

Original Article

New Diagnostic Framework for Transformer Health using IEEE C57.104 2019 with Gradient Boosting and PCA of DGA Data

ApolinarioAwit¹, Joseph Jay Brañanola², Maria Vina Presbitero³, Edizar Sablas⁴, Rannel Sinangote⁵,
John Kenneth Tabla⁶, Jestoni P. Tan⁷, Emerita M. Tan⁸, Chona R. Dagatan⁹

^{1,2,3,4,5,6,7,8}Electrical Engineering Department, Cebu Technological University - Danao Campus; Danao City, Cebu, Philippines.

⁹Computer Engineering Department, Cebu Technological University - Danao Campus; Danao City, Cebu, Philippines.

¹Corresponding Author : jestonicvi22@gmail.com

Received:25 August 2025

Revised:10 April 2026

Accepted:08 May 2026

Published: 27 June 2026

Abstract - Transformers play an indispensable role in any power system. The health condition of these devices should be the top priority. Early fault detection of these devices is essential to have sustainable power flow. There are many routinary transformer tests like winding test, furan analysis, insulation resistance tests, but dissolved gas analysis stands to be one of the most critical tests among others. This is to the fact that the DGA test can evaluate the major condition of the transformer. Dissolved Gas Analysis (DGA) methods, while widely used, often struggle with accuracy and scalability under complex fault scenarios. This paper proposed a novel ML-based DGA framework that integrates the IEEE standard with Principal Component Analysis (PCA) and Gradient Boosting Machine (GBM) to enhance transformer fault diagnosis. PCA captures 95% of the variance with five principal components. The framework showed a test accuracy of 87.5% and a cross-validation accuracy of 86.05%, outperforming traditional methods such as the Duval Triangle (83.08%) and IEC Ratio Method (82.05%), as well as other machine learning models, including Random Forest (77%) and Support Vector Machines (37%). These findings demonstrate the effectiveness of the framework as a soft sensor in Transformer diagnostics.

Keywords - Dissolved Gas Analysis, Gradient Boosting Machine, Transformer Health, Principal Component Analysis.

1. Introduction

Accurate and timely monitoring is essential for transformers to avert power outages and asset damage. Dissolved Gas Analysis (DGA) is a long-established pragmatic method for assessing transformer health by evaluating the concentrations of gases dissolved in insulating oil [1]. However, traditional interpretation techniques such as the Duval Triangle, Rogers Ratio, IEC Ratio, and the Doernenburg Method are often limited by human subjectivity and rigid thresholds [5, 6]. As power grids grow in complexity, the increasing necessity for scalable, automated diagnostic tools that can overcome the inconsistencies found in manual interpretation is limited, which have become a motivation for the integration of machine learning into transformer diagnostics to improve fault detection accuracy and adaptability over time [2]. Modernizing these diagnostics is often hindered by the inherent noise and high dimensionality of raw Dissolved Gas Analysis data, which frequently causes models to overfit rather than generalize to new transformer conditions. To mitigate this, Principal Component Analysis (PCA) has been shown to effectively reduce the dimensionality of high-dimensional DGA datasets

while preserving critical information. Principal Component Analysis (PCA) has been proven effective in reducing the number of dimensions in large DGA datasets while still preserving important information [3].

Given the inherent high dimensionality of these datasets, Principal Component Analysis (PCA) presents a valuable approach for improving both computational efficiency and model interpretability for engineering applications. The classification process needs to handle complex nonlinear relationships that exist in transformer gas patterns, despite dimensionality reduction making the data more manageable. The performance of Gradient Boosting Machines (GBMs) exceeds that of traditional methods because they use labeled data for their operation. Furthermore, Gradient Boosting Machines (GBMs), which are known for their ability to capture complex, non-linear relationships, are better than traditional methods because they learn directly from labeled data [4]. Despite the successes of these algorithms, many current machine learning models lack a validation step that considers the constraints of physical engineering.



The operational success of these algorithms has been achieved through their implementation in machine learning models, which still need validation and testing based on physical engineering constraints. The existing gap exists because the current studies have not studied dimensionality reduction together with class imbalance and standards-based validation in one comprehensive diagnostic system used for transformer fault detection [8, 19].

This research endeavors to solve existing deficiencies by proposing a diagnostic framework that synergizes Principal Component Analysis (PCA) and Gradient Boosting Machines (GBMs) with the IEEE C57.104-2019 standard for interpretability [7]. Through combining dimensionality reduction techniques with ensemble learning and using established industry benchmarks, this framework should become a more accurate and adaptable method for diagnosing faults.

The research employs a dataset comprising 318 Dissolved Gas Analysis samples, wherein the SMOTE-Tomek method is implemented to address intrinsic class imbalances. To validate the reliability of the findings, model performance is evaluated via stratified k-fold cross-validation. Overall, the system undergoes testing through its evaluation process, which assesses its performance against traditional diagnostic methods and current machine learning systems.

2. Literature Review

2.1. Dissolved Gas Analysis (DGA) Condition Assessment

Dissolved Gas Analysis (DGA) is a widely used technique for monitoring the condition of a transformer using insulating oil analysis. According to Patil et al. [8], this technique assists in the early detection of faults in a transformer. This occurs through the analysis of dissolved gases in insulating oil, which develop due to oil degradation. This process assists in the maintenance of a transformer.

Recent studies have revealed that it is a complex process to analyze the DGA technique because of insufficient data and a subjective approach. To resolve this problem, a machine learning approach has been adopted for the accurate determination of faults in a transformer. This process was proposed in a recent study by Raghuraman et al. [9]. Nanfak et al. [10] also highlighted that DGA analysis is a powerful technique for identifying insulation failure in a transformer. It is a challenging task to handle a complex fault using conventional DGA analysis.

2.2. Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) is regarded as a highly efficient machine learning technique with strong prediction capability. It was successfully used in the classification of transformer faults using DGA data, as confirmed in the latest literature. For example, Li et al. [4]

confirmed the capability of the GBM technique in outperforming traditional diagnostic techniques by learning nonlinear relationships. The capability of the GBM technique in handling a large number of data points. Therefore, the GBM technique is regarded as a highly efficient technique in dealing with large-scale power systems. In addition, the GBM technique is able to automate the diagnosis. The other advantages of the GBM technique are its capability in creating robust models using the Extreme Gradient Boosting (XGBoost) technique through the application of regularization. The efficiency of the GBM models was confirmed in the literature compared to traditional models in terms of accuracy, precision, and reliability [11].

2.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) has been used in machine learning models for data reduction and feature extraction. Gagne et al. [12] proved that PCA reduces data complexity by taking into account fundamental data features while maintaining crucial information. This enhances the performance of the machine learning model while reducing data analysis complexity. Kim et al. [13] proved that PCA enhances the interpretability of the machine learning model without losing crucial information. This is crucial when dealing with large amounts of data in transformer diagnostics.

Although PCA has been applied in medical diagnostics and engineering diagnostics, there is still a gap identified by Hasan and Abdulazeez [14] regarding the direct application of PCA in transformer diagnostics. This proves that PCA should be evaluated regarding its application in transformer fault prediction.

2.4. Machine Learning Approaches

Machine learning is an integral part of health monitoring and diagnosing the condition of a transformer. The importance of intelligent monitoring systems for health monitoring of a transformer using machine learning models. The reliability of a transformer is improved with intelligent monitoring systems. The application of hybrid models for health monitoring of a transformer is also presented in this study. Zivkovic et al. [16] presented the application of a hybrid model consisting of a convolutional neural network model and XGBoost model for health monitoring of a transformer.

2.5. Data Pre-processing and Feature Engineering

Data pre-processing and feature engineering are crucial in enhancing the efficiency of machine learning models in transformer diagnostic analysis. Potamias et al. [17] used a sliding time window technique in the diagnosis of transformer faults. Zhao et al. [18] developed a framework for data pre-processing, feature extraction, and classification. Yang et al. [19] developed a double-stacked autoencoder model for dealing with imbalanced DGA data in transformer diagnosis.

Privacy-preserving and secure machine learning models have been proposed for the protection of transformer data without compromising the efficiency of the models [13].

3. Materials and Methods

3.1. Study Design, Population, and Setting

The diagnostic reference used in the IEEE standard for fault labeling and benchmarking is the IEEE C57.104-2019 standard. The standard defines the threshold limits of the recognized gases used in the classification of the health conditions of the transformer. There are a total of 318 samples used in the research, which were collected from publicly

available datasets. There are 201 samples from [4] and 117 from the IEC TC10 database. The samples used in the research are from a wide range of conditions, verified through visual inspection.

Table 1 and Figure 1 show the number of samples in the data set before preprocessing for different types of faults. As shown in Figure 1, the data set is imbalanced since the number of High Energy Discharge (D2) faults is 41.2%, while the number of Partial Discharge (PD) faults is only 7.3%. This is a major weakness in the data set for accurate classification using machine learning techniques.

Table 1. Dataset distribution

Fault Type	No. of Datasets	Percentage
Partial Discharge (PD)	17	7.3%
Low Energy Discharge (D1)	52	22.3%
High Energy Discharge (D2)	96	41.2%
Thermal Fault (T1 & T2)	32	13.73%
Thermal Fault (T3)	36	15.45%

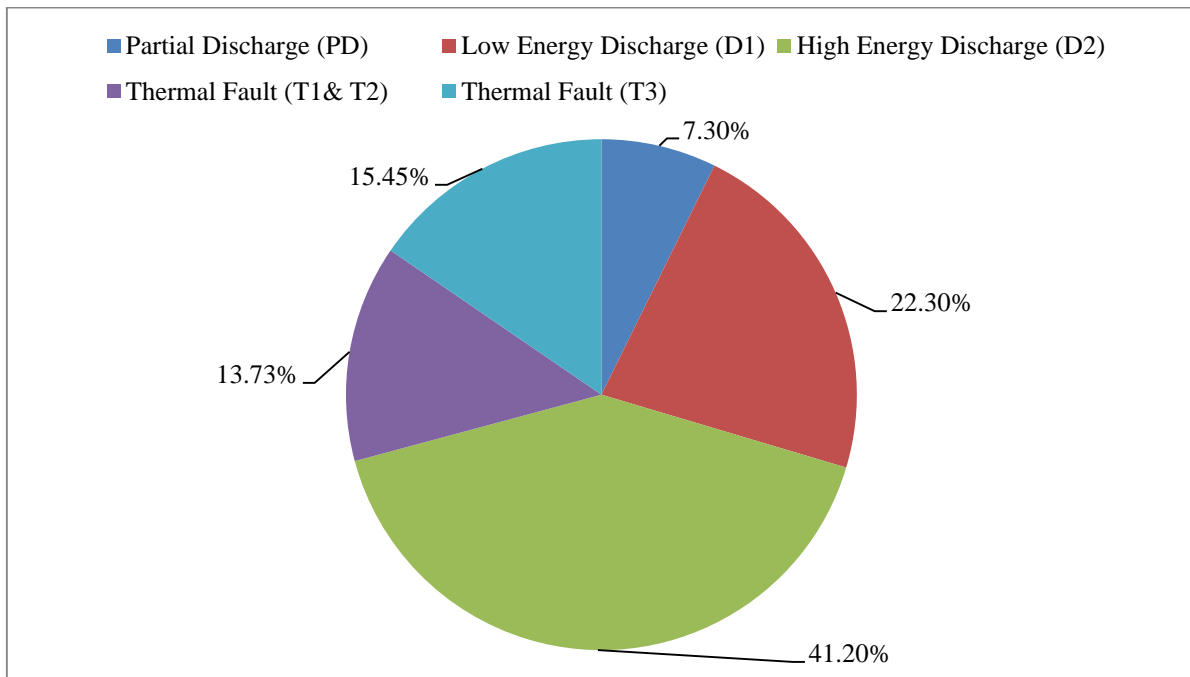


Fig. 1 Datasets Distribution Pie Chart

3.2. Preprocessing of the Dataset

The input variables of the dataset (in ppm) are Hydrogen (H2), Methane (CH4), Ethane (C2H6), Ethylene (C2H4), Acetylene (C2H2), Carbon Monoxide (CO), and Carbon Dioxide (CO2). These gases were used as predictors (independent variables). The output of the dataset is classified based on IEC 60599 into five categories: Partial Discharge (PD), Low Energy Discharge (D1), High Energy Discharge (D2), Low to Medium Temperature (T1/T2), and High Temperature Thermal Fault (T3).

3.2.1. SMOTE Tomek

The official dataset was imbalanced. This may affect the output of the framework. To address this challenge, SMOTE-Tomek was used—an oversampling and undersampling hybrid technique that combines: SMOTE (Synthetic Minority Oversampling Technique) for generating synthetic minority class examples and Tomek Links for removing borderline or overlapping majority class instances.

	H2	CH2	C2H2	C2H4	C2H6
count	233.000000	233.000000	233.000000	233.000000	233.000000
mean	5125.793991	3017.978541	1720.248927	3244.523605	1397.909871
std	11853.655268	8780.397705	6411.149691	10755.980574	7571.860238
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	210.000000	85.000000	11.000000	51.000000	11.000000
50%	960.000000	580.000000	190.000000	322.000000	49.000000
75%	4230.000000	1740.000000	769.000000	1560.000000	249.000000
max	92600.000000	64064.000000	57000.000000	95650.000000	72128.000000

	CO	CO2	Type
count	233.000000	233.000000	233.000000
mean	1157.789700	5718.480687	2.077253
std	2508.350772	15490.066080	1.126884
min	0.000000	0.000000	0.000000
25%	105.000000	413.000000	1.000000
50%	313.000000	1250.000000	2.000000
75%	829.000000	3923.000000	3.000000
max	15800.000000	103151.000000	4.000000

Fig. 2 Dataset statistics

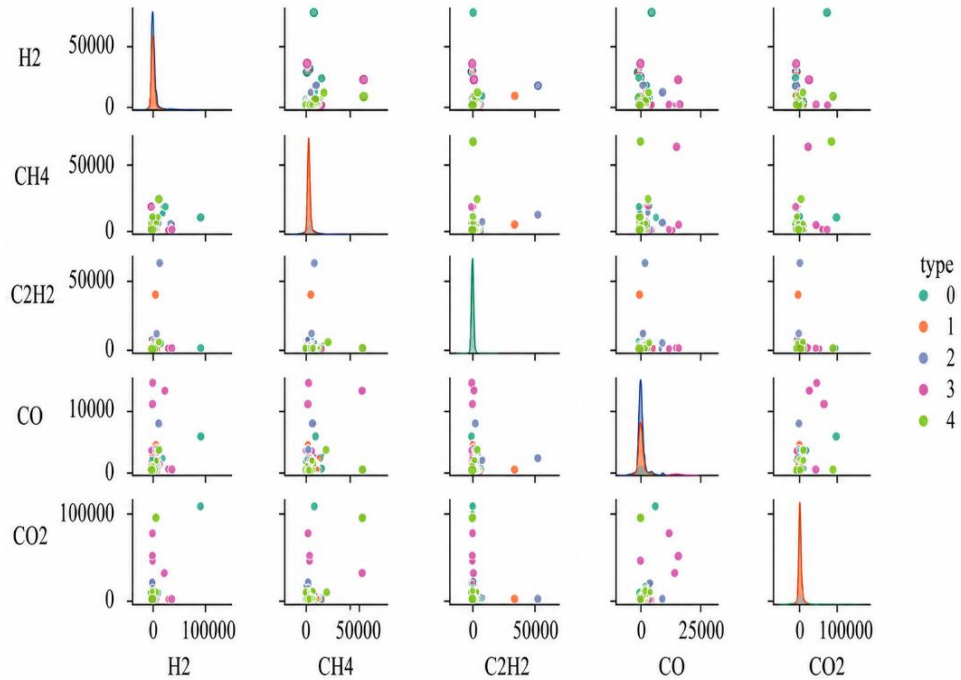


Fig. 3 Pairplot between features

3.2.2. Dimensionality Reduction with PCA

The curse of dimensionality exists in every machine learning endeavor. The study employs Principal Component Analysis to address this challenge. PCA is a statistical technique that determines high-variance principal components of the dataset, reducing memory usage and enhancing the speed of calculations. The first component has the highest information of the dataset (highest dispersion of datapoints), followed by the next principal component.

PCA constructs these components by using a standardization method and computing the covariance matrix. Consequently, get the eigenvectors ranking based on the eigenvalues of the covariance matrix. Those ranked eigenvectors will now be the principal components. As shown in Figure 4, the principal component is the purple line in the scatter plot, and it is the line on which the red dots, which represent the variance, are scattered (the highest variance value).

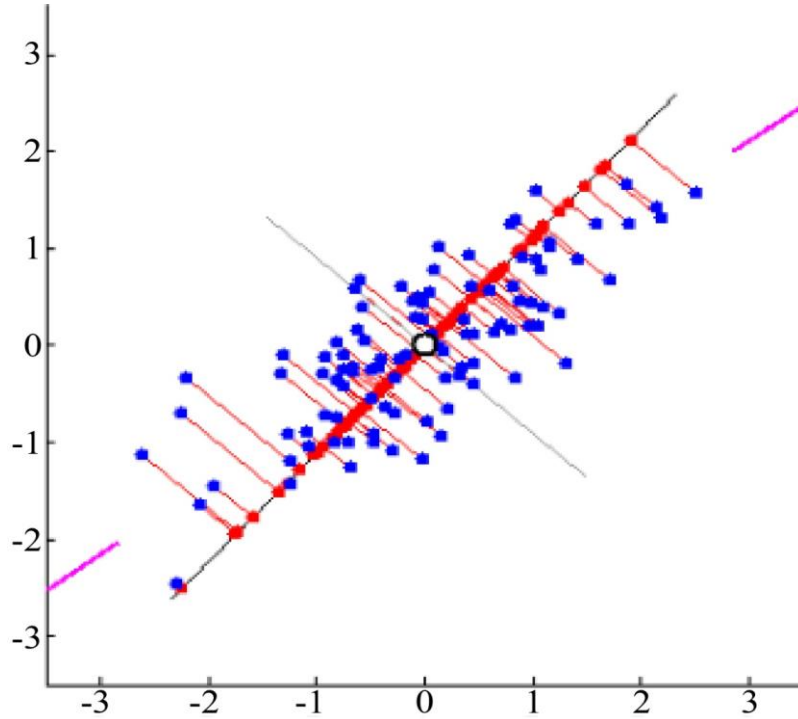


Fig. 4 PCA Illustration

3.3. Model Development and Training

3.3.1. Gradient Boosting Machine

Gradient Boosting Machine (GBM) is an ensemble sequential method known for handling non-linear relationships.

Unlike the Random Forest algorithm, which does majority voting and parallel computations, GBM is trained to correct the errors of the previous decision trees, and it reduces bias in the computations, as shown in Figure 5.

The advantages of Gradient Boosting Machines are numerous. These advantages include the ability of the model

to handle high-dimensional data, the model’s robustness when dealing with a limited dataset, and the model’s immunity from overfitting and noise. In addition, the model is able to provide insights into the importance of the features used in the model. As a result, the model is able to assist in the selection of features.

However, the model’s success is dependent on the tuning of the model’s parameters. These parameters include the number of trees in the model, the learning rate, and the depth of the trees. The search space for hyperparameters consisted of learning rate (0.01-0.20), maximum tree depth (3-10), and number of estimators (50- 300).

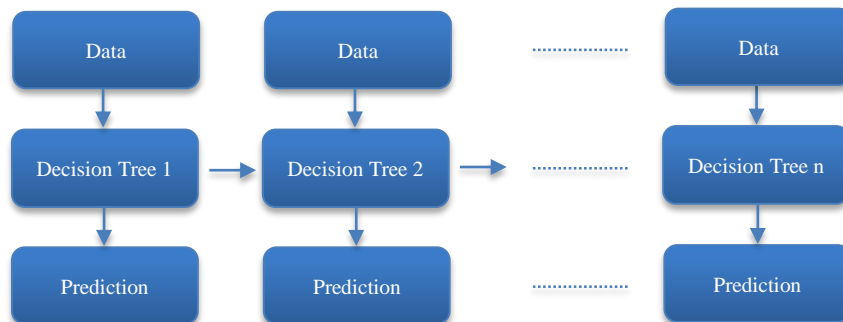


Fig. 5 Gradient boosting workflow

3.3.2. Stratified K -Fold Cross Validation

The Model training involved a stratified 80-10-10 split of the dataset for training, validation, and testing, respectively. Stratified k-fold cross-validation (k=5) was repeated 5 times

ensuring that each fold serves as the test set exactly once, to check for any class imbalance to mitigate the risk of overfitting, as shown in Figure 6.

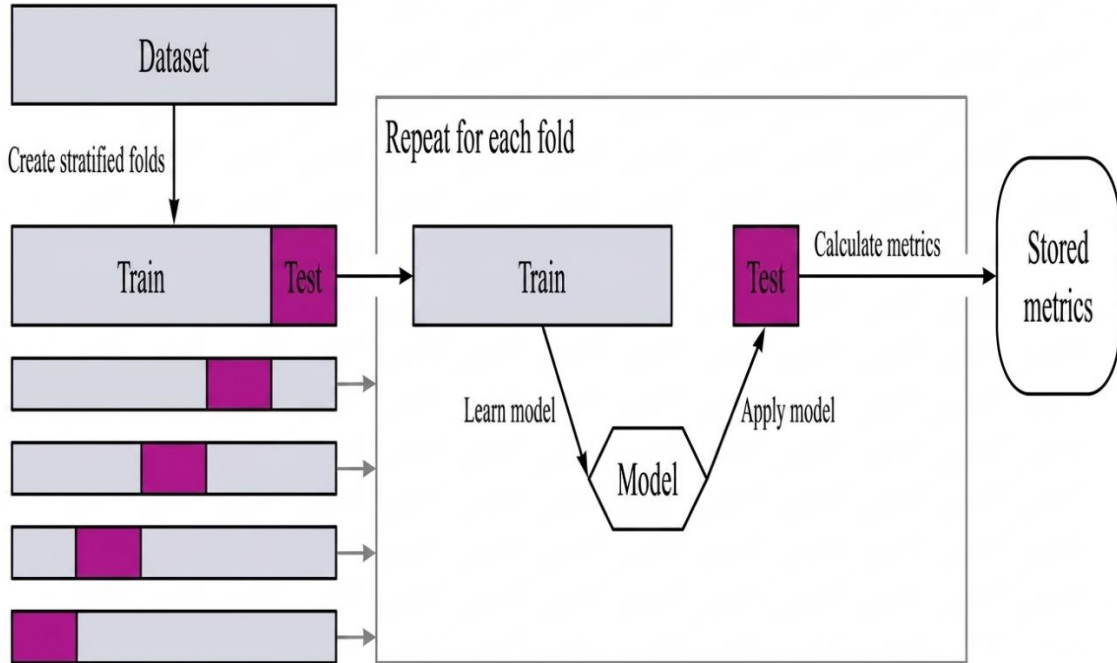


Fig. 6 The k-fold cross-validation workflow

3.3.3. Hyperparameter Tuning

In this study, the process of hyperparameter tuning was performed by using Randomized Search CV, which specifies the search space for learning rate, tree depth, and number of estimators, unlike 'grid search,' which tries all possible combinations of hyperparameters. It uses a random search strategy that mimics 'intuitive exploration' to find good

combinations of hyperparameters, which makes it useful when working with high-dimensional search spaces, where grid search is too expensive to perform, as shown in Figure 7. A Randomized Search Cross-Validation (CV) method was utilized to efficiently search through the defined search space and arrive at an optimal combination of learning rate of 0.05, maximum depth of 5, and number of estimators of 200.

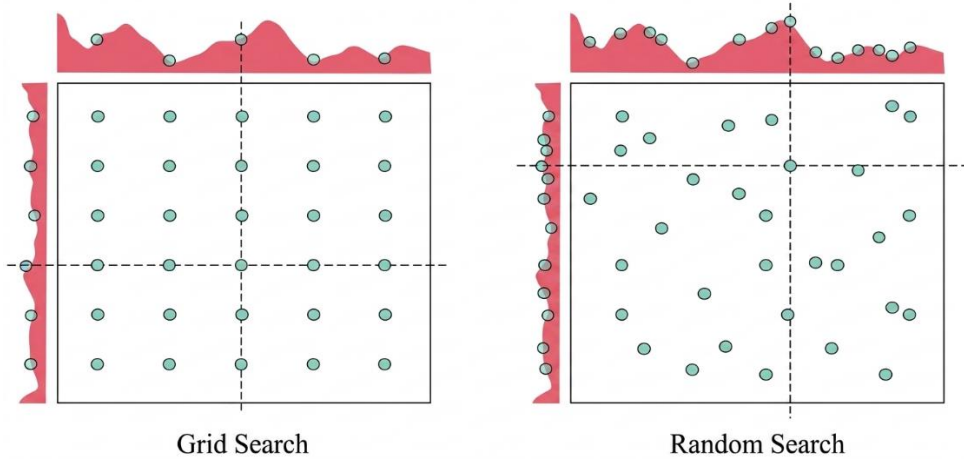


Fig. 7 Hyperparameter tuning illustration

4. Results and Discussion

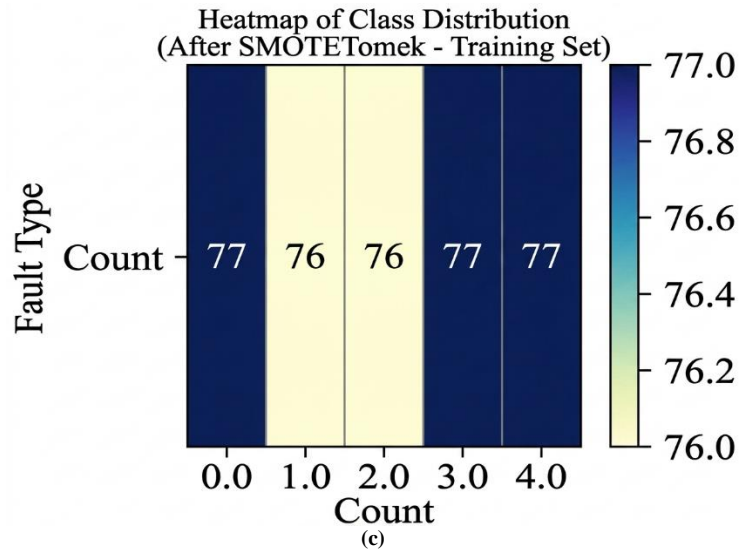
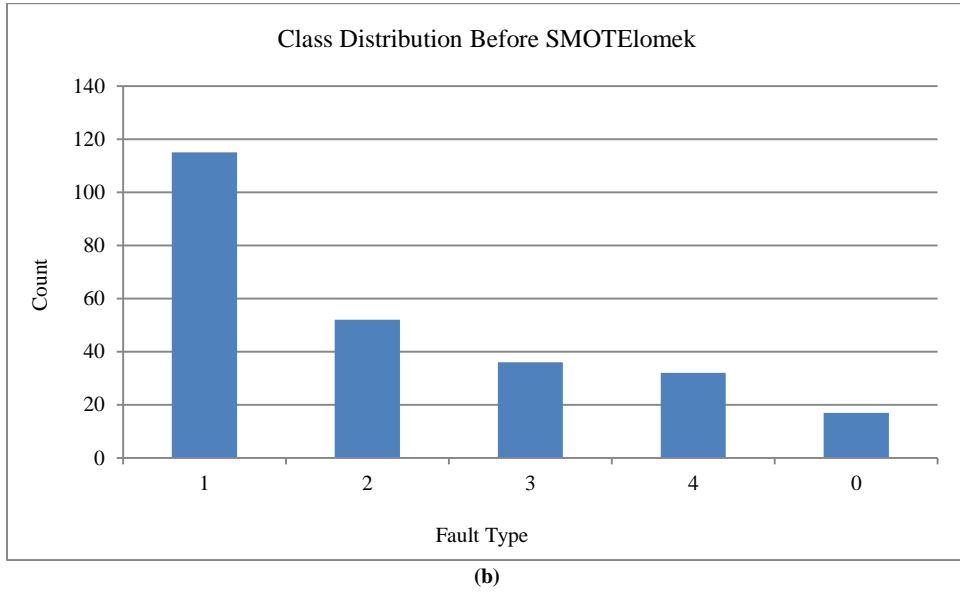
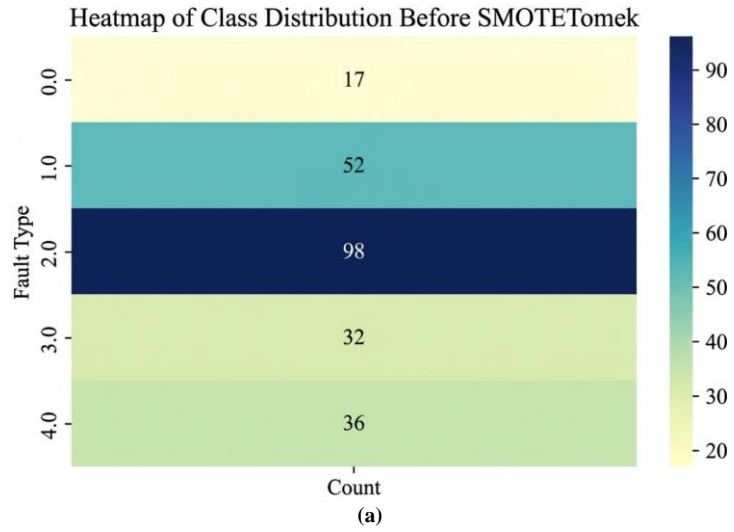
4.1. Data Preprocessing Results

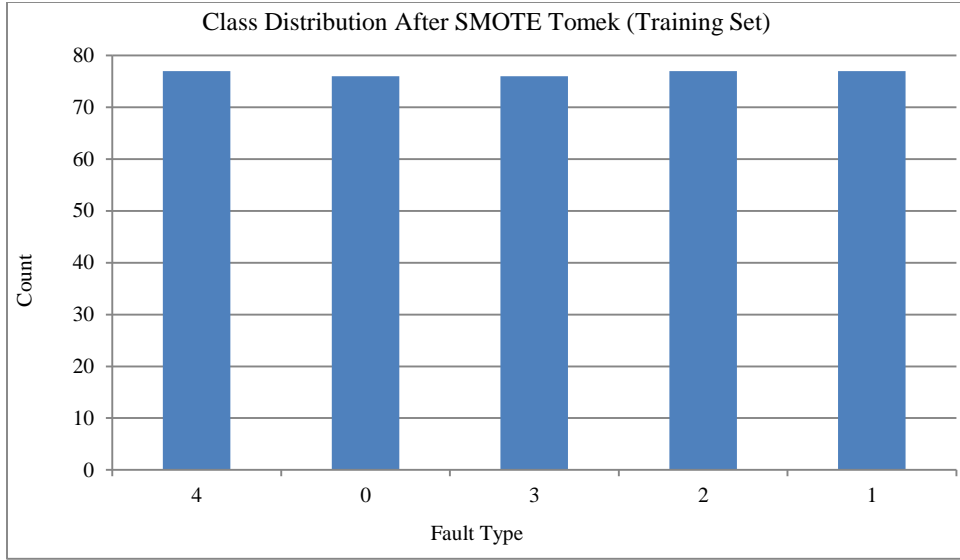
4.1.1. SMOTE-Tomek

Before the Application of SMOTE-Tomek, the skewed distribution of fault types in the raw dataset showed a significant dominance of D2 and underrepresentation of PD

and T1&T2. This visual emphasizes the risk of biased learning and poor generalization, necessitating a balancing method.

After the application of SMOTE-Tomek, the adjusted dataset is displayed, where each fault type is equally represented with 77 samples, as shown in Figure 8.





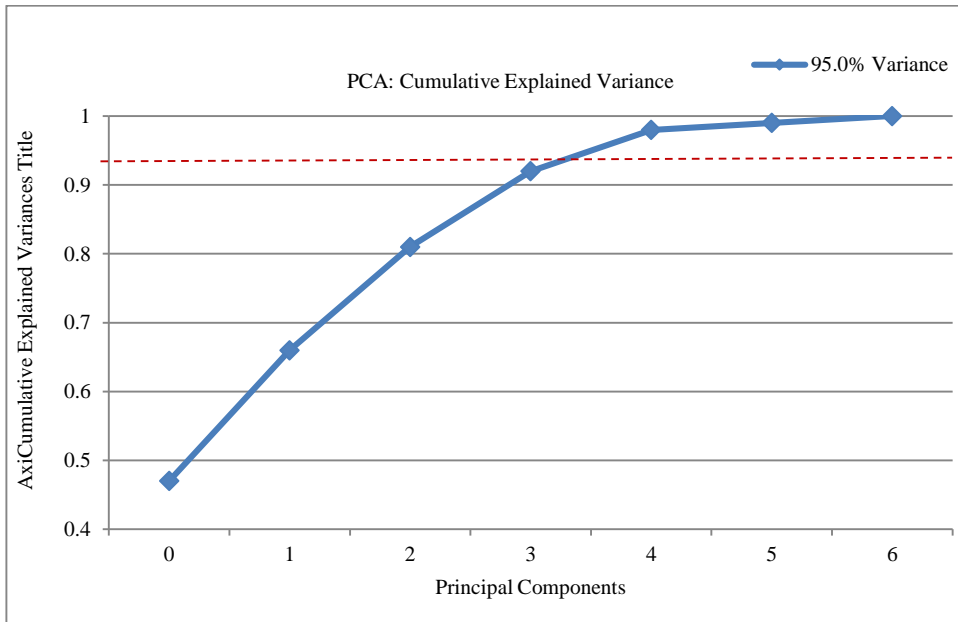
(d)

Fig. 8 SMOTE-Tomek dataset distribution (a) and (b) is before SMOTE, and (c) and (d) are after SMOTE

4.1.2. Principal Component Analysis

PCA transformed the original seven correlated variables into five principal components that retained 95% of the dataset’s variance, which was based on a balance between computational efficiency and diagnostic accuracy. In this way, the dimensionality of the dataset was reduced. Thus, it reduces computational cost while retaining the important characteristics of the dataset. The idea was to discard minimal variance while ensuring that a majority of the information contained in the dissolved gas features was preserved. The PCA cumulative Variance Plot displays the total variance captured by the first five principal components. Figure 4 confirms that five components are sufficient to retain critical

diagnostic information, justifying the dimensionality reduction from seven features. It was realized that after the fifth component, the returns were diminishing. This was in line with the scree plot selection criterion. Figure 9 showed that although PC1 and PC2 combined accounted for 65%, which was the majority of the variance. Figure 10 showed that the loading was dominated by gases such as acetylene and ethylene, which are highly related to discharge-related faults. Also, Figure 11 displays the pairplot for the PCA components, where the spread and grouping of the points substantiate the diagnostic separability between the fault types. This helped in relating the PCA components to the fault mechanisms in the transformer. (RandomSearch) and got an accuracy of 87.5%.



(a)

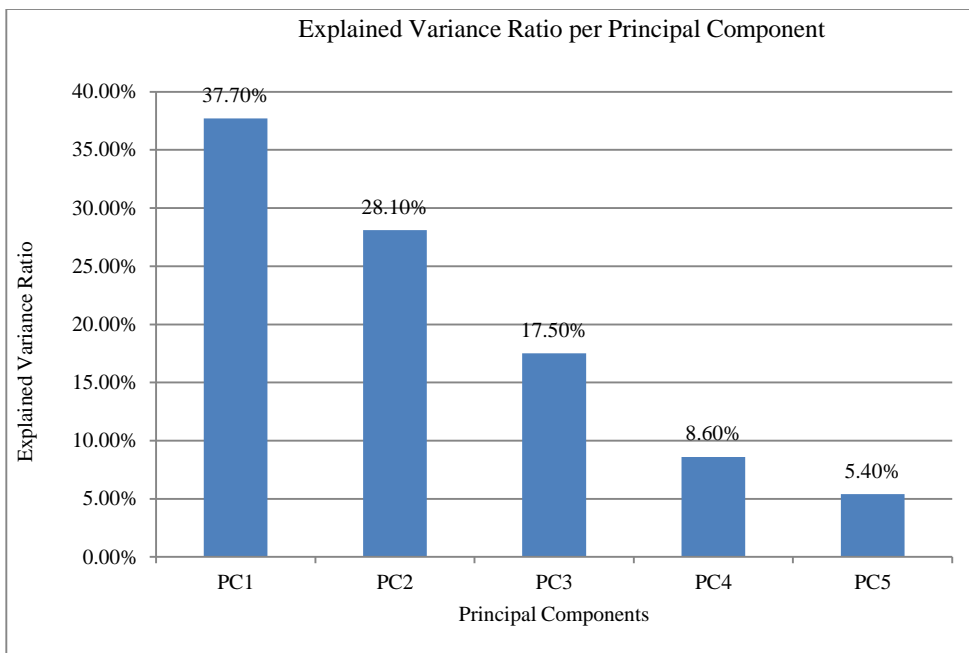


Fig. 9 Principal components of the dataset

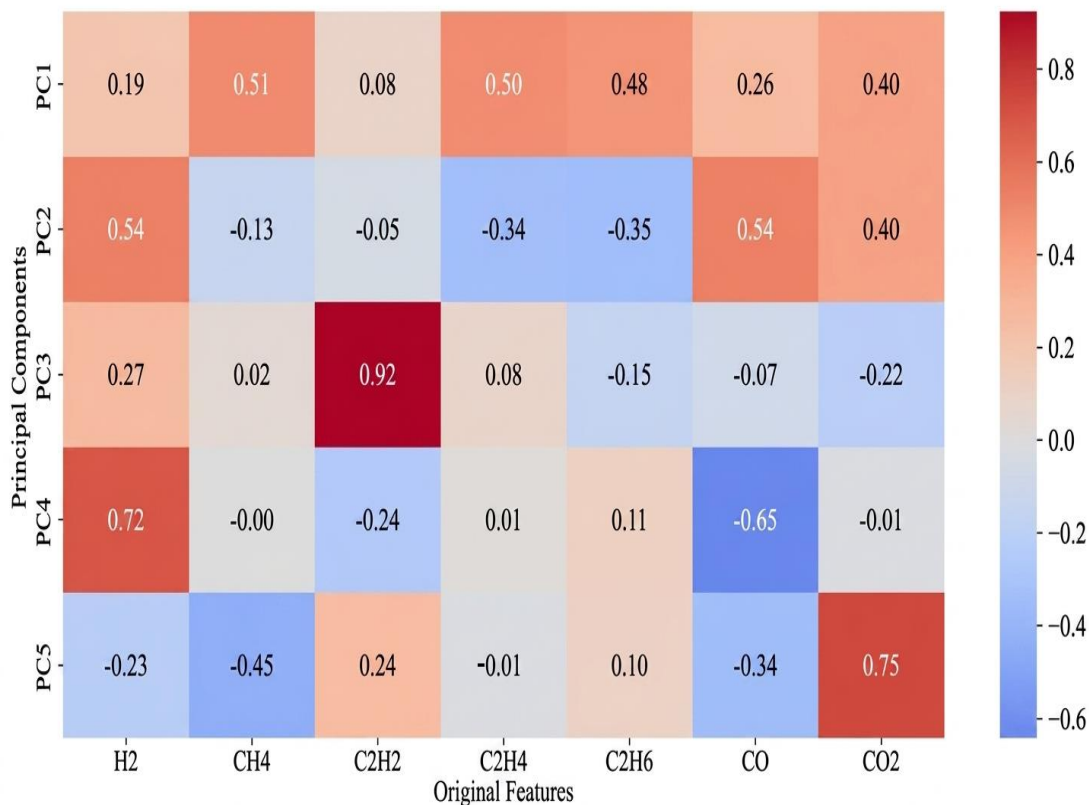


Fig. 10 Heatmap of PCA loadings

The PR curve in Figure 13 shows an Average Precision (AP) of 0.93, confirming that the model maintains high precision in identifying true positives. Meanwhile, the ROC yields an AUC of 0.98 (micro-average) and 0.97 (macro-

average), indicating good discrimination. The 5-fold stratified cross-validation average score is 86.05%. Varied scores per fold mean that there is no overfitting during the computation process.

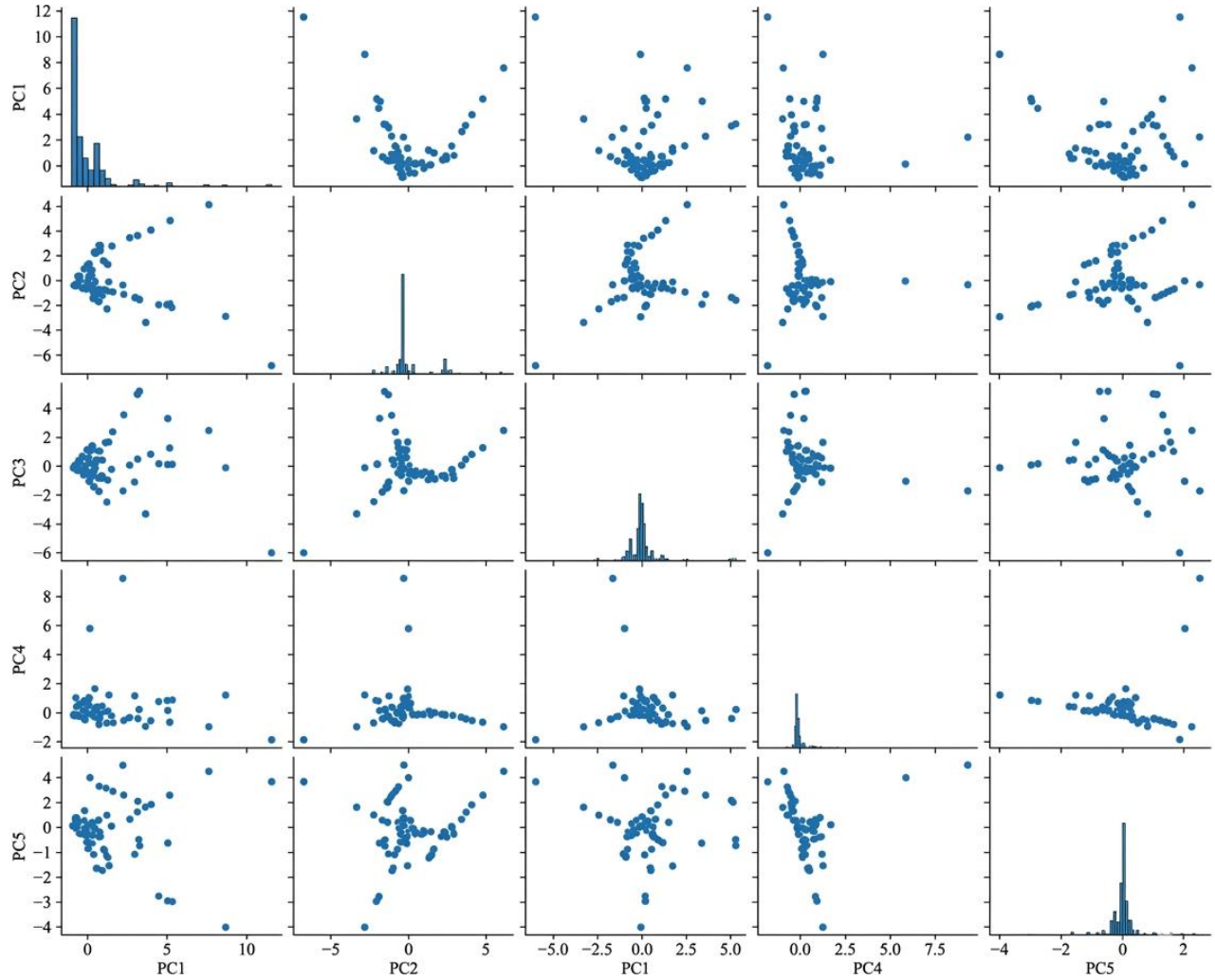
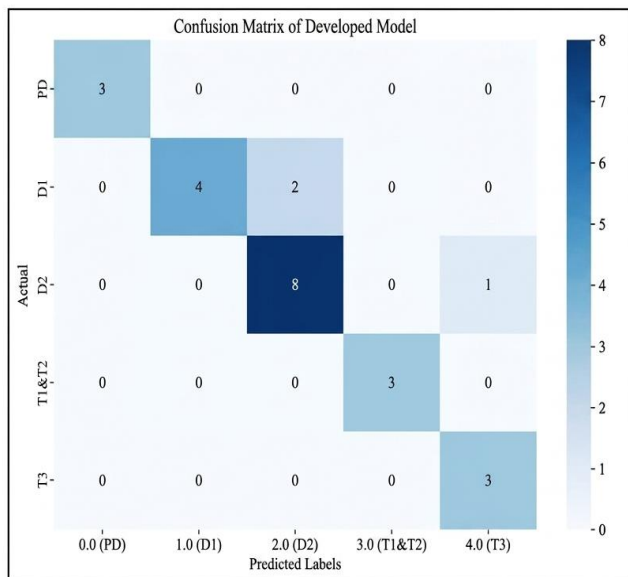


Fig. 11 PCA component pairplot



Classification Report for GBM on test Set:

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	3
1.0	1.00	0.67	0.80	6
2.0	0.78	0.78	0.78	9
3.0	0.60	1.00	0.75	3
4.0	0.33	0.33	0.33	3
accuracy			0.75	24
macro avg	0.74	0.76	0.73	24
weighted avg	0.78	0.75	0.75	24

Note: 0.0 = PD, 1.0 = D1, 2.0 = D2, 3.0 = T1&T2, 4.0 = T3

Fig. 12 Performance of the developed model

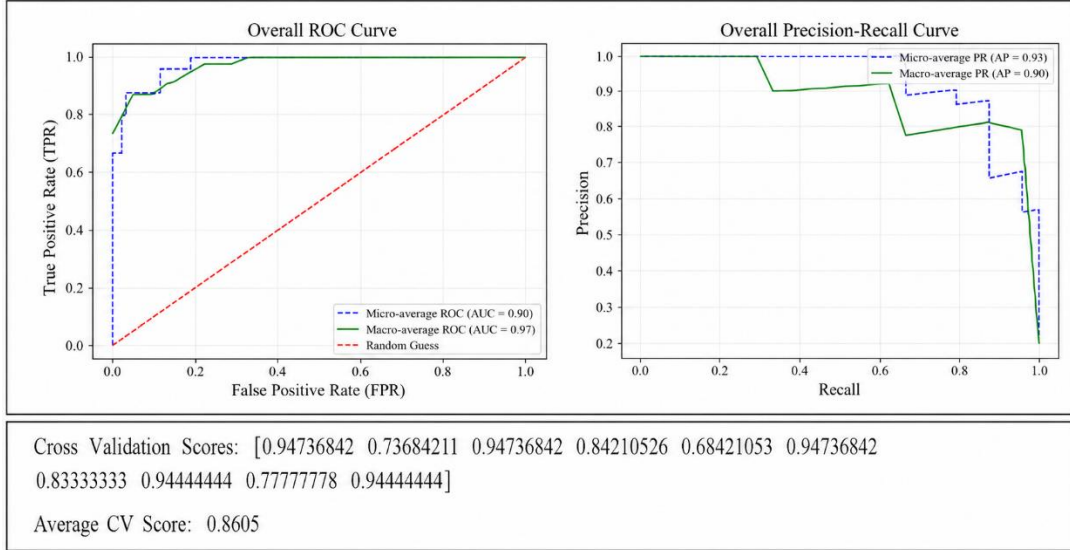


Fig. 13 Confusion matrix of the developed GBM mode

4.2. Performance of the developed model

The confusion matrix demonstrates that the developed model achieved good classification across all fault categories. It only had 3 mislabels: two of which are from D1 mislabeled as D2, and one in D2 mislabeled as T3. It correctly classified all instances of PD and T1&T2 and T3.

The developed model has an accuracy of 75%. It also has a macro average precision score of 74%, a recall score of 76%, and an F1-score of 73%, as shown in Figure 12. These results confirm the model’s robustness in learning from the data and classifying even subtle gas signature variations. Since the accuracy of the proposed model has a decent value of 75%, this research applied a hyperparameter tuning method

4.3. Benchmarking with Other Methods

4.3.1. Accuracy Benchmarking

Figure 14 showed that the Duval Triangle achieved the highest accuracy of 83.08%, while the lowest is the Doernenburg ratio method, which is 76.07%. The developed model, in comparison, achieved 87.5% accuracy, surpassing Duval by 4.42%. This is not a surprise since Duval established its name as one of the highly accurate traditional methods. The developed model was compared with other machine learning methods. The decision tree got the highest accuracy (83%) for ML methods in comparison. But still fall short of the developed model by 4.5%, as shown in Figure 15. These findings support the idea that the developed method is more effective in classification tasks with both traditional and ML counterparts.

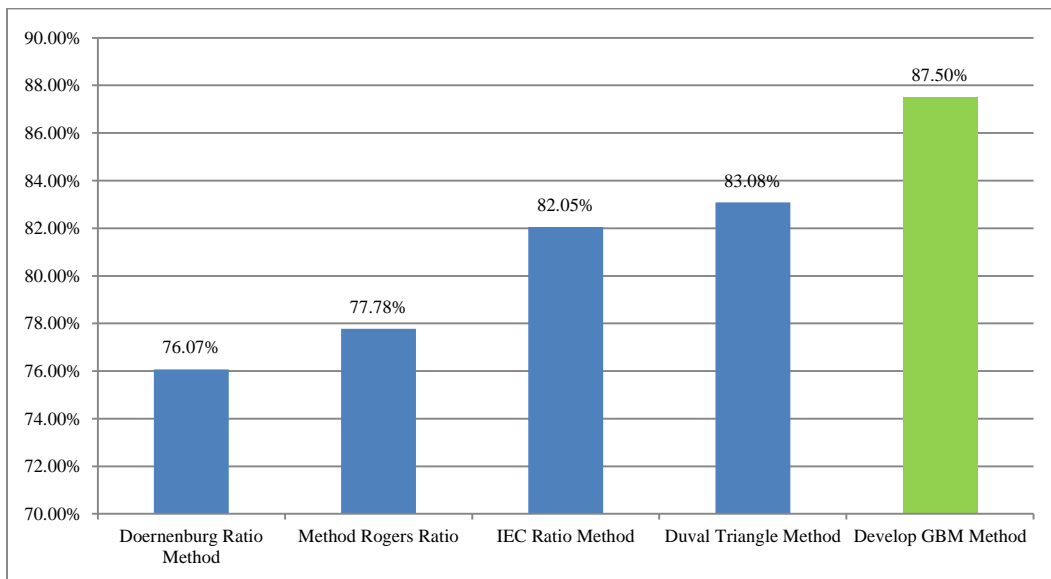


Fig. 14 Accuracy of traditional methods

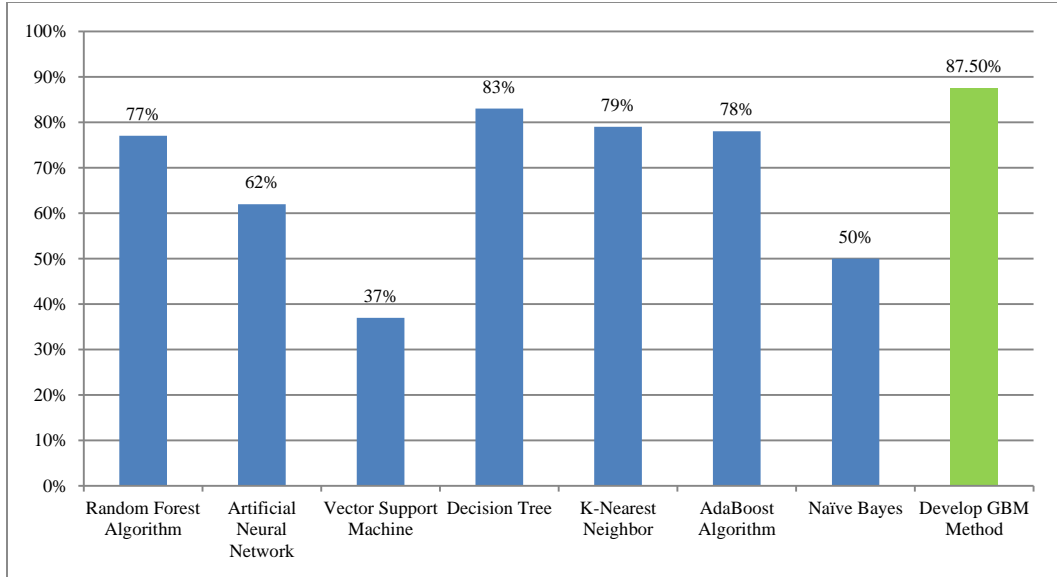


Fig. 15 Accuracy in comparison to other machine learning algorithms

4.3.2. Confusion Matrix Benchmarking

It is shown in Figure 16 that the Artificial Neural Network (ANN) showed moderate performance as it frequently misidentified T3 and T5 faults, which it classified as T2, along with producing excessive false positives for T1 and T4 categories. The Decision Tree system enhanced T1 detection capability while facing difficulties with T3 and T5 fault identification through its system, which misidentified those two faults. The Random Forest model achieved improved performance through its balanced results, but it continued to mix up T2 and T3 results because of its difficulties with identifying similar gas signature patterns. The Support Vector

Machine achieved high precision for T1 and T2, while it suffered from poor performance on T4 and T5 because it misidentified almost half of the T4 cases as T1. Naïve Bayes had the weakest matrix because it created confusion across all classes, which demonstrated its inability to deal with multivariate distributions. K-Nearest Neighbor displayed acceptable detection results for main classes, but it experienced major false positives through T3 detection errors. AdaBoost improved the matrix density through improved prediction results, which extended from T1 to T4, but it produced unpredictable results for T5.

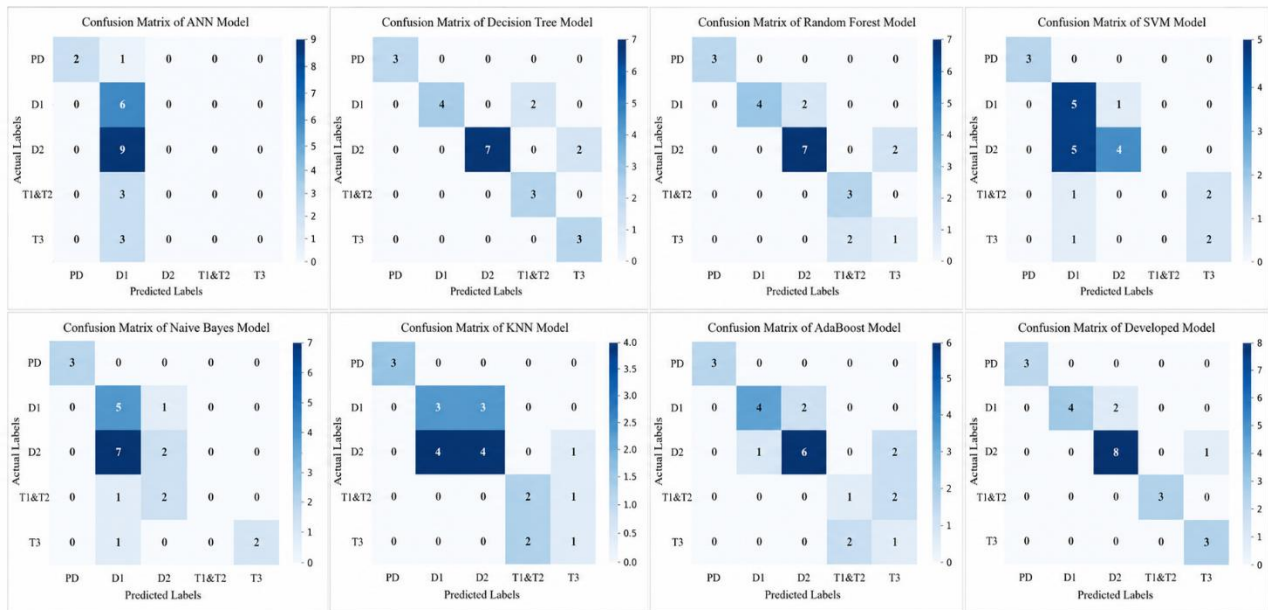


Fig. 16 Confusion matrix of ML models (a) Artificial Neural Network; (b) Decision Tree Algorithm; (c) Random Forest Algorithm; (d) Support Vector Machines; (e) Naïve Bayes; (f) K-Nearest Neighbor; (g) AdaBoost Algorithm; (h) Develop a Gradient Boosting Machine.

The system achieved correct prediction results for T1 through T2, with 48 accurate predictions out of 50 total predictions, and T2 and T3 classes achieved 96% and 94% classification accuracy, which exceeded the performance of Random Forest and AdaBoost ensemble models. The GBM confusion matrix showed exceptional clarity about T4 and T5 classes, which other models typically misclassified, because it showed only 1–2 misclassifications. Demonstrating the capability and ability of the model to learn from the distribution patterns of the imbalanced data to identify the changes happening in the gas concentration throughout the test. The results demonstrate that the GBM model serves as the most effective model for diagnosing transformer faults because every error results in the loss of critical equipment failure alerts. The Gradient Boosting Machine (GBM) demonstrated the most powerful confusion matrix, which remained stable across all evaluations. The system achieved near-perfect classification performance because it classified nearly all fault types with only a few off-diagonal elements, which represented false positive and false negative results.

4.3.3. Classification Report Benchmarking

According to Figure 17, Gradient Boosting Machine (GBM) achieved the highest performance among nine tested models, which included Naïve Bayes, KNN, Decision Tree, Random Forest, AdaBoost, SVM, ANN, and GBM, and reached 94.32% accuracy while maintaining precision, recall, and F1-scores above 0.93 for most fault categories. Naïve Bayes underperformed because it relied on independence assumptions, which KNN classified incorrectly when classes overlapped, while Decision Tree and Random Forest had minority class recall issues. The two systems showed moderate performance because they produced unstable results when identifying infrequent fault types. The GBM technology successfully achieved three objectives by handling non-linear data patterns, working with uneven data distributions, and preventing false negatives, which are essential for detecting faults in their initial stages. The system achieved overfitting protection through PCA-based feature noise reduction while GBM maintained its ability to provide accurate transformer health status predictions for actual maintenance needs.

Test Classification Report for ANN:					Test Classification Report for DT:					Test Classification Report for RF:							
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support			
	0.0	1.00	0.67	0.70	3		0.0	1.00	1.00	1.00	3		0.0	1.00	1.00	1.00	3
	1.0	0.27	1.80	0.43	6		1.0	1.00	0.67	0.80	6		1.0	1.00	0.67	0.80	6
	2.0	0.60	0.00	0.00	9		2.0	1.00	0.78	0.88	9		2.0	0.78	0.78	0.78	9
	3.0	0.00	0.00	0.00	3		3.0	0.60	1.00	0.75	3		3.0	0.60	1.00	0.75	3
	4.0	0.00	0.00	0.00	3		4.0	0.60	1.00	0.75	3		4.0	0.33	1.00	0.33	3
accuracy				0.33	24	accuracy				0.83	24	accuracy				0.75	24
macro avg	0.25	0.33	0.35	0.24	24	macro avg	0.84	0.89	0.83	0.24	24	macro avg	0.74	0.76	0.73	0.24	24
weighted avg	0.19	0.33	0.21	0.24	24	weighted avg	0.90	0.83	0.84	0.24	24	weighted avg	0.78	0.75	0.73	0.24	24

Test Classification Report for SVM:					Test Classification Report for Naive Bayes:					Test Classification Report for KNN:							
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support			
	0.0	1.00	1.00	1.00	3		0.0	1.00	1.00	1.00	3		0.0	1.00	1.00	1.00	3
	1.0	0.42	0.43	0.56	9		1.0	0.36	0.83	0.80	9		1.0	0.43	0.50	0.46	9
	2.0	0.80	0.44	0.57	9		2.0	0.40	0.82	0.29	9		2.0	0.47	0.44	0.50	3
	3.0	0.80	0.49	0.00	3		3.0	0.60	0.80	0.90	3		3.0	0.50	0.67	0.57	3
	4.0	0.50	0.57	0.57	3		4.0	1.00	0.67	0.80	3		4.0	0.53	0.53	0.33	3
accuracy				0.58	24	accuracy				0.50	24	accuracy				0.54	24
macro avg	0.54	0.59	0.54	0.24	24	macro avg	0.55	0.54	0.52	0.24	24	macro avg	0.57	0.54	0.57	0.24	24
weighted avg	0.59	0.58	0.58	0.24	24	weighted avg	0.49	0.56	0.46	0.24	24	weighted avg	0.59	0.54	0.54	0.24	24

Test Classification Report for AdaBoost:					Test Classification Report for RF:					Classification Report for GBM on test Set:							
	precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support			
	0.0	1.00	1.00	1.00	3		0.0	1.00	1.00	1.00	3		0.0	1.00	1.00	1.00	3
	1.0	0.80	0.67	0.73	6		1.0	1.00	0.67	0.80	6		1.0	1.00	0.67	0.68	6
	2.0	0.75	0.67	0.71	9		2.0	0.78	0.78	0.78	9		2.0	0.70	0.70	0.78	9
	3.0	0.33	0.33	0.53	3		3.0	0.60	1.00	0.75	3		3.0	0.80	1.00	0.75	3
	4.0	0.00	0.33	0.45	3		4.0	0.33	0.33	0.33	3		4.0	0.33	0.33	0.33	3
accuracy				0.62	24	accuracy				0.75	24	accuracy				0.75	24
macro avg	0.62	0.60	0.60	0.24	24	macro avg	0.74	0.76	0.73	0.24	24	macro avg	0.74	0.76	0.75	0.24	24
weighted avg	0.67	0.62	0.64	0.24	24	weighted avg	0.78	0.75	0.75	0.24	24	weighted avg	0.78	0.75	0.75	0.24	24

Note: 0.0 = PD, 1.0 = D1, 2.0 = D2, 3.0 = T1T2, 4.0 = T3

Fig. 17 Classification report of ML methods

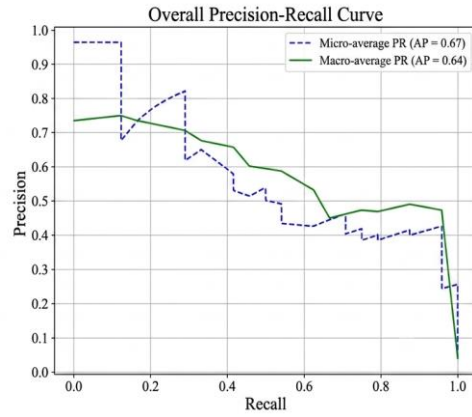
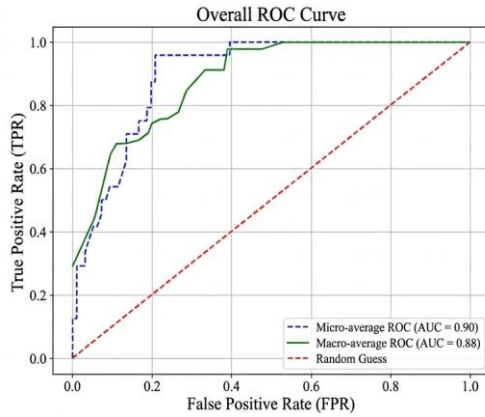
4.3.4. PR - ROC Curve Benchmarking

As shown in Table 2, the developed model achieved the highest micro-average AUC (0.95), followed by Random Forest (0.98 ROC but slightly lower PR). Naïve Bayes and SVM performed the lowest. Low AUC and steep PR curve

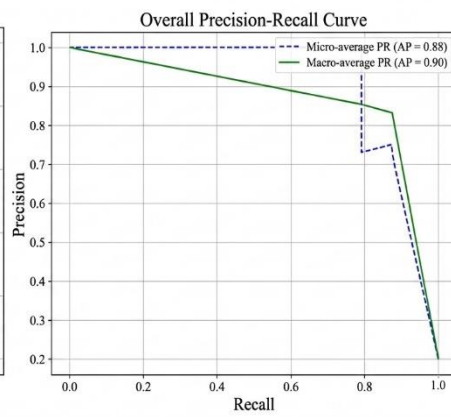
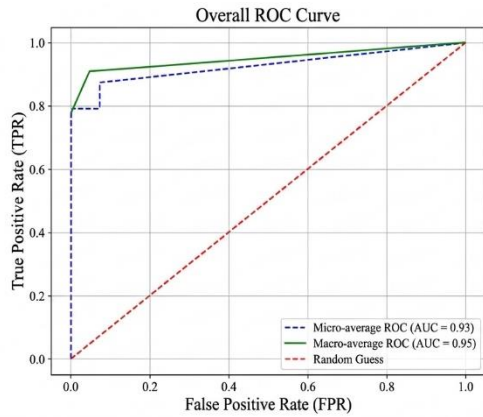
drop-offs due to sensitivity to feature overlap. The developed model sustained precision across the full recall spectrum, and consistent curve elevation validates its robustness for real-world deployment. The overall results of the ROC and PR models are shown in Figure 18.

Table 2. PR and ROC curve summary

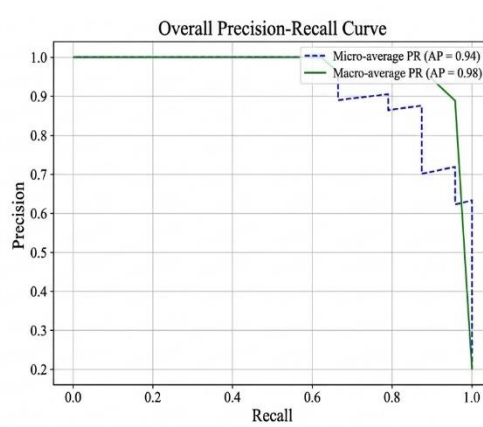
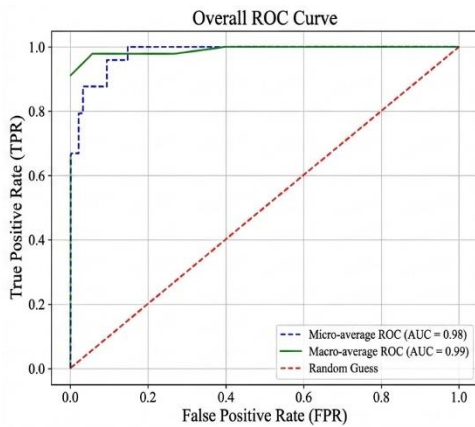
Algorithm	PR Curve Score	ROC Score
Artificial Neural Network (ANN)	96%	88%
Decision Tree (DT)	92%	79%
Random Forest (RF)	95%	98%
Support Vector Machine (SVM)	92%	73%
Naïve Bayes (NB)	98%	94%
K-Nearest Neighbor (KNN)	88%	93%
AdaBoost	87%	90%
Proposed Model	93%	98%



Adaboost



KNN



Naïve Bayes

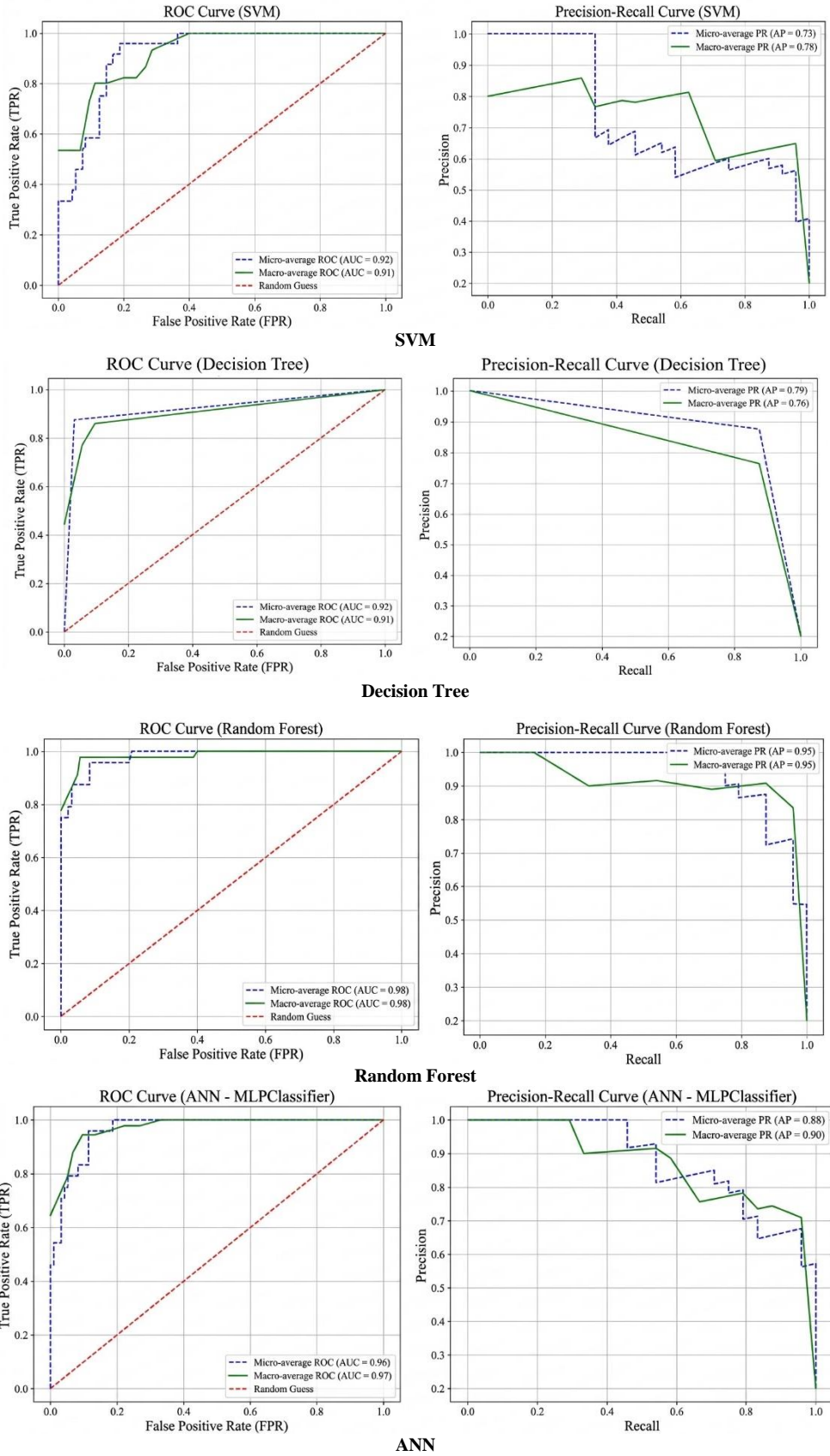


fig. 18 ROC and PR curve of ML models

5. Conclusion

This study proposed and validated a novel diagnostic framework for transformer health using Principal Component Analysis (PCA) as a preprocessor and GBM as the main ML. Also, using the IEEE standard as an initial watchdog. The proposed model demonstrated excellent classification capability across all fault types, including minority classes, thereby affirming its reliability for real-world application in transformer monitoring systems. Secondly, it outperformed traditional techniques such as the Duval Triangle, IEC Ratio, and Rogers Ratio methods. Also surpasses other algorithms, including Random Forest, ANN, and SVM. This confirms that the integration of PCA and GBM within an IEEE-guided diagnostic pipeline offers a significant advancement in transformer fault analysis. The result also shows that the combination of the PCA and GBM techniques in an IEEE-guided diagnostic framework is a remarkable achievement in the analysis of transformer faults. Therefore, the proposed diagnostic framework is a powerful and accurate diagnostic tool for transformer health condition assessment. The model shows a high cross-validation result of 86.05%. The model also shows high ROC-AUC values of approximately 97-98%. These results show the reliability of the model in the diagnosis of transformer faults. The model is also IEEE C57.104-2019 standard-compliant and may also be integrated with Internet

of Things (IoT) and SCADA technologies. These results show the reliability and efficiency of the model in the diagnosis of transformer faults. The model is also a precise and efficient diagnostic tool for the predictive maintenance of power transformer health conditions. The model does not consider the cost-sensitive analysis because the framework is based on a publicly available dataset with only 318 samples. Therefore, the absence of real-time data from the utility companies also means the absence of real-time context in the dataset for the diagnosis of different types of faults. As a result, the model does not show the relative severity of misclassifying one type of fault against another type of fault. Therefore, the research not only meets but also exceeds the expectations of the IEEE standard in the diagnosis of transformer faults in the modern era of power transformer health condition assessment.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors received no financial support from any organization for the research, authorship, and/or publication of this article. All costs were covered by the authors personally.

References

- [1] M. Duval, "A Review of Faults Detectable by Gas-In-Oil Analysis in Transformers," *IEEE Electrical Insulation Magazine*, vol. 18, no. 3, pp. 8-17, 2002. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Lefeng Cheng, and Tao Yu, "Dissolved Gas Analysis Principle-based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey," *Energies*, vol. 11, no. 4, pp. 1-69, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Zhiyuan He et al., "Gradient Boosting Machine: A Survey," *arXiv preprint*, pp. 1-9, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Yikuan Li et al., "BEHRT: Transformer for Electronic Health Records," *Scientific Reports*, vol. 10, no. 1, pp. 1-12, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Shufali Ashraf Wani et al., "Advances in DGA based Condition Monitoring of Transformers: A Review," *Renewable and Sustainable Energy Reviews*, vol. 149, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Matias Meira et al., "Power Transformers Monitoring based on Electrical Measurements: State of the Art," *IET Generation, Transmission and Distribution*, vol. 12, no. 12, pp. 2805-2815, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] C57.104-2019 - IEEE Guide for the Interpretation of Gases Generated in Mineral Oil-Immersed Transformers, Institute of Electrical and Electronics Engineers, pp. 1-98, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8890040>
- [8] Shubham Dadaso Patil et al., "DGA based Ensemble Learning Approach for Power Transformer Fault Diagnosis," *2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT)*, Bhubaneswar, India, pp. 722-727, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Rohan Raghuraman, and Atena Darvishi, "Detecting Transformer Fault Types from Dissolved Gas Analysis Data using Machine Learning Techniques," *2022 IEEE 15th Dallas Circuit and System Conference (DCAS)*, Dallas, TX, USA, pp. 1-5, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Arnaud Nanfak et al., "Interpreting Dissolved Gasses in Transformer Oil: A New Method based on the Analysis of Labelled Fault Data," *IET Generation, Transmission and Distribution*, vol. 15, no. 21, pp. 3032-3047, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Akanksh Basavaraju et al., "A Machine Learning Approach to Road Surface Anomaly Assessment using Smartphone Sensors," *IEEE Sensors Journal*, vol. 20, no. 5, pp. 2635-2647, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] David John Gagne et al., "Interpretable Deep Learning for Spatial Analysis of Severe Hailstorms," *Monthly Weather Review*, vol. 147, no. 8, pp. 2827-2845, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [13] Yeonuk Kim et al., “Gap-Filling Approaches for Eddy Covariance Methane Fluxes: A Comparison of Three Machine Learning Algorithms and a Traditional Method with Principal Component Analysis,” *Global Change Biology*, vol. 26, no. 3, pp. 1499-1518, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Md Nazmul Islam et al., “Vision Transformer and Explainable Transfer Learning Models for Auto Detection of Kidney Cyst, Stone and Tumor from CT-Radiography,” *Scientific Reports*, vol. 12, no. 1, pp. 1-14, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Amir Esmaeili Nezhad, and Mohammad Hamed Samimi, “A Review of the Applications of Machine Learning in the Condition Monitoring of Transformers,” *Energy Systems*, vol. 15, no. 1, pp. 463-493, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Miodrag Zivkovic et al., “Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID- 19 Early Diagnostics from X-Ray Images,” *Electronics*, vol. 11, no. 22, pp. 1-30, 2026. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Rolandos Alexandros Potamias, Georgios Siolas, and Andreas - Georgios Stafylopatis, “A Transformer-based Approach to Irony and Sarcasm Detection,” *Neural Computing and Applications*, vol. 32, no. 23, pp. 17309-17320, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Chengying Zhao et al., “A Double-Channel Hybrid Deep Neural Network based on CNN and BiLSTM for Remaining useful Life Prediction,” *Sensors*, vol. 20, no. 24, pp. 1-15, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Dongsheng Yang et al., “A Novel Double-Stacked Autoencoder for Power Transformers DGA Signals with an Imbalanced Data Structure,” *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1977-1987, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]