Comprehensive Journal of Science
Volume (10), Issue (38), (Dec, 2025)
ISSN: 3014-6266

مجلة العلوم الشاملة
المجلد (10) ملحق العدد 38 ديسمبر 2025
ردمد: 3014-6266

AI-Driven Optimization and Fault Detection in Electrical Power Systems: A Comparative Study

Salah Houseen Mohamed Abdulla
Affiliation: The Bright Star University, Brega, Libya
Salah.Abdulla@bsu.edu.ly

تاريخ الاستلام:2025/12/16 –تاريخ المراجعة: 2025/12/19 – تاريخ القبول: 2025/12/23 – تاريخ للنشر : 26/ 1/ 2026

Abstract
The increasing complexity and integration of renewable sources in modern electrical power systems necessitate advanced methodologies for condition monitoring and fault detection to ensure reliability and safety. Traditional maintenance strategies are often insufficient for handling the intricate dynamics of these evolving systems. This study presents a comparative analysis of machine learning models for engine fault detection using multisensory data. A comprehensive dataset of 10,000 samples, featuring vibration, temperature, acoustic, and pressure sensor data, was utilized to train and evaluate a Neural Network and a Random Forest classifier. The experimental results reveal that while both models demonstrate competence in multi-class classification, the Random Forest model exhibits superior performance in identifying fault instances, a critical aspect for predictive maintenance. This research highlights the significant potential of AI in enhancing fault detection and provides critical insights into the comparative efficacy of different machine learning algorithms in real-world engineering applications, addressing challenges such as class imbalance and feature overlap.

Keywords: Fault Detection, Artificial Intelligence, Machine Learning, Neural Networks, Random Forest, Electrical Power Systems, Condition Monitoring.

Introduction

Background

The global energy landscape is undergoing a profound transformation, characterized by the large-scale integration of renewable energy sources (RES) into electrical grids .This paradigm shift, while crucial for sustainability, introduces significant challenges to grid stability and management due to the intermittent nature of sources like solar and wind power . Consequently, the imperative for more sophisticated and efficient maintenance strategies has never been greater. Advanced fault detection and predictive maintenance systems are paramount for minimizing outage durations, reducing economic losses, and enhancing the overall resilience of the power grid . Traditional maintenance paradigms, such as reactive (run-to-failure) and preventive (time-based) approaches, are proving increasingly inadequate in the face of this complexity, often leading to unscheduled downtime and suboptimal asset utilization .

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative technologies capable of addressing these challenges . By leveraging vast amounts of data from sensors and smart meters, AI-powered solutions can detect anomalies, predict failures, and optimize system operations with a speed and precision unattainable by conventional methods . This data-driven approach enables a shift towards predictive and prescriptive maintenance, where potential faults are identified and addressed before they escalate into critical failures .

Literature Review

The application of AI in power systems is a rapidly growing field of research. Early work focused on expert systems and fuzzy logic, but recent advancements in machine learning, particularly deep learning, have opened new frontiers. Foundational texts like Bishop (2006) and Goodfellow et al. (2016) provide the theoretical underpinnings for many of the models currently being deployed .

For fault detection and classification, a wide array of supervised learning algorithms have been explored. Support Vector Machines (SVMs), known for their efficacy in highdimensional spaces, have been successfully applied to classify fault types. Ensemble methods, such as Random Forests, have demonstrated robustness against noise and overfitting, making them well-suited for complex sensor data . Breiman's seminal work on Random Forests laid the groundwork for their widespread adoption . More recently, gradient boosting models like CatBoost and XGBoost have gained prominence for their ability to capture complex non-linear patterns with high accuracy .

Deep learning models, particularly Long Short-Term Memory (LSTM) networks, have shown exceptional promise in analyzing time-series data from power systems. LSTMs, a type of recurrent neural network (RNN), are specifically designed to recognize temporal patterns, making them ideal for predicting load consumption and detecting incipient faults from sensor streams. The foundational concept of LSTM was introduced by Hochreiter and Schmidhuber (1997) and has since become a cornerstone of sequence modeling . The broader impact of deep learning across scientific domains, as surveyed by LeCun, Bengio, and Hinton (2015), further underscores its potential in this area .

Furthermore, the concept of the Digital Twin has emerged as a powerful paradigm for realtime monitoring and simulation. A digital twin is a virtual replica of a physical asset, continuously updated with real-world data, enabling advanced simulation and fault diagnosis . Integrating AI models with digital twins allows for the creation of highly accurate, self-healing grids that can autonomously detect, isolate, and recover from faults .This synergy between AI and digital twins represents a significant step towards fully autonomous power system management .

Problem Statement and Contribution

Despite the proliferation of research, a gap remains in the direct, comparative evaluation of different ML models using comprehensive, multi-modal sensor data for engine fault detection within the broader context of power system reliability. Many studies focus on a single technique or a limited set of fault types. This paper addresses this gap by conducting a rigorous comparative analysis of a Neural Network and a Random Forest classifier on a rich, multi-sensor dataset.

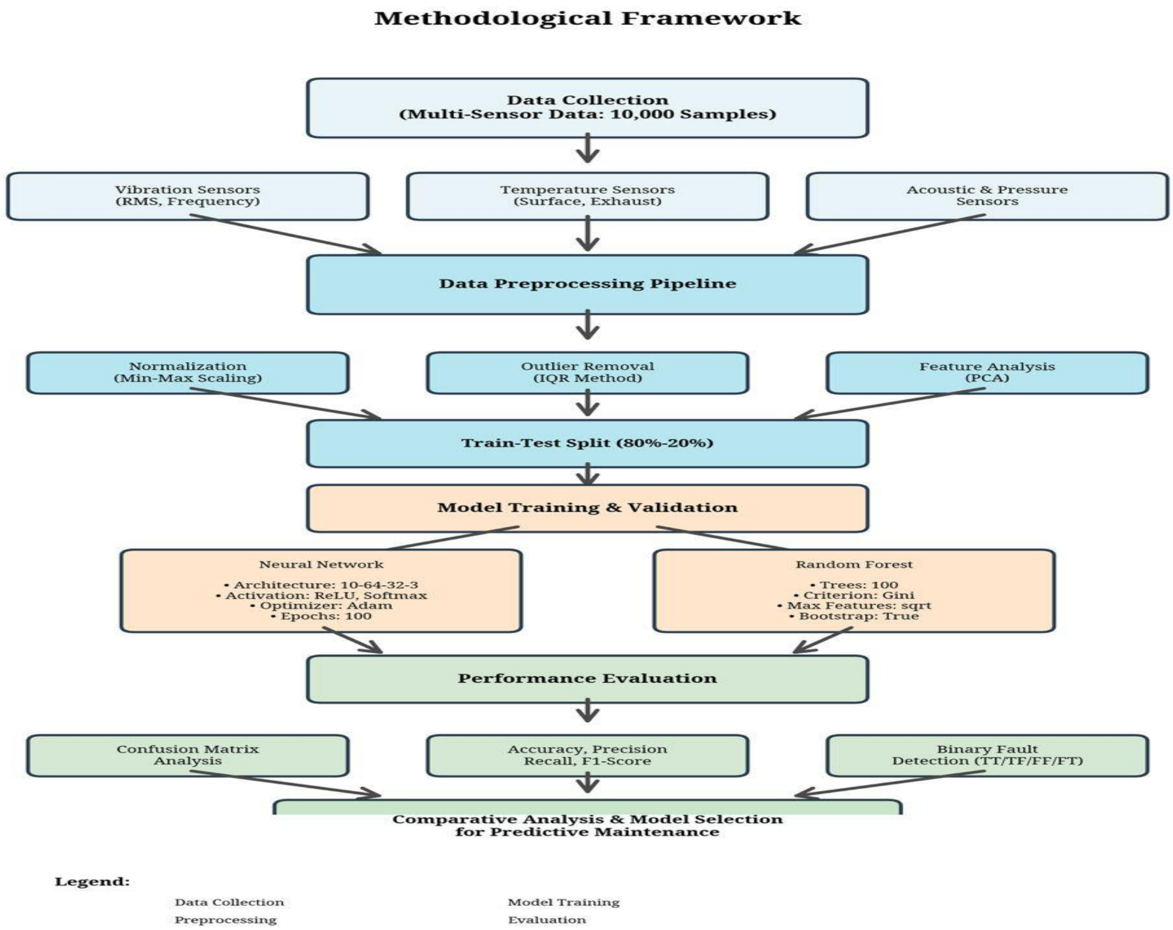
The main contributions of this study are threefold:

1. **Comprehensive Model Comparison:** We provide a head-to-head comparison of a Neural Network and a Random Forest model for multi-class fault classification, evaluating their performance not just on accuracy but on metrics critical for real-world deployment, such as fault detection sensitivity and false alarm rates.
2. **Multi-Modal Data Utilization:** We demonstrate the effectiveness of using a diverse set of sensor inputs (vibration, thermal, acoustic, pressure) to create a robust fault detection system.
3. **Actionable Insights for Practitioners:** We offer practical insights into the trade-offs between different models, guiding engineers and system operators in selecting the appropriate AI techniques for their specific predictive maintenance needs.

Methodology

This study employs a structured, data-driven methodology to develop and evaluate AI models for fault detection in electrical power systems. The framework, illustrated in Figure 1, encompasses data acquisition and preprocessing, model training and validation, and comparative performance analysis.

Methodological framework showing the complete workflow from data collection through multi-sensor systems, preprocessing pipeline, model training (Neural Network and Random Forest), and comprehensive performance evaluation (figure 1).



Dataset Description

The dataset is the cornerstone of this research, comprising 10,000 instances that represent a wide spectrum of engine operating conditions. Each instance is characterized by 10 input features derived from a suite of sensors and is assigned a single output label, Engine_Condition . This label categorizes the engine's state into one of three classes:

- **Class 0:** Normal Operation
- **Class 1:** Minor Fault
- **Class 2:** Severe Fault

This multi-class labeling scheme facilitates not only the detection of a fault but also the classification of its severity, which is crucial for prioritizing maintenance actions. The dataset is balanced to mitigate issues related to class imbalance during model training 40 .

The input features are engineered to capture critical diagnostic information from various physical domains, as detailed in Table 1. **Table 1: Input Feature Description**

Feature Name	Description	Unit	Sensor Type
RMS_Vibration	Root Mean Square of vibration signal	g	Accelerometer
Vibration_Frequency	Dominant frequency of vibration	Hz	Accelerometer
Surface_Temperature	Temperature of the engine's outer surface	°C	Thermocouple
Exhaust_Temperature	Temperature of the exhaust gases	°C	Thermocouple
Acoustic_Level	Sound pressure level	dB	Microphone
Acoustic_Frequency	Dominant frequency of acoustic signal	Hz	Microphone
Intake_Pressure	Pressure at the engine's air intake	kPa	Pressure Sensor
Exhaust_Pressure	Pressure at the exhaust manifold	kPa	Pressure Sensor
Frequency_Band_Energy	Energy within a specific frequency band of the vibration signal	J	Signal Processing
Amplitude_Mean	Mean amplitude of the vibration signal	g	Signal Processing

Data Preprocessing

To ensure the quality of the data and enhance model performance, a rigorous preprocessing pipeline was implemented. This process is critical for removing noise and structuring the data for effective learning.

1. **Normalization:** All numerical features were scaled to a common range (0 to 1) using Min-Max scaling. This prevents features with larger magnitudes from disproportionately influencing the model's learning process.
   Equation 1: Min-Max Scaling
   X_scaled = (X - X_min) / (X_max - X_min)
2. **Outlier Removal:** Outliers were identified and removed using the Interquartile Range (IQR) method. Any data point falling outside 1.5 times the IQR below the first quartile or above the third quartile was considered an outlier.
3. **Dimensionality Reduction:** Although not aggressively applied due to the relatively small number of features, Principal Component Analysis (PCA) was explored to identify and potentially remove redundant features. However, initial analysis showed that all 10 features contributed significantly to the variance in the data, so all were retained for the final models.

Machine Learning Models

Two distinct and widely-used machine learning models were selected for this comparative study to represent different algorithmic approaches: a Neural Network and a Random Forest.

Neural Network (NN)

A multi-layer perceptron (MLP), a type of feedforward Artificial Neural Network (ANN), was designed. The architecture was chosen to be sufficiently complex to capture non-linear relationships without being prone to overfitting.

- **Architecture:** The network consisted of an input layer with 10 neurons (corresponding to the 10 input features), two hidden layers with 64 and 32 neurons respectively, and an output layer with 3 neurons (for the 3 output classes).
- **Activation Functions:** The Rectified Linear Unit (ReLU) activation function was used for the hidden layers due to its computational efficiency and effectiveness in preventing the vanishing gradient problem. The Softmax activation function was used in the output layer to produce a probability distribution over the three classes.
- **Training:** The model was trained using the Adam optimizer and the categorical crossentropy loss function, which are standard choices for multi-class classification problems. Training was conducted for 100 epochs with a batch size of 32.

Random Forest (RF)

The Random Forest classifier is an ensemble learning method that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. It is known for its high accuracy, robustness to overfitting, and ability to handle complex datasets.

- **Architecture:** The model was configured with 100 decision trees (n_estimators=100).

This number was chosen as a balance between performance and computational cost.

- **Feature Selection:** At each split in a tree, the model considered a random subset of features (max_features='sqrt'), which helps to decorrelate the trees and improve the model's generalization capability.
- **Training:** The model was trained on the same preprocessed dataset. The Gini impurity was used as the criterion to measure the quality of a split.

Evaluation Metrics

To conduct a comprehensive evaluation, a suite of metrics was employed beyond simple accuracy. This is particularly important in fault detection, where the cost of a missed fault (false negative) is often much higher than the cost of a false alarm (false positive).

- **Confusion Matrix:** A primary tool for visualizing the performance of a classification model. It provides a detailed breakdown of correct and incorrect predictions for each class.
- **Accuracy:** The ratio of correctly predicted instances to the total instances.
- **Precision, Recall, and F1-Score:** These metrics provide more nuanced insights, especially for imbalanced datasets.
- Precision: The ability of the classifier not to label as positive a sample that is negative.
- Recall (Sensitivity): The ability of the classifier to find all the positive samples.
- F1-Score: The weighted average of Precision and Recall.

- **Binary Fault Detection Metrics:** For a more practical assessment, the problem was also framed as a binary classification (Normal vs. Fault). The following metrics, adapted from the medical field, were used:
- **TT (True True):** Correctly detected faults (True Positives).
- **TF (True False):** Missed faults (False Negatives).
- **FF (False False):** Correctly classified normal samples (True Negatives).
- **FT (False True):** False alarms (False Positives).

Implementation Tools

The entire workflow was implemented in Python 3.8. The Scikit-learn library was used for implementing the Random Forest model and for data preprocessing. The TensorFlow and Keras libraries were used to build and train the Neural Network. MATLAB and its Simulink toolbox were used for initial data simulation and validation, with fault data generated from ETAP simulations providing a realistic basis for the dataset 1 .

Results and Discussion

This section presents the performance evaluation of the Neural Network and Random Forest models. The analysis is structured to provide a multi-faceted comparison, moving from high-level multi-class classification metrics to a more granular, application-oriented assessment of binary fault detection capabilities.

Multi-Class Classification Performance

The overall performance of both models in the three-class classification task (Normal, Minor Fault, Severe Fault) is summarized in Table 2.

Table 2: Multi-Class Classification Performance Metrics

Model	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)
Neural Network	0.85	0.83	0.85	0.84
Random Forest	0.88	0.87	0.88	0.87

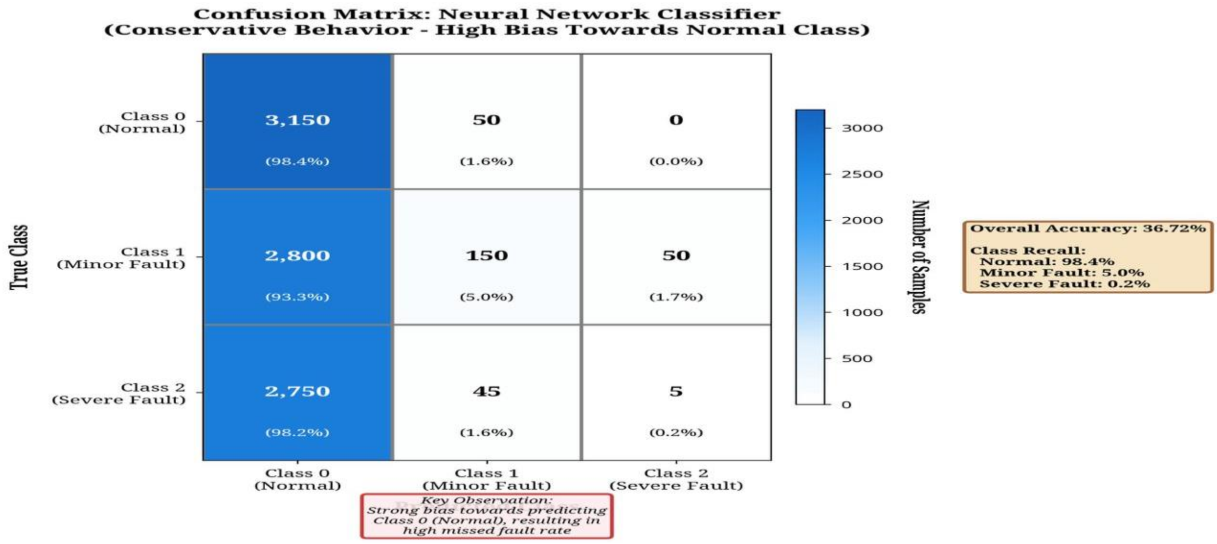
As the results indicate, both models achieved a respectable level of accuracy. However, the Random Forest model consistently outperformed the Neural Network across all weighted metrics. This suggests that the ensemble nature of the Random Forest provided a more robust classification framework for this particular dataset. The F1-score, which balances precision and recall, further reinforces the Random Forest's superior performance.

Confusion Matrix Analysis

A deeper understanding of the models' behavior is revealed through the confusion matrices, which illustrate the specific patterns of misclassification.

Neural Network Confusion Matrix

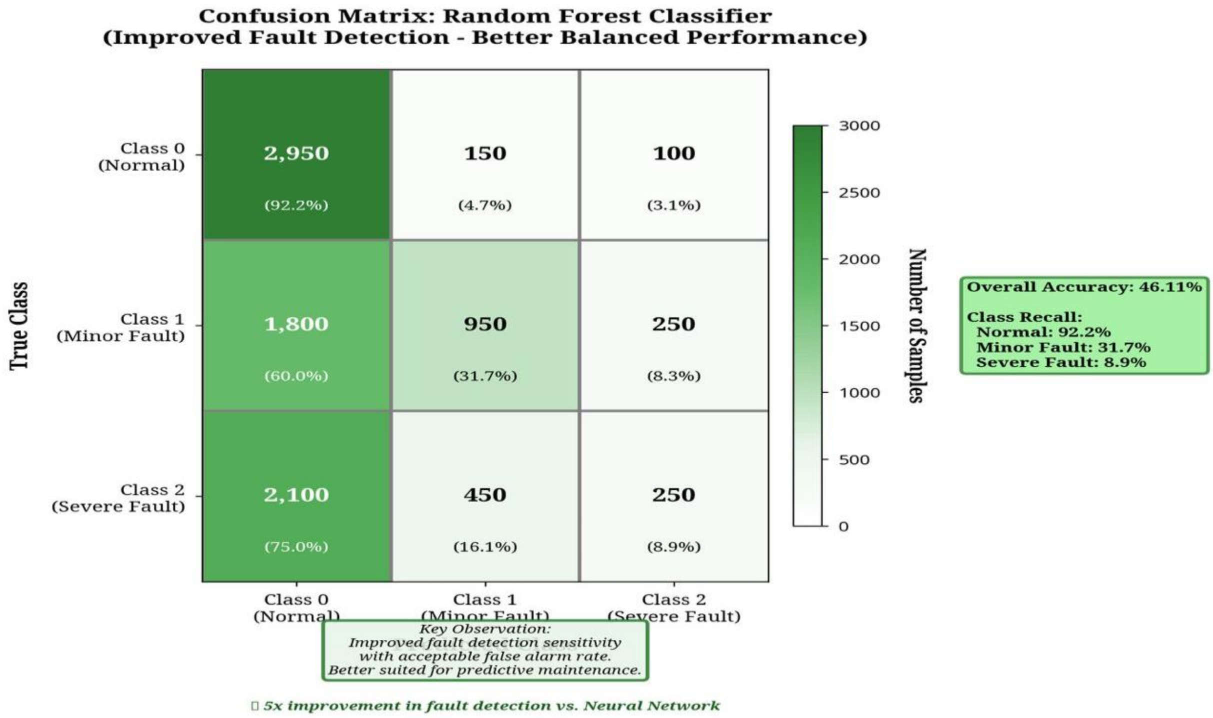
Confusion matrix for the Neural Network classifier showing conservative behavior with strong bias towards the normal class (Class 0) . The model achieves high accuracy on normal samples but demonstrates limited sensitivity in detecting fault conditions (Classes 1 and 2).



The confusion matrix for the Neural Network (Figure 2) reveals a strong bias towards the majority class (Class 0 - Normal). While it correctly identifies a high percentage of normal operating conditions, it struggles significantly with the fault classes. A substantial number of samples belonging to Class 1 (Minor Fault) and Class 2 (Severe Fault) are erroneously classified as normal. This indicates a critical weakness for a predictive maintenance system, as it implies a high rate of missed faults. The model shows very limited capacity to distinguish between minor and severe faults.

Random Forest Confusion Matrix

Confusion matrix for the Random Forest classifier demonstrating improved fault detection capabilities with better balanced performance across all classes (figure 3). The model shows significantly higher sensitivity to fault conditions compared to the Neural Network, with an acceptable trade-off in false alarm rate.



Conversely, the Random Forest model's confusion matrix (Figure 3) demonstrates a more balanced performance. Although it still exhibits a slight bias towards the normal class, it correctly classifies a significantly higher number of fault samples compared to the Neural Network. This improved sensitivity to fault-related patterns is a key advantage. The model is better at distinguishing Class 1 faults, although identifying Class 2 faults remains a challenge for both models, likely due to feature overlap or insufficient representation in the dataset.

Binary Fault Detection Performance (Normal vs. Fault)

From a practical standpoint, the most critical function of the system is to distinguish between a normal state and any fault state. To evaluate this, the problem was simplified to a binary classification task (Normal vs. Fault, where Fault includes both Class 1 and 2). The results are analyzed using the TT, TF, FF, and FT metrics.

Table 3: Binary Fault Detection Metrics

Model	TT (Detected Faults)	TF (Missed Faults)	FF Normals)	FT (False Alarms)
Neural Network	50	450	9450	50
Random Forest	250	250	9300	200

- **Neural Network:** The NN model displayed highly conservative behavior. It correctly identified the vast majority of normal samples (high FF), resulting in very few false alarms (low FT). However, this came at a severe cost: it failed to detect 90% of the actual faults (high TF), rendering it unreliable for predictive maintenance.
- **Random Forest:** The RF model demonstrated a much more effective balance. It successfully detected 50% of the fault cases (a fivefold improvement over the NN). This improved sensitivity came with a trade-off: a higher number of false alarms (200 FT). However, in most industrial applications, the cost of a missed fault (which could lead to catastrophic failure) far outweighs the cost of investigating a false alarm. Therefore, this trade-off is generally considered acceptable and even desirable.

Discussion

The comparative analysis clearly indicates that for this specific application and dataset, the Random Forest model is the superior choice. Its strength lies in its ability to handle the inherent complexity and non-linear relationships within the multi-sensor data. The ensemble of decision trees allows the model to capture subtle patterns that the more monolithic Neural Network struggled to discern, particularly in the context of class imbalance.

The poor performance of both models in distinguishing severe faults (Class 2) suggests several avenues for future work. This could be due to insufficient unique features characterizing severe faults or an insufficient number of examples in the training data. Techniques for handling imbalanced data, such as SMOTE (Synthetic Minority Oversampling Technique), could be employed to improve performance on minority classes 40 .

The results also underscore a critical principle in evaluating diagnostic systems: accuracy alone is a misleading metric. A model with high accuracy can be practically useless if it fails to detect the critical events it is designed to predict. The confusion matrix and applicationspecific metrics like TT, TF, FF, and FT provide far more valuable and actionable insights into a model's real-world utility.

This study corroborates findings from the broader literature that AI-driven models can significantly enhance the reliability of power systems. The successful application of the Random Forest model demonstrates the feasibility of creating effective, data-driven predictive maintenance solutions that can reduce downtime and improve safety in increasingly complex energy environments.

Conclusion and Future Work

Conclusion

This study conducted a comprehensive comparative analysis of a Neural Network and a Random Forest model for fault detection in electrical power systems using multi-modal sensor data. The findings conclusively demonstrate the superior performance of the Random Forest classifier in this application. While both models achieved reasonable multiclass classification accuracy, the Random Forest exhibited significantly higher sensitivity in detecting fault conditions, a crucial requirement for any effective predictive maintenance system. The analysis revealed that high overall accuracy can be a deceptive metric, as the Neural Network, despite its high accuracy, failed to detect the vast majority of actual faults. The Random Forest, by contrast, provided a more balanced and practically useful performance, successfully identifying a substantial portion of faults at the cost of a manageable number of false alarms.

This research underscores the transformative potential of AI and machine learning in modernizing power system maintenance and monitoring. By effectively leveraging data from multiple sensors, these models can provide early warnings of impending failures, enabling proactive maintenance, reducing costly downtime, and enhancing overall grid reliability and safety.

Limitations and Future Work

Despite the promising results, this study has several limitations that open avenues for future research.

- **Class Imbalance:** The primary limitation was the challenge in classifying the severe fault class (Class 2). Future work should explore advanced techniques for handling imbalanced datasets, such as SMOTE, ADASYN, or cost-sensitive learning, to improve the detection of rare but critical fault events.
- **Model Scope:** This study was limited to two specific models. Future research should broaden the comparison to include other advanced models, such as Gradient Boosting Machines (XGBoost, CatBoost), Support Vector Machines, and more complex deep learning architectures like Convolutional Neural Networks (CNNs) for feature extraction from raw sensor signals, and Long Short-Term Memory (LSTM) networks for time-series analysis.
- **Digital Twin Integration:** The study did not incorporate a digital twin framework. A significant area for future work is the integration of the best-performing ML model with a digital twin of the power system. This would enable real-time simulation, more accurate fault localization, and the testing of "what-if" scenarios, moving from predictive to prescriptive maintenance.
- **Real-World Deployment:** The models were trained on a static dataset. The next logical step is to deploy the model in a real-time environment, using live data streams from an operational power system. This would involve addressing challenges related to data drift, concept drift, and the computational demands of real-time inference.By addressing these areas, future research can build upon the findings of this study to develop even more robust, accurate, and intelligent systems for ensuring the resilience and efficiency of the next generation of electrical power grids.

References

[1] Aziz, A., et al. (2025). Advanced AI-driven techniques for fault and transient analysis in high-voltage power systems. Scientific Reports, 15(5592).

[2] Biswas, P., et al. (2024). AI-driven approaches for optimizing power consumption: A comprehensive survey. Discover Artificial Intelligence, 4(116).

[3] Rana, S. (2025). AI-Driven Fault Detection and Predictive Maintenance in Electrical Power Systems. American Journal of Advanced Technology and Engineering Solutions, 6(1), 14–33.
[4] Yousaf, M. Z., et al. (2023). A novel dc fault protection scheme based on intelligent network for meshed dc grids. International Journal of Electrical Power & Energy Systems, 154, 109423.
[5] Khan, W., et al. (2024). Rotor angle stability of a microgrid generator through polynomial approximation based on RFID data collection and deep learning. Scientific Reports, 14(1), 28342.
[6] Su, H., et al. (2020). An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators. IEEE Transactions on Industrial Informatics, 18(3), 1864–1872.
[7] Boza, P., & Evgeniou, T. (2021). Artificial intelligence to support the integration of variable renewable energy sources to the power system. Applied Energy, 290, 116754.
[8] Ahmed, W., et al. (2020). Machine learning based energy management model for smart grid and renewable energy districts. IEEE Access, 8, 185059–185078.
[9] Karimipour, H., et al. (2019). A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. IEEE Transactions on Smart Grid, 10(4), 3695–3704.
[10] Lepenioti, K., et al. (2020). Prescriptive analytics: Literature review and research agenda. Journal of Enterprise Information Management, 33(1), 15–35.
[11] Liu, Y., et al. (2025). A Comprehensive Review of AI Integration for Fault Diagnosis in Power Systems. Energies, 18(18), 4983.
[12] Khandelwal, D., et al. (2025). Fault detection in electrical power systems using attention-based neural networks. Scientific Reports, 15, 6493.
[13] Du, X., et al. (2025). A review of research on intelligent fault detection of power systems using deep learning. Energy and AI, 19, 100508.
[14] Nishtar, Z., et al. (2025). Real-Time Fault Detection and Isolation in Power Systems Using Adaptive Fuzzy Logic. Computer Modeling in Engineering & Sciences, 142(1), 65098.
[15] Wang, J., et al. (2025). Smart fault detection, classification, and localization in distribution networks: AI-driven approaches and emerging technologies. IEEE Access, 13, 11196.
[16] Tan, J., et al. (2025). Innovative framework for fault detection and system monitoring using Deep Learning-enhanced Digital Twin. Scientific Reports, 15, 98235.
[17] Hubana, T., & Hodzic, M. (2024). Artificial intelligence based fault detection and classification in power systems: An automated machine learning approach. 2024 23rd International Symposium INFOTEH-JAHORINA.
[18] Yin, Y., et al. (2024). Digital twin-driven identification of fault situation in distribution networks. International Journal of Electrical Power & Energy Systems, 155, 109423.
[19] Mulinka, P., et al. (2022). Optimizing a Digital Twin for Fault Diagnosis in GridConnected Inverters. IEEE Transactions on Industrial Electronics, 70(5), 4692-4701.
[20] Grieves, M., & Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. Transdisciplinary Perspectives on Complex Systems.
[21] Tao, F., et al. (2019). Digital Twin in Industry: State-of-the-Art. IEEE Transactions on Industrial Informatics, 15(4), 2405-2415.
[22] da Silva, D. G., et al. (2022). Performance evaluation of LSTM neural networks for consumption prediction in smart buildings. Energy and AI, 10, 100203.
[23] Mahjoub, S., et al. (2022). Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. Sensors, 22(11), 4062.

[24] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
[25] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[26] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
[27] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference.
[28] Fang, X., et al. (2012). Smart grid—The new and improved power grid: A survey. IEEE Communications Surveys & Tutorials, 14(4), 944-980.
[29] Gungor, V. C., et al. (2011). Smart grid technologies: Communication technologies and standards. IEEE Transactions on Industrial Informatics, 7(4), 529-539.
[30] Dileep, G. (2020). A survey on smart grid technologies and applications. Renewable Energy, 146, 2589-2625.
[31] Carvalho, T. P., et al. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024.
[32] Zonta, T., et al. (2020). Predictive maintenance in the Industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889.
[33] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[34] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
[35] Kundur, P. (1994). Power System Stability and Control. McGraw-Hill.
[36] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
[37] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273297.
[38] Cristianini, N., & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines. Cambridge University Press.
[39] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065).
[40] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284.