

<sup>1</sup>A.O. ABDULKAREEM, <sup>1</sup>Bilkisu JIMADA-OJUOLAPE, <sup>1</sup>Morufat O. BALOGUN, <sup>1</sup>Lambe Mutalub ADESINA

## COMPARATIVE ANALYSIS OF XGBOOST AND RANDOM FOREST ALGORITHMS FOR TRANSFORMER FAILURE PREDICTION

<sup>1</sup>Department of Electrical and Computer Engineering, Kwara State University, Maletе, NIGERIA

**Abstract:** In power systems, electrical power transformers are crucial for efficient electricity transmission. Transformer failures can cause significant disruptions, highlighting the need for effective maintenance strategies. This study explores predictive maintenance (PdM) for power grids using machine learning to identify transformer failures preemptively. It compares the effectiveness of Extreme Gradient Boosting (XGBoost) and Random Forest algorithms using operational and historical data. The research underscores the inefficiencies of traditional maintenance strategies and positions PdM as a proactive alternative for enhanced grid resilience and optimized maintenance schedules. Utilizing data from Kaggle, the study employs rigorous preprocessing and model development to evaluate both algorithms. Findings indicate that while both XGBoost and Random Forest show significant predictive capabilities, XGBoost outperforms Random Forest in accuracy and efficiency. This research highlights the potential of machine learning in improving transformer maintenance strategies and provides insights for future studies aimed at enhancing predictive models in electrical engineering. The results emphasize the importance of advanced analytical techniques to safeguard critical infrastructure and optimize maintenance practices in power systems

**Keywords:** Failure Prediction, Power Transformer, Machine Learning, Predictive Maintenance (PdM), Predictive Models, Random Forest, Transformer Failures, XGBoost

### 1. INTRODUCTION

Transformers are crucial for stepping down high-voltage electricity for safe delivery to homes and businesses. However, they are susceptible to malfunctions, posing significant challenges to grid reliability and stability (Hussain et al., 2021). Transformer faults can be categorized into internal, which account for 70%-80% of transformer faults and external faults as depicted in Figure 1 below (Hussain et al., 2021).

Traditional maintenance methods, based on fixed schedules or reactive repairs, are often inefficient (Rojek et al., 2023). Preventive maintenance involves regular, scheduled inspections and servicing to prevent equipment failures before they occur (Rojek et al., 2023). Corrective maintenance, on the other hand, involves repairing or replacing components after a failure has occurred (Rojek et al., 2023).

Rather than relying on preventive or corrective maintenance, a proactive strategy based on predictive maintenance can be employed. Predictive maintenance (PdM) uses machine learning to analyze historical data, such as load patterns, voltage levels, and frequency fluctuations, to predict potential transformer failures. This approach enables scheduled maintenance interventions, preventing failures and minimizing downtime, contributing to a more efficient and reliable transformer (Tianjin da xue et al., 2018). By tracking transformer data over time, predictive maintenance programs can identify trends and early indicators of equipment degradation, allowing for timely interventions. This proactive approach reduces unplanned outages and optimizes transformer performance.

Machine learning, combined with IoT, plays a pivotal role in predictive maintenance. IoT devices feed real-time data to centralized systems, allowing for precise maintenance schedules (Marcelino et al., 2021). Predictive maintenance, combined with machine learning, reduces breakdowns by 70%, increases productivity by 25%, and lowers maintenance costs by 25% (Rojek et al., 2023). This approach makes power system operators more proactive, efficient, and resilient. Machine learning algorithms like support vector machines (SVM), neural networks (NN), and decision trees (DT) have been used in PdM (Fei & Zhang, 2009; Y. Zhang et al., 1996). However, these often rely on direct sensor data, limiting their

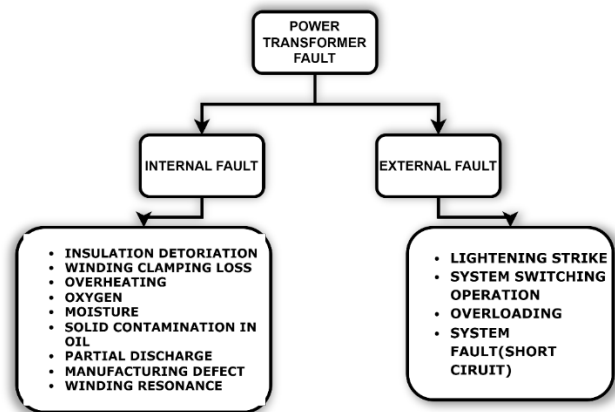


Figure 1: Electrical Faults of Transformer

applicability in scenarios with limited or unreliable sensor coverage. Among these algorithms, Extreme Gradient Boosting (XGBoost) and Random Forest stand out for failure prediction (Breiman, 2001; Zhang et al., 2019). XGBoost is known for its high performance and efficiency with large datasets, while Random Forest offers robustness and ease of interpretation. Therefore, this study compares XGBoost and Random Forest in transformer maintenance, evaluating their predictive capabilities using operational and historical data. Key performance metrics such as accuracy, precision, and recall are assessed to highlight the strengths and weaknesses of each algorithm. The rest of the article is structured as follows: Section 2 reviews related papers on the study. Section 3 explains the methodology used in the study. Section 4 presents the results and discussion of the study, and Section 5 provides the conclusion.

## 2. REVIEW OF RELATED WORKS

In a recent study, Wang et al., (2023) present a novel approach named TPE-XGBoost for diagnosing transformer faults using incomplete data. This methodology utilizes Bayesian optimization to fine-tune the hyperparameters of the XGBoost model, showcasing superior performance in comparison to alternative machine learning algorithms. A notable strength of this method lies in its capability to effectively handle incomplete datasets, as evidenced by its robust performance. However, the study highlights a limitation regarding reduced diagnostic accuracy when the rate of missing data exceeds 20%, emphasizing the necessity for further enhancement, particularly in scenarios with a high missing data rate exceeding 30%.

Chen et al., (2019) introduces a methodology for predicting transient stability status in power systems using the XGBoost model. Key features of the generator's state are extracted, and redundant ones are removed. The paper emphasizes the XGBoost model as a competitive technology for transient stability prediction due to its advantages as a tree structure model that does not require data normalization and can effectively handle missing values. Despite its advantages, the paper calls for more empirical validation and real-world application to fully evaluate its effectiveness.

The study by R. Zhang et al., (2019), explores the use of the XGBoost algorithm for diagnosing bearing faults in complex industrial environments. The research compares XGBoost with alternative tree models and highlights its superior performance in terms of both training time and accuracy. Notably, the paper emphasizes the importance of managing model complexity through regular coefficients and employing Bayesian optimization for parameter tuning. However, the study acknowledges limitations, including the need for high-quality data and challenges related to model generalizability and interpretability. Overall, the findings provide valuable insights into the potential of advanced machine learning techniques for industrial fault diagnosis.

In the realm of artificial intelligence (AI) applications within Industry 4.0, particularly focusing on its utilization in maintenance processes Rojek et al., (2023) focuses on the use of AI methods, particularly artificial neural networks (ANN), to enhance the supervision of machine failures and support their repair. It addresses the challenges associated with unbalanced training data in real industrial settings and emphasizes the limitations of using supervised machine learning models in such scenarios. The study also proposed future research directions to enhance AI-based maintenance solutions' predictive accuracy and utility in industry, emphasizing the practical challenges that need further investigation.

A study by Breviglieri et al., (2021) explored within an in-depth literature review centered on the application of deep learning models for predicting smart grid stability, with a specific emphasis on the Decentral Smart Grid Control (DSGC) system. The study highlights the challenges of integrating renewable energy sources into smart grids and underscores the significance of stability analysis in networked control systems. Acknowledging some limitations, such as the need for more generalization and extension of the analysis to larger grids with more than 10 users, the paper provides valuable perspectives on the complexities of smart grid stability prediction.

A novel unsupervised analysis method for anomaly detection in industrial machinery by Carratu et al., (2023) using electrical current values and power grid parameters. The framework combines machine learning algorithms and traditional analysis, with a focus on optimizing performance and execution time. It includes a technique for analyzing temporal dynamics based on short-time Fourier transform (STFT) to enhance detection accuracy. Results show exceptional performance, with zero false positives across



all datasets tested and less than 4% undetected outlier events, surpassing expert evaluations and other existing methodologies. However, the paper acknowledges dependence on specific features, highlighting a potential limitation for future research with diverse anomaly types.

### 3. METHODOLOGY

This section discusses the proposed framework of this study in detail. First, the data set is described including its method of preparation for use in the proposed comparison framework.

#### ■ DATASET

The Distributed Transformer Monitoring dataset was collected via Internet of Things (IoT) devices, the dataset spans from June 25th, 2019, to April 14th, 2020, with updates recorded at 15-minute intervals (Sreshta, 2020). It consists of 21,174 rows and 17 columns, with each row representing a unique observation and each column denoting a specific feature or attribute. The dataset encompasses both numerical and categorical variables, providing comprehensive insights into transformer health and performance.

#### ■ DATA PREPARATION

Pre-processing data is a critical step in transforming raw data into a machine-readable format, essential for effective utilization by machine learning models (Abbasi, 2021). This project involved several key steps to ensure data accuracy and uniformity, enhancing the models' ability to learn and make accurate predictions.

**Data Cleaning:** The first step in the data preparation process was data cleaning. This involved removing noise and inconsistencies from the dataset to ensure its integrity. Any missing values were identified and addressed to maintain the dataset's completeness.

**Handling Missing Values:** Handling missing data was a crucial part of the pre-processing phase. For most features, missing values were handled using interpolation, which estimates missing values based on the surrounding data points. For the dependent feature, missing data were handled using forward fill, which propagates the last observed value forward to fill in the gaps.

**Feature Selection:** Feature selection was performed to identify the most relevant attributes for the predictive models. This step was crucial not only for maintaining the data's accuracy and uniformity but also for removing multicollinearity identified from the correlation matrix. By eliminating highly correlated features, we ensured that the models, particularly XGBoost and Random Forest, did not suffer from multicollinearity, which could lead to overfitting.

Using Python, this project employed libraries such as Pandas for data manipulation, NumPy for numerical operations, and Scikit-learn for various pre-processing techniques. These tools provided comprehensive functionalities to handle the entire workflow, ensuring the data was in the best possible shape for analysis and model training.

**Data Splitting:** The final step in the data preparation process was splitting the data into training and test subsets. The data was divided into 80% for training and 20% for testing. The training set was used for fitting the models and tuning hyperparameters, while the test set provided an independent assessment of model performance. This approach facilitated robust and unbiased model evaluation.

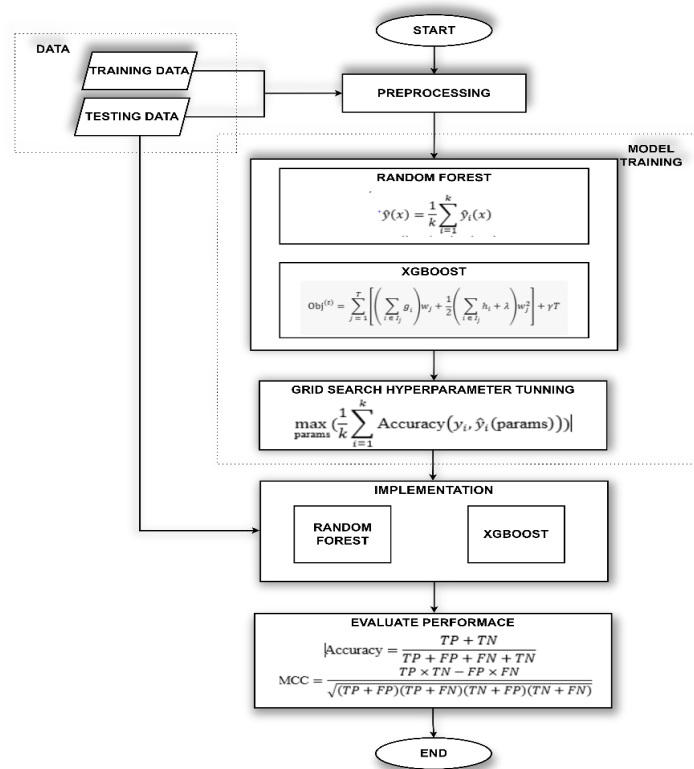


Figure 2: Study Framework

This preparation was key to developing reliable and efficient transformer failure prediction models and maintenance practices.

### **RANDOM FOREST ALGORITHM**

Random Forest is a robust ensemble learning technique that leverages the collective power of multiple decision trees to tackle complex problems in supervised learning (Wang et al., 2023). It's versatile and can be applied to both Classification and Regression tasks in Machine Learning. This ensemble learning method combines a multitude of sensor readings and historical trends, enhancing accuracy and resilience in pinpointing potential failures (Wang et al., 2023). As the name implies, a "Random Forest" is a classifier comprising numerous decision trees on various subsets of the given dataset. It averages the results to enhance the predictive accuracy of the dataset (Wang et al., 2023). The more trees in the forest, the higher the accuracy, which helps prevent overfitting. The below diagram explains the working of the Random Forest algorithm:

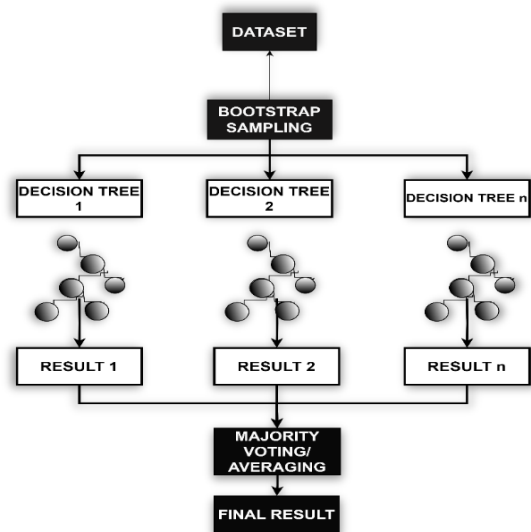


Figure 3: Random Forest Algorithm

In classification, the algorithm begins by randomly sampling subsets of the training data with replacement. For each subset, decision trees are constructed using sensor readings and historical trends. At each node, a subset of features is randomly selected, and the optimal feature and split point are chosen based on their ability to minimize the Gini impurity. Gini impurity, a measure of the uncertainty or impurity of a set of samples, is calculated as:

$$\text{Gini}(D) = 1 - \sum_{i=1}^C p_i^2 \quad (1)$$

Here,  $p_i$  is the probability of class  $i$  in node  $D$ .

In regression, each tree analyzes relationships between features and actual values to estimate the remaining lifespan of equipment. The final predicted value for a sample is the average of predictions by all the individual trees, calculated as:

$$\hat{y}(x) = \frac{1}{k} \sum_{i=1}^k \hat{y}_i(x) \quad (2)$$

Here,  $\hat{y}_i(x)$  represents the predicted value by the  $i^{\text{th}}$  decision tree for sample  $x$ , and  $k$  is the number of decision trees.

### **EXTRA GRADIENT BOOSTING (XGBOOST) ALGORITHM**

Gradient Boosting is a machine learning technique that builds a strong predictive model by combining the predictions of several weaker models.

It's particularly useful for regression and classification problems. XGBoost, or Extreme Gradient Boosting, is a scalable system for tree boosting developed by Chen and Guestrin. It uses Classification and Regression Trees (CART) as the base classifier and integrates them with gradient boosting (Chen et al., 2019). XGBoost adds a regularization term to the loss function, reducing model complexity and achieving a balance between accuracy and complexity. Each new CART is added by fitting the prediction residuals

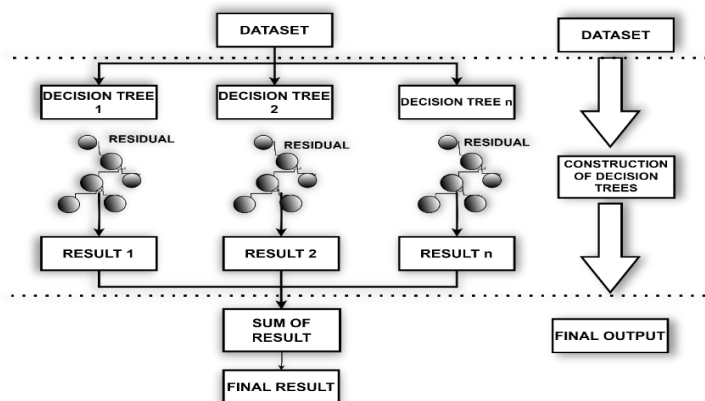


Figure 4: Extreme Gradient Boosting Algorithm



of the previous CART, and the accumulated prediction results of all CARTs yield the final model results (Wang et al., 2023). This makes XGBoost a highly efficient, flexible, and portable tool for machine learning tasks. The below diagram explains the working of the XGBoost algorithm:

The mathematical equations guiding the model learning process are as follows (Chen et al., 2019)

Objective function:

$$Obj(\Theta) = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))] + \Omega(f_t) \quad (3)$$

Regularization term:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (4)$$

Simplified objective function:

$$Obj^{(t)} = \sum_{i=1}^n [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (5)$$

Final Objective Function

$$Obj^{(t)} = \sum_{j=1}^T [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) w_j^2] + \gamma T \quad (6)$$

In these equations,  $Obj(\Theta)$  is the objective function to minimize, combining a loss term  $l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$  that measures the difference between predicted and actual values and a regularization term  $\Omega(f_t)$  that controls model complexity by penalizing the number of leaves  $T$  and leaf weights  $w$ . The simplified objective function uses gradients  $g_i$  and Hessians  $h_i$  to approximate the loss function, making optimization easier. The final objective function sums the contributions from all leaves, calculating scores based on the gradients and Hessians of instances assigned to each leaf, helping the algorithm make accurate predictions while avoiding overfitting. Where  $t$  is iteration value,  $x$  is inputted data,  $y$  is true label of dataset.

### ■ HYPERPARAMETER TUNNING AND FITTING

**Hyperparameter Tuning:** For both XGBoost and Random Forest classifiers, hyperparameters are tuned using grid search. This involves exploring a predefined set of hyperparameters to identify the best combination that maximizes model performance. Grid search is performed with cross-validation to ensure robust and unbiased evaluation of each hyperparameter combination.

**Model Fitting:** Once the optimal hyperparameters are identified, the best models are fitted on the training data. This step involves training the models using the entire training dataset to learn the underlying patterns and relationships within the data.

**Prediction:** After training, predictions are made on the test data using the best models.

### ■ MODEL PERFORMANCE EVALUATION

Performance evaluation involves using various metrics to assess the effectiveness of different machine learning algorithms (Abbasi, 2021). The metrics used in this research include the following:

**Confusion Matrix:** is a performance metric used for statistical classification. It consists of a table layout that allows visualization of an algorithm's accuracy and correctness (Abbasi, 2021).. It evaluates the predicted model against the actual class outcomes to see the number of correctly classified instances. Key terms include:

- **True Positive (TP):** Actual value is 1 (True) and predicted value is 1 (True).
- **True Negative (TN):** Actual value is 0 (False) and predicted value is 0 (False).
- **False Positive (FP):** Actual value is 0 (False) and predicted value is 1 (True).
- **False Negative (FN):** Actual value is 1 (True) and predicted value is 0 (False).

		ACTUAL	
		Positive(1)	Negative(0)
PREDICTED	Positive(1)	(TP) TRUE POSITIVE	(FP) FALSE POSITIVE
	Negative(0)	(FN) FALSE NEGATIVE	(TN) TRUE NEGATIVE

Figure 5: Confusion Matrix

**Accuracy** is the ratio of correct predictions to the total number of predictions (Abbasi, 2021; Mohammed, 2017). It is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (7)$$

**Precision** is the ratio of true positive predictions to the total number of positive predictions (Abbasi, 2021; Mohammed, 2017). It is calculated as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

**Recall** is the ratio of true positive predictions to the total number of actual positive (Abbasi, 2021; Mohammed, 2017). It is calculated as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

**Specificity** is the ratio of correctly predicted negative observations to all actual negatives (Abbasi, 2021; Mohammed, 2017). It is calculated as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (10)$$

**F1 score** is the harmonic mean of Precision and Recall, making it a better choice for evaluating imbalanced datasets (Abbasi, 2021). It is calculated as:

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

**Matthews Correlation Coefficient (MCC)**: is a measure of the quality of binary classifications, considering true and false positives and negatives (Chicco & Jurman, 2020). It is calculated as:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (12)$$

The **Area Under the Receiver Operating Characteristics (AUROC)** curve is used to visualize the performance of a classification model across all classification thresholds (Abbasi, 2021). These metrics provide a comprehensive evaluation of the machine learning models, ensuring their effectiveness and reliability in predicting transformer failures.

#### 4. RESULT AND DISCUSSION

A transformer dataset containing 21174 instances and 11 attributes was extracted from a CSV file and loaded into a Jupyter Notebook environment. After data collection, an exploratory data analysis (EDA) is carried out on the dataset to evaluate and classify the data's key features by means of visualizations, then data cleaning and preparation is carried out before the models are implemented.

Table 1: Datasets Description

	count	Mean	std	min	25%	50%	75%	max
OTI	21174	30.17786	11.961138	0	26	30	34	250
WTI	21174	0.259564	0.438406	0	0	0	1	1
ATI	21174	27.740059	5.750776	0	24	28	32	44
OLI	21174	69.661519	27.79258	36	40	64	100	100
OTI_A	21174	0.00477	0.068902	0	0	0	0	1
OTI_T	21174	0.00222	0.047063	0	0	0	0	1
MOG_A	21174	0.101681	0.302236	0	0	0	0	1
VL1	20652	241.023455	9.392606	0	235.8	242.4	247.3	261.2
VL2	20652	240.490538	9.784313	0	235.5	241.9	246.4	261.3
VL3	20652	239.923107	8.712857	0	235.5	241	245.1	261.3
IL1	20652	70.56517	42.963096	0	43.6	67.5	98.4	224.1
IL2	20652	56.522187	41.311167	0	28.7	48.7	80.3	253.6
IL3	20652	79.705825	45.816941	0	53.1	77.7	111.9	247.3
VL12	20652	363.387391	140.051283	0	397.4	416.2	427.1	446.5
VL23	20652	362.494984	139.488304	0	398.2	415.3	424.9	444.8
VL31	20652	363.871165	140.200657	0	399	416.5	427.3	447.3
INUT	20652	25.170497	15.705378	0	15.3	24.8	35.1	145.8

#### DESCRIPTIVE STATISTICS OF DATA

The dataset comprises 21,174 observations, detailing various parameters relevant to transformer failure prediction. Key features include oil temperature indicators (OTI, ATI, OLI), winding temperature indicator



(WTI), and electrical characteristics such as voltages (VL1, VL2, VL3) and currents (IL1, IL2, IL3). Descriptive statistics reveal that OTI has a mean of 30.18°C with a standard deviation of 11.96°C, ranging from 0 to 250°C, while ATI averages 27.74°C with a standard deviation of 5.75°C. OLI shows a mean value of 69.66 with a wider spread, indicating variability in oil levels. Electrical parameters, such as VL1, VL2, and VL3, show close mean values around 240V, with standard deviations near 9V, reflecting consistent voltage levels. Currents IL1, IL2, and IL3 display higher variability, with means around 70A, 56A, and 80A respectively. Additionally, the dataset includes binary variables like OTI\_A and OTI\_T, and measured values for MOG\_A, showcasing diverse data points crucial for predictive modeling.

### CORRELATION MATRIX

To further understand the data, the correlations between the features were checked to ensure that the correlation between them is not too high, making them suitable for machine learning algorithms and avoiding overfitting or underfitting the models. Figure 6 shows the heatmap of the features in the dataset. Notably, the variables VL1, VL2, VL3, and IL1, IL2, IL3 exhibit strong positive correlations with each other, indicating that they capture similar information about the system's state, which suggests potential redundancy. Moderate correlations are observed between OTI, ATI, and CI1 with other features, hinting at their unique contributions to the dataset. Understanding these correlations is crucial for effective feature selection and engineering, ensuring that the models, particularly XGBoost and Random Forest, do not suffer from multicollinearity, which could lead to overfitting. By strategically selecting and possibly reducing features, the project aims to improve the robustness and accuracy of the predictive models.

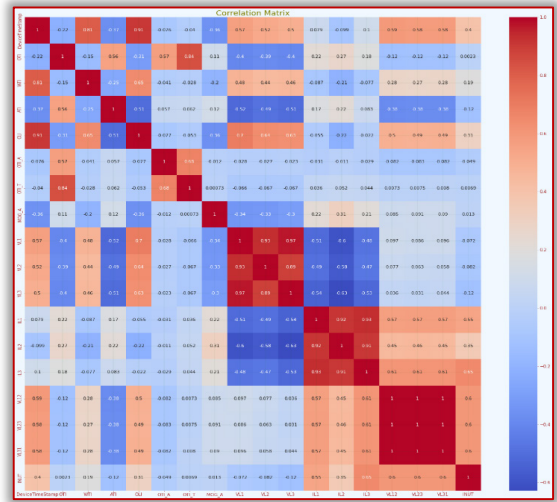


Figure 6: Correlation Matrix

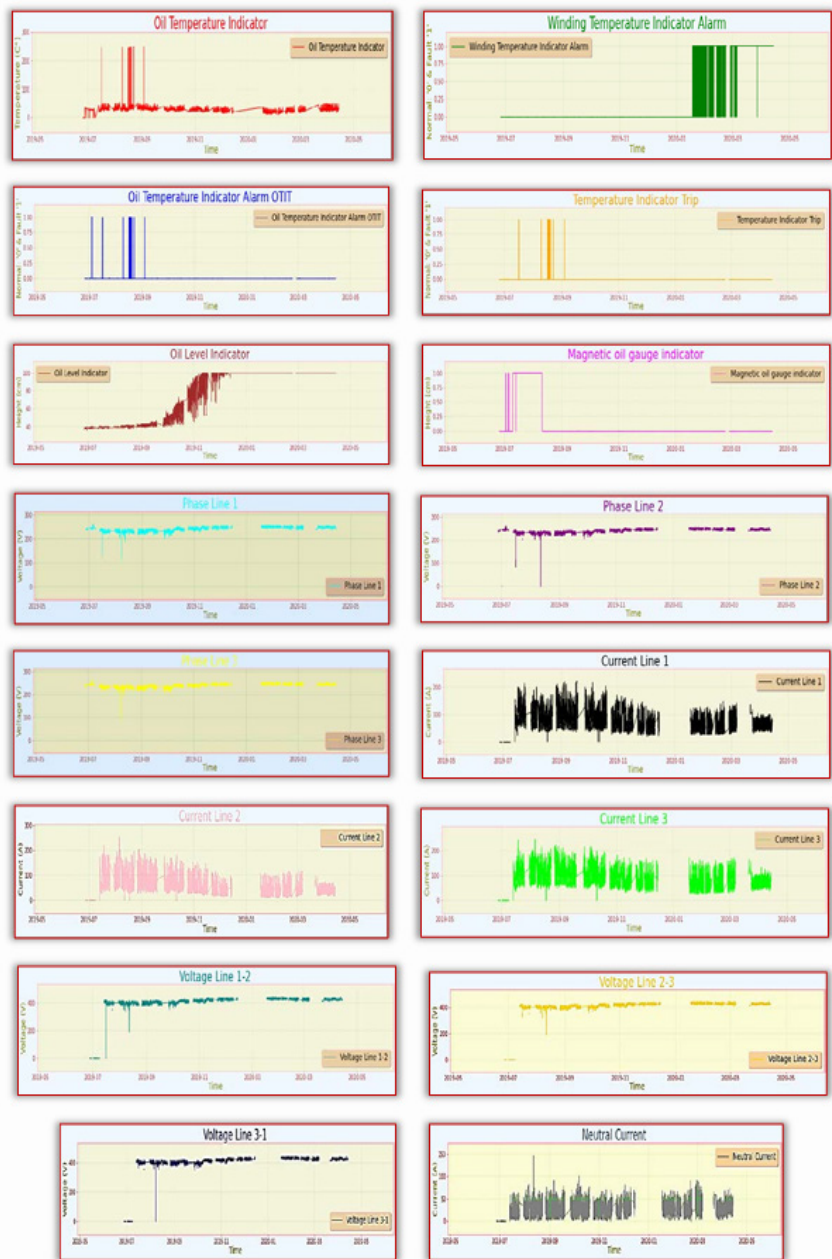


Figure 7: Data Visualization

## EXPLORATORY DATA ANALYSIS

Visualizations relating to the distribution of the data to be used for the modeling are provided below.

## RANDOM FOREST ALGORITHM

The Random Forest model demonstrates strong performance at lower thresholds, achieving an accuracy of 98.43%, F1 score of 98.45%, recall of 98.43%, and precision of 98.49% at a threshold of 0.5. These metrics slightly decrease to 97.99% accuracy, 97.94% F1 score, 97.99% recall, and 97.95% precision at a threshold of 0.7. However, a substantial performance drop is observed at a threshold of 0.9, with accuracy falling to 91.74%, F1 score to 89.01%, recall to 91.74%, and precision to 92.43%. While the ROC AUC consistently maintains a high value of 99.68% across all thresholds, indicating good overall discriminative power, the MCC (Matthews Correlation Coefficient) undergoes a dramatic decline from 99.55% at 0.5 to 88.46% at 0.7 and further to 39.84% at 0.9. This suggests a significant reduction in the model's predictive ability at higher thresholds.

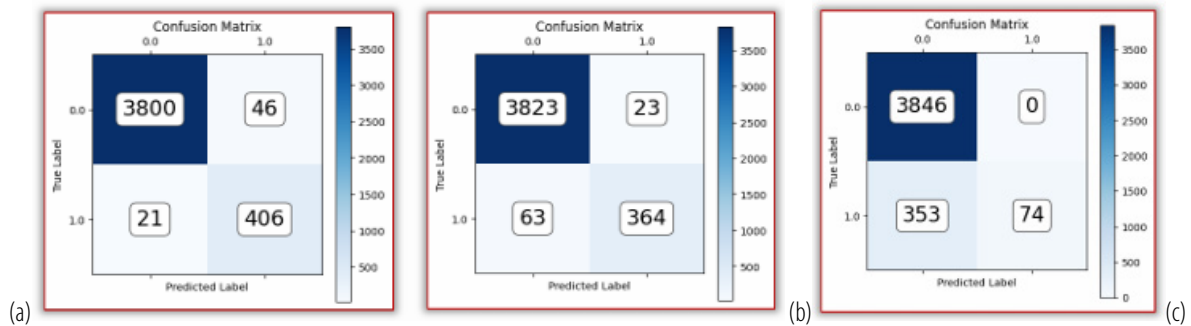


Figure 8: RF CM (a) at 0.5 (b) at 0.7 (c) at 0.9

Table 2: Random Forest Performance

Random forest performance summary at different threshold			
Measure	Value at 0.5	Value at 0.7	Value at 0.9
Accuracy	98.43%	97.99%	91.74%
F1 score	98.45%	97.94%	89.01%
Recall	98.43%	97.99%	91.74%
precision	98.49%	97.95%	92.43%
Roc	99.68%	99.68%	99.68%
MCC	99.55%	88.46%	39.84%

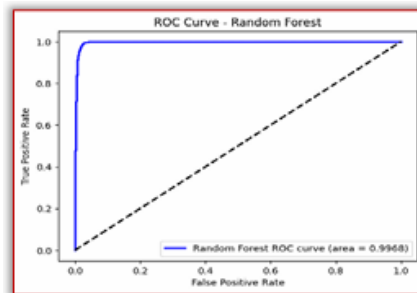


Figure 9: RF ROC

## XGBOOST ALGORITHM

The XGBoost model demonstrates consistent performance across different thresholds, maintaining high accuracy, F1 score, recall, and precision levels. At a threshold of 0.5, the model achieves an accuracy of 98.5%, F1 score of 98.51%, recall of 98.50%, and precision of 98.53%. While these metrics slightly decline to 98.36% accuracy, 98.34% F1 score, 98.36% recall, and 98.34% precision at a threshold of 0.7, and further to 97.52% accuracy, 97.40% F1 score, 97.52% recall, and 97.50% precision at a threshold of 0.9, the overall performance remains strong.

Table 3: XGBoost Performance at different thresholds

Measure	Value at 0.5	Value at 0.7	Value at 0.9
Accuracy	98.5%	98.36%	97.52%
F1 score	98.51%	98.34%	97.40%
Recall	98.50%	98.36%	97.52%
precision	98.53%	98.34%	97.50%
Roc	99.74%	99.74%	99.74%
MCC	91.83%	90.76%	85.50%

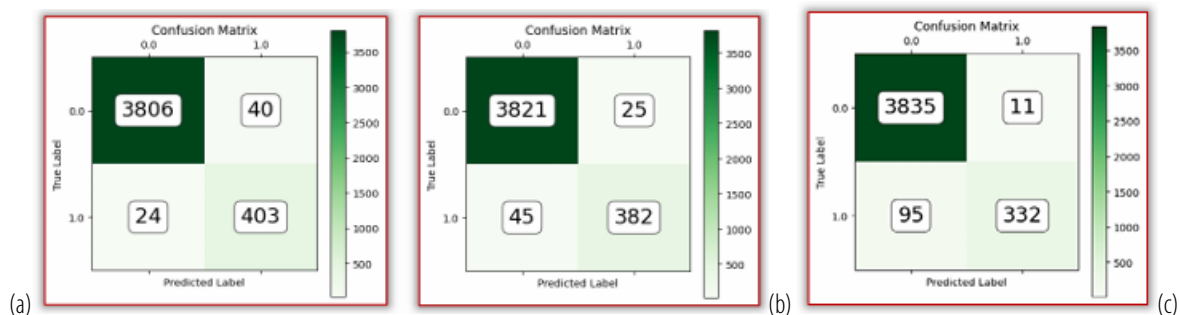


Figure 10: XGBoost CM (a) at 0.5 (b) at 0.7 (c) at 0.9



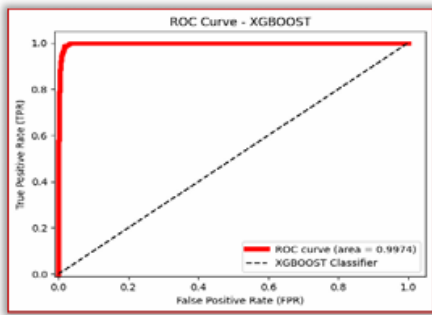


Figure 11: XGBoost ROC

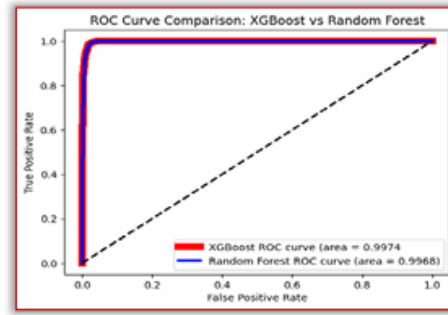


Figure 12: ROC Curve Comparison

The ROC AUC consistently stays at 99.74% across all thresholds, indicating excellent discriminative power. However, the MCC (Matthews Correlation Coefficient) shows a gradual decrease from 91.83% at 0.5 to 90.76% at 0.7 and 85.50% at 0.9, suggesting a slight reduction in the model's predictive ability as the threshold increases.

### MODEL COMPARISON

Accuracy, Confusion Matrix, AUROC, Precision, Recall and MCC are metrics used to evaluate the performance of both models. Both Random Forest and XGBoost models achieve high accuracy, F1 score, recall, and precision at a threshold of 0.5. However, the performance of the Random Forest model drops significantly at higher thresholds (0.7 and 0.9), while XGBoost maintains a more consistent performance across all thresholds. This is reflected in the MCC metric, which shows a much sharper decline for Random Forest compared to XGBoost.

Overall, XGBoost appears to be a more robust model, as it is less sensitive to the choice of the threshold and delivers consistently good performance across different thresholds. However, if a high true positive rate is critical and a higher false positive rate is acceptable, then the Random Forest model might be a good choice at a lower threshold. The ROC comparison of the models.

Table 4: Model Comparison

Measure \ Model	Random Forest	XGBoost	Random Forest	XGBoost	Random Forest	XGBoost
Threshold	0.5		0.7		0.9	
Accuracy	98.43	98.5	97.99	98.36	91.74	97.52
F1 score	98.45	98.51	97.94	98.34	89.01	97.4
Recall	98.43	98.5	97.99	98.36	91.74	97.52
Precision	98.49	98.53	97.95	98.34	92.43	97.5
ROC AUC	99.68	99.74	99.68	99.74	99.68	99.74
MCC	99.55	91.83	88.46	90.76	39.84	85.5

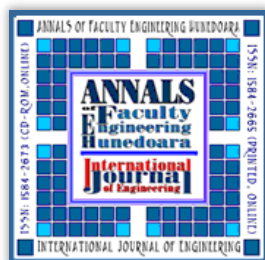
## 5. CONCLUSION

This study has demonstrated the feasibility and effectiveness of utilizing machine learning algorithms, specifically Random Forest and XGBoost, for predicting transformer failures using operational and historical data. The analysis revealed the superior performance of XGBoost across all evaluated metrics. While both models achieved high accuracy, precision, recall, and F1-score, XGBoost consistently outperformed Random Forest, particularly at higher thresholds. The significantly higher MCC value for XGBoost reinforces its superior predictive capability. These findings suggest that XGBoost is a more robust and reliable model for transformer failure prediction. By effectively leveraging its strengths, utilities can significantly enhance grid reliability, optimize maintenance schedules, and reduce operational costs associated with unplanned outages. Future research should explore additional algorithms and incorporate a broader range of data sources to further enhance predictive model robustness and improve electrical grid operations.

### References

- [1] Abbasi, J. A. (2021). Predictive Maintenance in Industrial Machinery using Machine Learning.
- [2] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1)
- [3] Breviglieri, P., Erdem, T., & Eken, S. (2021). Predicting Smart Grid Stability with Optimized Deep Models. *SN Computer Science*, 2(2)
- [4] Carratu, M., Gallo, V., Iacono, S. Dello, Sommella, P., Bartolini, A., Grasso, F., Ciani, L., & Patrizi, G. (2023). A Novel Methodology for Unsupervised Anomaly Detection in Industrial Electrical Systems. *IEEE Transactions on Instrumentation and Measurement*, 72

- [5] Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C. H., & Liu, R. (2019). XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System. *IEEE Access*, 7, 13149–13158
- [6] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1)
- [7] Fei, S. wei, & Zhang, X. bin. (2009). Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications*, 36(8)
- [8] Hussain, M. R., Refaat, S. S., & Abu-Rub, H. (2021). Overview and Partial Discharge Analysis of Power Transformers: A Literature Review. In *IEEE Access* (Vol. 9, pp. 64587–64605). Institute of Electrical and Electronics Engineers In
- [9] Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, 22(3), 341–354
- [10] Mohammed, S. (2017, November 11). Performance Metrics for Classification problems in Machine Learning. *Medium.Com*. <https://medium.com/>
- [11] Rojek, I., Jasiulewicz-Kaczmarek, M., Piechowski, M., & Mikołajewski, D. (2023). An Artificial Intelligence Approach for Improving Maintenance to Supervise Machine Failures and Support Their Repair. *Applied Sciences (Switzerland)*, 13(8)
- [12] Sreshta, P. (2020). Distributed Transformer Monitoring. *Distributed Transformer Monitoring*. <https://www.kaggle.com/datasets/sreshta140/ai-transformer-monitoring>
- [13] Tianjin da xue, Zhongguo dian ji gong cheng xue hui (Beijing, C., Guo jia dian wang gong si (China), IEEE Power & Energy Society, Institution of Engineering and Technology, International Council on Large Electric Systems, Institute of Electrical and Electronics Engineers, & International Conference on Electricity Distribution. Chinese National Committee, organizer. (2018). 2018 China International Conference on Electricity Distribution : proceedings : 17-19 September 2018, Tianjin, China.
- [14] Wang, T., Li, Q., Yang, J., Xie, T., Wu, P., & Liang, J. (2023). Transformer Fault Diagnosis Method Based on Incomplete Data and TPE-XGBoost. *Applied Sciences (Switzerland)*, 13(13)
- [15] Zhang, R., Li, B., & Jiao, B. (2019). Application of XGboost Algorithm in Bearing Fault Diagnosis. *IOP Conference Series: Materials Science and Engineering*, 490(7)
- [16] Zhang, Y., Ding, X., Liu, Y., & Griffin, P. J. (1996). An artificial neural network approach to transformer fault diagnosis. *IEEE Transactions on Power Delivery*, 11(4)



ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665

copyright © University POLITEHNICA Timisoara, Faculty of Engineering Hunedoara,  
5, Revolutiei, 331128, Hunedoara, ROMANIA

<http://annals.fih.upt.ro>