Rev. Roum. Sci. Techn.– Électrotechn. et Énerg.
Vol. **71**, *2*, pp. 247–252, Bucarest, 2026

FAULT DIAGNOSIS IN POWER TRANSFORMERS USING RESAMPLED DISSOLVED GAS ANALYSIS AND MACHINE LEARNING

SUDHAKAR THAVASILINGAM[1], SELLIGOUNDANUR SUBRAMANIAM SIVARAJU [2]

Keywords: Power transformer; Hybrid sampling; Support vector machine (SMVM); Dissolved gas analysis (DGA); Imbalanced dataset.

Fault diagnosis in power transformers is essential for ensuring operational reliability and minimizing maintenance downtime. This paper presents a hybrid machine learning framework that leverages advanced resampling and optimization techniques to enhance fault classification accuracy using dissolved gas analysis (DGA). To address class imbalance in the dataset, resampling methods such as the synthetic minority oversampling technique SMOTE), edited nearest neighbors (ENN), and adaptive synthetic sampling (ADASYN) are employed. Three classifiers, namely nonlinear support vector machine (NLSVM), linear discriminant analysis (LDA), and artificial neural networks (ANN) are evaluated, with hyperparameter tuning performed using gradient-based optimizer function (GFO) and firefly optimization (FFO). Among them, the FFO-optimized NLSVM combined with SMOTE-ENN achieved the highest performance, with a recognition rate of 98.7%. The results, validated through confusion matrices and ROC analysis, demonstrate the robustness of the proposed approach. This framework provides an effective and reliable tool for condition-based maintenance, enabling precise multi-class fault identification in power transformers.

1. INTRODUCTION

Power transformers (PTs) are crucial components of electricity distribution and transmission networks and may fail during operation due to internal arcing, partial discharge, short circuits, aging, and degradation of paper-oil insulation. Numerous recent studies have established fault diagnosis techniques using DGA in PT for condition monitoring [1-4]. DGA is widely used because thermal and electrical stressors cause gases like CH_4, C_2H_6, C_2H_2, C_2H_4, CO, CO_2, and H_2 to dissolve in oil, indicating defects that shorten transformer life and cause malfunction or outage. DGA data can also be interpreted using the Doerneburg ratio, IEC ratio, Key gas, Duval triangle, and Mansour Pentagon. However, these conventional methods require multiple test results and are affected by uncertainty in weight determination [5].

Deep learning enables the use of both labeled and unlabeled DGA data [6–8]. While effective, accurate diagnosis requires large, balanced datasets, yet most DGA datasets are imbalanced since inspection is costly, and fault occurrence is random. Resampling methods can address imbalance effectively [9]. With DGA datasets, support vector machine, extreme learning machine, radial basis function, and back-propagation neural networks have been applied, with the synthetic minority oversampling technique (SMOTE) used to reduce imbalance (misclassified and synthetic, MAS; misclassified initial samples, MIS)

$$IR = \frac{Quantity\ of\ MAS}{Quantity\ of\ MIS}. \quad (1)$$

A highly imbalanced data issue arises when the imbalance ratio (IR) is high, given by eq. (1). Misdiagnosis of the minority class is critical since it is financially sensitive. Treating imbalance can be done at the data level (oversampling, undersampling, hybrid), algorithm level (new or modified algorithms), or hybrid approaches [10,11]. This paper addresses class imbalance in DGA datasets using oversampling, under-sampling, and hybrid sampling before classifier training.

Most existing works either solve imbalance using traditional sampling only [12–17] or apply optimization without advanced resampling [16,17]. Hybrid schemes like SMOTE-ENN remain poorly explored with optimization for multi-class transformer fault classification. This research proposes a hybrid diagnostic model combining SMOTE-ENN and firefly optimization-optimized NLSVM, achieving 98.7% diagnostic accuracy and improving fault classification robustness for transformer maintenance. In [18], the authors proposed this novel methodology for fault detection and diagnosis of photovoltaic systems that employs Fisher's linear discriminant (FLD) with Mahalanobis distance. The functioning operational data, used in FLD, is projected in a low-dimensional space while maximizing the separation between healthy and faulty conditions. Moreover, in [19], the cuckoo-optimized modular neural network (COMNN) is proposed to detect and classify a crack in the blades of a wind turbine. Besides, the method is created using the capability of a piezoelectric accelerometer to calculate the vibration response of the blade while energizing it. Cuckoo optimization is also employed to initialize and then re-adjust the weight vector of the Modular Neural Network. In conclusion, the experimentation findings reveal that COMNN detects and categorizes faults accurately within an acceptable time frame on the wind turbine's blades.

In [20], two fault location methodologies were presented: the Atom search optimization metaheuristic approach (ASO) and machine learning (ML) with cubic spline models. The performance of both approaches was estimated by considering different fault types, fault distances, and fault resistance.

2. MATERIALS AND METHOD

2.1 DATASET DESCRIPTION

The dataset, compiled from multiple literature sources [13–17], consists of 592 DGA records classified into six fault types per IEC 60599 and IEEE C57.104: partial discharge (PD), low energy discharge (LED), high energy discharge (HED), low thermal fault (LTF), medium thermal fault (MTF), and high thermal fault (HTF). The classification of faults based on IEEE C57.104 and IEC 60599 standards is presented in Table 1, while Table 2 shows the distribution.

Table 1

Classification of faults based on IEEE C57.104 and IEC 60599 standards.

Shortenings	Common Faults in PT
PD (1)	Partial discharge
LED (2)	Low energy discharge
HED (3)	High energy discharge
LTF (4)	Low thermal fault T<300ºC
MTF (5)	Medium thermal fault 300ºC < T < 700ºC
HTF (6)	High thermal fault T>300ºC

[1] Department of Electrical and Electronics Engineering, PSR Engineering College, Sivakasi, Tamil Nadu, India.
[2] Department of Electrical and Electronics Engineering, R V S College of Engineering and Technology, Coimbatore, Tamil Nadu, India.
E-mails: sudhakar@psr.edu.in, sssivaraju@rvsgroup.com

DOI: 10.59277/RRST-EE.2026.2.13

A "No-fault" class was added using 145 DGA records from healthy transformers verified by field inspection, bringing the dataset to 737 samples across seven classes. To ensure reproducibility, the dataset was split using stratified sampling: 196 samples for testing (28 per fault type, 29 for "No-fault") and the rest for training.

Table 2

Distribution of gathered information following references to the faults in PT.

	Fault Types					
References	PD	LED	HED	LTF	MTF	HTF
[13]	0	32	32	0	0	0
[14]	16	35	15	29	19	30
[15]	32	51	74	85	41	56
[16]	7	2	2	0	5	4
[17]	0	7	18	0	0	0
Total	55	127	141	114	65	90

A "No-fault" class was added using 145 DGA records from healthy transformers verified by field inspection, bringing the dataset to 737 samples across seven classes. To ensure reproducibility, the dataset was split using stratified sampling: 196 samples for testing (28 per fault type, 29 for "No-fault") and the rest for training.

The six fault types, defined in IEC 60599, arise from issues such as loose connections, poor cooling, short circuits, and conductor or winding overheating. Table 2 shows their distribution, which is highly imbalanced and reduces classifier accuracy. To address this, the dataset was balanced using three resampling approaches: under-sampling, oversampling, and hybrid sampling. Figure 1 shows the proposed workflow for PT fault recognition.

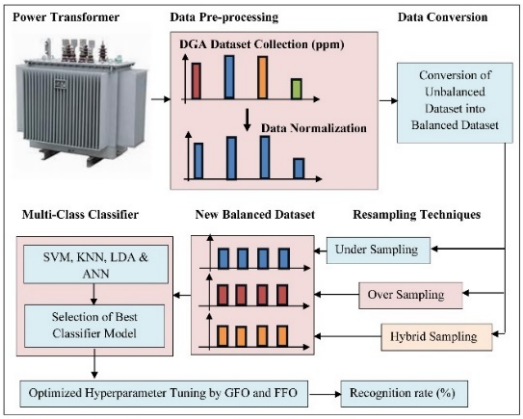


Fig. 1 – Flow diagram of proposed work for fault recognition.

2.2 GREY WOLF OPTIMIZATION TECHNIQUES

Grey wolf optimization is a metaheuristic algorithm proposed by Mirjalili and Lewis in 2014. It is a well-known nature-inspired algorithm that has evolved to handle complex, non-linear, high-dimensional problems. It delivers the best optimal solution to issues with more uncertainty. The stochastic nature of the algorithm initiates the optimization of derivative-free search spaces and avoids local maxima, which is highly useful in solving complex real-world problems. The central theme of this algorithm is a social hierarchy and the hunting of grey wolves [21].

Grey wolves live in closely organized packs. The average pack size ranges from 5 to 12. The wolves are divided into four groups based on their roles and contributions to the pack. First comes the alpha wolf, the most dominant, and it ranks top among its subordinates. It performs decision-making activities about hunting, sleeping places, and the time to wake up. It manages the pack in a highly organized and disciplined way. The pack members pay due respect to the alpha leader by holding their tails down. Beta Wolf shares the next position on the ranking ladder and is considered an equal performer to Alpha. It helps the alpha wolf in decision-making activities and gives feedback to the alpha about the pack's behaviour. It replaces the alpha's position in the event of death and during old age.

Delta Wolf occupies the third position on the ladder. They act as scouts, sentinels, elders, hunters, and caretakers. Scouts warn the pack in case of danger and guard the territorial boundaries. Sentinels ensure the safety of the pack and protect it from danger. The mathematical model for encircling the prey is as follows,

$$\vec{R} = \left|\vec{G}\vec{W}_p - \vec{W}(t)\right|, \quad (2)$$

$$\vec{W}(t+1) = \left|\vec{W}_p(t) - \vec{G}\vec{H}\right|, \quad (3)$$

where t determines the current iterations, $\vec{W}_p$ represents the position of prey, $\vec{W}$ characterize the position of the grey wolf. The coefficient vector of GFO is represented as $\vec{G}$ &$\vec{H}$. A mathematical model for hunting is as follows,

$$\vec{W}_1 = \left|\vec{W}_\alpha - H_1\vec{W}(t)\right|, \quad (4)$$

$$\vec{W}_2 = \left|\vec{W}_\beta - H_2\vec{W}(t)\right|, \quad (5)$$

$$\vec{W}_3 = \left|\vec{W}_\delta - H_3\vec{W}(t)\right|. \quad (6)$$

The updated grey wolf position is

$$\vec{W}(t+1) = \frac{\vec{W}_1 + \vec{W}_2 + \vec{W}_3}{3}, \quad (7)$$

where $\vec{W}_\alpha$, $\vec{W}_\beta$, $\vec{W}_\delta$ is a position vector of α , β , δwolves, and $G_1, G_2, G_3, H_1, H_2, H_3$ is a coefficient vector for GFO.

The hunting process involves searching for the prey, encircling, and attacking. Initially, the wolves wait for the target to reach their territorial boundary. They encircle them with other wolves and choose the weak target. They chase the target until it stops running. Finally, the leader attacks the prey and completes the hunting process.

2.3 FIREFLY OPTIMIZATION TECHNIQUES

The firefly algorithm is a metaheuristic algorithm proposed by Xin-She Yang in 2007. It is based on the flashing behavior of fireflies. The bright flashes they emit serve as a means of communication, attraction, and warning to predators. Drawing inspiration from past endeavors, Yang developed this approach based on the assumption that every firefly is unisexual, meaning that each can attract other fireflies and that each firefly's attraction is directly correlated with its brightness. As a result, brighter fireflies attract less bright ones to approach them. In addition, if there are no fireflies that are brighter than a particular firefly, then it moves at random. The objective function in the firefly algorithm determines the light intensity level. Individual fireflies must make random movements within the population to optimize the fitness function.

Fireflies' algorithm execution is controlled by three parameters: attractiveness, randomization, and absorption. The attractiveness parameter is defined using exponential functions based on the light amount between two fireflies. FFO fitness value and its light intensity are determined by

$$F(y_i) = \sum_{i=1}^{m} (y_i - 1)^2 , \quad (8)$$

$$I(d) = I_0 e^{-\gamma d^2} , \quad (9)$$

where I_0 represents the light intensity of the initial firefly, d is the distance between the two fireflies, and γ represents the optical absorption coefficient. The attractiveness between any two fireflies is directly proportional to the intensity of their brightness.

$$\beta(d) = \beta_0 e^{-\gamma d^2}, \quad (10)$$

where β_0 is the absorbance at $d = 0$. The random walk occurs when the attractiveness parameter is set to zero, corresponding to the randomization parameter chosen by the Gaussian distribution principle. Conversely, absorption parameters impact attractiveness parameter values, which vary from zero to infinity. Additionally, the movement of the firefly appears to be a random walk in the situation of convergence to infinity. The pseudocode for the proposed model is given as follows:

Input: D_train, D_val, bounds for (C, Conv), optimizer, N, T

Output: (C_best, Conv_best)

Initialize N random candidates (C[i], Conv[i]) within bounds

Repeat t = 1..T:

For each candidate i:

Train SVM on D_train using (C[i], Conv[i])

fitness[i] = score on D_val

best = argmax(fitness)

For each candidate i:

(C[i], Conv[i]) = OptimizerUpdate((C[i], Conv[i]), (C[best], Conv[best]), t)

Clamp (C[i], Conv[i]) to bounds

Return the candidate with the highest fitness

This proposed work shows SVM hyperparameters with ranges and optimal values chosen by FFO in Table 3.

Table 3

SVM hyperparameters range.

Hyperparameters	Description	Minimum	Maximum	Optimal Value
C	Intricacy Constant	-1	100	42
Conv	Convergence Epsilon	0.001	0.1	0.05

Regulating the trade-off between minimizing the classification error and maximizing the margin is the regularisation parameter (C). Greater values of C enable more intricate decision limits, which could improve the training data fit but increase the risk of overfitting. Although a broader margin is encouraged by lower values of C, underfitting may occur.

Finding the ideal variance ratio to bias is aided by tuning C. Optimizing hyperparameters can also help make the SVM model more straightforward. The decision boundaries become more logical and indicate the underlying patterns in the data by determining the ideal settings.

3. RESULT AND DISCUSSION

In this proposed work, the DGA dataset is balanced using three different resampling approaches, and the resampled data are used as input for three multi-class classification models to identify the types of faults occurring in power transformers. The machine learning models used in this experiment are NLSVM, LDA, and ANN. The classification results of these models are presented in Tables 4-6, while Table 7 summarizes the overall performance comparison. Resampling was performed only on the training set; the test set was kept unchanged for unbiased evaluation.

Confusion matrices are used to evaluate the classification performance for each balanced dataset, helping to identify which classifier performs best with each sampling technique. A total of 196 samples is used for testing, comprising 28 samples for each fault type and 29 samples for the no-fault class.

Table 4

Effect of resampling techniques on various classifier performances without hyperparameter tuning.

Support Vector Machine without Hyperparameter Tuning						
	Under-sampling		Oversampling		Hybrid sampling	
Parameters	RUS	TL	ADASYN	BI-SMOTE	SMOTE-ENN	SA
Accuracy	0.781	0.760	0.824	0.714	0.852	0.789
Precision	0.781	0.746	0.814	0.743	0.821	0.799
Recall	0.792	0.754	0.817	0.784	0.834	0.754
F1-Score	0.788	0.762	0.817	0.748	0.870	0.748
RoC	0.812	0.841	0.846	0.733	0.921	0.795
Linear discriminant analysis						
	Under-sampling		Oversampling		Hybrid sampling	
Parameters	RUS	TL	ADASYN	BI-SMOTE	SMOTE-ENN	SA
Accuracy	0.765	0.744	0.801	0.698	0.799	0.805
Precision	0.772	0.715	0.800	0.714	0.801	0.819
Recall	0.784	0.705	0.812	0.700	0.822	0.822
F1-Score	0.754	0.741	0.799	0.727	0.847	0.854
RoC	0.801	0.792	0.827	0.734	0.887	0.899
Artificial neural network						
Parameters	RUS	TL	ADASYN	BI-SMOTE	SMOTE-ENN	SA
Accuracy	0.711	0.722	0.711	0.684	0.741	0.721
Precision	0.688	0.714	0.721	0.699	0.744	0.705
Recall	0.674	0.714	0.701	0.674	0.722	0.699
F1-Score	0.687	0.701	0.701	0.677	0.702	0.701
RoC	0.654	0.671	0.721	0.690	0.754	0.700

Several performance metrics, such as accuracy, precision, recall, and F1-score, are derived from the confusion matrices. These metrics provide a quantitative assessment of the model's ability to distinguish between different fault classes.

From Table 4, hybrid sampling SMOTE – ENN attains an accuracy rate of 85% using the NLSVM classifier without tuned hyperparameters. LDA attains 80% accuracy because it assumes that the data are distributed according to Gaussian distributions, which may not necessarily be true in real-world scenarios. LDA's sensitivity to outliers stems from its attempt to reduce variance within each class. However, knowing its assumptions and limits is crucial when using it on real-world datasets.

Table 5

Effect of resampling techniques on SVM classifier performance with hyperparameter tuning using the GFO algorithm.

Support Vector Machine with Hyperparameter Tuning using GFO						
	Under-sampling		Oversampling		Hybrid sampling	
Parameters	RUS	TL	ADASYN	BI-SMOTE	SMOTE-ENN	SA
Accuracy	0.875	0.834	0.879	0.790	0.923	0.820
Precision	0.899	0.855	0.897	0.824	0.912	0.835
Recall	0.915	0.867	0.897	0.845	0.894	0.871
F1-Score	0.942	0.899	0.912	0.856	0.921	0.871
RoC	0.927	0.912	0.922	0.899	0.954	0.894

ANN achieves the highest accuracy rate of 84% due to many hyperparameters, including learning rate, batch size, number of layers, and number of neurons per layer, which must be tuned when using artificial neural networks (ANNs). Determining the optimal set of hyperparameters can be difficult, and it frequently necessitates much experimentation. ANNs are prone to overfitting, particularly in intricate models and sparse training data. When a model becomes overfitted, it

performs poorly on unknown data because it has learned to memorize the training set instead of generalizing it. A significant volume of labeled training data is usually needed for ANNs, intense neural networks, to learn efficiently. Table 5 shows the NLSVM classifier with hyperparameter tuning using GFO.

NLSVM with hyperparameters tuned using GFO attains the highest accuracy rate for SMOTE-ENN of 92%. SVM works by utilizing a kernel function, usually the RBF, to translate the input space into a high-dimensional feature space, enabling the algorithm to produce non-linear decision boundaries. Compared to other complicated models, such as neural networks, RBF SVM is less prone to overfitting, mainly when the ratio of training samples to features is minimal. The robustness of the model is enhanced by the kernel parameter (gamma), which controls the smoothness of the decision boundary, and the regularisation parameter (C) in SVM, which aids in controlling overfitting. To improve generalization performance, SVMs seek to identify the ideal hyperplane that maximizes the margin between classes. Table 6 shows the accuracy rate of NLSVM with hyperparameters tuned by FFO.

The strongest scores are obtained with SMOTE-ENN under FFO tuning (Accuracy 0.987, Recall 0.988, ROC 0.991), which suggests that hybrid sampling improves minority-class learning while maintaining low false positives. The ENN cleaning step likely reduces borderline ambiguities introduced during oversampling, enabling the RBF-SVM to form a more reliable margin.

Table 6

Effect of resampling techniques on SVM classifier performance with hyperparameter tuning using the firefly algorithm.

Support Vector Machine with Hyperparameter Tuning using FFO						
	Under-sampling		Oversampling		Hybrid sampling	
Parameters	RUS	TL	ADASYN	BI-SMOTE	SMOTE-ENN	SA
Accuracy	0.897	0.911	0.922	0.857	0.987	0.875
Precision	0.821	0.924	0.947	0.867	0.979	0.899
Recall	0.922	0.934	0.967	0.899	0.988	0.912
F1-Score	0.960	0.944	0.957	0.901	0.976	0.954
RoC	0.944	0.955	0.964	0.934	0.991	0.922

NLSVM with RBF kernel attains an accuracy rate of 99% using the FFO optimization algorithm. A subset of the training samples, or support vectors, helps define the decision boundary in most cases when RBF SVM generates sparse solutions. This characteristic, particularly for large-scale datasets, makes the model computationally quicker and more memory-efficient during the training and inference stages. Despite having relatively few training datasets, RBF SVM can produce good results. It is beneficial when getting a lot of complex or expensive labeled data.

Figure 2 represents the classification result of various classifiers used in this work. Spider plots can be used to evaluate the accuracy and precision of various classifiers used in this proposed work with one another across a range of parameters. The shape's outward extensions indicate strengths, and its inner contractions indicate weaknesses. Figure 2(b) and 2(d) represent the inner contractions for RUS and TL, respectively. From the investigation, SMOTE-ENN accuracy provides a higher value than other classifiers.

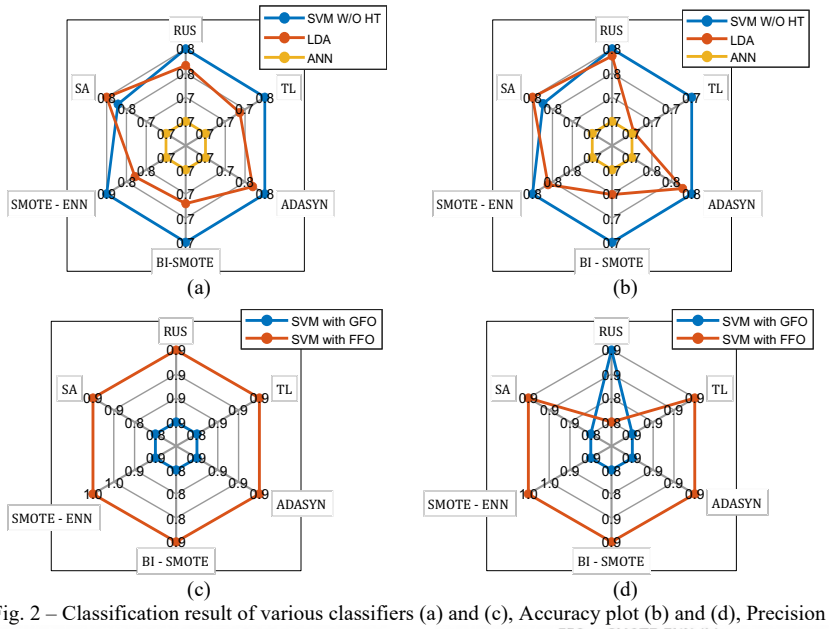
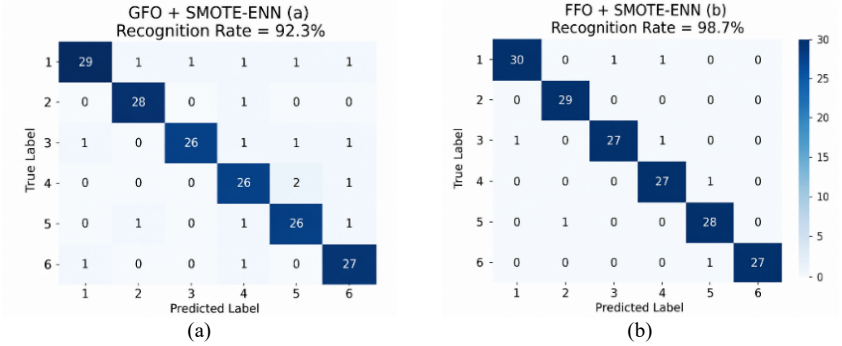
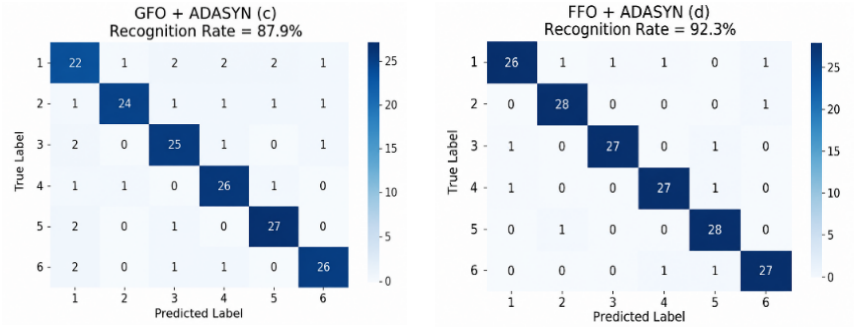


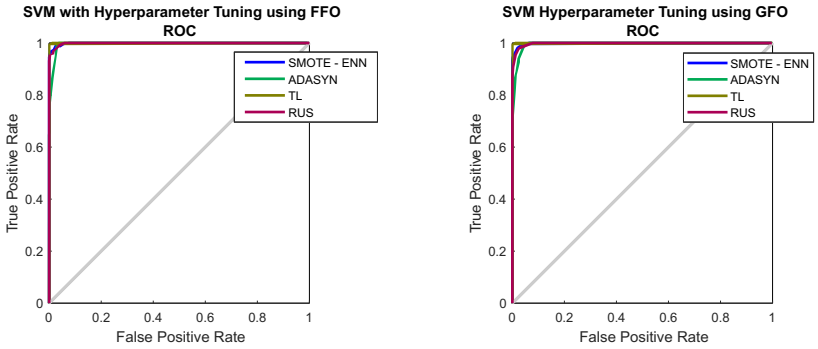
Fig. 2 – Classification result of various classifiers (a) and (c), Accuracy plot (b) and (d), Precision plot.





(c) (d)

Fig. 3 – Confusion matrices for NLSVM under different sampling and optimization configurations: (a) GFO with SMOTE-ENN, (b) FFO with SMOTE-ENN, (c) GFO with ADASYN, and (d) FFO with ADASYN.



(a) (b)

Fig. 4 – ROC curves for the classifier-tuned hyperparameters using GFO and FFO.

Figure 3 shows the confusion matrix for NLSVM with hyperparameters tuned using GFO and FFO. FFO provides a higher recognition rate of 98.7% for hybrid sampling SMOTE-ENN. Figure 4 represents the ROC curves for GFO and FFO. The ROC curve (F+) plots the true-positive rate (T+) on the y-axis against the false-positive rate (F+) on the x-axis, as shown in Fig. 4.

Table 7

Comparison of the recent state-of-the-art transformer fault diagnosis algorithms.

References	Dataset Types	Accuracy Rate
[13]	Balanced	92%
[14]	Balanced	95%
[15]	Imbalanced	88%
[16]	Balanced	98%
Proposed Work	Imbalanced	98.6%

ROC determines the brief evaluation of the model's quality. The model is better if it is closer to 1. T (+), T (-), F(+), and F(-) are all included in a confusion matrix that offers a thorough analysis of the model's estimates. ROC makes it possible to thoroughly assess the model's performance over various error types and classes. Generally, models with higher Accuracy values perform better than those with lower Accuracy values. When choosing the optimal model, it's crucial to consider additional elements, including the application environment, computing capacity, and interpretability. ROC curves are especially helpful when assessing models on unbalanced datasets with unequal class representation. Thus, the proposed work performs better and has less computation time than other approaches.

The DGA dataset used in this work is inherently imbalanced because fault occurrence is rare and inspection is costly, meaning minority classes have fewer samples and are more likely to be misdiagnosed, which is financially sensitive. In such settings, standard learners tend to favor majority classes, and borderline minority samples can be treated as noise. The SMOTE-ENN pipeline addresses both issues simultaneously: SMOTE increases minority representation by synthesizing additional samples in sparse regions, while ENN serves as a cleaning step that removes samples inconsistent with their local neighborhoods, helping reduce the influence of overlapping or noisy instances (a common issue in multi-class imbalanced data).

This is particularly beneficial for NLSVM with an RBF kernel because SVM decision boundaries are sensitive to support vectors near class borders; removing ambiguous points after oversampling yields a cleaner margin and improves generalization.

In addition, the performance gain depends strongly on selecting suitable hyperparameters (e.g., regularization and kernel-related parameters). Table 6 and the confusion-matrix comparison show that FFO-tuned NLSVM with SMOTE-ENN achieves the highest recognition rate (98.7%) and strong ROC behavior, indicating improved separability across classes even under imbalance. This can be attributed to FFO's effective exploration of the hyperparameter space, which helps locate a parameter setting that balances margin maximization and misclassification tolerance on the resampled (cleaned) training data. Overall, SMOTE-ENN improves the training distribution and label consistency, while FFO improves the model configuration, and their combination produces the most stable class boundaries for DGA fault discrimination.

The proposed framework can be incorporated into an online DGA monitoring pipeline to support condition-based maintenance by providing fast, automated fault identification once the model is trained offline, since real-time operation mainly involves feature extraction and an NLSVM inference step. However, practical deployment should consider several limitations: field DGA streams may include sensor noise,

missing values, calibration drift, or sampling delays that shift feature distributions; the training dataset (compiled from limited labeled records) may not fully represent all utility operating conditions, transformer designs, or oil types, so periodic validation and retraining with local data may be necessary; SMOTE-ENN improves learning during training but the deployed stream remains naturally imbalanced, requiring careful alarm thresholds to avoid excessive false alerts; and real-world adoption may be constrained by edge computing capacity and the need for interpretability and operator trust, suggesting the value of confidence reporting, logging, and simple explanation mechanisms alongside the model.

4. CONCLUSION

This study investigated the impact of undersampling, oversampling, and hybrid sampling techniques on power transformer fault diagnosis using imbalanced DGA datasets. Results demonstrated that class imbalance significantly reduces classifier performance, which can be improved through appropriate resampling and optimization strategies. Among the evaluated models, NLSVM, LDA, and ANN, the FFO-tuned NLSVM combined with SMOTE-ENN achieved the highest classification accuracy. This confirms the effectiveness of combining hybrid sampling with hyperparameter optimization. The proposed framework enhances diagnostic reliability and supports condition-based maintenance, offering a practical tool for improving the operational efficiency of power transformers.

ACKNOWLEDGEMENTS

The author thanks the supervisor for guidance and support.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

SUDHAKAR THAVASILINGAM: Conceptualization of the research, methodology development, overall paper structure, data analysis and interpretation, drafting the manuscript.

SELLIGOUNDANUR SUBRAMANIAM SIVARAJU: Supervision of the project, reviewing and editing the manuscript, and providing critical feedback on theoretical and experimental results.

Received on 22 August 2025

REFERENCES

1. A. Christina, M.A. Salam, Q.M. Rahman, F. Wen, S.P. Ang, W. Voon, *Causes of transformer failures and diagnostic methods – A review*, Renew. Sustain. Energy Rev., **82**, pp. 1442–1456 (2017).
2. A. Moeini, M. Dabbaghjamanesh, T. Dragičević, J.W. Kimball, J. Zhang, *Machine learning technique for low-frequency modulation techniques in power converters*, Control of Power Electronic Converters and Systems, Academic Press, pp. 149–167 (2021).
3. Y. Wang, X. Zeng, Z. Dong, Y. Huang, *Novel protection scheme of stator single-phase-to-ground fault for power formers*, Int. J. Electr. Power Energy Syst., **53**, pp. 321–328 (2013).
4. J. Zhao, Y. Xu, F. Luo, Z. Dong, Y. Peng, *Power system fault diagnosis based on history-driven differential evolution and stochastic time domain simulation*, Inf. Sci., **275**, pp. 13–29 (2014).
5. L. Wang, X. Zhao, J. Pei, G. Tang, *Transformer fault diagnosis using continuous sparse autoencoder*, SpringerPlus, **5**, *1*, pp. 1–10 (2016).
6. J. Dai, H. Song, G. Sheng, X. Jiang, *Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network*, IEEE Trans. Dielectr. Electr. Insul., **24**, *5*, pp. 2828–2835 (2017).
7. S. Kim, S.H. Jo, W. Kim, J. Park, J. Jeong, Y. Han et al., *A semi-supervised autoencoder with an auxiliary task (SAAT) for power transformer fault diagnosis using dissolved gas analysis*, IEEE Access, **8**, pp. 178295–178310 (2020).
8. Y. Cui, H. Ma, T. Saha, *Improvement of power transformer insulation diagnosis using oil characteristics data preprocessed by SMOTEBoost technique*, IEEE Trans. Dielectr. Electr. Insul., **21**, *5*, pp. 2363–2373 (2014).
9. G. Du, J. Zhang, M. Jiang, J. Long, Y. Lin, S. Li et al., *Graph-based class-imbalance learning with label enhancement*, IEEE Trans. Neural Netw. Learn. Syst., pp. 1–12 (2021).
10. M. Koziarski, M. Woźniak, B. Krawczyk, *Combined cleaning and resampling algorithm for multi-class imbalanced data with label noise*, Knowl.-Based Syst., **204**, pp. 1–12 (2020).
11. S.M.A. Lopes, R.A. Flauzino, R.A.C. Altafim, *Incipient fault diagnosis in power transformers by data-driven models with an over-sampled dataset*, Electr. Power Syst. Res., **201**, pp. 1–10 (2021).
12. H.A. Illias, W.Z. Liang, *Identification of transformer fault based on dissolved gas analysis using hybrid support vector machine-modified evolutionary particle swarm optimisation*, PLoS ONE, **13**, *1*, pp. 1–18 (2018).
13. E. Li, L. Wang, B. Song, *Fault diagnosis of power transformers with membership degree*, IEEE Access, **7**, pp. 28791–28798 (2019).
14. S.I. Ibrahim, S.S.M. Ghoneim, I.B.M. Taha, *DGALab: An extensible software implementation for DGA*, IET Gener. Transm. Distrib., **12**, *18*, pp. 4117–4124 (2018).
15. B. Zeng, J. Guo, W. Zhu, Z. Xiao, F. Yuan, S. Huang, *A transformer fault diagnosis model based on hybrid grey wolf optimizer and LS-SVM*, Energies, **12**, *21*, pp. 1–10 (2019).
16. G. Lv, H. Cheng, H. Zhai, L. Dong, *Fault diagnosis of power transformer based on multi-layer SVM classifier*, Electric Power Systems Research, **75**, *1*, pp. 9–15 (2005).
17. M.S. Kraiem, F. Sánchez-Hernández, M.N. Moreno-García, *Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association models*, Appl. Sci., **11**, pp. 1–15 (2021).
18. P. Babu, C. Columbus, S.R. Lakshmi, J. Chithambaram, *Wind turbine fault modeling and classification using cuckoo-optimized modular neural networks*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **68**, *4*, pp. 369–374 (2023).
19. A. Saidi, A. Draoui, T. Abdelouahed, *Fault detection and diagnosis in photovoltaic power systems using Fisher random matrix approach*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **70**, *4*, pp. 513–518 (2025).
20. K. Guerraiche, A.B. Abbou, L. Dekhici, *Intelligent fault detection and location in electrical high-voltage transmission lines*, Rev. Roum. Sci. Techn. – Électrotechn. et Énerg., **69**, *3*, pp. 269–276 (2024).
21. S. Saravanan, R.S. Kumar, P. Balakumar, *Binary firefly algorithm-based reconfiguration for maximum power extraction under partial shading and machine learning approach for fault detection in solar PV arrays*, Appl. Soft Comput., **154**, pp. 1–15 (2024).