comparative performance of HGB, HGB-GA, HGB-PSO, and the classifier of [26].

Furthermore, confusion matrices for all four classifiers (GB, XGB, HGB, and HGB-PSO) are provided in Figures 7–10.

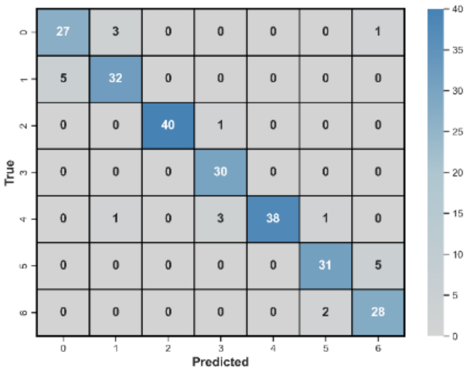

Figure 7. Confusion matrix of the GB classifier.

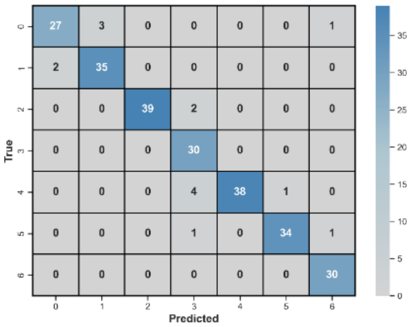

Figure 8. Confusion matrix of the XGB classifier.

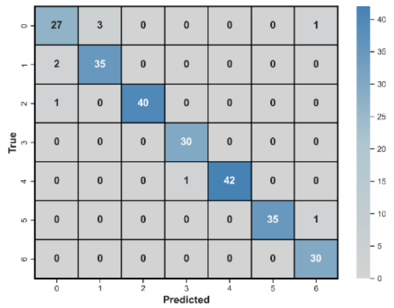

Figure 9. Confusion matrix of the HGB classifier.

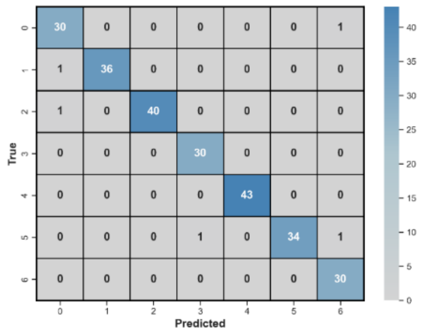

Figure 10. Confusion matrix of the HGB-PSO classifier.

The confusion matrices confirm that HGB-PSO yields the lowest misclassification rate and the highest diagnostic accuracy for the seven fault types in oil-immersed power transformers.

5. Conclusion

This work has presented a novel hybrid framework for transformer fault prediction that couples Histogram-based Gradient Boosting (HGB) with Particle Swarm Optimization (PSO). By leveraging five key dissolved gas features and seven Duval-Pentagon fault categories, we demonstrated that HGB significantly outperforms conventional GB and XGB models, achieving up to 96.37% accuracy and 99.83% AUC. PSO-tuned HGB further improves performance by approximately 2% over standard HGB and a state-of-the-art LightGBM ensemble benchmark. Confusion-matrix analysis confirmed that the optimized model yields the lowest misclassification rates across all seven fault types. These findings underscore the efficacy of metaheuristic tuning in enhancing ensemble learners for DGA-based diagnosis. Future work may explore the use of k-fold cross-validation in combination with advanced deep learning techniques to further enhance the robustness of transformer fault diagnosis.

Algorithm 1. PSO-Based Hyperparameter Optimization for HGB.

Input: Preprocessed training and validation datasets $\mathcal{D}_{train}, \mathcal{D}_{val}$:
Search bounds for $learning\_rate, max\_iter$, max_depth, and PSO parameters: ω, c_1, c_2
Output: Optimized hyperparameters: $\theta^* = (learning\_rate^*, max\_iter^*, max\_depth^*)$

1. Initialize a swarm of N with random positions within defined bounds and zero initial velocities.
2. For each particle i:
   - Evaluate $f_i = Accuracy\ (HGB(\theta_i, \mathcal{D}_{train}), \mathcal{D}_{val})$
   - Set $pbest_i = \theta_i, f_{pbest_i} = f_i$
3. Determine $gbest$ from $\{pbest_i\}$ with the highest f_{pbest}
4. For each iteration $t = 1$ to T:
   - For each particle i:
     - Update velocity v_i using Eq. (9)
     - Update position θ_i using Eq. (12), clip values to predefined bounds
     - Re-evaluate new fitness f_i: update $pbest_i$ and $gbest$ if improved
5. Return $\theta^* = gbest$ as the final optimized hyperparameter set.

References

[1] X. Zheng, "Intelligent Fault Diagnosis of Power Transformer base on Fuzzy Logic and Rough Set Theory," *7th World Congress on Intelligent Control and Automation*, pp. 6858 - 6862, 2008.

[2] A. Nanfak, E. Samuel, I. Fofana, F. Meghnefi, M. G. Ngaleu, and C. H. Kom, "Traditional fault diagnosis methods for mineral oil-immersed power transformer based on dissolved gas analysis: Past, present and future," *IET Nanodielectrics*, vol. 7, no. 3, pp. 97–130, Apr. 2024.

[3] C. Aj, M. A. Salam, Q. M. Rahman, F. Wen, S. P. Ang, and W. Voon, "Causes of transformer failures and diagnostic methods – A review," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1442–1456, Jul. 2017.

[4] Y. Zhang, Y. Tang, Y. Liu, and Z. Liang, "Fault diagnosis of transformer using artificial intelligence: A review," *Frontiers in Energy Research*, vol. 10, Art. no. 1006474, 2022.

[5] S. R. Al-Sakini, G. A. Bilal, A. T. Sadiq, and W. A. K. Al-Maliki, "Dissolved gas analysis for fault prediction in power transformers using machine learning techniques," *Applied Sciences*, vol. 15, no. 1, p. 118, 2025.

[6] S.A.Wani, S.A. Khan, G. Prashal, and D. Gupta. Smart Diagnosis of Incipient Faults Using Dissolved Gas Analysis-Based Fault Interpretation Matrix (FIM). *Arab J Sci Eng*, 44, 6977–6985 (2019).

[7] Y. Liu, B. Song, L. Wang, J. Gao, and R. Xu, "Power transformer fault diagnosis based on dissolved gas Analysis by Correlation Coefficient-DBSCAN," *Applied Sciences*, vol. 10, no. 13, p. 4440, Jun. 2020.

[8] N. Suwarno, H. Sutikno, R. A. Prasojo, and A. Abu-Siada, "Machine learning based multi-method interpretation to enhance dissolved gas analysis for power transformer fault diagnosis," *Heliyon*, vol. 10, no. 4, p. e25975, Feb. 2024.

[9] N. Manisha, K. Kaur, N. K. Sharma, J. Singh, and D. Bhalla, "Performance Assessment of IEEE/IEC Method and Duval Triangle technique for Transformer Incipient Fault Diagnosis," *IOP Conference Series Materials Science and Engineering*, vol. 1228, no. 1, p. 012027, Mar. 2022.

[10] A. Abu-Siada, "Improved Consistent Interpretation Approach of Fault Type within Power Transformers Using Dissolved Gas Analysis and Gene Expression Programming," *Energies*, vol. 12, no. 4, p. 730, Feb. 2019.

[11] C. Guo, Q. Zhang, R. Zhang, X. He, Z. Wu, and T. Wen, "Investigation on gas generation characteristics in transformer oil under vibration," *IET Generation Transmission & Distribution*, vol. 16, no. 24, pp. 5026–5040, Oct. 2022.

[12] G. K. Irungu, A. O. Akumu, and J. L. Munda, "A new fault diagnostic technique in oil-filled electrical equipment; the dual of Duval triangle," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 6, pp. 3405–3410, Dec. 2016.

[13] A. Khanna and P. Bisht, "Rogers ratio test for fault diagnosis of transformer using dissolved gas analysis," *Materials Today Proceedings*, vol. 71, pp. 243–246, Jan. 2022.

[14] J.-Y. Lim, D.-J. Lee, and P.-S. Ji, "Fault diagnosis of power transformer using confidence weight based fusion method," *24th International Conference on Electrical Machines and Systems (ICEMS)*, vol. 71, pp. 1–4, Aug. 2017.

[15] E. Moradi, "A Data-Driven based Robust Multilayer Perceptron Approach for Fault Diagnosis of Power Transformers," *20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Babol, Iran, Feb. 2024.

[16] A. Nanfak, A. Hechifa, S. Eke, A. Lakehal, C. H. Kom, and S. S. M. Ghoneim, "A combined technique for power transformer fault diagnosis based on k-means clustering and support vector machine," *IET Nanodielectrics*, vol. 7, no. 3, pp. 175–187, Jul. 2024.

[17] R. A. Prasojo, M. A. A. Putra, Ekojono, M. E. Apriyani, A. N. Rahmanto, S. S. M. Ghoneim, K. Mahmoud, M. Lehtonen, and M. M. F. Darwish, "Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique," *Electric Power Systems Research*, vol. 220, p. 109361, Apr. 2023.

[18] A. Abdo, H. Liu, H. Zhang, J. Guo, and Q. Li, "A new model of faults classification in power transformers based on data optimization method," *Electric Power Systems Research*, vol. 200, p. 107446, Jul. 2021.

[19] X. Lv, F. Liu, M. Jiang, F. Zhang, and L. Jia, "Fault diagnosis of power transformers based on dissolved gas analysis and improved LightGBM hybrid integrated model with dual-branch structure," *IET Electric Power Applications*, Dec. 2024.

[20] S. A. M. Abdelwahab, I. B. M. Taha, R. Fahim, and S. S. M. Ghoneim, "Transformer fault diagnose intelligent system based on DGA methods," *Scientific Reports*, vol. 15, p. 8263, 2025.

[21] A. Kirkbas, A. Demircali, S. Koroglu, and A. Kizilkaya, "Fault diagnosis of oil-immersed power transformers using common vector approach," *Electric Power Systems Research*, vol. 184, p. 106346, Apr. 2020.

[22] D. Zou, Z. Li, H. Quan, Q. Peng, S. Wang, Z. Hong, W. Dai, T. Zhou, and J. Yin, "Transformer fault classification for diagnosis based on DGA and deep belief network," *Energy Reports*, vol. 9, pp. 250–256, Oct. 2023.

[23] I. B. M. Taha, S. Ibrahim, and D.-E. A. Mansour, "Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in

measurements," *IEEE Access*, vol. 9, pp. 111162–111170, Jan. 2021.

[24] L. Tightiz, M. A. Nasab, H. Yang, and A. Addeh, "An intelligent system based on optimized ANFIS and association rules for power transformer fault diagnosis," *ISA Transactions*, vol. 103, pp. 63–74, Mar. 2020.

[25] L. Wang, T. Littler, and X. Liu, "Dynamic incipient fault forecasting for power transformers using an LSTM model," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 30, no. 3, pp. 1353–1361, Mar. 2023.

[26] E. Moradi, "Accuracy enhancement of fault diagnosis for power transformers with a hybrid approach integrating robust and tree-based algorithms," *Majlesi Journal of Electrical Engineering*, vol. 19, no. 2, 2025.

[27] P. A. R. Azmi, M. Yusoff, and M. T. M. Sallehud-Din, "Improving transformer failure classification on imbalanced DGA data using data-level techniques and machine learning," *Energy Reports*, vol. 13, pp. 264–277, Dec. 2024.

[28] M. Demirci, H. Gözde, and M. C. Taplamacioglu, "Improvement of power transformer fault diagnosis by using sequential Kalman filter sensor fusion," *International Journal of Electrical Power & Energy Systems*, vol. 149, p. 109038, Feb. 2023.

[29] J. Liu, Z. Zhao, Y. Zhong, C. Zhao, and G. Zhang, "Prediction of the dissolved gas concentration in power transformer oil based on SARIMA model," *Energy Reports*, vol. 8, pp. 1360–1367, Mar. 2022.

[30] B. C. Mateus, J. T. Farinha, and M. Mendes, "Fault detection and prediction for power transformers using fuzzy logic and neural networks," *Energies*, vol. 17, no. 2, p. 296, Jan. 2024.

[31] T. Kari, W. Gao, D. Zhao, K. Abiderexiti, W. Mo, Y. Wang, and L. Luan, "Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm," *IET Generation Transmission & Distribution*, vol. 12, no. 21, pp. 5672–5680, Sep. 2018.

[32] S. A. Gamel, S. S. M. Ghoneim, and Y. A. Sultan, "Improving the accuracy of diagnostic predictions for power transformers by employing a hybrid approach combining SMOTE
and DNN," Computers & Electrical Engineering, vol. 117, p. 109232, Apr. 2024.

[33] I. B. M. Taha and D.-E. A. Mansour, "Novel Power Transformer fault diagnosis using optimized machine learning methods," *Intelligent Automation & Soft Computing*, vol. 28, no. 3, pp. 739–752, Jan. 2021.

[34] D. A. Mansour, "Development of a new graphical technique for dissolved gas analysis in power transformers based on the five combustible gases," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 5, pp. 2507-2512, Oct. 2015.

[35] G. Xu, M. Zhang, W. Chen, and Z. Wang, "Transformer fault diagnosis utilizing feature extraction and ensemble learning model," *Information*, vol. 15, no. 9, p. 561, Sep. 2024.

[36] Ch. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," *Decision Analytics Journal*, vol. 6, p. 100164, Jan. 2023.

[37] C. Nkikabahizi, W. Cheruiyot, and A. Kibe, "Chaining Zscore and feature scaling methods to improve neural networks for classification," *Applied Soft Computing*, vol. 123, p. 108908, May 2022.

[38] E. Moradi, "Comparative analysis of Tree-Based machine learning algorithms on thyroid disease prediction using ROS technique and hyperparameter optimization," *Journal of Artificial Intelligence and Data Mining*, vol. 12, no. 4, pp. 511–520, 2024.

[39] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in Neurorobotics*, vol. 7, Jan. 2013.

[40] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[41] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD *Int. Conf. Knowl. Discovery Data Mining (KDD)*, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

[42] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4514–4523, Oct. 2020.

[43] Y. Zhang, X. Li, and Z. Wang, "A histogram-based gradient boosting approach for predictive maintenance in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 19, no. 3, pp. 2456–2465, Mar. 2023.

[44] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Proc. 31st *Int. Conf. Neural Inf. Process. Syst*. (NIPS), Long Beach, CA, USA, Dec. 2017, pp. 3149–3157.

[45] A. A. Yaghoubi, P. Karimi, E. Moradi, and R. Gavagsaz-Ghoachani, "Implementing engineering education based on posing a riddle in field of instrumentation and artificial intelligence". 2023 *9th International Conference on Control, Instrumentation and Automation (ICCIA)*, 2023, pp. 1–5.

[46] E. Moradi, " Leveraging Bayesian Optimization and Multilayer Artificial Neural Network (MLANN) for Fault Prediction in Oil-Immersed Transformers,"*e-Prime- Advances in Electrical Engineering, Electronics and Energy*, vol. 12, 2025.

[47] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," Proc. *IEEE Int. Conf. Neural Networks*, Perth, Australia, 1995, pp. 1942–1948.

[48] S. Rezashoar, and A. A. Rassafi, "Analyzing the Performance of the Red Deer Optimization Algorithm in Comparison to Other Metaheuristic Algorithms", *Journal of AI and Data Mining*, vol. 13, no. 1, pp. 53-61, 2025.

[49] M. Dehbozorgi, M, P. Shamsinejadbabaki, and E. Ashoormahani, "Better Neighbors, Longer Life: an Energy Efficient Cluster Head Selection Algorithm in Wireless Sensor Networks based on Particle Swarm Optimization", *Journal of AI and Data Mining*, vol. 11, no. 3, pp. 443-451, 2023.

[50] H. Kalani, and E. Abbasi, "Classification of sEMG Signals for Diagnosis of Unilateral Posterior Crossbite in Primary Dentition using Fast Fourier Transform and Logistic Regression", *Journal of AI and Data Mining*, vol. 10, no. 2, pp. 151-158, 2022.

[51] Z. Hassani, and M. Alambardar Meybodi, "Hybrid Particle Swarm Optimization with Ant-Lion Optimization: Experimental in Benchmarks and Applications", *Journal of AI and Data Mining*, vol. 9, no. 4, pp. 583-595, 2021.

[52] T. Sibalija, "Particle Swarm Optimisation in Designig Parameters of Manufacturing Processes: A Review (2008–2018) ". Applied Soft Computing, vol. 84, pp. 1-33, 2019.

یک تکنیک نوین پیش‌بینی خطا در ترانسفورماتورهای روغنی مبتنی بر گرادیان بوستینگ پیشرفته و بهینه‌سازی ازدحام ذرات (PSO)

الهه مرادی*

گروه مهندسی برق، واحد یادگار امام خمینی(ره) شهرری، دانشگاه آزاد اسلامی، تهران، ایران.

ارسال ۲۰۲۵/۰۵/۱۰؛ بازنگری ۲۰۲۵/۰۸/۱۷؛ پذیرش ۲۰۲۵/۰۸/۲۳

چکیده:

پیش‌بینی خطا در ترانسفورماتورهای قدرت برای حفظ قابلیت اطمینان عملیاتی و کاهش اختلالات سیستم بسیار مهم است. در سال‌های اخیر، با بهره‌گیری از داده‌های آنالیز گازهای محلول (DGA)، روش‌های مبتنی بر هوش مصنوعی برای ارتقای عملکرد پیش‌بینی مورد استفاده قرار گرفته‌اند. در این مقاله، یک چارچوب نوین یادگیری ماشین معرفی می‌شود که در آن روش تقویت گرادیان هیستوگرامی (HGB) با الگوریتم فراابتکاری بهینه‌سازی ازدحام ذرات (PSO) برای تنظیم ابرپارامترها ادغام شده است تا استحکام طبقه‌بندی‌کننده را تضمین نماید. ارزیابی روش پیشنهادی در دو مرحله انجام شده است: در مرحله نخست، عملکرد روش‌های تقویت گرادیان (GB)، تقویت گرادیان شدید (XGBoost) و HGB مورد ارزیابی قرار گرفته‌اند که HGB را به عنوان موثرترین روش نشان داد؛ در مرحله دوم، الگوریتم PSO برای بهینه‌سازی ابرپارامترهای HGB به‌کار گرفته شد که منجر به بهبود بیشتر عملکرد گردید. نتایج تجربی نشان می‌دهد که مدل ترکیبی HGB-PSO به دقت ۹۷٫۸۵ درصد، صحت98.9 درصد، فراخوانی 97.33 درصد و امتیاز F1 معادل ۹۸٫۹۹ درصد دست یافته است. تمامی شبیه‌سازی‌ها و تحلیل‌های مقایسه‌ای با روش‌های پیشرفته موجود در محیط پایتون پیاده‌سازی شده و تحلیل ماتریس اغتشاش برای ارزیابی جامع عملکرد پیش‌بینی مورد استفاده قرار گرفته است. این یافته‌ها حاکی از آن است که روش ترکیبی HGB-PSO به دقت و استحکام بالاتری در پیش‌بینی خطای ترانسفورماتور دست می‌یابد.

کلمات کلیدی: ترانسفورماتور قدرت، یادگیری ماشین، تقویت گرادیان، بهینه‌سازی ازدحام ذرات (PSO)، پیش‌بینی خطا، آنالیز گازهای محلول (DGA).