14th Annual International Conference on Industrial Engineering and Operations Management, Dubai, United Arab Emirates (UAE), February 12-14, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/AN14.202400032
Published: February 12, 2024

Explainable AI Enhanced Fault Diagnosis of Power Transformer: Unveiling the Black Box

Jeetesh Sharma, Murari Lal Mittal and Gunjan Soni
Department of Mechanical Engineering
Malaviya National Institute of Technology
Jaipur, 302017, India
2019rme9528@mnit.ac.in, mlmittal.mech@mnit.ac.in, gsoni.mech@mnit.ac.in

Abstract

Substantial capital investments in vital assets, particularly power transformers, necessitate the application of precise diagnostics. These diagnostics are crucial for assessing performance, identifying potential issues, and ensuring these assets' long-term operational and maintenance efficiency. The primary objective is proactively mitigating asset failure risk and the subsequent need for costly replacements. In recent years, significant progress has been made in developing AI models for fault classification, primarily leveraging machine learning methodologies. However, a notable characteristic of many machine learning approaches is their inherent black-box nature, which limits their interpretability. The opacity of these models necessitates adopting Explainable Artificial Intelligence (XAI) techniques to elucidate their decision-making processes. In this study, we have explored the application of various machine learning algorithms, including Support Vector Machines (SVM), k-nearest Neighbors (KNN), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Artificial Neural Networks (ANN), for fault classification. Among these models, the Random Forest algorithm yielded the most promising results. We applied XAI approaches to enhance our understanding of its decision-making mechanisms and facilitate better-informed decision-making.

Keywords
Fault Diagnosis, Machine Learning, Explainable Artificial Intelligence (XAI), Power Transformer

1. Introduction
Fault detection and diagnosis in transformers are crucial for maintaining and managing power systems. Transformers are critical for power distribution, and any problem can cause power disruption, equipment damage, and human safety risks (Sharma et al. 2011; Faiz and Soleimani 2018). Therefore, it is essential to find and understand the issues to prevent failures to keep the power system running smoothly. Earlier, Transformer Fault Diagnosis depended on visual examinations, electrical assessments, and hands-on approaches (Stone 2005). These approaches require a significant amount of time and money, and they may not always be successful in identifying potential problems or assessing a transformer's condition. AI approaches have been recognized as a promising strategy in transformer fault diagnosis. These approaches can analyze complex datasets and identify trends and abnormalities that may indicate transformer faults (Li et al. 2022; Wang et al. 2023).

For better accuracy in fault diagnosis, complex machine-learning approaches are frequently used. As the complexity of machine learning algorithms increases, they tend to behave like "black boxes," making them less transparent. When we give inputs to the algorithm, it provides the outputs, but without a clear understanding, the algorithm's inner workings remain hidden (Matzka 2020). In this situation, Explainable AI (XAI) becomes helpful. XAI improves the transparency and interpretability of machine learning models. In the context of power transformer fault diagnosis, XAI is also crucial. It gives us confidence in the model's decisions and helps us understand why the model chose a particular choice. This transparency is critical because it helps us detect and treat transformer issues early on,

improving power system reliability. As a result of using XAI, power transformer fault classification can be improved by making it more accurate, transparent, and proactive in maintenance.

1.1 Objectives

The research's objectives unfold as follows:

- To develop a precise fault diagnosis system for power transformers using machine learning algorithms.
- Implement Explainable AI (XAI) techniques to enhance interpretability and assess their feasibility in the context of power transformer fault diagnosis.

The article is organized as follows: Section 2 begins with a relevant literature review. The dataset is explained in detail in Section 3. The methodology is outlined in Section 4, which discusses several machine learning models and XAI methodologies. Results and discussion are presented in Section 5. In Section 6, the study concludes with a summary of the essential findings and their implications.

2. Literature Review

Artificial intelligence algorithms have found widespread use in transformer fault diagnostics, producing impressive results. These methods are appropriate for dealing with complex issues such as transformer fault diagnosis because they address broader problem descriptions, which usually lack specific structural information. AI examines sensor data for patterns and irregularities that may indicate a problem. Many studies have found that machine learning, in particular, is effective at identifying transformer faults. Tamilselvan and Wang (2013) proposed an approach for health diagnosis using a deep belief network (DBN) with multiple sensors. Fei and Zhang (2009) introduced a combination of SVM and genetic algorithm (GA) for power transformer fault diagnosis. GA was used to optimize the hyper-parameters of SVM. Yang et al. (2019) proposed a combination of probabilistic neural network (PNN) and bat algorithm (BA) to improve the diagnosis performance of power transformers. Malik et al. (2020) introduced an intelligent classifier for detecting early-stage faults in power transformers based on fuzzy reinforcement learning (RL).

Li et al. (2018) utilized the Cuckoo Search (CS) algorithm to enhance the performance of fault diagnostics models for power transformers by optimizing the Back-propagation (BP) neural network. Islam et al. (2017) employed the k-nearest neighbor (KNN) algorithm to index the three nearest clusters from an unfamiliar data point related to a transformer. This approach facilitates cluster voting to determine one or more fault categories. Luo et al. (2022) introduced a fault diagnosis approach for power transformer fault detection, which relies on canonical Variate Analysis and support Vector Machine (CVA-SVM). Prasojo et al. (2023) created an accurate model for fault identification using machine learning. This model utilizes the random forest algorithm along with the synthetic minority over-sampling technique (SMOTE) preprocessing method. Li et al. (2023) introduced a novel method for diagnosing faults in dry-type transformers using vibration signals. An enhanced Convolutional Neural Network (CNN) model was formulated to recognize faults in transformer images. Subsequently, the CNN model, as proposed, was trained and tested using the gathered data, and its optimal structure and hyper parameters were determined.

3. Dataset description

Power transformers are critical in the electrical power system, facilitating efficient energy transmission and distribution. Although power transformers are renowned for their reliability, they remain susceptible to failures from various factors within the transformer or external influences. These potential contributors to severe transformer breakdown can generally be categorized into two significant groups:

1. **Mechanical Failure:** This encompasses issues related to the physical components and structure of the transformer.

2. **Dielectric Failure:** These failures are linked to problems in the insulating materials used within the transformer.

A comprehensive dataset was collected using Internet of Things (IoT) devices to research power transformer health monitoring and fault detection (shreshta140, n.d.). The dataset covers the period from June 25, 2019, to April 14, 2020, with data updates every 15 minutes. Measurements from various sensors that monitor vital parameters such as phase voltages, current, temperature, and other critical operational variables are included in the dataset. To aid clarity, the abbreviations of these sensors, as well as their full names, are listed in table 1 below:

The phase line is critical for monitoring the electrical potential or voltage at the transformer's first phase, which is necessary for maintaining a balanced and stable power supply. The current line is responsible for measuring current levels. It assists in assessing the transformer's electrical load, which is critical for ensuring that it operates within its designed capacity and does not overload, which can result in overheating and damage. The oil temperature indicator is a sensor that measures the temperature of the insulating oil in the transformer. Oil temperature monitoring is critical because it can indicate abnormal heating within the transformer, which could be caused by overloading or malfunctioning components. The oil level indicator monitors the insulating oil level within the transformer. Maintaining the oil level is essential for effective cooling and insulation. Deviations in oil levels can signify leaks or other problems that need attention. An oil temperature indicator alarm is associated with the oil temperature indicator. It is triggered when the measured oil temperature surpasses a predefined threshold, alerting operators to potential overheating issues. The oil temperature indicator trip mechanism is related to the oil temperature indicator. When activated, it indicates that the oil temperature has exceeded a critical limit, which may necessitate shutting down the transformer to prevent further damage. These sensors collectively contribute to the ongoing health monitoring of the transformer, helping to prevent faults, improve safety, and ensure reliable power distribution.

Table 1. Sensor abbreviations and descriptions

Sensor Abbreviation	**Full Sensor Name**
VL1	Phase Line 1
VL2	Phase Line 2
VL3	Phase Line 3
IL1	Current Line 1
IL2	Current Line 2
IL3	Current Line 3
VL12	Voltage Line 1 2
VL23	Voltage Line 2 3
VL31	Voltage Line 3 1
INUT	Neutral Current
OTI	Oil Temperature Indicator
WTI	Winding Temperature Indicator
ATI	Ambient Temperature Indicator
OLI	Oil Level Indicator
OTI_A	Oil Temperature Indicator Alarm
OTI_T	Oil Temperature Indicator Trip
MOG_A	Magnetic Oil Gauge Indicator

The dataset comprises sensor readings, including 'WTI,' 'OTI_A,' 'OTI_T,' and 'MOG_A.' These sensors provide binary indications, with '0' representing normal data and '1' indicating faulty data. To transform the problem into a binary classification task, a new column, "Faulty Transformer," was introduced. If any of the four sensors detect a fault, the entry is marked as "Yes," signifying a faulty transformer. Conversely, if all four sensors report normal behavior, the label is "No," indicating a healthy transformer. Subsequently, the original four sensor columns were removed. Following this initial data manipulation, standard data preprocessing techniques, such as standard scaling, were applied to prepare the dataset for model input. This meticulous data preparation and transformation process enhances the dataset's suitability for robust model training and analysis within the context of transformer health diagnosis.

4. Methodology

This section describes the methodology used to analyze and diagnose power transformer health using the AI Transformer Monitoring Dataset. As described in the preceding section, the dataset provides valuable insights into transformer performance through a series of sensor readings with binary indications of normal and faulty states. We begin with precise data preprocessing to ensure that the dataset is suitable for training and evaluating machine learning

models. Standard scaling was applied to normalize the feature values and bring them to a uniform scale, minimizing the impact of varying measurement units and enhancing model convergence.

Given the binary nature of the sensor data, we transformed the problem into a binary classification task. Four key sensors ('WTI,' 'OTI_A,' 'OTI_T,' and 'MOG_A') were selected. For each data point, if these sensors reported a '1,' indicating a fault, the corresponding label was "Yes" to signify a faulty transformer. Conversely, if all four sensors registered '0,' exhibiting normal behavior, the label was marked as "No" to denote a healthy transformer. These labels were used as the ground truth for model training and evaluation.

4.1 Machine Learning Models: We leveraged a selection of machine learning algorithms to perform the transformer health diagnosis. The following models were employed:

- Support Vector Machine (SVM): SVM is a robust classification algorithm known for its versatility in handling linear and non-linear data. It was applied to classify transformers as "Faulty" or "Healthy" based on the transformed labels.

- Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to improve classification accuracy. It was utilized to capture complex relationships within the dataset.

- k-Nearest Neighbors (KNN): KNN is a simple yet effective classification algorithm that assigns labels based on the majority class of its nearest neighbors. It was employed to determine transformer health.

- eXtreme Gradient Boosting (XGBoost): XGBoost is a gradient-boosting algorithm known for its high performance. It was harnessed to model the transformer health classification task.

- Artificial Neural Network (ANN): ANN, a deep learning approach, was applied to capture intricate patterns in the data by utilizing multiple layers and neurons.

4.2 Model Evaluation: Various evaluation metrics were used to assess model performance, such as accuracy, precision, recall, and F1-score.

4.3 Explainability Methods - LIME and SHAP: Two XAI approaches are used to interpret models' decision-making processes and recognize the key factors influencing their predictions: LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations). LIME provides local, instance-specific explanations, whereas SHAP offers a global view of feature importance. These explainability tools are critical in simplifying the models' inner workings, giving transparency and valuable insights for model validation and improved decision support.

The methodology described above employs a variety of machine learning algorithms to provide a comprehensive approach to transformer health diagnosis. These models and detailed preprocessing steps are critical in accurately classifying transformers as faulty or healthy, contributing to the power transformer health assessment field.

5. Results and Discussion

In this section, we present the results of our transformer fault classification models and delve into the insights provided by local and global explainability techniques. Our methodology incorporated various machine learning algorithms, with the Random Forest model emerging as the top performer, as evident in Figure 1. The precision, recall, and F1-score values for the "No Fault" and "Fault" categories are notably high, reflecting the model's accuracy and effectiveness. The "No Fault" category demonstrates a precision of 0.96, a recall of 0.92, and an F1-score of 0.94. The "Fault" category exhibits a precision of 0.86, a recall of 0.93, and an F1-score of 0.90. Overall, the model achieved an impressive accuracy of 92%, further reinforcing its ability to discern between healthy and faulty transformers accurately. These results indicate the model's robustness in transformer fault classification, holding significant promise for real-world power system maintenance and reliability applications.

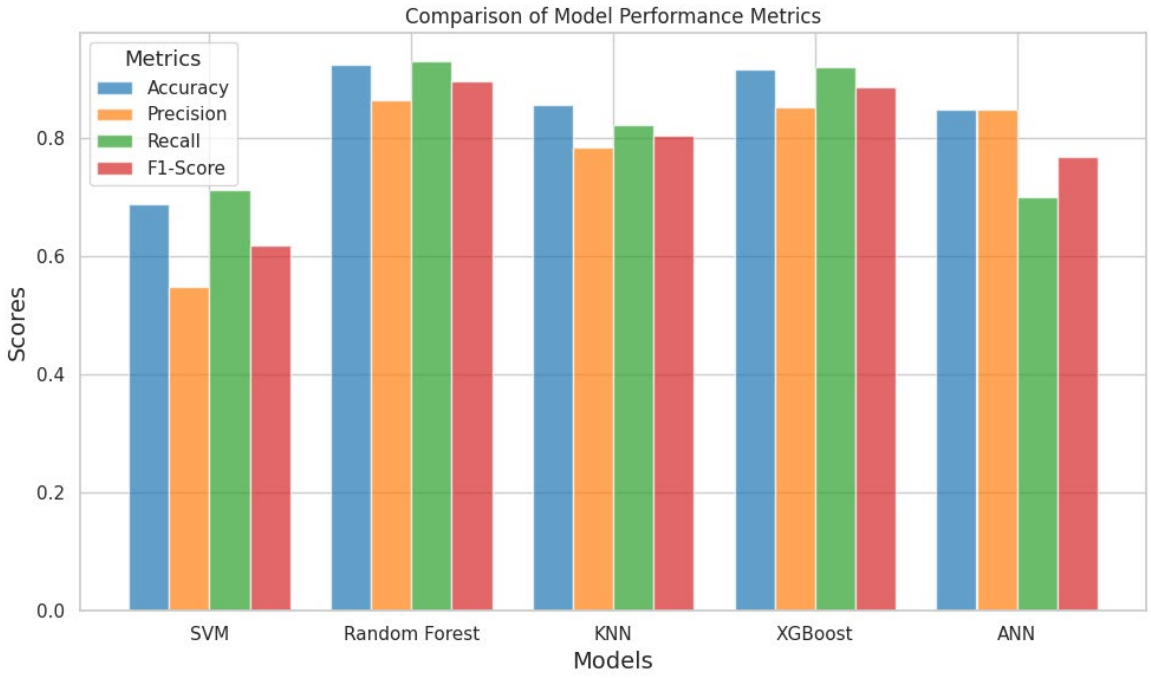

Figure 1. Models performance metrics comparison

5.1 Local explanation using LIME

We offer a detailed glimpse into the decision-making process of the Random Forest model through a local explanation using a random sample. The local explanation provides insights into the model's prediction probabilities for both "No Fault" and "Fault" labels, shedding light on the key features and their respective values that drive these predictions. For example, for a specific data point, the model assigns a probability of 0.82 for "No Fault" and 0.18 for "Fault," guided by particular feature values in Figure 2.

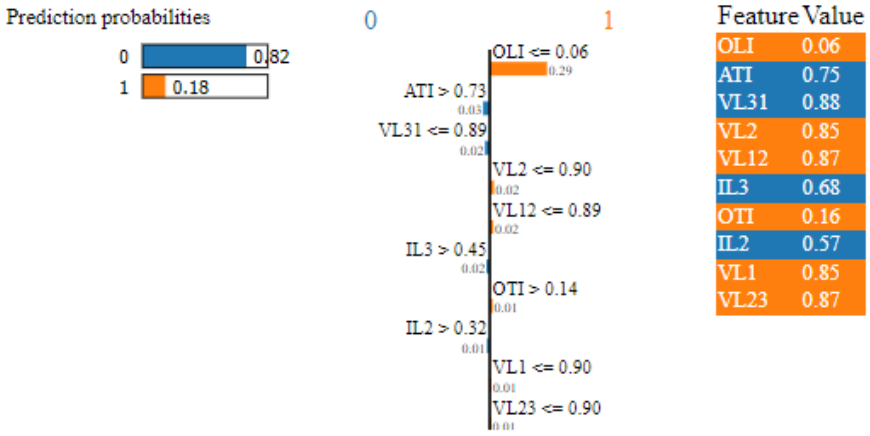

Figure 2. Local explanation results for a sample id

5.2 Global explanation using SHAP

The SHAP (SHapley Additive exPlanations) method is implemented for a comprehensive understanding of feature importance. SHAP offers a global view of feature importance, revealing the impact of each feature on model predictions in Figure 3. Notably, the features 'OLI' (Oil Level Indicator), 'VL23' (Voltage Line 2-3), and 'OTI' (Oil Temperature Indicator) displayed the highest mean SHAP values. These features are pivotal in the model's decision-making process, indicating their significance in transformer fault classification.

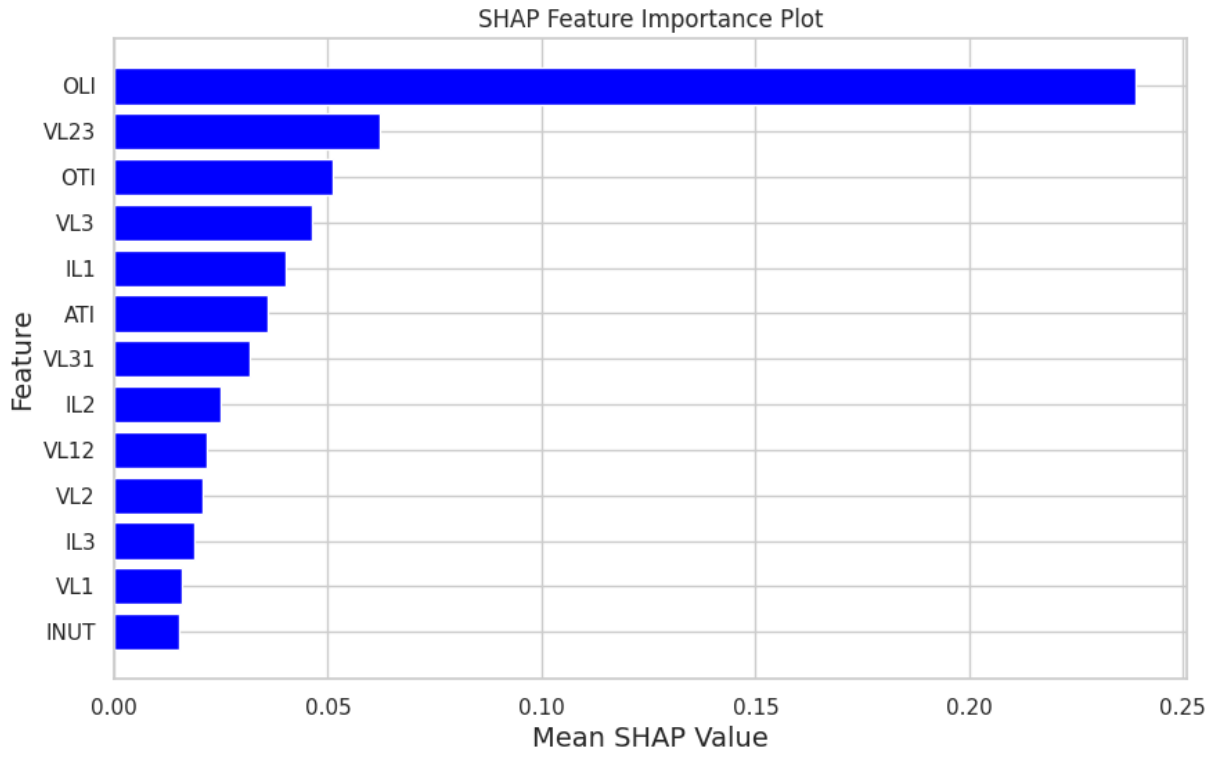

Figure 3. Global feature importance based on the SHAP values

To provide visual insights into the impact of feature values, a BeeSwarm SHAP plot is generated, as shown in Figure 4. In this plot, blue signifies low feature values, while red corresponds to high ones. The x-axis depicts SHAP values. This visualization enables a clear understanding of how feature values influence model predictions, highlighting the varying degrees of importance. In the real world, features such as 'OLI,' 'VL23,' and 'OTI' hold substantial importance in transformer fault classification. For instance, 'OLI' is critical for monitoring oil levels ensuring proper cooling and insulation. 'VL23' is instrumental in assessing voltage stability, while 'OTI' provides insights into oil temperature, which can indicate potential overheating.

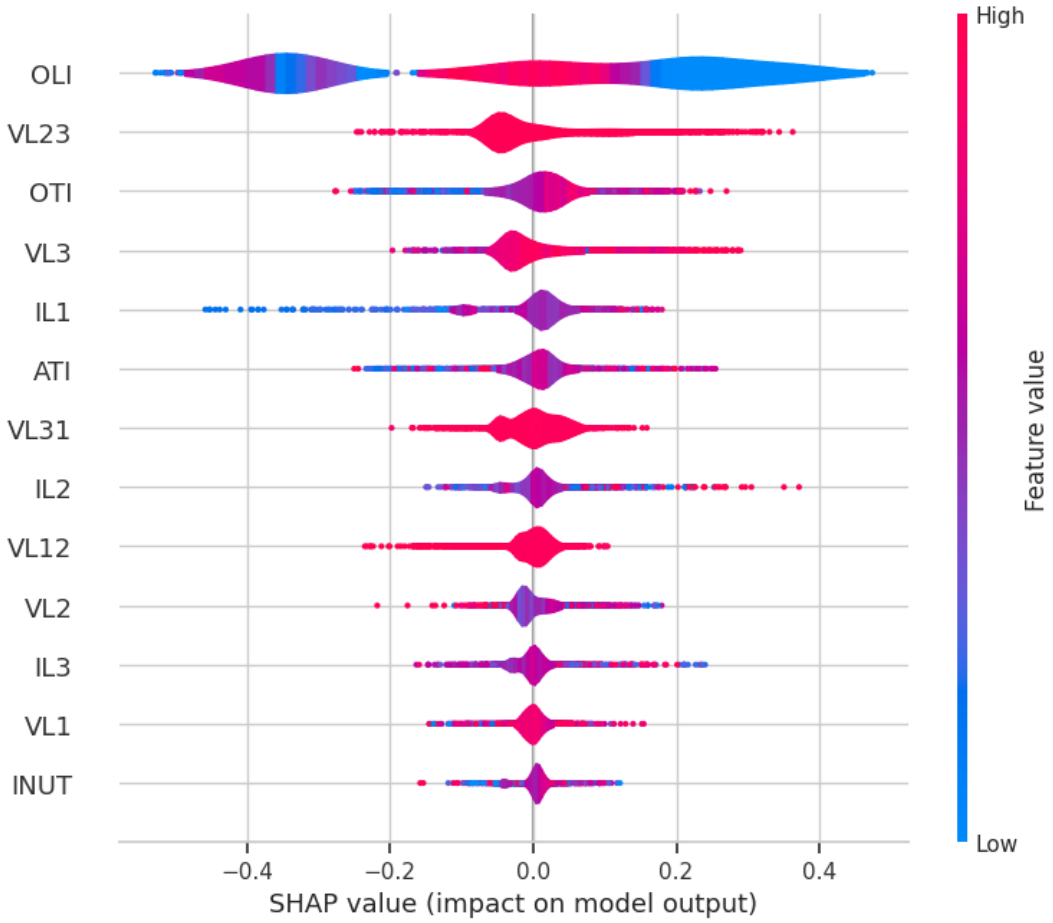

Figure 4. SHAP Beeswarm plot

6. Conclusion

In this study, we have explored the application of various machine learning algorithms for transformer fault diagnosis, explicitly focusing on enhancing the interpretability of the algorithms. The results demonstrate that the random forest excelled compared to other algorithms, demonstrating its effectiveness in accurately classifying transformer health. The high precision, recall, and F1-score values for both the "No Fault" and "Fault" categories underscore the robustness of the Random Forest model. Through local explanations, we gained valuable insights into the model's decision-making process for individual data points. Specific features and their values were examined to understand their influence on predictions. Furthermore, we employed global explanation techniques such as SHAP (SHapley Additive exPlanations) to understand feature importance at a broader scale. Notably, features like 'OLI' (Oil Level Indicator), 'VL23' (Voltage Line 2-3), and 'OTI' (Oil Temperature Indicator) were identified as significant factors in transformer fault classification. XAI techniques like SHAP and LIME enhance the transparency of complex machine learning processes, fostering trust and confidence in AI-driven decision-making. Applying XAI techniques is vital for ensuring the reliability and safety of power systems.

In conclusion, our research provides a data-driven approach to preventing failures and improving power transformer reliability. Machine learning algorithms, interpretability techniques, and domain-specific knowledge work well to enhance transformer fault diagnosis and maintenance practices. As the power sector continues to evolve, integrating AI and XAI will be pivotal in ensuring the stability and efficiency of power distribution systems. As we look to the future, several exciting avenues exist for further research and development in transformer fault diagnosis. One of the promising directions is the exploration of ensemble models and hybrid approaches. These ensemble models can effectively capture diverse patterns in the data, leading to improved diagnostic accuracy. A wealth of XAI techniques

beyond SHAP and LIME warrant exploration, such as Integrated Gradients, DeepLIFT, and Counterfactual Explanations, each offering unique insights into the model's behavior and opportunities for enhancement.

References

Faiz, J. and Soleimani, M., Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis. *IEEE Transactions on Dielectrics and Electrical Insulation*, *25*(5), pp.1798-1806, 2018.

Fei, S.W. and Zhang, X.B., Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications*, *36*(8), pp.11352-11357, 2009.

Islam, M.M., Lee, G. and Hettiwatte, S.N., A nearest neighbour clustering approach for incipient fault diagnosis of power transformers. *Electrical Engineering*, *99*, pp.1109-1119, 2017.

Li, A., Yang, X., Dong, H., Xie, Z. and Yang, C., Machine learning-based sensor data modeling methods for power transformer PHM. *Sensors*, *18*(12), p.4430, 2018.

Li, C., Chen, J., Yang, C., Yang, J., Liu, Z. and Davari, P., Convolutional Neural Network-Based Transformer Fault Diagnosis Using Vibration Signals. *Sensors*, *23*(10), p.4781, 2023.

Li, Y., Xu, Y., Li, X., Li, R., Lin, J. and Zhang, G., Addressing imbalance of sample datasets in dissolved gas analysis by data augmentation: Generative adversarial networks. *IET Generation, Transmission & Distribution*, *16*(22), pp.4494-4504, 2022.

Luo, L., Li, Y., Shi, Y., Han, T., Yang, W., Jin, X. and Han, D., A Fault Diagnosis Method for Power Transformer Using Canonical Variate Analysis and Support Vector Machine. In *International Conference in Communications, Signal Processing, and Systems* (pp. 138-146), 2022.

Malik, H., Sharma, R. and Mishra, S., Fuzzy reinforcement learning based intelligent classifier for power transformer faults. *ISA transactions*, *101*, pp.390-398, 2020.

Matzka, S., Explainable artificial intelligence for predictive maintenance applications. In *2020 third international conference on artificial intelligence for industries (ai4i)* (pp. 69-74). IEEE, 2020.

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

Sharma, N.K., Tiwari, P.K. and Sood, Y.R., Review of artificial intelligence techniques application to dissolved gas analysis on power transformer. *International Journal of Computer and Electrical Engineering*, *3*(4), pp.577-582, 2011.

Sreshta140, *AI Transformer Monitoring*, from https://www.kaggle.com/datasets/sreshta140/ai-transformer-monitoring, 2020, Accessed on september 25, 2023.

Stone, G.C., Partial discharge diagnostics and electrical equipment insulation condition assessment. *IEEE Transactions on Dielectrics and Electrical Insulation*, *12*(5), pp.891-904, 2005.

Tamilselvan, P. and Wang, P., Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, *115*, pp.124-135, 2013.

Wang, L., Littler, T. and Liu, X., Dynamic Incipient Fault Forecasting for Power Transformers using an LSTM Model. *IEEE Transactions on Dielectrics and Electrical Insulation, 2023*.

Yang, X., Chen, W., Li, A., Yang, C., Xie, Z. and Dong, H., BA-PNN-based methods for power transformer fault diagnosis. *Advanced engineering informatics*, *39*, pp.178-185, 2019.

Biographies

Jeetesh Sharma obtained his Bachelor of Technology in Mechanical Engineering from the Indian Institute of Information Technology, Design and Manufacturing, Jabalpur, India, in 2016 and his Master of Technology in Industrial Engineering from Malaviya National Institute of Technology Jaipur, India, in 2019. He is a doctoral student at Malaviya National Institute of Technology Jaipur, India. His research areas include Predictive maintenance, Explainable Artificial Intelligence, and Machine Learning.

Dr. Murari Lal Mittal obtained his Bachelor of Engineering in Mechanical Engineering from the University of Rajasthan, India, in 1988, his Master of Engineering in Production Engineering from MS University of Baroda, India, in 1991, and PhD in Industrial Engineering from Indian Institute of Technology Delhi, New Delhi, India, in 2006. He is a Professor at the Department of Mechanical Engineering, Malaviya National Institute of Technology Jaipur. His areas of interest include Operations Management, Project Management, Scheduling, Artificial Intelligence, Machine

Learning and Multi-agent Systems. He has published more than 100 research papers in journals and conference proceedings.

Dr. Gunjan Soni is an Assistant Professor in the Department of Mechanical Engineering at the Malaviya National Institute of Technology (MNIT) in Jaipur, India, with a Ph.D. from BITS, Pilani, and a Master of Technology from IIT Delhi. His expertise lies in supply chain management, risk analysis, production planning, machine learning, and artificial intelligence. Currently, his research interests are focused on developing AI-based predictive maintenance models for engineering systems. Additionally, he is actively engaged in exploring Industry 4.0 interventions in supply chain management and AI applications in manufacturing systems, further contributing to advancements in these domains.