Transformer Maintenance Strategies: A K-Means Based Approach for 33 kV DTs

Kittisak Chaisuwan[1], Paradon Boonmeeruk[2] and Kiattisak Wongsopanakul[2,*]

[1] Provincial Electricity Authority, Chatuchak, Bangkok, 10900, Thailand
[2]Department of Electrical & Biomedical Engineering, Prince of Songkla University, Hatyai, Songkhla, 90110, Thailand.

*Corresponding Email : kiattisak.w@psu.ac.th

Received August 2, 2023, Revised November 20, 2024, Accepted December 15, 2024, Published June 25, 2025

Abstract. *The distribution transformer (DT) is crucial for connecting utility providers to consumers, and its failure can disrupt the distribution network's reliability. The Provincial Electricity Authority (PEA) in Thailand manages a large number of transformers, necessitating efficient maintenance planning to prevent DT failures. This paper introduces a method for classifying the condition of 33 kV DTs without pre-existing cluster data, utilizing the K-means clustering algorithm on data from 150 samples. The dataset includes 7 features from DT annual maintenance records and the Geographic Information System (GIS) of PEA Southern Area 3. Key factors identified are insulation between high voltage and ground, high-low voltage, and low voltage-ground. The method categorizes DT conditions into three clusters: "poor," requiring urgent action; "risk," requiring close monitoring; and "normal," requiring routine maintenance. Validation with K-Nearest Neighbors yields an accuracy of 96.67%, demonstrating the effectiveness of the proposed classification method.*

Keywords:

Distribution transformer (DT) condition assessment, preventive maintenance planning, K-means clustering algorithm, transformer insulation analysis, Provincial Electricity Authority (PEA)

1. Introduction

The distribution transformer (DT) serves as the primary link between utility providers and consumers, and its failure significantly impacts the reliability of the distribution network. Monitoring the health of transformers for external and internal factors can extend their operational lifespan and maintain network stability [10]. Short-circuits (SC) in transformer windings can lead to very high temperatures, resulting in conductor melting failures [11]. Additionally, factors such as high temperatures, overloads, unbalanced loads, insulation issues, and transformer inrush currents contribute to DT degradation [12].

Preventive maintenance plays a crucial role in transformer upkeep, aiming to minimize DT failures and facilitate effective maintenance planning. This requires tools, methods, or expert knowledge for condition classification. The Provincial Electricity Authority (PEA), Thailand's largest electric power service provider, manages the most extensive network of DT installations. Thus, it is essential to classify DT conditions to enhance reliability and reduce failures. Traditionally, condition classification focused on power transformers in substations, while DTs typically underwent only preliminary checks, often insufficient for accurate condition assessment, leading to frequent trans-former breakdowns.

Efficient and uninterrupted electricity distribution within a network requires the effective classification of abnormal events in power transformers for maintenance planning. Previous research has highlighted the application of Machine Learning (ML) models and mathematical approaches for analyzing, predicting, and classifying transformer abnormalities to optimize maintenance schedules, thereby enhancing electricity distribution efficiency. For instance, M. Abdillah et al. [2] introduced a Prognostics Health Management (PHM) system utilizing a kernel extreme learning machine (K-ELM) to evaluate the health of power transformers, employing two distinct datasets. The first dataset (Set-1) pertains to parameters assessing transformer efficiency, while the second dataset (Set-2) focuses on overall health assessment. The PHM system demonstrated an assessment accuracy of 68.67% with Dataset-1 and 93.61% with Dataset-2. However, the system's reliance on K-ELM necessitates a substantial amount of comprehensive data for model training, and the limited scope of the datasets used for testing constrains the ability to fully assess the health of power transformers. A. J. Patil et al. [5] proposed a computational algorithm to develop a Fuzzy Logic model aimed at reducing dependence on experts and enhancing accuracy in transformer health assessment. The development process of the Fuzzy Logic model comprises five steps: (1) Fuzzy Logic for Condition Indicator Score (CIS), utilized for oil turbidity analysis, power factor analysis, efficiency, and life expectancy analysis of transformers; (2) Computation of Primitive Health Index (PHI), which involves general health analysis of the transformer and decision-making conditions of the Fuzzy Logic model; (3) Testing the decision-making conditions of the model; (4) Adjustment of PHI, where decision-making conditions are refined post-testing; and (5) Final Health Index (HI), determining the transformer's health status. The HI categorizes transformer health into three groups: Good (HI range: 8-10), Fair (HI range: 4-7), and Poor (HI range: 0-3). The limitations of this model include the necessity for expert involvement in the initial step, and the accuracy of the health assessment heavily relies on the precision of the input data. Furthermore, the model is not suitable for real-time monitoring analysis. D. Granados-Lieberman et al. [1] present a system for real-time detection of short-circuits in power trans-formers using a Harmonic Phasor Measurement Unit (HPMU) and a Fuzzy Logic model. The operation of the short-circuit detection system is divided into five steps: (1) Data is read from the HPMU, an extension of the PMU, which includes magnitude, phase, and frequency information; (2) Real-time harmonic data measured in the first step is sampled to prepare for classification by the Fuzzy Logic model; (3) The sampled harmonic signal data is classified to detect short-circuit abnormalities in the power transformer; (4) Identification of short-circuit abnormalities within the power transformer after classification by the Fuzzy Logic model; and (5) Validation of the event analysis performed by the Fuzzy Logic model. The system's test results show it can detect minor harmonic signal changes indicating short-circuit abnormalities more effectively than traditional methods and supports real-time monitoring of such abnormalities. Limitations include the dependency of the Fuzzy Logic model's accuracy on the quantity of data used for its decision-making and the complexity of developing or customizing the system parameters to fit the model. X. Zhao et al. [3] proposed a system for real-time analysis of irreversible deformations of power transformer internal components caused by mechanical and electromotive forces (winding deformations). The analysis method utilizes V-I Lissajous Patterns, which describe the relationship between the voltage drop across the transformer terminals and the current flowing through them. These characteristics enable the identification of abnormal winding deformations within the transformer. The proposed system demonstrates high sensitivity to minor winding deformations and offers high repeatability. However, it requires high-precision data for accurate internal fault analysis. Additionally, the system's effectiveness is limited by the complexity of model development, as the accuracy of fault analysis depends on the fine-tuning of the V-I Lissajous Patterns model, a process that is both complicated and time-consuming. W. Wattakapaiboon et al. [8] presented the Health Index (HI) method for evaluating maintenance planning conditions of transformers in the industrial sector and power distribution systems. This method was developed to reduce the number of parameters used for testing and maintenance planning. The results indicate that the HI method can reduce the evaluation parameters from 24 to 15, with a deviation in evaluation scores of approximately 7% compared to the original method. Additionally, the method is user-friendly for field applications. However, by reducing the number of evaluation parameters, the method may not cover certain conditions and has limited verification of evaluation accuracy.

This study introduces a method for classifying the condition of 33 kV distribution transformers (DTs) without pre-existing cluster data, employing the K-means clustering algorithm, a widely-used technique in machine learning. The K-means algorithm is favored for its ease of application, customization, and efficiency, particularly in scenarios involving large datasets. It also requires less processing time compared to other clustering algorithms [15], [16]. The study is based on 150 samples and 7 features extracted from the annual maintenance records of DTs in PEA, Southern Area 3. Geographic Information System (GIS) software, which facilitates data management, analysis, and visualization in diagrammatic or three-dimensional formats, is utilized to assist in maintenance planning and enhance operator comprehension [17]. The primary objective is to optimize maintenance planning and reduce DT failures.

2. Comparison of General Methods for Fault Classification in Transformers

Table 1 presents a comparison of methods used for fault classification in power transformers. Previous studies have primarily focused on assessing transformer conditions or detecting faults using chemical tests such as Dissolved Gas Analysis (DGA), Furan testing, breakdown voltage (B.V.), and oil quality. While these methods are accurate and widely adopted in the industry, they involve high costs for sample collection and testing. On the other hand, electrical tests, including V-I characteristics and insulation tests, offer real-time data crucial for transformer failure analysis but require extensive datasets for algorithm development and real-time signal detection. Much of the existing research has been dedicated to the deterioration assessment of power transformers, often through costly chemical testing or real-time data collection using signal detectors. These methods may not be cost-effective for routine transformer condition assessment. Therefore, this study explores an alternative approach to assessing or classifying distribution trans-formers using annual test results, incorporating both chemical and electrical tests. This approach is intended to provide a more cost-effective solution for maintenance planning by offering a comprehensive method for trans-former condition assessment.

Table 1 Comparison of General Methods for Fault Classification in Transformers

Author	Contributions	Methods	Results	Advantages	Limitations
A. J. Patil et al. [5]	- Fuzzified approach for Health Index (HI) determination in transformer diagnostic tests. - Incorporates decision logic for Tier-2 test selection and Primitive Health Index (PHI) adjustment.	- Fuzzy logic applied for Condition Indicator Score (CIS) computation and Health Index (HI) determination. - Decision logic utilized for Tier-2 test selection based on Tier-1 results. - Distinct fuzzy models developed for each Diagnostic Test (DT) and Se-verity Indicator (SI) in HI computation.	- Fuzzy model enhances Health Index (HI) computation using diagnostic test results. - Decision logic selects non-routine tests based on routine test outcomes. - Fuzzy model accounts for maintenance history and expected lifespan in HI computation.	- Fuzzy model reduces reliance on diagnostic experts. - Fuzzy logic effectively manages uncertainty in condition indicators.	- Traditional tech-niques are limited in computing health indices. - Fuzzy models eli-minate the reliance on diagnostic ex-perts for assess-ments.
C.-T. Lee [18]	- Developed FLCDT to classify abnormal defects in castresin transformers. - FLCDT outperforms CART and See5 in classification precision.	- Ultrasound, acoustic emission, UHF antenna, electrical contact, optical, and radio frequency sensing. - Fuzzy Logic Clustering Decision Tree (FLCDT). - See5 and CART software packages for classification comparison.	- FLCDT outperformed CART and See5 in classification precision. - FLCDT compared with See5 and CART software packages.	- Achieves higher classification accuracy with lower computational complexity. - Integrates hierarchical clustering with decision tree for enhanced classification results. - Outperforms CART and See5 in classification precision.	- Relies on expert judgments for partial dis-charge (PD) classification and defect level deter-mination. - Challenges in deter-mining the optimal composition level for wavelet analysis. - Clusters of PD patterns close in fractal map may lead to misidentification.
M. Abdillah et al. [2]	- K-ELM method applied for power transformer health assessment in engineering field.	- Kernel Extreme Learning Machine (K-ELM) - Support Vector Machine (SVM) - Least-Squares Support Vector Machine (LS-SVM)	- Proposed PHM system using K-ELM achieved 100% accuracy in transformer health assessment. - Testing phase accuracy for K-ELM was 68.67%. - Set-1 dataset yielded 68.67% accuracy in transformer health assessment. - Set-2 dataset achieved 93.61% accuracy in transformer health assessment.	- K-ELM PHM system outperforms SVM, LS-SVM, and ELM in accuracy. - Faster learning algorithm with superior generalization and accuracy rates. - Achieves 100% health assessment accuracy for power transformers.	- Limited comparison with other machine learning models for transformer health assessment.

Table 1 Comparison of General Methods for Fault Classification in Transformers (continue)

Author	Contributions	Methods	Results	Advantages	Limitations
X. Zhao et al. [3]	- New algorithm developed to eliminate load dynamics' effect on transformer V-I characteristics. - Sensitivity and reliability of the V-I technique assessed through practical measurements. - Assessment conducted on the effect of substation noise. - Data acquisition system developed for capturing V-I characteristics. - Hardware developed for real-time monitoring of transformer winding conditions.	- V-I Lissajous pattern method utilized for transformer fault detection. - Compensation technique proposed to eliminate load influence on the measured ellipse.	- V-I Lissajous pattern method validated through practical measurements. - New compensation approach proposed to eliminate transformer loading variation effects. - Lissajous graphical method validity verified in a practical environment.	- Real-time detection of winding deformations in power transformers. - Feasibility of V-I Lissajous pattern method for practical applications. - Compensation approach effectively eliminates loading effects on measured ellipses. - Perfect alignment of ellipses across varying load magnitudes.	- Load dynamics influence V-I characteristics. - Digital filtering necessary due to high background noise. - Impact of load and voltage imbalance on measured ellipses.
M. Raichura et al. [20]	- CNN-XGBoost proposed for transformer fault classification with high accuracy. - Performance comparison with RVM and HE-ELM techniques.	- CNN-XGBoost integration for classification. - Relevance Vector Machine (RVM) and hierarchical ensemble of ELM.	- Proposed technique achieves 99.95% accuracy in classifying internal faults. - XGBoost classifier demonstrates high precision in distinguishing transformer operational conditions. - CNN-XGBoost method attains 100% accuracy in identifying magnetizing inrush conditions.	- Achieves 99.95% classification accuracy within 22 ms. - Efficient feature extraction using 1D CNN and XGBoost combination. - Validated through real-time hardware for practical implementation.	- RVM method requires longer computational time with larger datasets. - CNN-XGBoost technique not compared with SVM for classification accuracy.
A. S. Mogos [19]	- Hybrid one-class deep SVDD method for predicting distribution transformer failures. - Employs SMOTE for data preprocessing and mRMR for feature selection. - Reduces costs using historical maintenance data and risk index.	- Introduction of a novel hybrid one-class classification method. - Utilization of mRMR for enhancing classification accuracy. - Application of Synthetic Minority Oversampling Technique (SMOTE) to address data imbalance.	- mRMR algorithm selects 13 important features for classification. - Exclusion of low-impact features enhances model accuracy and performance metrics. - Simulation results demonstrate improved recall, F1, and F2 scores with mRMR. - Analysis conducted using real-world data from 15,066 transformers in Colombia.	- Applicable to health assessment of various power system equipment. - Utilizes feature importance scores in the selection process. - Incorporates mutual feature information for both continuous and discrete variables.	- Less important features impact classification accuracy, necessitating mRMR feature selection. - Synthetic data is commonly used in research but may not accurately represent real-world scenarios.

Table 1 Comparison of General Methods for Fault Classification in Transformers (continue)

Author	Contributions	Methods	Results	Advantages	Limitations
R. M. A. Velásquez et al. [4]	- Mathematical modeling and optimization techniques discussed.	- Mathematical modeling techniques. - Computational simulations conducted. - Statistical analysis performed. - Data visualization methods applied.	- Novel algorithm for image segmentation presented. - Superior accuracy and efficiency compared to existing methods.	- Gene identification linked to therapeutic resistance. - Model training enhanced using OpenVC dataset.	- Limited data availability. - Potential bias in analysis. - Small sample size constraints.
W. Wattakapaiboon et al. [8]	- New HI table for transformer condition assessment featuring simplified parameters. - Visual inspection rating incorporated into the HI table for enhanced clarity.	- Dissolved Gas Analysis (DGA) employed for detecting incipient faults in transformers. - Health Index (HI) method used for transformer condition assessment and maintenance planning. - Visual inspections, oil quality tests, and power factor tests are also utilized.	- 7% difference in scores between the conventional and new HI tables. - Similar test result interpretation observed in most cases.	- New HI table requires fewer testing parameters for transformer condition assessment. - Reduces operational costs and is practical for transformer maintenance planning. - Employs simplified testing parameters for evaluating transformer conditions.	- Furanic compound testing was discontinued for the new HI table. - Ambiguous parameters in the conventional HI table were consolidated into visual inspec-tion.

3. Methodology

The data used in the condition classification analysis comprises numerical values and includes 150 samples with 7 features derived from the annual maintenance results of distribution transformers (DT) and the Geographic Information System (GIS) of PEA, Southern Area 3. This study employs recursive feature elimination (RFE) for feature selection and a K-means clustering algorithm for feature classification. A detailed explanation of all features is provided in Table 2. The visual inspection feature (F3) pertains to the external inspection of DTs conducted with the naked eye, encompassing 7 specific items outlined in Table 3. The input data, initially unclustered, is classified based on the 7 features using the K-means clustering algorithm, as illustrated in Fig. 1. From the dataset of 150 samples and 7 features, correlation analysis is performed using the correlation equation, a widely used technique to identify relationships between features, offering a highly general and flexible data analytic approach [7]. The Pearson correlation equation is particularly popular for this purpose. This study employs all features to analyze the data for relationship values, as shown in equation (1).

$$\rho_{(x_i y_i)} = \frac{\sum_{i=0}^{n} (x_i - \overline{x_i})(y_i - \overline{y}_i)}{\sqrt{\sum_{i=0}^{n} i(x_i - \overline{x_i})^2} - \sqrt{\sum i(y_i - \overline{y}_i)^2}} \tag{1}$$

Where $\rho_{(x_i y_i)}$ is Correlation between feature x_i and y_i

The significance of each feature can be determined using Recursive Feature Elimination (RFE). RFE is a widely used machine learning technique for feature selection due to its ease of configuration and effectiveness. It identifies the most or least relevant features in a training dataset for predicting the target variable. The method ranks features by evaluating the impact of removing one feature at a time on the objective function. The iterative procedure of RFE involves the following steps [14][13]:

1. Train the classifier (optimize the weights w_i with respect to J).
2. Compute the ranking criterion for all features ($DJ(i)$ or $(w_i)^2$).
3. Remove the feature with the lowest ranking criterion.

$$w_i = (\mu_i(+) - \mu_i(-)) / (\sigma_i(+) + \sigma_i(-)) \tag{2}$$

Where μ_i and σ_i are the mean and standard deviation of the gene expression values of gene i for all the patients of class (+) or class (−), i = 1,... n.

$$DJ(i) = (1/2)\frac{\partial^2 J}{\partial w_i^2}(Dw_i)^2 \tag{3}$$

Where Dw_i is w_i corresponds to removing feature i, J is cost function, which $J = \sum_{x \in X} \|W.X - y\|^2$ for the mean squared error and $J = (1/2)\|w\|^2$ for support vector machines (SVMs).

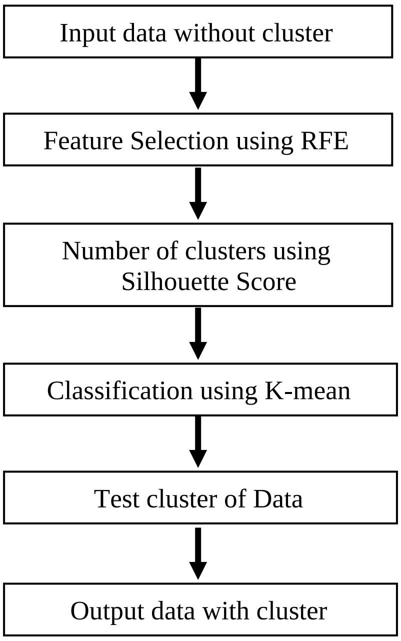

Fig. 1 Operation of the K-means clustering algorithm.

Once the important features are identified, the next step is to determine the appropriate number of clusters, which can be achieved using the Silhouette score. This technique helps fine-tune the optimal number of clusters by measuring how similar instances are to their own cluster compared to other clusters. The Silhouette score ranges from -1 to 1, with higher scores indicating that instances are more similar to their own cluster and less similar to other clusters, as described by Equation (4) [6].

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{4}$$

Where $s(i)$ is silhouette score at i, $b(i)$ are least dissimilarity mean of i with other clusters, $a(i)$ are dissimilarity mean of i with all instance in the cluster.

K-means Clustering (K-means) is a method in unsupervised learning for data mining, aimed at statistical classification by creating clusters such that data within the same cluster are highly similar, while data in different clusters are distinctly different. The computation begins with the dataset, where k is a hyperparameter specified at the start. It represents the number of cluster centroids and is determined based on the appropriate number of clusters identified using the Silhouette score.

The center of k points $x_i = (x_i^1, x_i^2, ..., x_i^r)(i = 1, 2, ..., k)$ is defined as $c = (c^1, c^2, ..., c^r)$ where c^i is mean of $x_i^1, x_i^2, ..., x_i^r$ This method obtains set M of k centroids by the iteration of a starting set M_0 of k points of r-dimensional Euclidean or R^r. The method consists of the following step [9]:

1. Start with the initial list of centers $M_0 = (m_1, m_2, ..., m_k)$, where m_i are distinct points in an r-dimensional real distance space R^r. (Initial step)
2. II. Let $X = (x_1, x_2, ..., x_n)$ be the dataset. If for a fixed x_i $d(x_i, m_j)$ is minimum for $j = p, x_i$ is assigned the value m_p. Thus each x_i is assigned one of the k elements of M_0. (Assignment step)
3. Let S_p^1 be the set of points of X, which are assigned the center m_p. Let m_p^1 is the center of the points belonging to S_p. Where $M_1 = (m_1^1, m_2^1, ..., m_k^1)$ is the updated list of centers and the 1st iterate of M_0. We repeat the steps to get a sequence (M_i) of iterates which converges to the optimal list M of centroids in a finite number of steps. (Update step)

4. Results and Discussion

This research demonstrates the correlation between each factor, the importance of each factor, the quantification of clusters, and the classification of 150 samples. The results of the feature correlation analysis indicate that four factors positively affecting transformer condition are oil breakdown voltage (F1), high voltage–ground insulation (F4), high voltage–low voltage insulation (F5), and low voltage–ground insulation (F6). Three factors negatively affecting transformer condition are age (F2), visual inspection (F3), and load (F7). The highest positive correlation is between F5 and F6, and F4, while F2 and F6 exhibit the highest negative correlation, as shown in Fig. 2.

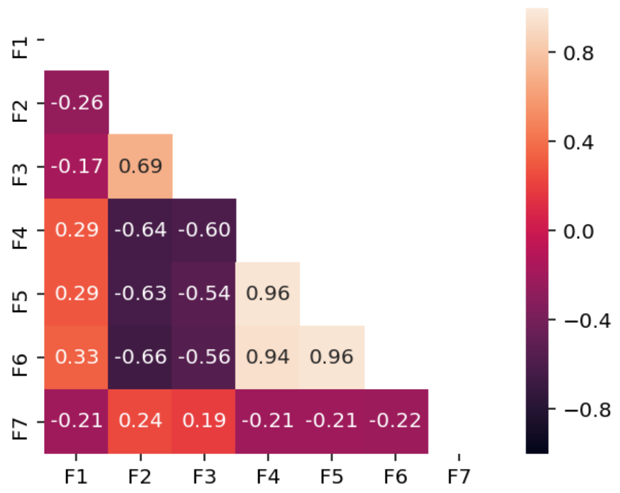

Fig. 2 Feature correlation

Table 2 Feature information

Feature	Information
F1	Oil breakdown voltage (kV/2.5 mm.)
F2	Age (years)
F3	Visual inspection (Number of defects)
F4	High Voltage – ground Insulation (Mohm)
F5	High – Low Voltage Insulation (Mohm)
F6	Low Voltage – ground Insulation (Mohm)
F7	Load (%)

Table 3 Visual inspection feature

Feature	Information
1	Gaskets, Seals
2	Oil leaks
3	Oil Level
4	Bushing Condition
5	Oil Tank Corrosion
6	Main Tank Corrosion
7	Grounding

The silhouette score at 7 features shows the best silhouette score is 0.3234 at 2 clusters, and the inferior score is 0.2873 at 3 clusters. This paper used 2 clusters due to the resolution of the clusters in the data shown in Fig 3. The proportion variance explained an increase in features affecting the data. This paper used 5 features for data explanation, accounting for more than 98%. All features are prioritized using RFE, and the 5 most important features are selected for determining the number of clusters, as shown in Fig. 4.RFE identifies the 5 most important features based on the ranking criterion. The results indicate that high-low voltage insulation (F5) is the most important feature, followed by low voltage–ground insulation (F6), high voltage–ground insulation (F4), age (F2), and oil breakdown voltage (F1). These features are utilized to determine the optimal number of clusters, as shown in Fig. 5. The results from the K-means classification, shown in Fig. 7, are divided into three clusters as follows: Cluster Poor: This cluster comprises distribution transformers (DTs) that have been in operation for a very long time. Additionally, the insulation resistance of the windings in this cluster has decreased.

The silhouette score with 5 features is 0.3416, higher than the score with 7 features, which is 0.2873. The 150 samples are grouped into clusters using the K-means algorithm, as shown in Fig. 6. The characteristics of this cluster are as follows: oil breakdown voltage (F1) ranges from 15.6 to 53.3 kV/2.5 mm, age (F2) ranges from 8 to 44 years, high voltage–ground insulation (F4) ranges from 250 to 1,200 MΩ, high voltage–low voltage insulation (F5) ranges from 210 to 1,080 MΩ, and low voltage–ground insulation (F6) ranges from 100 to 600 MΩ. The average values for each feature are shown in Table 4.

Cluster Risk: This cluster includes transformers that have been in service for a long time. While the insulation test results are satisfactory, the oil quality of these DTs is low, placing them at risk of failure. The characteristics of this cluster are as follows: oil breakdown voltage (F1) ranges from 15.9 to 47 kV/2.5 mm, age (F2) ranges from 14 to 44

years, high voltage–ground insulation (F4) ranges from 1,000 to 1,600 MΩ, high voltage–low voltage insulation (F5) ranges from 1,000 to 1,600 MΩ, and low voltage–ground insulation (F6) ranges from 500 to 1,400 MΩ. The average values for each feature are shown in Table 5.

Cluster Normal: This cluster comprises DTs that have not been in service for a long time. The insulation test results are favorable, and the oil quality is high. The characteristics of this cluster are as follows: oil breakdown voltage (F1) ranges from 18.1 to 60.8 kV/2.5 mm, age (F2) ranges from 1 to 37 years, high voltage–ground insulation (F4) ranges from 1,600 to 2,000 MΩ, high voltage–low voltage insulation (F5) ranges from 1,600 to 2,000 MΩ, and low voltage–ground insulation (F6) ranges from 1,400 to 2,000 MΩ. The average values for each feature are shown in Table 6.

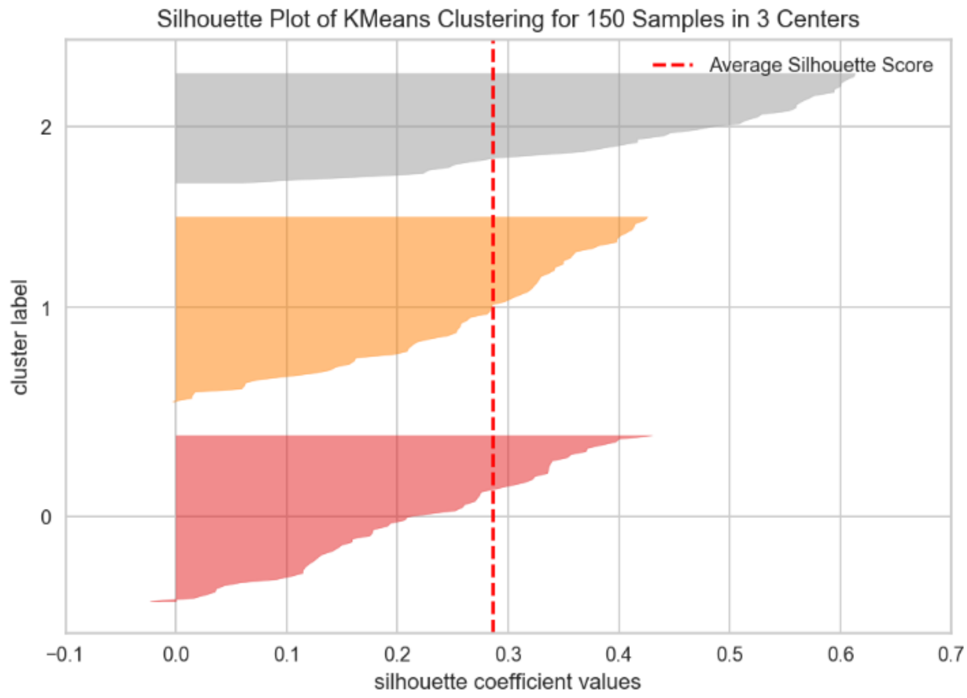

Fig. 3 The silhouette score at 7 features

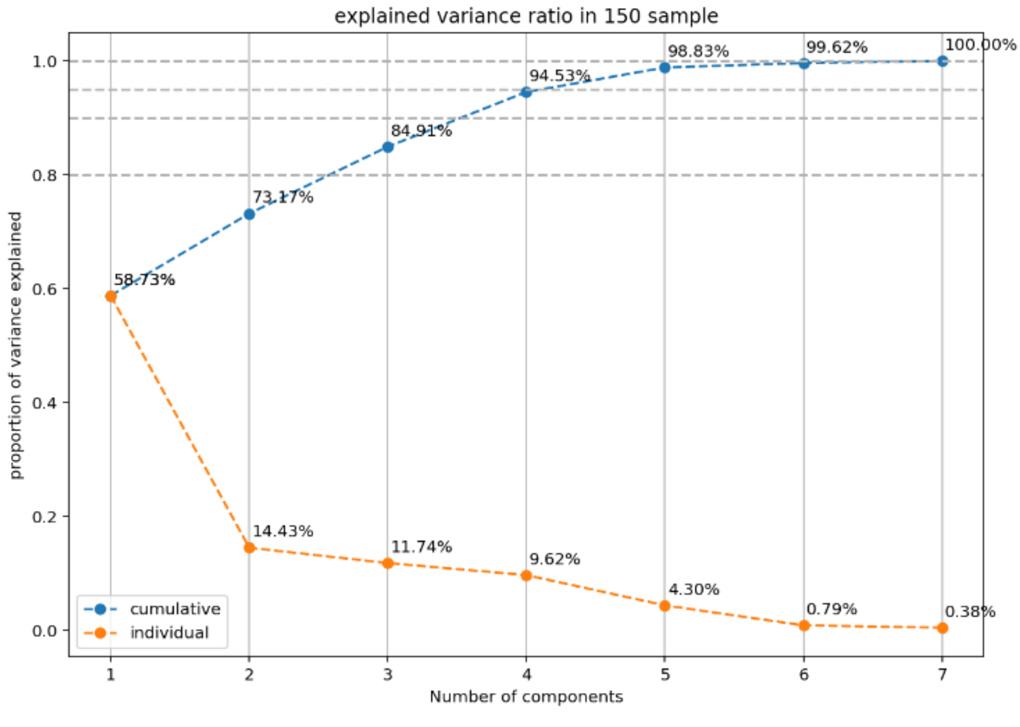

Fig. 4 Proportion variance explained.

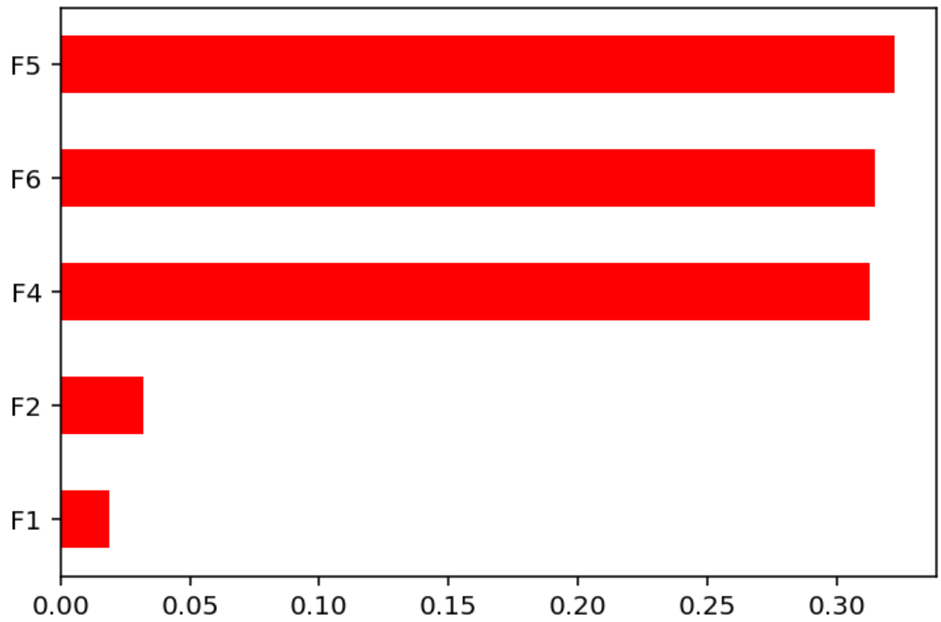

Fig. 5 The 5 most important features

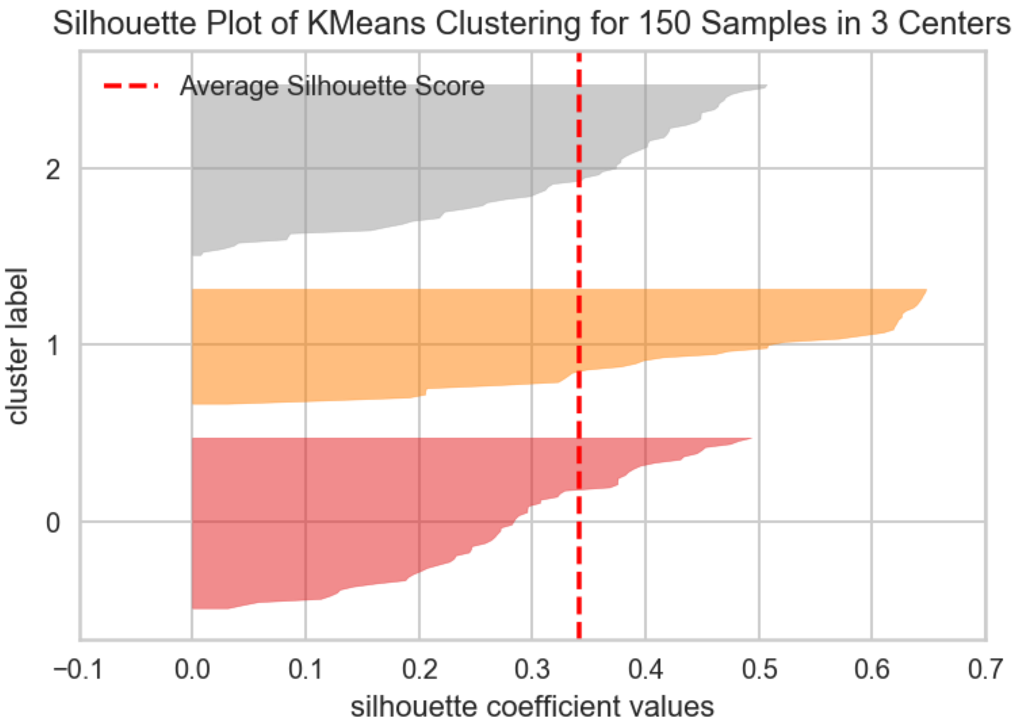

Fig. 6 The silhouette score at 5 features

Data defining clusters with K-means can establish distances between clusters and ascertain accuracy using K-Nearest Neighbors (kNN). The dataset is divided into 120 samples (20%) for training and 30 for testing out of 150 samples. The mean accuracy of the five randomized tests is 96.67%, as depicted in Table 6.

Each cluster obtained from K-means can be utilized to construct a decision tree, facilitating the classification of transformer conditions, as depicted in Fig. 8.

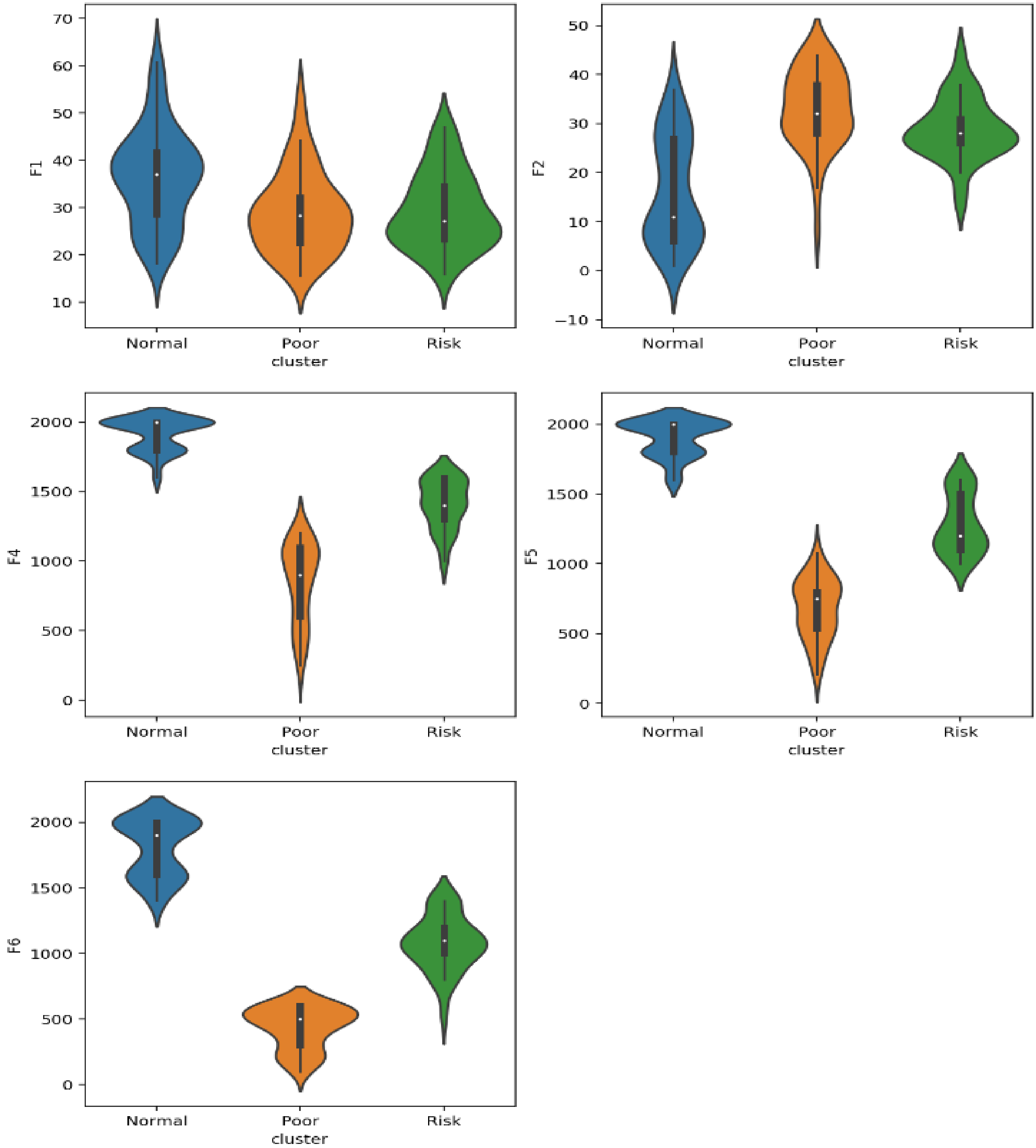

Fig. 7 The value of each feature in the cluster

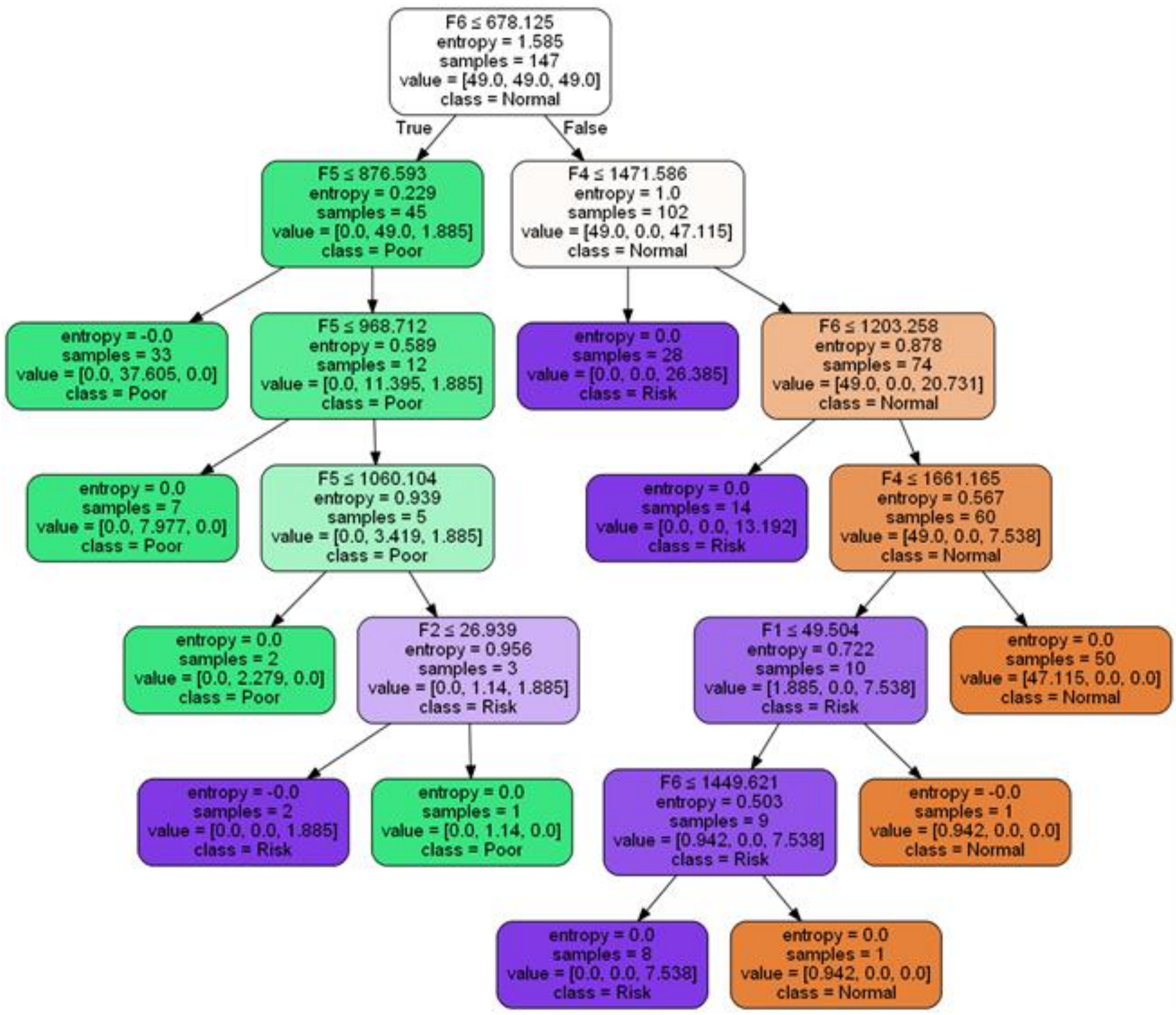

Fig. 8 Decision tree

Table 4 Each feature average (Poor)

Feature	average value
F1	28.5696
F2	33.09
F4	856.20
F5	680.43
F6	437.34

Table 5 Each feature average (Risk)

Feature	average value
F1	29.44
F2	28.81
F4	1,411.13
F5	1,283.02
F6	1,077.36

Table 6 Each feature average (Normal)

Feature	average value
F1	36.55
F2	16.30
F4	1915.09
F5	1,900.00
F6	1,800.00

Table 7 Accuracy results

Test 1	Test 2	Test 3	Test 4	Test 5	Average
100%	96.67%	96.67%	93.33%	96.67%	96.67%

5. Conclusions and Recommendations

The research on distribution transformer (DT) condition classification using the proposed method elucidated that the most influential factors affecting transformer condition are the insulation between high voltage and ground (F4), high-low voltage (F5), and low voltage and ground (F6) of the DT. This method effectively categorizes DT conditions into three clusters. While oil breakdown voltage (F1) and age (F2) may exhibit trends less pronounced than insulation values (F4, F5, F6), their trends still align with the overall pattern. The decrease in oil breakdown voltage with increasing age signifies deteriorating transformer condition. Consequently, maintenance results of DT can be compared with cluster classifications for condition assessment. The maintenance approach should prioritize repair or maintenance for the "poor" cluster, exercise vigilance or perform maintenance for the "risk" cluster following the "poor" cluster and maintain routine maintenance or perform maintenance for the "normal" cluster.

Despite achieving the research objectives, certain limitations persist. Firstly, the study is constrained by the limited number of transformer features examined, typical of distribution transformers which have fewer monitoring factors compared to power transformers. Secondly, the distribution transformers in Thailand operate at various voltage levels (e.g., 19 kV, 22 kV, and 33 kV), potentially. leading to differing classification values. Lastly, the geographical location of the transformer installation, whether seaside or urban, may also impact classification outcomes.

References

[1] D. Granados-Lieberman, J. R. Razo-Hernandez, V. Venegas-Rebollar, J. C. Olivares-Galvan, and M. Valtierra-Rodriguez, "Harmonic PMU and fuzzy logic for online detection of short-circuited turns in transformers," Electric Power Systems Research, vol. 190, p. 106862, 2021.

[2] M. Abdillah et al., "Prognostics Health Management (PHM) System for Power Transformer Using Kernel Extreme Learning Machine (K-ELM)," in Proceedings of the 2020 International Conference on Engineering and Information Technology for Sustainable Industry, 2020, pp. 1-6.

[3] X. Zhao et al., "Experimental evaluation of transformer internal fault detection based on V–I characteristics," IEEE Transactions on Industrial Electronics, vol. 67, no. 5, pp. 4108-4119, 2019.

[4] R. M. A. Velásquez, J. V. M. Lara, and A. Melgar, "Converting data into knowledge for preventing failures in power transformers," Engineering Failure Analysis, vol. 101, pp. 215-229, 2019.

[5] A. J. Patil, A. Singh, and R. Jarial, "A novel fuzzy based technique for transformer health index computation," in 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 2019: IEEE, pp. 1-6.

[6] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Comparative Analysis of Electricity Consumption at Home through a Silhouette-score prospective," in 2019 21st International conference on advanced communication technology (ICACT), 2019: IEEE, pp. 589-591.

[7] R. A. Prasojo, K. Diwyacitta, and H. Gumilang, "Correlation of transformer paper deterioration to oil characteristics and dissolved gases," in 2017 International Conference on High Voltage Engineering and Power Systems (ICHVEPS), 2017: IEEE, pp. 40-45.

[8] W. Wattakapaiboon and N. Pattanadech, "The new developed Health Index for transformer condition assessment," in 2016 International Conference on Condition Monitoring and Diagnosis (CMD), 2016: IEEE, pp. 32-35.

[9] R. Kumari, M. Singh, R. Jha, and N. Singh, "Anomaly detection in network traffic using K-mean clustering," in 2016 3rd international conference on recent advances in information technology (RAIT), 2016: IEEE, pp. 387-393.

[10] A. A. Nelson, G. C. Jaiswal, and S. Ballal, "Economical aspects of remote conditon monitoring system for distribution transformer," in 2014 international conference on power, automation and communication (INPAC), 2014: IEEE, pp. 45-49.

[11] A. J. Ghanizadeh and G. Gharehpetian, "ANN and cross-correlation based features for discrimination between electrical and mechanical defects and their localization in transformer winding," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 21, no. 5, pp. 2374-2382, 2014.

[12] M. Abdel-Hafez and A. M. Gaouda, "A new look at classification of transformer normal and abnormal currents," in Melecon 2010-2010 15th IEEE Mediterranean Electrotechnical Conference, 2010: IEEE, pp. 830-834.

[13] X. Zeng, Y.-W. Chen, and C. Tao, "Feature selection using recursive feature elimination for handwritten digit recognition," in 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009: IEEE, pp. 1205-1208.

[14] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," Machine learning, vol. 46, pp. 389-422, 2002.

[15] I. Sarker, "Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2: 160," ed: ed, 2021.

[16] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Electronics, vol. 9, no. 8, p. 1295, 2020.

[17] A. D. Ashkezari, N. Hosseinzadeh, A. Chebli, and M. Albadi, "Development of an enterprise Geographic Information System (GIS) integrated with smart grid," Sustainable Energy, Grids and Networks, vol. 14, pp. 25-34, 2018.

[18] C.-T. Lee and S.-C. Horng, "Abnormality detection of cast-resin transformers using the fuzzy logic clustering decision tree," Energies, vol. 13, no. 10, p. 2546, 2020.

[19] A. S. Mogos, X. Liang, and C. Y. Chung, "Distribution Transformer Failure Prediction for Predictive Maintenance Using Hybrid One-Class Deep SVDD Classification and Lightning Strike Failures Data," IEEE Transactions on Power Delivery, vol. 38, no. 5, pp. 3250-3261, 2023.

[20] M. Raichura, N. Chothani, and D. Patel, "Efficient CNN-XGBoost technique for classification of power transformer internal faults against various abnormal conditions," IET Generation, Transmission & Distribution, vol. 15, no. 5, pp. 972-985, 2021.