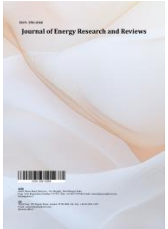

Journal of Energy Research and Reviews

Volume 18, Issue 4, Page 14-24, 2026; Article no.JENRR.155065
ISSN: 2581-8368

Enhancing Power System Reliability with Machine Learning-enabled Fault Management

Rahman Md. Mahidur [a], Nayem Zannatun [a], Hossain Md. Sajjad [a] and Shahiduzzaman Kazi Md. [a*]

[a] Department of Electrical and Electronics Engineering, Faculty of Science and Engineering, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/jenrr/2026/v18i4504

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/155065

Original Research Article

Received: 11/01/2026
Published: 25/03/2026

Abstract

Aims: To develop and evaluate a machine learning-based framework for the accurate detection, classification, and localization of faults in a power transmission system, with the goal of enhancing system reliability and reducing maintenance delays.
Study Design: Simulation-based experimental modeling study.
Place and Duration of Study: Department of Electrical and Electronics Engineering, Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh, Bangladesh. *(Note: The exact duration/dates of the study are not specified in the provided text).*
Methodology: A 4-bus power transmission system was simulated using MATLAB Simulink to generate a diverse, realistic dataset of 4,440 samples. This included 4,000 non-fault samples and 440 fault samples representing 11 distinct fault types across four buses. Four machine learning models—Support Vector

**Corresponding author: E-mail: shahiduzzaman@jkkniu.edu.bd;*

Cite as: *Mahidur, R. M., Zannatun, N., Md. Sajjad, H., & Kazi Md., S. (2026). Enhancing Power System Reliability with Machine Learning-enabled Fault Management. Journal of Energy Research and Reviews, 18(4), 14–24. https://doi.org/10.9734/jenrr/2026/v18i4504*

Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and K-Nearest Neighbors (KNN)—were trained and evaluated. Their performance was assessed using standard metrics, including accuracy, precision, recall, and F1-score.

Results: The Random Forest model achieved the highest performance, recording an overall accuracy of 96%. It proved to be the most effective model for handling complex fault patterns with minimal misclassifications, consistently outperforming the SVM (95% accuracy), KNN (94% accuracy), and LSTM (93% accuracy) models.

Conclusion: The proposed framework provides a robust, reliable tool for power system engineers to quickly and accurately identify, classify, and localize transmission line faults. Utilizing this approach can improve predictive maintenance, drastically reduce system downtime, and enhance overall power grid safety and efficiency.

Keywords: Fault detection; power system faults; machine learning; random forest; fault classification; predictive maintenance.

1. Introduction

Fault detection must be done correctly and on time to ensure that power systems operate safely and consistently. Failures may have a negative impact not just on the system and crew's safety, but also on revenue. Faults in power systems are unavoidable (Ghaderi et al., 2017). As a result, it is critical to participate in initiatives that support defect detection and classification.

An electric power system network (EPSN) comprises three main zones: generation, transmission, and distribution (Morais et al., 2010). Due to rising electricity consumption, the EPSN is becoming more advanced and vulnerable to power interruptions or malfunctions (Arouche Freire et al., 2019). Transmission lines (TL) account for 80% of power system problems across the EPSN because they are exposed to a variety of environmental, animal, and human contacts (Aleem et al., 2015; Chen et al., 2020). Most of the TL's failures are short circuit faults (Chen et al., 2020). When faults occur, the voltage and current signals of a three-phase TL deviate from their reference values, which can lead to serious consequences if not corrected promptly (Saber et al., 2018). As a result, fault analysis has become a valuable research tool for power engineers.

During daily operation, many short-circuit faults occur, which are usually classified as symmetrical or unsymmetrical. Symmetrical or balanced faults, which keep the system balanced, are made up of triple line-to-ground (LLLG) and triple line (LLL) faults, which have a low occurrence rate but are the most severe type of short-circuit fault due to their large impact and damage to system equipment. Asymmetrical or unbalanced faults that cause the electrical system to become unbalanced during the fault are double-line-to-ground (LLG), line-to-ground (LG), and line-to-line (LL) faults. Although less severe than balanced ones, they have a higher occurrence probability of 0.80 due to single-line-to-ground faults (Belagoune et al., 2021; He et al., 2014). Short circuit issues have several negative effects on electrical distribution systems and equipment. Some negative effects include reduced device lifespan, increased power losses, and excess heat generated by cables, wires, insulators, and transformers. When a short circuit occurs in a transmission line, the first thing protection relays do is cause a power outage, which interrupts power delivery to consumers. Thus, it is essential to use methods or technologies that can quickly and accurately assess problem locations (Fathabadi, 2016). Fault detection, diagnosis, and localization are crucial for maintaining the continuous and reliable operation of power systems. Advancements in signal processing, artificial intelligence, and machine learning have enabled researchers to develop more comprehensive and specialized approaches to fault protection, particularly in modern smart grid and energy systems (Al-Shourbaji & Alameen, 2025).

Conventional fault diagnosis techniques often struggle with limited data availability, delayed response, and diminished reliability under complex fault conditions (Raza et al., 2020). To address these limitations, machine learning algorithms have attracted significant attention for power system fault detection and classification; however, many existing approaches still struggle to achieve both accurate classification and precise localization across diverse operating conditions (Terron-Santiago et al., 2021). Machine learning frameworks are based on the concept that systems should be trained using statistical data and mathematical models to detect fault patterns with little human intervention. Recent studies have also shown that hybrid optimization-based machine learning models can significantly improve prediction accuracy in complex energy systems (Raidid et al., 2023).

As a result of advances in the integration of intelligent electronics into smart grids, the use of cutting-edge machine learning algorithms with large datasets is becoming increasingly important. This will open the way for the adoption of precise and reliable machine learning architectures for detecting abnormal situations (Raja et al., 2022; Khan et al., 2023). Several studies have explored the suitability of neural networks and other machine learning models for power system fault diagnosis. These techniques have demonstrated the ability not only to detect faults but also to determine their type and severity, thereby improving the overall reliability of the system (Belagoune et al., 2021; Shakiba et al., 2023; Sahoo & Samal, 2023; Shakiba et al., 2022).

This work proposes a machine learning-based framework for fault detection, classification, and localization in transmission lines. The approach leverages inherent patterns in voltage and current signals, enabling data-driven models to distinguish between normal and faulty conditions with high accuracy. Improving the system can lead to cost and time savings. The dataset for this study was created using MATLAB Simulink to design a 4-bus power system. It includes statistical methods for phase voltage and current in transmission lines under both fault and non-fault conditions. The dataset contains 4,440 samples: 1,000 Non-Fault (NF) samples and 10 samples for each of the 11 fault classes across four buses. Faults are rare in practical settings, as demonstrated by the dataset used in this study. Out of 4,440 samples, 4,000 are in no-fault conditions, whereas only 440 are in fault conditions, with 10 instances of each fault category. This approach is more representative of real-world settings, making the model more practical.

Unlike many previous studies that rely on large but often impractical datasets, this work emphasizes a realistic data distribution that reflects actual power system conditions, where fault events are relatively rare, thereby improving practical applicability and generalizability. This motivation guides the design and development of the proposed framework. The main contributions of this research are listed below:

i. Gather and prepare data from the simulation model.
ii. Differentiate between fault and non-fault conditions in transmission lines.
iii. Identify the primary failure types in power system transmission lines.
iv. Validate and assess the defect detection method.
v. Identify the problematic buses to locate the fault source.
vi. Demonstrate the potential of machine learning algorithms in analyzing large datasets to detect patterns and anomalies indicating faults in transmission lines.

2. Literature Review

Our research focuses on detecting, classifying, and localizing faults in power transmission lines using four different machine learning models. There are many AI-based systems for detecting faults in power transmission lines. In 2011, a pattern recognition approach was introduced for fault localization in electric power systems. The paper analyzed fault components and sections in big power systems. The technology is designed only for single-line-to-ground and three-phase short-circuit faults (Zhang et al., 2011). In 2016, this study (Chawardol & Sheikh, 2016) proposed using back-propagation-based neural network propagation to classify transmission line faults and identify fault zones. It improves accuracy in recognizing zones. The accuracy of classifying symmetrical and unsymmetrical defects is not good. In 2019, an integrated network with dispersed generators was proposed to detect faults in power systems through machine learning techniques. This research combines the Discrete Wavelet Transform (DWT) and the Support Vector Machine (SVM) to detect and classify faults in integrated power systems, achieving excellent results (Moloi et al., 2019). In 2021, a machine learning strategy was developed that uses Long Short-Term Memory (LSTM) and synchronized phasor measurements to detect faults quickly and efficiently, even in the presence of communication network delays (Rafique et al., 2021). In 2022, a hybrid method combining Wavelet Transform (WT) and Convolutional Neural Networks (CNN) was developed for defect detection and classification. Simulations show high accuracy and processing speed (Bon & Dai, 2022). In 2023, a strategy was presented that uses a MATLAB Simulink dataset to achieve high efficiency. Compared to typical ANN models, this approach reduces both computational complexity and processing time. The technique increased detection and classification accuracy (Goni et al., 2023).

3. Methodology

3.1 Proposed Framework

This study focuses on detecting, characterizing, and localizing faults in electric transmission lines. Due to difficulties in acquiring real-time failure data from operating power systems, a simulation-based approach was used. The proposed framework includes simulation, preprocessing, model training, and prediction stages, as shown in Figure 1. The generation of a robust and diverse dataset serves as the foundation for this investigation. To simulate real-world operational conditions, a four-bus power system is modeled in Simulink. Various fault types, including single-, two-, and three-phase faults, are deliberately incorporated to ensure the dataset covers a wide range of fault types and conditions. Non-fault (healthy) states are also included to help the model distinguish between normal and faulty states. Each fault type is designed to mimic realistic situations, providing a comprehensive representation of potential faults in power transmission networks. The created dataset includes time-series data for the three-phase voltages (Va, Vb, Vc) and currents (Ia, Ib, Ic), which are critical for properly diagnosing and localizing system defects. Before training machine learning models, the raw dataset is preprocessed to improve its quality and applicability. Preprocessing involves eliminating unwanted noise to provide clean, relevant signals, standardizing voltage and current values to maintain uniformity across all characteristics, and encoding fault types into category labels for simple interpretation by classifiers. To enable fair model evaluation, the dataset is separated into three subsets: training, validation, and testing. These preprocessing stages ensure the data is properly prepared for effective training while preserving its fundamental properties.

The preprocessed dataset is then used to train four distinct machine learning models, including Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), and K-Nearest Neighbors. Each model has distinct strengths. SVM is a supervised learning model that classifies data points using hyperplanes. It works especially well with small-to-medium datasets. Random Forest is an ensemble learning method that builds multiple decision trees to improve predictive accuracy while minimizing overfitting. LSTM, a recurrent neural network (RNN) designed for time-series data, captures temporal dependencies and is particularly useful for analyzing voltage and current signals. KNN, a non-parametric method, classifies data points based on their proximity to labeled neighbors.

Once trained, the models are applied to incoming time-series data reflecting real-time voltage and current signals from the electrical grid. These signals are input into the trained network, which classifies fault types and localizes defects to specific buses in the system. The classification task accurately identifies fault types, whereas localization enables a quick response by pinpointing where the problem occurs in the transmission network. The performance of each classifier is measured using common measures such as accuracy, precision, recall, and F1-score. In addition, confusion matrices are used to investigate misclassification patterns, providing more information about each model's strengths and limitations. A comparative analysis of the models demonstrates their usefulness for defect detection and localization, underscoring the proposed framework's suitability for real-world applications.

The proposed framework seamlessly integrates simulation, preprocessing, machine learning, and fault analysis into a single system. This approach delivers robust performance in detecting and localizing faults in electrical transmission lines by leveraging a diverse dataset and multiple classifiers. This framework gives a solid structure for real-world applications focused on improving power system reliability and fault management.

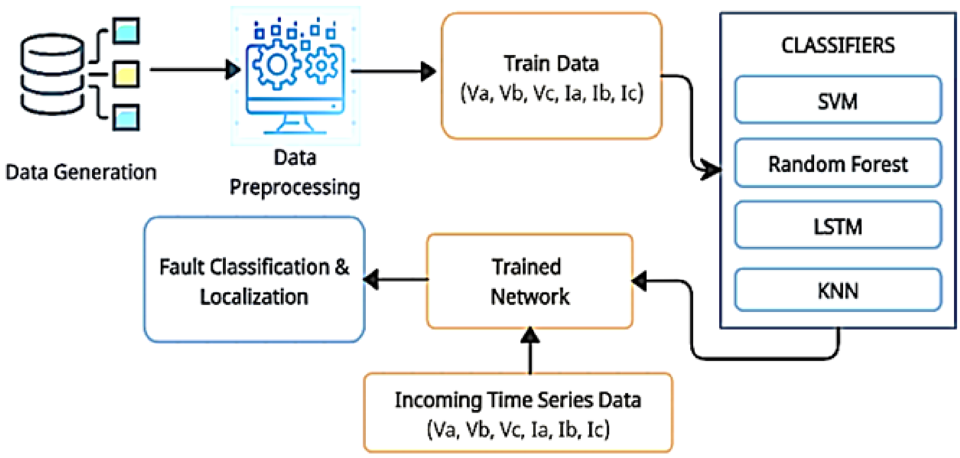

Fig. 1. Proposed framework for transmission line fault classification and localization

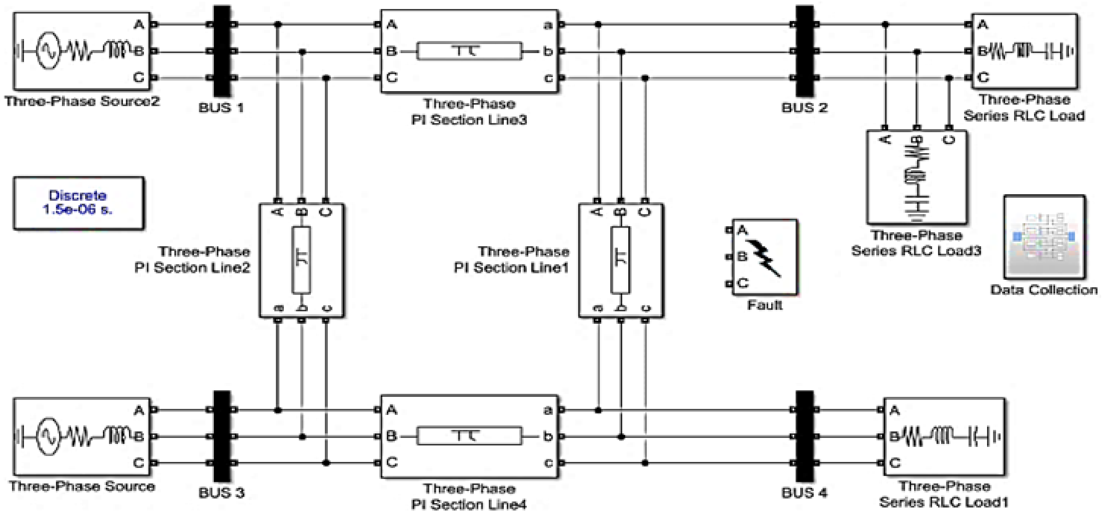

Fig. 1. Simulink model of the 4-bus system

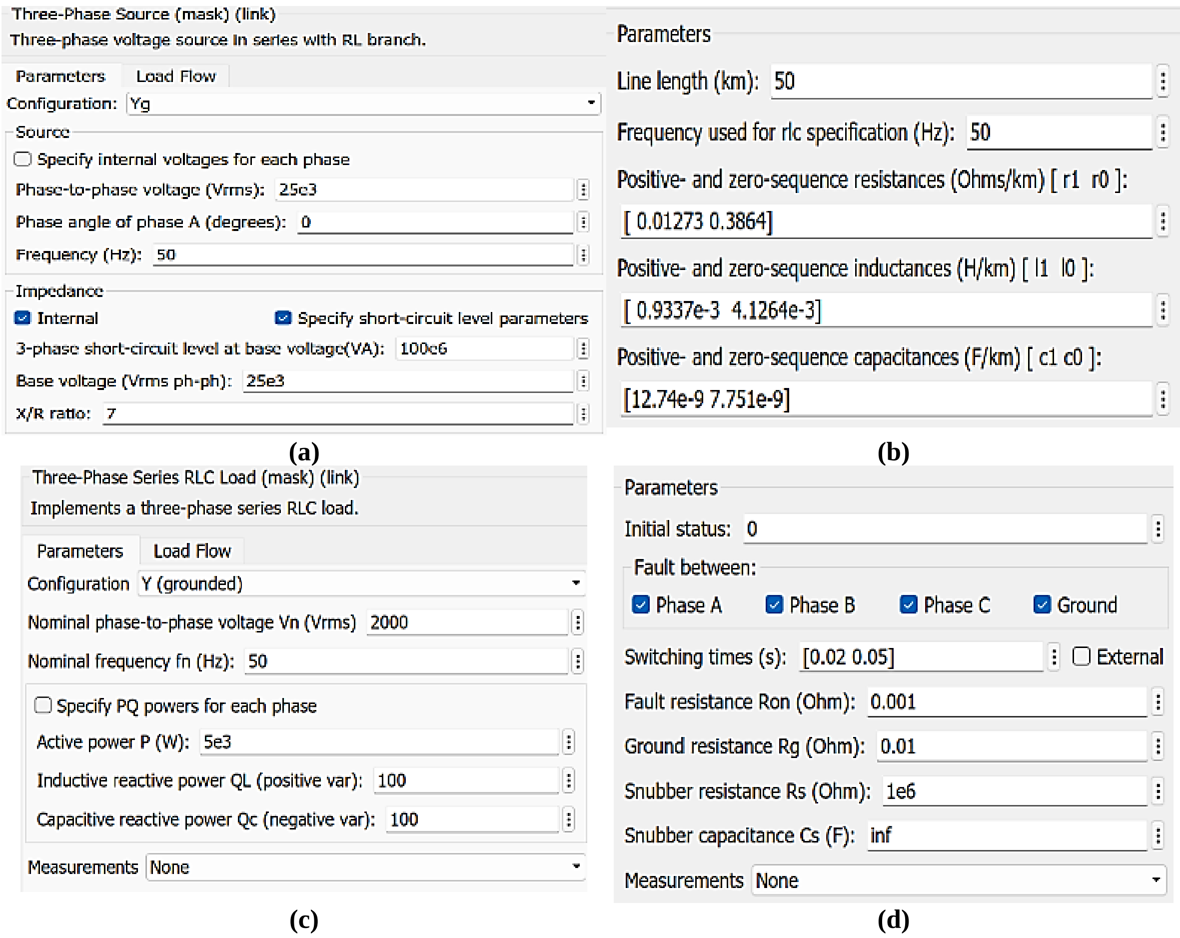

Fig. 2. Dataset development and validation

Transmission line simulation involves determining essential characteristics, including resistance, inductance, capacitance, and conductance, before constructing the model in software such as MATLAB/Simulink. This study employs a four-bus system. The 4-bus system was chosen for this study because it is simple yet effective for simulating key operational characteristics of larger electrical networks. This system offers a simple yet sufficiently complex framework for investigating transmission line behavior, making it excellent for preliminary research and validation. The 4-bus system, consisting of 2 generators, 3 loads, and 4 pi-section transmission lines, balances computational efficiency with the ability to simulate realistic power flow and failure scenarios. The generators are 25KV and 100MVA. The load values are 10 kW, 5 kW, and 5 kW, with a voltage of 2 kV. The section lines are 50km each.

Furthermore, its small size enables extensive parameter tuning and failure simulation without increasing computational resources, ensuring that the focus remains on developing and proving the fault detection system. The 4-bus system's standardized structure also improves reproducibility and allows for direct comparisons with other studies. By choosing this system, we ensure that the dataset created captures a wide range of failure conditions while also leaving room to expand the technique to more complex systems in future work. Figures 2 and 3 illustrate the structure and parameter details of this system, respectively. As shown in the Simulink models, a three-phase fault has been created using the three-phase fault block. The failure is simulated at 0.02s using 0.001Ω and 0.01Ω resistances, as shown in Figure 3(d).

3.2 Dataset Preprocessing

Dataset preprocessing is a critical step to get raw data suitable for machine learning algorithms. This study's data was obtained from a CSV file. The dataset originally comprised redundant and unnecessary columns, such as 'A', 'B', 'C', 'G', and ' Fault & Bus (F&B)', which were eliminated to simplify the workflow. This stage eliminated extraneous complexity, leaving only the essential features required for analysis. The data were then divided into two main subsets based on the 'Fault' column. The first subset, called "No Fault (NF)," contained data samples with the 'Fault' label 'NF'. The second subset contained "Fault" data, with the 'Fault' label identifying specific fault kinds such as 'AG', 'BG', and others.

To achieve a fair distribution across all buses, the data was divided by bus number (1, 2, 3, 4). For the NF subset, the first 200 samples per bus were assigned to the training set, while the remaining samples were used for testing. Similarly, for the Fault subset, 50% of each fault type (only 5 samples) on each bus was chosen at random for the training set, with the rest used for testing. This approach ensured diversity and balance across fault categories in both the training and testing datasets.

3.3 Dataset Description

The dataset utilized in this work was specifically constructed for fault detection and classification in a three-phase electrical system. It contains 4,440 instances with 28 features, including three-phase voltages (Va, Vb, Vc) and currents (Ia, Ib, Ic) from four buses (Bus 1–4), along with phase indicators and target labels for faults. Each bus includes data for healthy and fault conditions, covering 11 distinct fault classes. These characteristics ensure the dataset is both comprehensive and representative of real-world conditions while remaining manageable for machine learning model training. Figure 4 shows the prepared dataset of a four-bus system.

The phase indication columns A, B, C, and G indicate binary values about the involvement of specific phases or ground in the fault conditions. The "Fault" column represents a categorical label for the kind of fault condition. Table 1 summarizes the types of faults and their corresponding labels used in the dataset.

	A	B	C	D	E	F	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	Va1	Vb1	Vc1	Ia1	Ib1	Ic1	Ic2	Va	Vb	Vc3	Ia3	Ib3	Ic3	Va4	Vb4	Vc4	Ia4	Ib4	Ic4	A	B	C	G	Fault	Bus	F&B
2	0.02001	-0.0128	-0.85918	0.87197	-0.08626	0.7226	-1	1	-0	-1	1	-0	0.72248	-0.63623	-0.0496	-0.84036	0.88996	-0.03875	-0.65653	0	0	0	0	NF	1	NF
3	0.02004	-0.00338	-0.86395	0.86733	-0.09365	0.7256	-1	1	-0	-1	1	-0	0.72544	-0.6318	-0.04018	-0.84543	0.88562	-0.03139	-0.66049	0	0	0	0	NF	1	NF
4	0.02007	0.00604	-0.86865	0.86261	-0.10103	0.7285	-1	1	0	-1	1	-0	0.72834	-0.62732	-0.03076	-0.85043	0.88119	-0.02403	-0.6644	0	0	0	0	NF	1	NF
5	0.0201	0.01546	-0.87327	0.85781	-0.10841	0.7313	-1	1	0	-1	1	-0	0.73117	-0.62277	-0.02134	-0.85536	0.8767	-0.01667	-0.66825	0	0	0	0	NF	1	NF

Fig. 4. A snapshot of prepared dataset from 4 Bus system

Table 1. Fault types and corresponding labels

Fault Type	Fault Lebel	Phase A	Phase B	Phase C	Ground
Phase A to Ground	AG	1	0	0	1
Phase B to Ground	BG	0	1	0	1
Phase C to Ground	CG	0	0	1	1
Phases A, B, Ground	ABG	1	1	0	1
Phases B, C, Ground	BCG	0	1	1	1
Phases A, C, Ground	ACG	1	0	1	1
Phase A and Phase B	AB	1	1	0	0
Phase B and Phase C	BC	0	1	1	0
Phase A and Phase C	AC	1	0	1	0
All Three Phases	ABC	1	1	1	0
No Fault	NF	0	0	0	0

The "Bus" column specifies the bus number on which the fault occurred, whereas "F&B" combines the fault type and bus location into a composite target. This dataset contains operational data of the electrical system in both healthy and faulty conditions. Each row represents one time-stamped instance of the system operating and captures the dynamic variation in voltage and current waveforms under different scenarios.

The data has been normalized to a consistent range for continuous variables such as voltages and currents to ensure the dataset is of good quality and useful. The categorical labels have been encoded for compatibility with machine learning algorithms. Additionally, balancing has been performed to make the dataset representative of all fault categories, thereby avoiding model bias.

The structure of this dataset is comprehensive, with rich temporal-spatial features, making it highly suitable for precise fault classification and localization in a three-phase electrical system.

3.4 Computational Complexity of Models

The computational complexity of the machine learning models used in this study is as follows:

- **SVM:** Training complexity is $O(n^3)$ for large datasets, and prediction requires $O(n_s \cdot d)$ per instance.
- **Random Forest (RF):** Training complexity is $O(t \cdot n \cdot \log n)$, prediction is $O(t \cdot \log n)$.
- **KNN:** Training is negligible; prediction complexity is $O(n \cdot d)$.
- **LSTM:** Complexity depends on sequence length TTT, hidden size hhh, and input dimension d, typically $O(n \cdot T \cdot h^2)$ for training.

Among these, Random Forest provides a good balance between accuracy and computational cost, making it suitable for near real-time fault detection in power systems.

4. Results and Discussion

4.1 Experimental Setup

The experimental setting was intended to simulate fault scenarios in a controlled environment, yielding a broad, representative dataset for training and testing fault classification and localization models. A four-bus power system was modeled and simulated with MATLAB/Simulink, a powerful tool for accurately mimicking real-world power system dynamics. The system consists of a power supply, transmission lines, load locations, and a variety of fault-inducing components that simulate different fault conditions.

The power system simulation was set up to run under both healthy and defective conditions. Faults were deliberately created at various locations in the system, such as busbars and transmission lines, to simulate real-world events. The injected faults are single-phase (AG, BG, CG), two-phase (AB, BC, AC), and three-phase (ABC), with varying durations and fault resistances. These criteria were carefully picked to ensure that the dataset was comprehensive and reflected a wide range of operational situations. The simulation was conducted at multiple time intervals to capture both transient and steady-state responses, enabling extensive analysis of voltage and current data.

Data acquisition was an essential component of the experimental setup. Voltage (Va, Vb, Vc) and current (Ia, Ib, Ic) waveforms for all three phases were recorded for each bus in each simulated scenario. These time-series signals were recorded at a high sampling rate to ensure no important information was lost. The dataset also contained characteristics such as fault kind, fault location, and time of occurrence, making it ideal for training machine learning models for fault classification and localization. The simulation environment was outfitted with noise injection to model real-world measurement uncertainties, such as sensor errors and communication noise. This phase improved the dataset's realism, making the trained models more resistant to real-world settings. To improve data quality and ensure compliance with machine learning algorithms, preprocessing techniques such as signal filtering and normalization were applied to the raw data. To confirm the framework's performance, the experimental setup was intended to include previously unknown fault scenarios and novel operational conditions during the testing phase. This method ensured that models trained on simulated data were evaluated for their generalizability and ability to diagnose and localize defects in unknown settings effectively. The entire setup was methodically crafted to align with the goals of fault classification and localization, while allowing for the simulation of additional situations or changes to system parameters for future research. A visual representation of our experimental setup is the same as Figure 1.

4.2 Performance Evaluation Matrices

To evaluate the effectiveness of different models of machine learning in fault detection and localization, four classification algorithms—Support Vector Machine (SVM), Random Forest (RF), Long Short-Term Memory (LSTM), and K-Nearest Neighbors (KNN) were evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. These evaluations provide a thorough understanding of each model's ability to classify different fault conditions in the transmission system appropriately.

These metrics give an extensive overview of the models' accuracy, precision, recall, and robustness in identifying and classifying faults in the 4-bus power system. A confusion matrix for each model was created to investigate misclassifications and the model's ability to differentiate between different fault types.

4.3 Comparative Analysis of Models

The classification results for each model are summarized in TABLE 2, which highlights overall accuracy and macro and weighted averages for precision, recall, and F1-score.

Table 2. Performance metrics comparison of different models

Model	Accuracy	Precision	Recall	F1-Score
SVM	95%	85%	79%	79%
Random Forest	96%	88%	82%	84%
LSTM	93%	68%	69%	66%
KNN	94%	81%	74%	75%

Table 2 compares the classification performance of four models, SVM, Random Forest, LSTM, and KNN, using accuracy, macro average precision, recall, and F1-score. Random Forest is the best-performing model, achieving the highest accuracy (96%) and strong macro averages across all measures, indicating that it effectively tackles both common and complex defects. SVM achieves 95% accuracy, but has lower recall (79%) and F1-score (79%), indicating difficulty with more complex errors such as ABCG and ABC. KNN achieves high accuracy (94%), but low recall (74%) and F1-score (75%), indicating limits in differentiating less common fault types. LSTM, while reaching 93% accuracy, performs poorly in precision (68%), recall (69%), and F1-score (66%), indicating an incapacity to classify complex errors correctly. In conclusion, while Random Forest provides the most balanced and stable performance, SVM and KNN have flaws, and LSTM's fault classification power has to be significantly improved.

4.4 Confusion Matrix Analysis

Confusion matrices provide information on how successfully each model classifies various fault types. TABLE 3 summarizes the misclassification tendencies detected in the models.

Table 3. Misclassification trends observed in different models

Model	Fault Types with High Misclassification	Best Performing Fault Types
SVM	ABCG, ABC	AB, AC, NF
Random Forest	ABCG, ABC	AB, AC, NF, BG
LSTM	ABC, BC, BCG	AB, AC, BG, NF
KNN	AG, BC, BCG	AB, AC, BG, NF

Table 3 shows that the Support Vector Machine (SVM) misclassified ABCG and ABC faults, indicating difficulty in identifying more complex fault categories. However, it did admirably with smaller flaws like AB and AC. Random Forest, on the other hand, had the fewest misclassification errors while maintaining excellent precision for both common and complicated fault types. The Long Short-Term Memory (LSTM) model struggled to accurately distinguish between ABC and BC problems, resulting in lower recall. Finally, the K-Nearest Neighbors (KNN) model misclassified AG and BCG faults, underscoring its limitations in handling less common fault patterns.

The evaluation findings show that Random Forest outperforms other models in fault classification and localization because it can handle complex fault patterns while maintaining high accuracy. SVM also performs well, especially in classifying non-fault circumstances and some simple faults. LSTM and KNN, while successful, struggle to identify specific defect types, leading to lower recall. Overall, Random Forest is the best model for detecting faults in a 4-bus power system, as it balances accuracy, precision, recall, and computational efficiency. These findings highlight the potential of machine learning algorithms to improve the reliability and safety of power transmission systems.

5. Conclusions

The research successfully achieved its goal by developing and evaluating a machine-learning-based framework for fault classification and localization in power systems. By simulating a 4-bus power system, a realistic, representative dataset was created, which proved crucial for training robust models. Among the four models tested, the Random Forest algorithm demonstrated superior performance, achieving 96% accuracy and effectively handling complex, varied fault patterns.

The findings confirm the significant potential of machine learning to enhance power system reliability and safety. By providing a framework that accurately and efficiently identifies, classifies, and localizes faults, this study offers a practical, valuable tool for power system engineers. The use of a dataset with a realistic fault distribution contributes to the originality of this work and ensures the models are more adaptable to real-world conditions. Ultimately, the proposed framework has the potential to reduce system downtime, enhance predictive maintenance, and improve the overall stability and safety of electrical grids.

However, the study has some limitations: it relies on a simulated 4-bus system and a limited number of fault instances per class, which may not fully capture the complexity of larger, real-world power grids. Environmental factors and communication delays in live systems are also not considered. Future work should extend the framework to larger networks, incorporate real-time sensor data, and explore hybrid approaches to enhance performance under dynamic conditions.

Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Competing Interests

Authors have declared that they have no known competing financial interests or non-financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aleem, S. A., Shahid, N., & Naqvi, I. H. (2015). Methodologies in power systems fault detection and diagnosis. *Energy Systems*, 6(1), 85–108. https://doi.org/10.1007/s12667-014-0129-1

Al-Shourbaji, I., & Alameen, A. (2025). Optimizing solar radiation prediction with ANN and explainable AI-based feature selection. *Technologies*, 13(7), 263. https://doi.org/10.3390/technologies13070263

Arouche Freire, J. C., Garcez Castro, A. R., Homci, M. S., Meiguins, B. S., & De Morais, J. M. (2019). Transmission line fault classification using hidden Markov models. *IEEE Access*, 7, 113499–113510. https://doi.org/10.1109/ACCESS.2019.2934938

Belagoune, S., Bali, N., Bakdi, A., Baadji, B., & Atif, K. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement*, 177, 109330. https://doi.org/10.1016/j.measurement.2021.109330

Belagoune, S., Bali, N., Bakdi, A., Baadji, B., & Atif, K. (2021). Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement*, 177, 109330. https://doi.org/10.1016/j.measurement.2021.109330

Bon, N. N., & Van Dai, L. (2022). Fault identification, classification, and location on transmission lines using combined machine learning methods. *International Journal of Engineering and Technology Innovation*, 12(2), 91–109. https://doi.org/10.46604/ijeti.2022.7571

Chawardol, P., & Sheikh, H. (2016). Transmission line fault classification and fault zone identification using back-algorithm based neural network propagation. *International Journal of Engineering Research and Technology*, 5(12). https://doi.org/10.17577/IJERTV5IS120272

Chen, K., Hu, J., Zhang, Y., Yu, Z., & He, J. (2020). Fault location in power distribution systems via deep graph convolutional networks. *IEEE Journal on Selected Areas in Communications*, 38(1), 119–131. https://doi.org/10.1109/JSAC.2019.2951964

Fathabadi, H. (2016). Novel filter-based ANN approach for short-circuit faults detection, classification and location in power transmission lines. *International Journal of Electrical Power & Energy Systems*, 74, 374–383. https://doi.org/10.1016/j.ijepes.2015.08.005

Ghaderi, H., Ginn, L., & Mohammadpour, H. A. (2017). High impedance fault detection: A review. *Electric Power Systems Research*, 143, 376–388. https://doi.org/10.1016/j.epsr.2016.10.021

Goni, Md. O. F., et al. (2023). Fast and accurate fault detection and classification in transmission lines using extreme learning machine. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 3, 100107. https://doi.org/10.1016/j.prime.2023.100107

He, Z., Lin, S., Deng, Y., Li, X., & Qian, Q. (2014). A rough membership neural network approach for fault classification in transmission lines. *International Journal of Electrical Power & Energy Systems*, 61, 429–439. https://doi.org/10.1016/j.ijepes.2014.03.027

Khan, M. A., Asad, B., Vaimann, T., Kallaste, A., Pomarnacki, R., & Hyunh, V. K. (2023). Improved fault classification and localization in power transmission networks using VAE-generated synthetic data and machine learning algorithms. *Machines*, 11(10), 963. https://doi.org/10.3390/machines11100963

Moloi, K., Hamam, Y., & Jordaan, J. A. (2019). Fault detection in power system integrated network with distribution generators using machine learning algorithms. In *2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI)* (pp. 18–22). IEEE. https://doi.org/10.1109/ISCMI47871.2019.9004304

Morais, J., Pires, Y., Cardoso, C., & Klautau, A. (2010). A framework for evaluating automatic classification of underlying causes of disturbances and its application to short-circuit faults. *IEEE Transactions on Power Delivery*, 25(4), 2083–2094. https://doi.org/10.1109/TPWRD.2010.2052932

Rafique, F., Fu, L., & Mai, R. (2021). End-to-end machine learning for fault detection and classification in power transmission lines. *Electric Power Systems Research*, 199, 107430. https://doi.org/10.1016/j.epsr.2021.107430

Raidid, M. A. M., et al. (2023). Improving prediction of solar radiation using cheetah optimizer and random forest. *PLOS ONE*, 18(5), e0285414. https://doi.org/10.1371/journal.pone.0285414

Raja, H. A., et al. (2022). Signal spectrum-based machine learning approach for fault prediction and maintenance of electrical machines. *Energies*, 15(24), 9507. https://doi.org/10.3390/en15249507

Raza, A., Benrabah, A., Alquthami, T., & Akmal, M. (2020). A review of fault diagnosing methods in power transmission systems. *Applied Sciences*, 10(4), 1312. https://doi.org/10.3390/app10041312

Saber, A., Emam, A., & Elghazaly, H. (2018). A backup protection technique for three-terminal multisection compound transmission lines. *IEEE Transactions on Smart Grid*, 9(6), 5653–5663. https://doi.org/10.1109/TSG.2017.2693322

Sahoo, A. K., & Samal, S. K. (2023). Online fault detection and classification of 3-phase long transmission line using machine learning model. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 6(1), 135–146. https://doi.org/10.1007/s41939-022-00132-x

Shakiba, F. M., Azizi, S. M., Zhou, M., & Abusorrah, A. (2023). Application of machine learning methods in fault detection and classification of power transmission lines: A survey. *Artificial Intelligence Review*, 56(7), 5799–5836. https://doi.org/10.1007/s10462-022-10296-0

Shakiba, F. M., Shojaee, M., Azizi, S. M., & Zhou, M. (2022). Real-time sensing and fault diagnosis for transmission lines. *International Journal of Network Dynamics and Intelligence*, 36–47. https://doi.org/10.53941/ijndi0101004

Terron-Santiago, C., Martinez-Roman, J., Puche-Panadero, R., & Sapena-Bano, A. (2021). A review of techniques used for induction machine fault modelling. *Sensors*, 21(14), 4855. https://doi.org/10.3390/s21144855

Zhang, Y.-G., Wang, Z.-P., Zhang, J.-F., & Ma, J. (2011). Fault localization in electrical power systems: A pattern recognition approach. *International Journal of Electrical Power & Energy Systems*, 33(3), 791–798. https://doi.org/10.1016/j.ijepes.2011.01.018

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/155065