Applied Science, Computing and Energy, 2026, 4(4), 612-631

AI-Based Fault Detection and Classification in Distribution Networks Using SVM and Random Forest

Balogun Monsurat O., Akomolafe Tosin O., Lambe Adesina M., Abdul-Waheed Musa, Ogunbiyi Olamilekan and Jimada-Ojuolape Bilkisu

Received: 19 March 2026/Accepted: 12 June 2026/Published: 20 June May 2026

Abstract: *Fault detection, classification, and localisation in radial distribution networks remain challenging. This is due to non-uniform loading, high resistance (R)/reactance (X) ratios, variable fault resistance, and measurement noise. These issues are especially pronounced in developing power systems such as Nigeria's. To address these challenges, this study proposes a data-driven fault diagnosis framework. The framework integrates Support Vector Machines (SVMs) for fault-type classification and Random Forest (RF) models for faulted-line identification and fault-distance estimation. The models are benchmarked against a conventional impedance-based method on the IEEE 69-bus radial test system, with clustered loads and heterogeneous line parameters. Electrical characteristics, including phase voltages, phase currents, voltage imbalance, current imbalance, and harmonic distortion indices, were considered for feature extraction and model building. To increase the method's robustness, data augmentation using Gaussian noise and z-score normalisation was applied, resulting in 4,896 samples. The experiments showed that the SVM classifier achieved a fault classification accuracy of 92.44%. The Random Forest model achieved a fault classification accuracy of 99.32%. The accuracy of fault zone determination reached 99.93%, and the accuracy of exact faulted-line identification reached 100%. For fault distance estimation, the Random Forest regression model achieved an RMSE of 11.78%, significantly outperforming traditional impedance-based methods. This model demonstrates significant robustness under noisy operational conditions. It offers a realistic way of managing faults intelligently in modern-day distribution networks*

.

Keywords: Fault Detection, Support Vector Machine, Random Forest, IEEE 69-Bus System, Distribution Network.

Monsurat O. Balogun*
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: monsurat.balogun@kwasu.edu.ng
https://orcid.org/0000-0002-5395-5375

Tosin O. Akomolafe
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: tosinakomolafe16@gmail.com
https://orcid.org/0009-0009-2251-9819

Adesina M. Lambe
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: lambe.adesina@kwasu.edu.ng

Musa Abdul-Waheed
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: abdulwaheed.musa@kwasu.ed.ng
https://orcid.org/0000-0001-8371-2465

Olamilekan Ogunbiyi
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: olalekan.ogunbiyi@kwasu.edu.ng
https://orcid.org/0000-0003-2050-8728

Bilkisu Jimada-Ojuolape
Department of Electrical Engineering, Kwara State University, Malete, Kwara State, Nigeria.
Email: bilksu.jimada@kwasu.edu.ng
https://orcid.org/0000-0001-8592-8438

1.0 Introduction

Electric power distribution networks constitute the final stage of the electrical power delivery chain, conveying electricity from transmission substations to residential, commercial, industrial, and institutional consumers. The reliability, security, and operational efficiency of these networks directly influence the quality and continuity of electricity supply, making distribution systems critical components of modern power infrastructures (Assis *et al*., 2022;

Also available @ https://doi.org/10.5281/zenodo.21469139

Gönen, 2024). In developing economies such as Nigeria, where electricity demand continues to outpace generation and network expansion, the distribution sector experiences significant operational challenges that adversely affect economic productivity, industrial development, and social welfare (Sambo *et al*., 2010; Singh *et al*., 2026).

Distribution networks are considerably more vulnerable to faults than transmission systems because they consist of extensive overhead feeders exposed to adverse weather conditions, vegetation encroachment, equipment ageing, insulation deterioration, lightning strikes, animal interference, and accidental human activities (Chen *et al*., 2016; De La Cruz *et al*., 2023). These disturbances frequently result in short circuits, conductor breakages, and high-impedance faults, causing service interruptions, equipment damage, voltage instability, and substantial financial losses to utilities and consumers alike (Thomas & Fung, 2022; Singh *et al*., 2026). In addition to economic losses, prolonged outages reduce customer satisfaction, weaken industrial competitiveness, and compromise the resilience of critical infrastructures such as hospitals, communication systems, and water supply facilities.

The challenges associated with fault management are particularly severe in radial distribution systems, which remain the predominant configuration in many developing countries due to their relatively low construction and maintenance costs (Turan Gönen, 2024). Unlike meshed transmission systems, radial feeders provide only a single path for power delivery, implying that a single fault may interrupt electricity supply to a large number of downstream customers until the fault is accurately located, isolated, and repaired (Saha *et al*., 2002; Mora-Florez *et al*., 2008). The situation becomes even more complicated in distribution feeders characterized by high resistance-to-reactance (R/X) ratios, clustered load centres, varying conductor parameters, distributed loads, and rapidly changing operating conditions, all of which reduce the effectiveness of conventional protection and fault-location techniques (Ohiero *et al*., 2022; De La Cruz *et al*., 2023).

Nigeria's electricity distribution networks exemplify many of these operational challenges. Most feeders are weakly monitored, manually operated, and lack sufficient real-time measurement infrastructure to support rapid fault diagnosis. Consequently, utilities frequently depend on manual feeder patrols and customer complaints to identify fault locations, resulting in prolonged outage durations and poor reliability performance (Akanni et al., 2023; Balogun et al., 2019; Ogunbiyi et al., 2020). These operational limitations contribute significantly to high values of the System Average Interruption Duration Index (SAIDI) and System Average Interruption Frequency Index (SAIFI), which remain among the major reliability concerns in many developing power systems.

Traditional fault detection and localisation methods employed in distribution networks primarily rely on impedance calculations, travelling-wave analysis, protective relays, sectionalizers, fault indicators, and manual inspections (De La Cruz et al., 2023; Mora-Florez et al., 2008; Saha et al., 2002). Although these techniques have been widely adopted by utilities, their performance deteriorates under practical operating conditions because of inaccurate line parameters, unknown fault resistance, measurement uncertainties, load variability, and network reconfiguration (Chen *et al*., 2020; Omeje *et al*., 2025). High-impedance faults are particularly difficult to detect because they produce current magnitudes comparable to normal load currents, thereby escaping conventional overcurrent protection devices (Ahmadi *et al*., 2022).

Recent advances in sensing technologies, communication systems, and intelligent grid monitoring have significantly transformed fault diagnosis in electric power systems. Modern smart grids increasingly incorporate Phasor Measurement Units (PMUs), micro-PMUs, Intelligent Electronic Devices (IEDs), Supervisory Control and Data Acquisition (SCADA) systems, and Advanced Metering Infrastructure (AMI) to provide synchronized, high-resolution measurements that enhance fault detection, classification, and localisation (Chen *et al*., 2016; De La Cruz *et al*., 2023). However, widespread deployment of these technologies remains economically challenging in many developing countries because of high installation costs, inadequate communication infrastructure, and limited investment in grid modernization (Hairi *et al*., 2022; Akanni *et al*., 2023). Consequently, there is a growing need for intelligent computational techniques capable of extracting useful diagnostic information from relatively limited electrical measurements available in conventional distribution systems.

Machine learning has emerged as one of the most promising computational paradigms for addressing complex nonlinear problems encountered in modern power systems. Unlike analytical approaches that require detailed mathematical modelling of network parameters, machine learning algorithms learn complex relationships directly from historical or simulated operational data, thereby enabling accurate prediction, classification, and decision-making under uncertain operating conditions (Najafzadeh *et al*., 2024; Olojede *et al*., 2025). These techniques have demonstrated remarkable capabilities in handling noisy datasets, nonlinear feature interactions, changing load conditions, and incomplete measurements that commonly characterize practical distribution networks.

Among the various machine learning algorithms, Support Vector Machines (SVMs) have attracted considerable attention because of their strong theoretical foundations, excellent generalisation capability, and robustness when handling high-dimensional feature spaces (Gururajapathy *et al*., 2018; Liu *et al*., 2021). SVM classifiers determine optimal separating hyperplanes that maximise the margin between different fault classes, thereby improving classification performance even when training samples are relatively limited. Their effectiveness has been demonstrated in numerous fault diagnosis applications involving single-line-to-ground, line-to-line, double-line-to-ground, and three-phase faults under varying operating conditions (Ahmadi *et al*., 2022; Wang *et al*., 2025).

Random Forest (RF), an ensemble learning algorithm based on multiple decision trees, has similarly gained significant popularity because of its ability to handle nonlinear relationships, noisy datasets, missing values, and large feature spaces without requiring extensive parameter tuning (Tun *et al*., 2021; Adugna *et al*., 2022). Random Forest models combine predictions from multiple independent trees through ensemble voting or averaging, resulting in improved predictive accuracy, reduced overfitting, and enhanced robustness. These characteristics make Random Forest particularly suitable for complex power system applications involving fault classification, fault localisation, condition monitoring, and predictive maintenance.

The integration of SVM and Random Forest algorithms therefore provides an attractive data-driven framework for intelligent fault diagnosis in distribution networks. While SVM offers excellent discrimination among different fault categories, Random Forest possesses superior capability for feature selection, faulted-line identification, regression-based fault distance estimation, and handling heterogeneous datasets. Combining these complementary algorithms enables the development of robust hybrid diagnostic frameworks capable of simultaneously performing fault detection, classification, localisation, and distance estimation under practical operating conditions.

Research on intelligent fault diagnosis has expanded considerably over the past decade owing to increasing interest in smart grids, digital substations, and artificial intelligence applications in power systems. Conventional analytical methods have progressively been complemented by machine learning, deep learning, and hybrid computational approaches that exploit measured electrical quantities such as voltages, currents, harmonics, travelling waves, and synchrophasor data to improve diagnostic accuracy (Chen *et al*., 2016; De La Cruz *et al*., 2023).

Early investigations demonstrated the limitations of impedance-based fault location techniques in practical distribution systems. Mora-Florez *et al*. (2008) compared several impedance-based algorithms and showed that their performance deteriorates significantly under varying load conditions, parameter uncertainties, and fault resistance. Similarly, Chen *et al*. (2020) employed Deep Graph Convolutional Networks for fault location and demonstrated that graph-based learning models substantially improve localisation accuracy compared with conventional impedance methods. More recently, Omeje *et al*. (2025) proposed an enhanced fault location methodology that significantly improved localisation precision in transmission networks by accounting for parameter uncertainties, illustrating the continuous evolution of intelligent fault diagnosis techniques.

Support Vector Machine-based approaches have consistently demonstrated excellent classification performance in distribution networks. Gururajapathy *et al*. (2018) developed a Directed Acyclic Graph Support Vector Machine (DAGSVM) integrated with Support Vector Regression (SVR) for simultaneous fault classification, localisation, and fault resistance estimation using three-phase voltage sag measurements. Their model achieved perfect fault classification accuracy while maintaining acceptable localisation performance on an actual Malaysian distribution network.

Building upon this concept, Manojna, *et al*. (2021) introduced a hybrid XGBoost–SVM framework in which XGBoost ranked important electrical features before multi-class SVM performed fault section identification on the IEEE 34-bus distribution network. The proposed framework achieved an impressive classification accuracy exceeding 99%, demonstrating the benefits of combining feature selection with robust machine learning classifiers.

Micro-PMU-based monitoring has also enhanced SVM applications. Mirshekali *et al*. (2022) extracted harmonic characteristics using Clarke transformation and Fast Fourier Transform (FFT) from synchronized micro-PMU measurements and employed SVM classifiers for fault section identification. Their approach achieved classification accuracies between 94% and 98%, confirming the capability of synchronized measurements to improve machine learning performance.

Wavelet-based feature extraction techniques have equally shown promising results. Mamuya *et al*. (2020) employed Discrete Wavelet Transform (DWT) to generate statistical features from three-phase current signals before training Multilayer Perceptron (MLP) and Extreme Learning Machine (ELM) models. Their methodology achieved near-perfect fault classification and localisation accuracy in radial distribution systems, demonstrating the importance of effective signal preprocessing for intelligent fault diagnosis.

Recent investigations have increasingly shifted towards deep learning architectures capable of automatically learning complex nonlinear feature representations. Hosseini *et al*. (2024) proposed a Black Widow Optimization-enhanced Bidirectional Long Short-Term Memory (BiLSTM) model incorporating Empirical Mode Decomposition (EMD) for fault detection in modified IEEE 69-bus systems with high penetration of electric vehicle chargers. Their framework achieved detection accuracies exceeding 98% while maintaining high robustness under noisy operating conditions, highlighting the growing applicability of deep learning for future smart distribution systems.

The increasing diversity of machine learning algorithms illustrates the transition from conventional model-based protection strategies toward intelligent data-driven decision-making frameworks capable of adapting to increasingly dynamic and uncertain distribution network operating conditions.

In addition to deep learning models, ensemble learning techniques have gained increasing attention because of their robustness against noisy measurements, nonlinear feature interactions, and overfitting. Random Forest (RF), in particular, has demonstrated superior predictive performance in several engineering applications owing to its ability to combine multiple decision trees into a single robust classifier. Tun *et al*. (2021) demonstrated the effectiveness of a hybrid Random Forest–Support Vector Machine framework for fault detection and diagnosis in heating, ventilation, and air-conditioning (HVAC) systems, highlighting the complementary strengths of both algorithms in improving diagnostic accuracy under uncertain operating conditions. Similarly, Adugna *et al*. (2022) compared Random Forest and Support Vector Machine classifiers for regional land cover mapping and reported that Random Forest consistently exhibited greater robustness to noisy and heterogeneous datasets. Although these studies were conducted outside electrical power systems, they provide strong evidence that ensemble learning techniques are highly suitable for complex classification problems involving nonlinear data distributions.

Within power system applications, Random Forest has shown considerable promise for intelligent fault diagnosis. Okumus *et al*. (2025) developed an ensemble learning framework that combined Neural Network Ensembles (NNE) with Random Forest using stacked and conformal averaging techniques for short-circuit fault location in distribution networks. Their model achieved average fault location errors of approximately 0.22% and maximum errors of only 2.25%, demonstrating the effectiveness of ensemble learning for precise fault localisation using limited measurement points.

Recent reviews further confirm the growing importance of artificial intelligence in power system protection. Wang et al. (2025) comprehensively reviewed AI-driven approaches for fault detection, classification, and localisation in modern distribution networks, covering Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Transformers, and Large Language Models (LLMs). The review concluded that future intelligent protection systems should integrate adaptive learning algorithms with synchronized measurements from PMUs and micro-PMUs to support autonomous and self-healing power grids.

Similarly, Mbey *et al*. (2023) investigated deep learning and neuro-fuzzy algorithms for smart distribution grids and demonstrated that hybrid intelligent systems substantially improved fault detection and classification performance compared with conventional protection approaches. Ponukumati et al. (2022). employed wavelet decomposition with several machine learning classifiers on the IEEE 33-bus distribution system and reported that cubic SVM achieved the highest classification accuracy among the evaluated algorithms. These findings further validate the capability of intelligent learning algorithms to capture complex fault signatures in radial distribution systems.

Machine learning applications have also expanded beyond fault classification to include predictive maintenance, reliability assessment, and asset management. Olojede *et al*. (2025) highlighted the increasing adoption of machine learning for predictive maintenance and condition monitoring in modern power grids, emphasizing its potential for reducing maintenance costs and improving asset reliability. Likewise, Najafzadeh *et al*. (2024) proposed optimized machine learning algorithms for fault detection, classification, and localisation along transmission and distribution lines, demonstrating significant improvements over traditional analytical techniques.

The reliability implications of efficient fault diagnosis have equally attracted considerable attention. Reliability studies conducted by Ogunbiyi *et al*. (2020) and Balogun *et al*. (2019) on Nigerian distribution networks reported poor reliability indices primarily due to prolonged fault detection and restoration times. Their findings indicate that improving fault diagnosis directly contributes to reductions in outage duration and enhancement of service reliability. In a broader context, Assis *et al*. (2022) demonstrated that reliability-oriented planning significantly improves overall power system performance by minimizing operational costs while maintaining adequate system security.

Conventional distribution network protection methods remain indispensable but continue to exhibit limitations under practical operating conditions. Hairi et al. (2022) reviewed overcurrent protection strategies for radial distribution systems and observed that protection coordination becomes increasingly challenging as network complexity increases. Furthermore, Chen *et al*. (2016) and De La Cruz *et al*. (2023) emphasized that conventional fault location techniques experience substantial performance degradation in the presence of uncertain line parameters, changing load conditions, distributed generation, measurement errors, and communication constraints.

High-impedance faults remain among the most challenging disturbances encountered in modern distribution systems because their fault currents are often insufficient to trigger conventional protection devices. Ahmadi *et al*. (2022) addressed this challenge using Support Vector Machine-based detection algorithms capable of distinguishing high-impedance faults from normal operating conditions with significantly improved accuracy. Similarly, Abubakar and Abdulkareem (2022) critically reviewed fault detection, classification, and localisation techniques and concluded that conventional analytical methods alone cannot adequately address the increasing complexity of modern distribution networks.

The availability and quality of training datasets have become equally important considerations in machine learning-based fault diagnosis. Zhang *et al*. (2025) demonstrated that synthetic sample augmentation significantly enhances data-driven fault detection performance by improving model generalisation under limited training data conditions. Their study showed that data augmentation techniques increase classification robustness without requiring additional field measurements. Likewise, Noshad *et al*. (2019) demonstrated that Random Forest classifiers maintain high detection accuracy even under noisy sensing environments, confirming the resilience of ensemble learning approaches.

From an operational perspective, improving fault diagnosis has significant economic implications. Thomas & Fung (2022) quantified substantial downstream supply-chain losses resulting from power disturbances, while Singh *et al*. (2026) demonstrated that prolonged outages adversely affect regional economic growth and industrial productivity. These findings reinforce the importance of rapid fault detection and restoration in ensuring reliable electricity supply and sustainable economic development.

Overall, the literature demonstrates remarkable progress in the application of artificial intelligence for fault detection, classification, and localisation in electrical distribution systems. Machine learning models consistently outperform conventional impedance-based approaches in terms of classification accuracy, robustness to noise, nonlinear modelling capability, and computational efficiency. Nevertheless, several important research challenges remain unresolved.

Despite the substantial progress achieved in intelligent fault diagnosis, several limitations remain evident in the existing body of literature. First, many published studies have focused on relatively small benchmark systems such as the IEEE 11-, 33-, and 34-bus feeders (Manojna, *et al*., 2021; Mamuya *et al*., 2020; Ponukumati *et al*., 2022), limiting confidence in their scalability to larger and more practical distribution networks.

Second, although deep learning techniques have demonstrated excellent predictive performance (Hosseini *et al*., 2024; Wang *et al*., 2025), they generally require extensive computational resources, large labelled datasets, and significant training time, making their implementation challenging for many distribution utilities in developing countries. Consequently, there remains a need for computationally efficient machine learning models that provide high diagnostic accuracy without imposing excessive computational or data requirements.

Third, several previous investigations concentrated primarily on either fault classification or fault localisation rather than developing integrated frameworks capable of simultaneously performing fault type identification, faulted-line determination, and fault distance estimation (Gururajapathy *et al*., 2018; Najafzadeh *et al*., 2024). An integrated intelligent framework offers substantial operational advantages by reducing the number of independent diagnostic modules required during system operation.

Fourth, numerous studies assume the availability of synchronized measurements from PMUs or micro-PMUs installed throughout the network (Mirshekali *et al*., 2022; Wang *et al*., 2025). Such measurement infrastructure remains scarce in many developing countries, including Nigeria, where distribution systems continue to rely on conventional monitoring equipment and limited communication facilities (Akanni *et al*., 2023; Ogunbiyi *et al*., 2020).

Fifth, relatively few studies explicitly evaluate model robustness under realistic operating conditions characterised by measurement noise, heterogeneous line parameters, clustered loading, varying fault resistance, and high R/X ratios that typify practical radial distribution systems in developing economies (De La Cruz *et al*., 2023; Abubakar & Abdulkareem, 2022). Addressing these practical operating conditions is essential for improving the real-world applicability and reliability of intelligent fault diagnosis systems.

Finally, although ensemble learning algorithms such as Random Forest have demonstrated exceptional predictive capabilities in numerous engineering applications, comparatively limited studies have explored their combined utilisation with Support Vector Machines for comprehensive fault diagnosis on the IEEE 69-bus distribution system using extensive feature engineering, data augmentation, and comparative evaluation against conventional impedance-based fault location methods.

The aim of this study is to develop a robust data-driven fault diagnosis framework for radial distribution networks by integrating Support Vector Machine and Random Forest algorithms for intelligent fault detection, fault classification, faulted-line identification, and fault distance estimation.

To achieve this aim, the study pursues the following objectives:

- To develop Support Vector Machine and Random Forest models for accurate fault type classification in distribution networks.
- To develop a Random Forest-based model for faulted-line identification and fault distance estimation.
- To extract informative electrical features from phase voltages, phase currents, voltage imbalance, current imbalance, and harmonic distortion indices for machine learning model development.
- To enhance model robustness through z-score normalisation and Gaussian noise-based data augmentation.
- To evaluate the developed framework using the IEEE 69-bus radial distribution network under diverse fault scenarios.
- To compare the performance of the proposed machine learning framework with conventional impedance-based fault location techniques.

This study contributes to the growing body of knowledge on intelligent power system protection by proposing an integrated machine learning framework capable of performing multiple fault diagnosis tasks within a unified computational architecture. Unlike many previous studies that focus on individual diagnostic functions, the proposed framework combines Support Vector Machine and Random Forest algorithms to achieve fault classification, faulted-line identification, and fault distance estimation simultaneously.

The findings are expected to provide practical benefits for electricity distribution companies by improving fault diagnosis accuracy, reducing outage duration, enhancing restoration efficiency, and strengthening overall network reliability. Accurate and rapid fault localisation can significantly reduce operational costs associated with manual feeder patrols, minimize customer interruption times, and improve utility reliability indices such as SAIDI and SAIFI.

The study is particularly relevant to developing countries where extensive deployment of advanced monitoring infrastructure remains economically constrained. By relying primarily on commonly measured electrical quantities and computational intelligence techniques, the proposed framework offers a cost-effective alternative for intelligent fault management without requiring widespread installation of expensive PMUs or micro-PMUs.

Furthermore, the investigation demonstrates the applicability of data augmentation and advanced feature engineering for improving machine learning robustness under noisy operating environments. The incorporation of Gaussian noise augmentation enhances the generalisation capability of the developed models, thereby increasing their suitability for practical field implementation where measurement uncertainty is unavoidable.

Finally, the comparative evaluation against conventional impedance-based fault location techniques provides valuable insights into the advantages and limitations of intelligent data-driven approaches for future smart distribution networks. The outcomes of this study therefore support ongoing efforts toward grid digitalisation, self-healing distribution systems, and the broader implementation of artificial intelligence for modern power system operation and protection.

3.0 Method and Material

This study proposes a machine learning-based fault diagnosis framework that integrates Support Vector Machine (SVM) and Random Forest (RF) algorithms for intelligent fault classification, faulted-line identification, and fault distance estimation in radial distribution networks. As shown in Fig. 2, the methodology consists of five sequential stages: (i) data acquisition and simulation, (ii) feature extraction and engineering, (iii) data preprocessing and augmentation, (iv) machine learning model development and training, and (v) performance evaluation. The entire framework was implemented in Python 3.11 using the Scikit-learn library on a workstation equipped with an Intel® Core™ i5-8265U processor (1.60–1.80 GHz), 16 GB RAM, a 64-bit operating system, and a 238 GB solid-state drive. The proposed framework was designed to ensure robust performance under varying loading conditions, measurement uncertainty, and diverse fault scenarios.. The IEEE 69-bus radial distribution network is a widely accepted benchmark for evaluating fault detection and distribution system analysis algorithms. The network operates at a nominal voltage of 12.66 kV with a total real and reactive load of approximately 3.80 MW and 2.69 MVAr, respectively. It comprises 69 buses connected through 68 feeder sections with heterogeneous line impedances and radial topology, making it suitable for evaluating intelligent fault diagnosis under realistic operating conditions.

Fig. 1 illustrates the radial topology of the IEEE 69-bus system, including the main feeder and lateral branches.

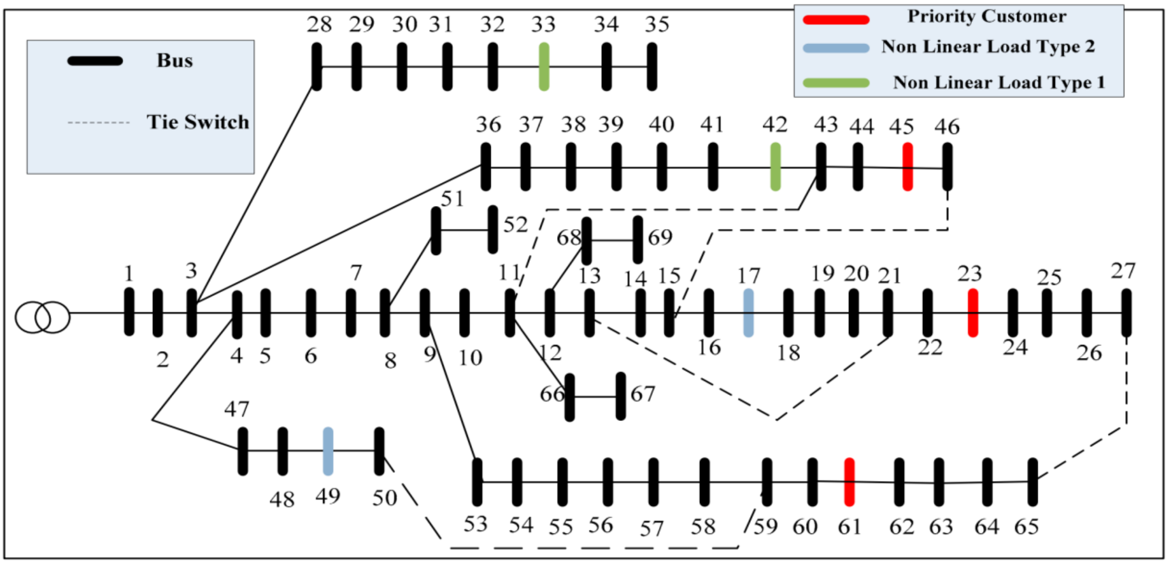

Fig. 1. IEEE 69-bus radial distribution network.

3.2 Dataset Generation

The fault diagnosis dataset was generated by performing extensive simulations on the IEEE 69-bus radial distribution network under a wide range of operating and fault conditions. A total of 816 distinct fault scenarios were created by systematically varying the fault type, faulted line, fault location (distance from the sending end), fault resistance, fault inception angle, and system loading condition. Four fault categories were considered, namely single-line-to-ground (SLG), line-to-line (LL), double-line-to-ground (DLG), and three-phase (LLL) faults. To ensure that the developed machine learning models were capable of generalising to practical operating conditions, simulations were conducted under light, nominal, and peak loading levels while considering different combinations of fault locations, fault resistances, and fault inception angles.

For each simulated fault scenario, three-phase voltage and current measurements were recorded and subsequently processed to generate the machine learning dataset. The measured electrical variables included the per-unit phase voltages ((Va), (Vb), and (Vc)), per-unit phase currents ((Ia), (Ib), and (Ic)), percentage voltage unbalance, percentage current unbalance, voltage total harmonic distortion (THDV), and current total harmonic distortion (THDI). These electrical characteristics were selected because they contain significant information regarding the occurrence, type, severity, and location of faults within radial distribution networks. The extracted measurements constituted the input features used for fault classification, faulted-line identification, and fault distance estimation.

Table 1 presents a representative sample of the generated fault diagnosis dataset. Each row corresponds to an individual simulated fault scenario, while the columns contain the measured electrical features used for subsequent machine learning model development.

Data preprocessing is carried out to eliminate real-world measurement uncertainty, and this is done by introducing controlled stochastic noise into the feature matrix as shown in equation 1:

$$\mathbf{X}_{\text{noisy}} = \mathbf{X} \odot (1+\epsilon) \quad (1)$$

where: $\odot$ denotes element-wise multiplication. $\epsilon \sim \mathcal{N}(0, \sigma^2)$. $\sigma = 0.05$represents 5% measurement noise

where N is a normally distributed random variable. The feature vector for each fault scenario was defined as:

$$X = [Va, Vb, Vc, Ia, Ib, Ic, VU, IU, THDV, THDI] \quad (2)$$

where: Va, Vb, Vc are the phase voltages., Ia, Ib, Ic are the phase currents, VU is the voltage unbalance percentage, IU is the current unbalance percentage, THDV is the voltage total harmonic distortion. And THDI is the current total harmonic distortion.

Gaussian noise augmentation was applied to increase the dataset size and improve model robustness Zhang *et al*., 2025). The dataset was expanded from 816 samples to 4,896 samples. The augmented dataset was obtained using:

$$X_{aug} = X \times (1 + \epsilon) \quad 3$$

where: $\mathrm{X_{aug}}$ = augmented feature vector. X = original feature vector.

ϵ = Gaussian random noise

$\epsilon \sim \mathcal{N}(0, \sigma^2)$

$\sigma = 0.03$represents 3% measurement noise

Table 1. Sample of the Generated Fault Diagnosis Dataset

SID	Vaa	Vb	Vc	Ia	Ib	Ic	VU	IU	THDV	THDI
1	0.913	0.895	0.905	1.451	1.432	1.585	5.43	5.38	4.84	10.20
2	0.824	0.822	0.841	1.398	1.296	1.510	3.55	5.69	4.95	5.49
3	0.823	0.812	0.810	1.221	1.194	1.291	5.30	2.87	4.75	10.06
4	0.838	0.814	0.812	1.132	1.157	1.185	3.86	1.52	3.10	9.14
5	0.872	0.872	0.878	1.680	1.516	1.623	5.59	3.97	2.08	10.41
6	0.914	0.917	0.933	1.686	1.586	1.695	4.18	1.81	2.57	7.63
7	0.909	0.925	0.924	1.569	1.572	1.658	3.00	7.63	2.81	5.98
8	0.856	0.834	0.856	1.679	1.674	1.621	5.26	5.36	2.92	9.24
9	0.939	0.931	0.916	1.382	1.345	1.277	5.38	7.90	3.94	9.26
10	0.934	0.918	0.898	1.596	1.542	1.558	4.85	2.65	7.04	4.07
11	0.936	0.914	0.951	1.230	1.144	1.194	2.49	4.05	3.68	5.92
12	0.933	0.942	0.909	1.373	1.323	1.237	1.94	3.61	4.61	6.84
13	0.917	0.926	0.938	1.207	1.256	1.117	3.84	5.42	4.89	4.65
14	0.887	0.852	0.882	1.633	1.708	1.515	2.08	4.67	7.32	11.47
15	0.859	0.876	0.870	1.525	1.489	1.449	2.19	7.31	3.27	9.82

Abbreviations: SID = Scenario Identification Number; Va, Vb, and Vc = per-unit phase voltages of phases A, B, and C, respectively; Ia, Ib, and Ic = per-unit phase currents of phases A, B, and C, respectively; VU = voltage unbalance (%); IU = current unbalance (%); THDV = voltage total harmonic distortion (%); THDI = current total harmonic distortion (%); p.u. = per unit.

3.3 Data Preprocessing

To ensure equal contribution of features during model training, feature normalisation was performed using the z-score normalisation technique. This was necessary to make all features represented in standard form.

$$x_{k,i}^{\text{norm}} = \frac{x_{k,i} - \mu_i}{\sigma_i} \qquad (4)$$

where: μ and σ are the mean and standard deviation of each feature.

The dataset was divided into training and testing subsets (70/30%). The training set was used for the model's development, while the testing set was reserved for independent performance evaluation. This procedure ensured an unbiased assessment of the model's generalisation capability.

$$\mathrm{X} = \mathrm{X}_{\text{train}} \cup \mathrm{X}_{\text{test}} \qquad (5)$$

Development of the Machine Learning Models

SVM was employed for multiclass fault classification using a one-vs-all strategy. The decision function is given by:

$$\mathrm{w}^T\mathrm{x} + b = 0 \qquad m \qquad (6)$$

Where w is the weight vector, and b is the bias term.

Random Forest was used for:

- Fault type classification
- Fault zone identification
- Exact faulted-line identification
- Fault distance estimation

Each Random Forest consists of multiple decision trees constructed using bootstrap sampling. The final prediction is obtained through majority voting:

$$\hat{y} = \text{mode}\{h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_T(\mathbf{x})\} \qquad 7$$

where $h_t(\cdot)$denotes the output of the t^{th}decision tree.

The ensemble model combines multiple decision trees to improve prediction accuracy and reduce overfitting.

The regression model predicts the relative fault location using:

$$\hat{d} = f(\mathbf{x}) \qquad (8)\_$$

where:$\hat{d}$ is the estimated fault distance (% of line length) and (X) = feature vector.

Performance Evaluation Metrics

The classification models were evaluated using:

Accuracy is used to calculate the ratio of accurate forecasts to total predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (9)$$

where: TP = True Positives, TN = True Negatives, FP = False Postives and FN = False Negative

Precision is used to calculate the ratio of all correctly predicted positive samples to genuine positive samples. It is modelled on:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (10)$$

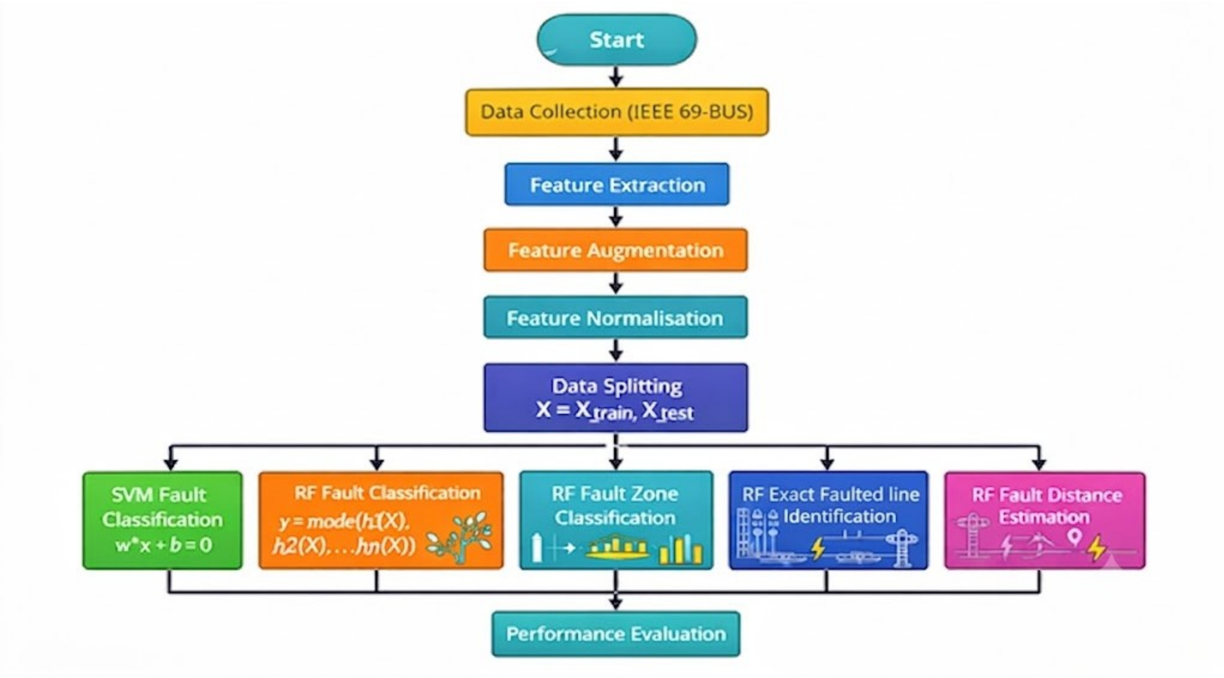

Fig. 2: Flow Diagram of the Proposed Machine Learning Model

Recall is the amount of true positive outcomes divided by the total of false negative and true positive results. It is modelled on:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

F1-score: it is a performance metric commonly used in binary classification problems. It is the harmonic mean of precision and recall, which measures the model's accuracy and completeness.

$$\text{FI} = 2\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

The fault-distance estimation model is a mathematical, analytical, or machine learning algorithm designed to calculate the actual physical distance from a substation point to a fault in a distribution network. The model was evaluated using quantitative performance metrics that compare the predicted fault distance with the actual fault location. These metrics included the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and estimation accuracy. In addition, computational efficiency, convergence characteristics, and robustness under varying fault types, fault resistances, loading conditions, and noise levels were assessed to determine the reliability and practical applicability of the model for real-time fault-location estimation in electrical distribution networks**.**Mean Absolute Error is defined as the average sum of the absolute differences between the actual value and the predicted value, serving as a measure of how well a model fits the data.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|d_i - \hat{d}_i\right| \tag{13}$$

Root Mean Square Error measures the average difference between a statistical model's predicted values and the actual values.

$$\text{RMSE}_{\text{ML}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \hat{d}_i)^2} \tag{14}$$

Coefficient of Determination (R^2) measures how well a statistical model predicts an outcome. The outcome is represented by the model's dependent variable. R^2 is calculated using:

$$1 - \frac{\sum(d_i-\hat{d}_i)^2}{\sum(d_i-\bar{d})^2} \tag{15}$$

These metrics provide a comprehensive assessment of model accuracy and reliability.

4.0 Results and Discussion

The IEEE 69-bus distribution network exhibits significant spatial variability in both load demand and line parameters. As illustrated in Fig. 3, active power demand ranges from lightly loaded buses to heavily loaded sections concentrated around the mid-feeder, resulting in uneven current distribution and heightened vulnerability to voltage drops and fault impacts. Similarly*,* Fig. 4 reveals substantial variation in reactive power demand across the network, which influences voltage sensitivity and **alters** fault signatures.at different locations. Furthermore, Fig. 5 indicates that line resistance values are highly non-uniform, with certain feeder sections exhibiting significantly higher resistance than others. These variations affect fault current magnitudes and diminish the effectiveness of conventional impedance-based fault location methods. The combined heterogeneity of load distribution and network parameters underscores the complexity of fault behaviour in practical distribution systems and emphasises the necessity for machine-learning-based approaches capable of accurately capturing nonlinear relationships between electrical measurements and fault locations.

4.1 Fault Classification Performance

As illustrated in Fig. 6, the confusion matrix indicates that most classification errors occurred between line-to-line (LL) and double line-to-ground (DLG) faults. This outcome is expected because both fault types involve multiple phase interactions and can exhibit similar electrical characteristics under certain network operating conditions, making them more difficult to distinguish. Despite these challenges, the SVM classifier achieved high classification performance, demonstrating its capability to establish robust decision boundaries within the engineered feature space.

The few remaining misclassifications reflect the inherent complexity of multiclass fault classification in distribution networks, where fault characteristics are influenced by variations in fault location, impedance, loading conditions, and system topology. Nevertheless, the overall classification accuracy confirms that the selected electrical features provide sufficient discriminatory information for reliable separation of the different fault categories. This performance was further enhanced through effective feature normalization, Gaussian noise-based data augmentation, and the use of the radial basis function (RBF) kernel to model nonlinear class boundaries. The obtained results are comparable to, and in some cases exceed, those reported in previous studies on SVM-based fault classification in power distribution systems, thereby demonstrating that SVM remains a reliable and practical approach for accurate fault-type identification in modern distribution networks.

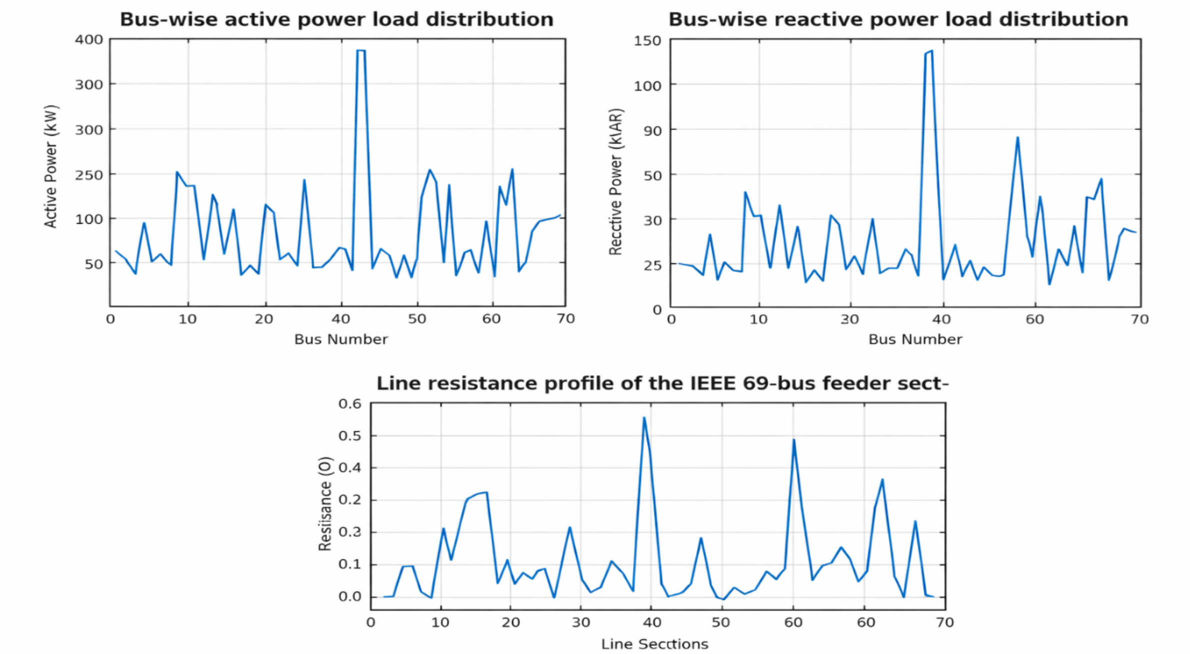

Fig. 3: Bus-wise active power load distribution of the IEEE 69-bus distribution system, F. 4: Bus-wise reactive power load distribution of the IEEE 69-bus distribution system and Fig. 5: Line resistance profile of the IEEE 69-bus feeder section

From Fig. 7, a Random Forest classifier was trained on the augmented dataset to perform fault type classification and provide a comparison with the SVM model. The confusion matrix results indicate strong overall classification performance across the four fault categories (DLG, LL, LLL, and SLG). The highest classification accuracy was observed in *Fig.* 6. Confusion matrix for fault-zone classification using Random Forest

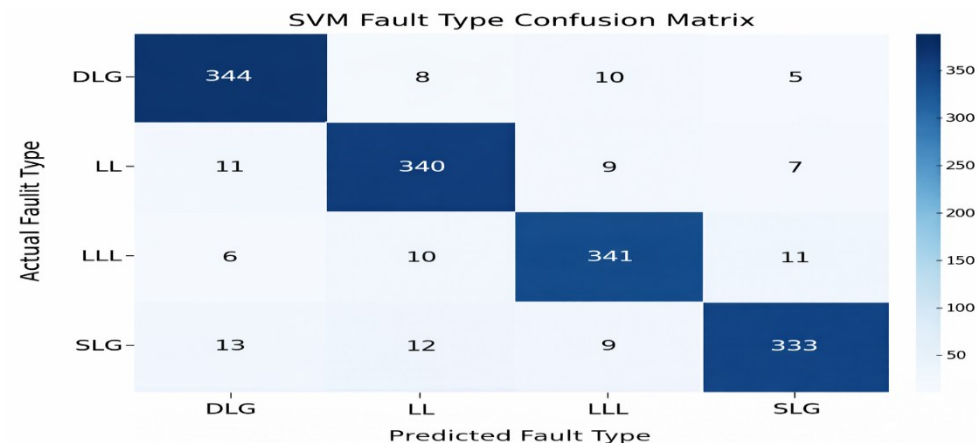

Fig. 6: Confusion matrix of SVM-based fault type classification

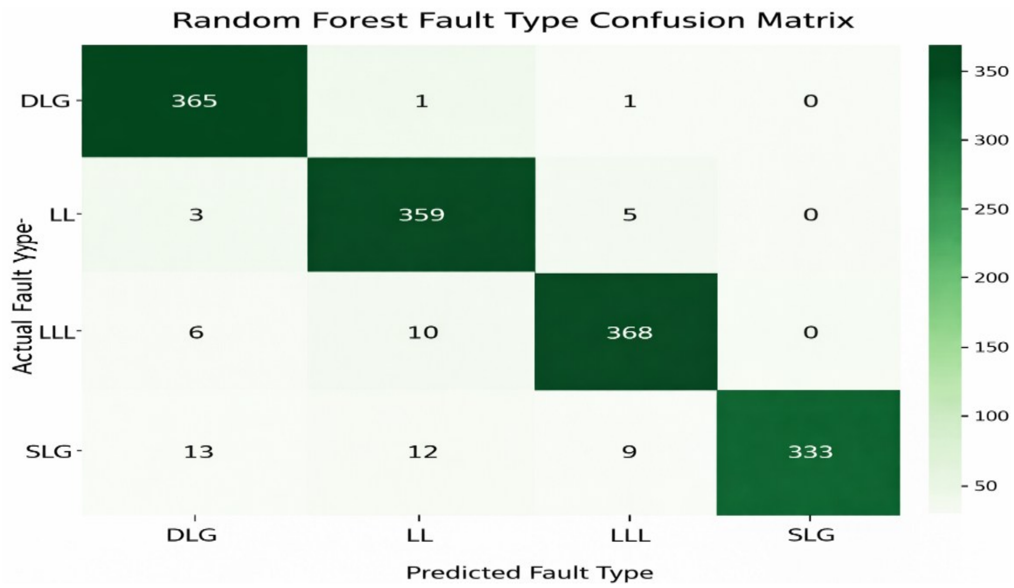

Fig. 7: Confusion matrix of RF-based fault type classification

4.2 Performance Comparison of SVM and Random Forest Models for Fault Classification

Table 2 presents the comparative performance of the Support Vector Machine (SVM) and Random Forest (RF) classifiers for fault-line classification. The results show that the Random Forest model consistently outperformed the SVM across all evaluation metrics. Specifically, the RF classifier achieved an overall accuracy of 99.32%, with precision, recall, and F1-scores exceeding 99.0%, demonstrating excellent balance, robustness, and generalisation across all fault categories. In comparison, the SVM classifier achieved an accuracy of 92.44%, with precision, recall, and F1-scores exceeding 91.0%. Although the SVM provided satisfactory classification performance, it exhibited relatively lower recall for single line-to-ground (SLG) faults and greater confusion between line-to-line (LL) and double line-to-ground (DLG) faults due to the similarity of their electrical signatures.

The confusion matrices further highlight the superior discriminative capability of the Random Forest model. The SVM classifier recorded 111 misclassifications, corresponding to an error rate of 7.56%, with most errors occurring between electrically similar fault types, particularly LL and DLG faults. In contrast, the Random Forest model produced only 10 misclassifications, demonstrating its greater ability to distinguish between fault categories under varying operating conditions.

Table 2. Comparison of the Performance of SVM and Random Forest Models for Fault-Line Classification

Metric	SVM (%)	Random Forest (%)
Accuracy	92.44	99.32
Precision	>91.0	>99.0
Recall	>91.0	>99.0
F1-score	>91.0	>99.0

Figure 8 illustrates the confusion matrix for fault-zone classification using the Random Forest model. The matrix shows that the overwhelming majority of test samples were correctly assigned to their respective fault zones, with only a few instances of misclassification between neighbouring zones. The model achieved an overall fault-zone classification accuracy of 99.93%, confirming the effectiveness of the engineered feature set and the zoning strategy in capturing location-dependent fault characteristics. This level of performance indicates that the proposed approach can accurately localise faults within predefined network zones, thereby reducing the fault search area and enabling faster fault isolation, improved maintenance planning, and enhanced operational reliability in power distribution networks.

4.3 Faulted-Line Identification Using RF

4.3.1 Performance of RF on Fault-Zone Localisation

The Random Forest fault-zone classifier achieved 99.93% accuracy as shown in Table 3.

The extremely high performance indicates that fault-zone identification can effectively narrow maintenance search areas before detailed line-level analysis is performed.

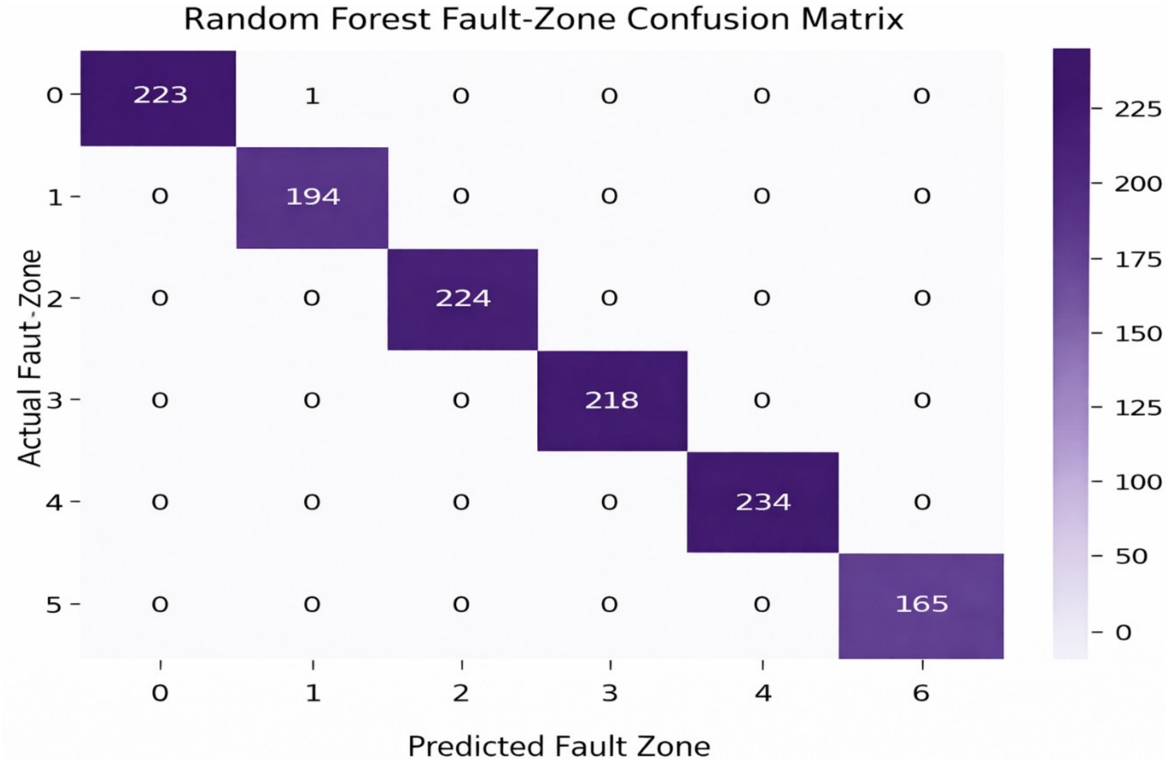

Fig. 8:. Confusion matrix for fault-zone classification using Random Forest

In terms of system operation, fault-zone classification provides several benefits, which include faster fault isolation, reduced maintenance search time, improved crew dispatch efficiency and reduced outage duration. This intermediate localisation stage is particularly valuable in developing-country environments where field resources are limited.

Table 3: Fault-Zone Classification Performance

Metrics	Values (%)
Accuracy	99.93
Precision	99.90
Recall	99.90
F1-Score	99.90

4.3.2 Performance of RF on Exact Faulted-Line Identification

One of the most significant contributions of this research is the achievement of near-perfect faulted-line localisation. The RF classifier was trained to identify the exact faulted line among (68) possible feeder sections. The faulted-line identification model achieved 100% accuracy, with all test samples correctly assigned to their respective feeder sections; this can be seen in *Fig.* 9. This exceptional performance as indicated in table 4 demonstrates the ability of the engineered features to effectively capture location-specific electrical characteristics despite challenges such as measurement uncertainty, load variations, and the similarity of fault signatures between neighbouring feeder sections. Compared with existing fault-location studies (Liu et al., 2021; Mirshekali et al., 2022), which typically report localisation accuracies between 85% and 98%, the proposed framework provides a significant improvement, highlighting its effectiveness for accurate fault localisation in distribution networks.

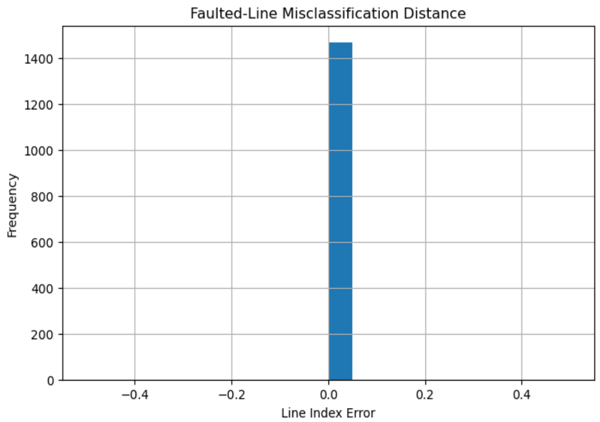

Fig. 9. Exact Faulted-Line Identification Using Random Forest

Table 4: Exact Faulted-Line Identification Performance

Metrics	Value
Number of classes	68
Test samples	1469
Accuracy	100(%)
Misclassification distance	0

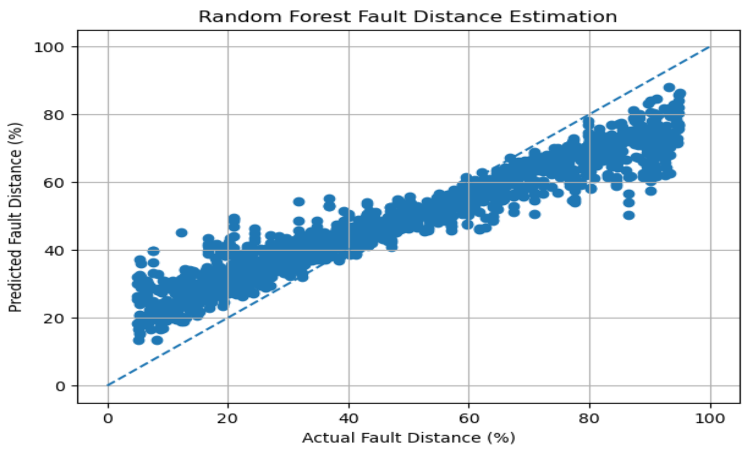

Fig. 10. Actual versus predicted fault distance using Random Forest regression.

4.4 Fault Distance Estimation

Beyond fault-line identification, the framework estimates the fault distance along the identified line using Random Forest regression. From *Fig.* 10, the scatter plot of actual versus predicted fault distances shows a strong positive correlation, with most points clustering closely around the ideal 45-degree reference line. The strong correlation indicates that the RF model successfully captures the nonlinear relationship between electrical measurements and fault distance despite heterogeneous feeder characteristics. While the model performs very well across the majority of the 0–100% range, some dispersion is observed at the extreme ends (near the sending and receiving ends of the lines), which is common in fault location problems due to reduced voltage sensitivity at distant points. Based on Table 5, the RF regressor achieved strong predictive performance across the entire feeder length. The coefficient of determination indicates that approximately 80% of the variability in fault distance is detected by the model.

Table 5: Fault Distance Estimation Performance

Metric	Value
MAE	9.67%
RMSE	11.78%
R^2	0.80

4.5 Comparison with Impedance-Based Fault Location

The Random Forest regression model achieved a Root Mean Square Error (RMSE) of 11.78% and a Mean Absolute Error (MAE) of 9.67% as shown in table 6. This represents a substantial improvement of more than 80% reduction in estimation error compared to the conventional impedance-based method, which typically yields RMSE values between 58% and 60% on the same IEEE 69-bus system under realistic conditions. The comparative fault-distance estimation results presented in *Fig.* 11 highlight a clear performance contrast between the classical impedance-based method and the proposed ML-based approach. The significant performance gap highlights the limitations of traditional analytical methods in modern radial distribution networks, which are characterised by heterogeneous line parameters, clustered loads, and measurement uncertainties conditions typical of Nigerian distribution feeders.

In contrast, the machine learning approach, supported by comprehensive feature engineering and data augmentation, demonstrates strong adaptability and robustness. The low RMSE of 11.78% indicates that the proposed framework can provide practically useful fault location estimates, potentially reducing the physical search area for maintenance crews and shortening outage restoration times.

This result confirms that the AI-based framework offers a much more reliable alternative to conventional impedance-based techniques for intelligent fault management in radial distribution systems.

Table 6: **Comparison of Fault Distance Estimation Methods**

Method	RMSE (%)
Impedance-Based	58–60
Random Forest Regression	11.78

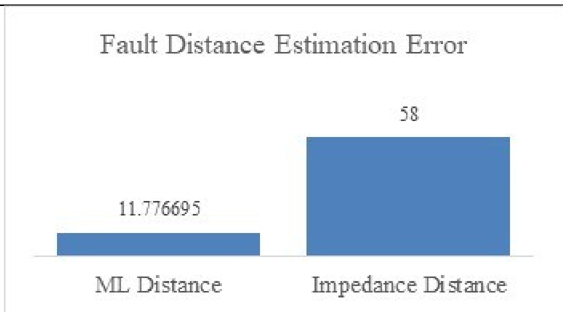

Fig. 11. Overall performance comparison of fault distance estimation errors.

4.6 Feature Importance Analysis

One of the advantages of Random Forest models is their ability to quantify feature importance. From Fig. 12, it was discovered that current and voltage-difference features dominate the ranking. These variables directly capture asymmetrical behaviour introduced by faults and therefore provide strong localisation information.

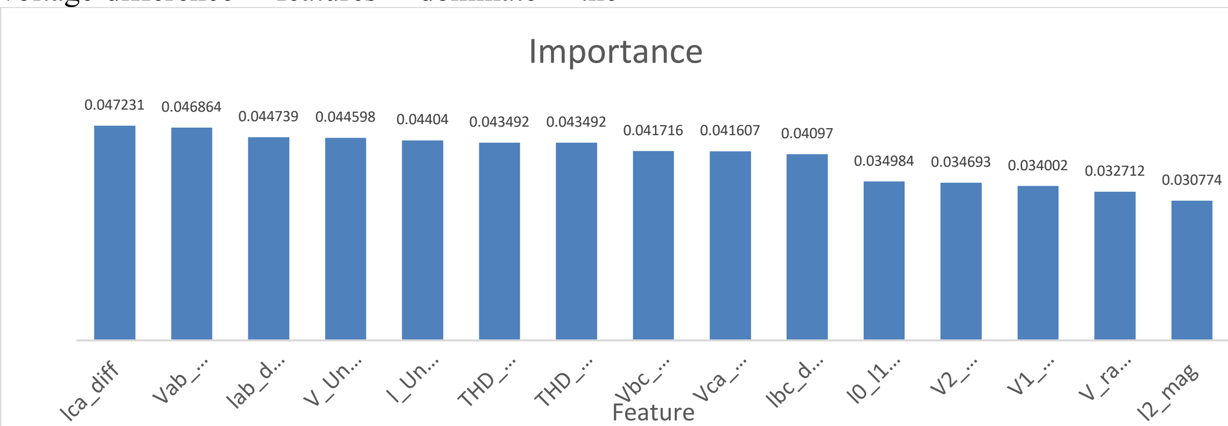

Fig. 12: **Top-Ranked Features for Faulted-Line Identification**

Voltage and current unbalance percentages ranked among the most influential variables. This finding confirms that the model is exploiting the imbalance created by fault conditions, consistent with traditional protection theory. Voltage and current THD features also contributed significantly. The results indicate that harmonic distortion contains valuable information for distinguishing fault conditions. Sequence-component ratios ranked highly among important predictors. This observation validates decades of protection-system practice in which symmetrical component analysis forms the basis of fault diagnosis. The strong alignment between machine-learning feature rankings and established power-system theory enhances confidence in the interpretability of the proposed model.

4.7 Feature Trends and Interpretability

The trends of the normalised voltage and current features across multiple fault scenarios, as illustrated in Fig. 13, show that the features are centred around zero with amplitudes typically spanning approximately −2 to +2 in normalised units. This indicates effective feature scaling and standardisation, which is beneficial for stable and unbiased machine-learning training. Across the scenario index range (0–900), noticeable fluctuations and intermittent peaks are observed in both voltage and current features, reflecting the dynamic response of the network under different fault types, fault locations, and loading conditions.

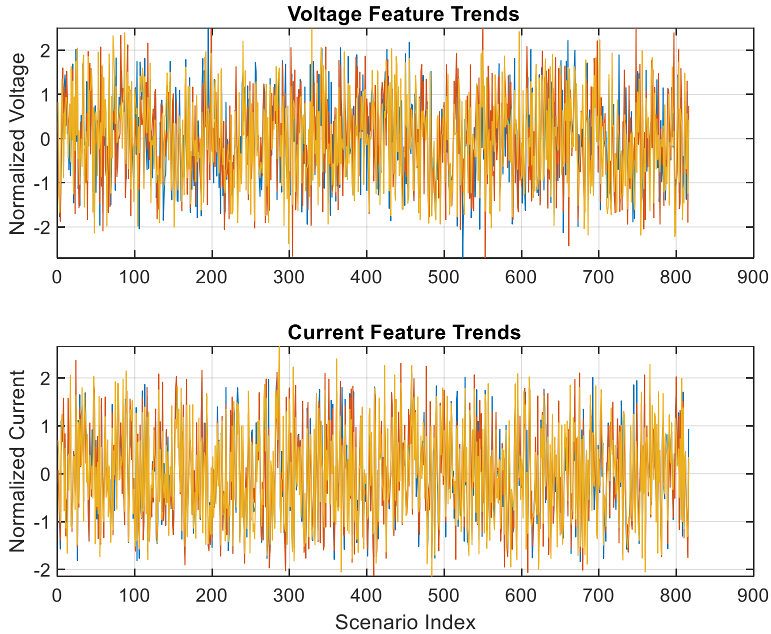

Fig. 13: Voltage and Current Feature Trends

Distinct variability patterns can be observed between voltage- and current-related features, with current features exhibiting slightly higher dispersion and more frequent extreme deviations, consistent with the stronger sensitivity of current magnitudes to fault severity and fault resistance. The presence of recurrent peaks and troughs across scenarios provides discriminative signatures that support fault type separation in the learning space. Moreover, the inclusion of physically meaningful engineered features such as apparent impedance, voltage drop, and current unbalance enhances the interpretability of the ML models by directly linking classification and regression decisions to underlying electrical phenomena. This physical grounding of the feature set contributes to improved generalisation and robustness of the proposed intelligent fault diagnosis framework under realistic operating conditions.

Based on the entire study, , the proposed machine learning framework consistently outperformed conventional impedance-based techniques across all fault diagnosis tasks. The combination of comprehensive feature engineering, Gaussian-noise-based augmentation, and Random Forest learning produced highly accurate fault classification, localisation, and regression models, demonstrating the feasibility of intelligent fault diagnosis for practical radial distribution networks.

5. 0 Conclusion

This paper presented a machine-learning framework for fault detection and localisation in radial distribution networks using Support Vector Machine and Random Forest algorithms. The IEEE 69-bus test system was used to generate a comprehensive dataset containing multiple fault scenarios. Results showed that the Random Forest model achieved superior performance, obtaining 99.32% fault classification accuracy, 99.93% fault zone identification accuracy, and 100% exact faulted-line identification accuracy. In addition, the proposed method (RF) estimated fault distance with an RMSE of 11.78%, significantly outperforming conventional impedance-based techniques (≈58–60%). The findings demonstrate that machine learning provides a reliable, accurate, and robust alternative to traditional fault diagnosis methods, particularly in distribution systems characterised by high R/X ratios, load variability, and measurement uncertainty.

6.0 References

Abubakar, J., & Abdulkareem, A. (2022). Critical review of fault detection, fault classification and fault location techniques for transmission network. *Journal of Engineering Science and Technology Review, 15*(2), 156–166. https://doi.org/10.25103/jestr.152.18

Adugna, T., Xu, W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine classifiers for regional land cover mapping using coarse-resolution FY-3C images. *Remote Sensing, 14*(3), Article 574. https://doi.org/10.3390/rs14030574

Ahmadi, A., Aghajari, E., & Zangeneh, M. (2022). High-impedance fault detection in power distribution grid systems based on support vector machine approach. *Electrical Engineering, 104*, 2593–2606. https://doi.org/10.1007/s00202-022-01544-1

Akanni, S., Melodi, A., & Ale, T. (2023). Analysis of the causes of outages in the distribution system of Lagos State, Nigeria. *FUOYE Journal of Engineering and Technology, 8*(1). https://doi.org/10.46792/fuoyejet.v8i1.964

Arsoniadis, C. G., & Nikolaidis, V. (2024). A machine learning-based fault location method for power distribution systems using wavelet scattering networks. *Sustainable Energy, Grids and Networks*. https://doi.org/10.1016/j.segan.2024.101551

Assis, F. A., Leite da Silva, A. M., Resende, L. C., Moura, R. A. R., & Schroeder, M. A. O. (2022). Generation maintenance scheduling with renewable sources based on production and reliability costs. *International Journal of Electrical Power & Energy Systems, 134*, Article 107370. https://doi.org/10.1016/j.ijepes.2021.107370

Balogun, M. O., Ahmed, M. O., Ajani, A. A., & Olaoye, H. (2019). Reliability assessment of 33 kV feeder (A case study of Transmission Company of Nigeria, Ganmo Work Centre). *Nigerian Journal of Technology*, *(1)*, 77–83.

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, pp. 119–131. https://doi.org/10.1109/JSAC.2019.2951964

Chen, K., Huang, C., & He, J. (2016). Fault detection, classification and location for transmission lines and distribution systems: A review on the methods. *High Voltage,* 1, 1, pp. 25–33. https://doi.org/10.1049/hve.2016.0005

De La Cruz, J., Gómez-Luna, E., Ali, M., Vasquez, J. C., & Guerrero, J. M. (2023). Fault location for distribution smart grids: Literature overview, challenges, solutions, and future trends. *Energies,* 16, 5, 2280. https://doi.org/10.3390/en16052280

Gönen, T. (2024). *Electric power distribution engineering*. CRC Press.

Gururajapathy, S. S., Mokhlis, H., & Bin Illias, H. A. (2018). Classification and regression analysis using support vector machine for classifying and locating faults in a distribution system. *Turkish Journal of Electrical Engineering & Computer Sciences, 26*(6), 3044–3056. https://doi.org/10.3906/elk-1711-194

Hairi, M. H., Kamarudin, M. N., Isira, A. S. M., Mohamed, M. F. P., & Sobri, S. A. (2022). Overcurrent Protection of Radial

Distribution Network. *International Journal of Electrical Engineering and Applied Sciences (IJEEAS)*, *5*(2), 2600–9633. https://doi.org/10.54554/ijeeas.2022.5.02.002

Hosseini, A., Taheri, B., Sadeghi, S., & Nasiri, A. (2024). A deep learning model for fault detection in distribution networks with high penetration of electric vehicle chargers. *e-Prime – Advances in Electrical Engineering, Electronics and Energy, 10*, Article 100845. https://doi.org/10.1016/j.prime.2024.100845

Ponukumati, B. K., Sinha, P., Maharana, M. K., Kumar, A. V. P., & Karthik, A. (2022). An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning. *Engineering, Technology & Applied Science Research*, *12*(4), 8972–8977. https://doi.org/10.48084/etasr.5107

Liu, K., Kang, T., Ye, X., Bai, M., & Fan, Y. (2021). A fault location method of distribution network based on XGBoost and SVM algorithm. *IET Cyber-Physical Systems: Theory & Applications,* 6, 4, pp. 254–264. https://doi.org/10.1049/cps2.12022

Mamuya, Y. D., Lee, Y. D., Shen, J. W., Shafiullah, M., & Kuo, C. C. (2020). Application of machine learning for fault classification and location in a radial distribution grid. *Applied Sciences,* 10, 14, 4965. https://doi.org/10.3390/app10144965

Manojna, S. S., Nikhil, A., Kumar, A., & Amrit, P. (2021). Fault detection and classification in power system using machine learning. In *Proceedings of the International Conference on Intelligent Systems, Smart Technologies and Management (ICOSEC)* (pp. 1801–1806). IEEE. https://doi.org/10.1109/ICOSEC51865.2021.9591972

Mbey, C. F., Foba Kakeu, V. J., Boum, A. T., & Souhe, F. G. Y. (2023). Fault detection and classification using deep learning method and neuro-fuzzy algorithm in a smart distribution grid. *Journal of Engineering,* https://doi.org/10.1049/tje2.12324

Mirshekali, H., Dashti, R., Keshavarz, A., & Shaker, H. R. (2022). Machine learning-based fault location for smart distribution networks equipped with micro-PMU. *Sensors, 22*(3), Article 945. https://doi.org/10.3390/s22030945

Mora-Florez, J., Meléndez, J., & Carrillo-Caicedo, G. (2008). Comparison of impedance-based fault location methods for power distribution systems. *Electric Power Systems Research, 78*(4), 657–666. https://doi.org/10.1016/j.epsr.2007.05.010

Najafzadeh, M., Pouladi, J., Daghigh, A., Beiza, J., & Abedinzade, T. (2024). Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. *International Journal of Computational Intelligence Systems,* 17, 1, https://doi.org/10.1007/s44196-024-00434-7

Ogunbiyi, O., Akomolafe, T., & Adesina, L. (2020). Reliability Indices of Electric Power Distribution System in Ilorin, Nigeria. *Nigeria Research Journal of Engineering and Environmental Sciences*, *5*(1), 491–498.

Ohiero, P., Obio, E., & Yaabari, N. (2022). Technical power losses evaluation in electricity distribution network. ***(Journal title is missing.)*** **Volume 1**, 198–207.

Okumus, H., Nuroglu, F. M., & McLoone, S. (2025). Ensemble learning for short circuit fault location estimation in distribution networks. *Engineering Applications of Artificial Intelligence, 159*. https://doi.org/10.1016/j.engappai.2025.111640

Olojede, D., King, S., & Jennions, I. K. (2025). Application of machine learning in power grid fault detection and maintenance. *Energy Informatics,* 8. https://doi.org/10.1186/s42162-025-00574-w

Omeje, O., Bankole, O., Okojie, D., & Eya, C. (2025). Advanced method for precise fault location in transmission networks. *Journal of King Saud University – Engineering Sciences,* 37. https://doi.org/10.1007/s44444-025-00013-x

Saha, M., Das, R., Verho, P., & Novosel, D. (2002). *Review of fault location techniques for distribution systems*. Conference: Power Systems and Communications Infrastructures for the future At: Beijing, China

Sambo, Abubakar Sani, Garba, Bashiru, Zarma, Ismaila Haliru, & Gaji, Muhammed Musa. *Electricity Generation and the Present Challenges in the Nigerian Power Sector.* WEC.

Singh, S., Keeler, Z., & Ewing, B. (2026). The distributional effects of power outages on regional economies: Evidence from dynamic panel data and machine learning models. *The Annals of Regional Science,* 75, 2, https://doi.org/10.1007/s00168-026-01500-y

Thomas, D., & Fung, J. (2022). Measuring downstream supply chain losses due to power disturbances. *Energy Economics,* 114, 106314. https://doi.org/10.1016/j.eneco.2022.106314

Tun, W., Wong, J. K. W., & Ling, S. H. (2021). Hybrid Random Forest and Support Vector Machine modeling for HVAC fault detection and diagnosis. *Sensors,* 21, 24. 8163. https://doi.org/10.3390/s21248163

Wang, J., Mokhlis, H., Nadzirah Mansor, N., Azil Illias, H., Ramasamy, A. K., Wu, X., & Wang, S. (2025). Smart Fault Detection, Classification, and Localization in Distribution Networks: AI-Driven Approaches and Emerging Technologies. *IEEE Access*, 13, pp. 141664–141693.https://doi.org/10.1109/ACCESS.2025.3596791

Zhang, G., Tong, X., & Lu, X. (2025). An enhanced data-driven fault detection scheme in power systems through synthetic samples augmentation. *IET Generation, Transmission & Distribution, 19*(1). https://doi.org/10.1049/gtd2.70151

Noshad, Z., Javaid, N., Saba, T., Wadud, Z., Saleem, M. Q., Alzahrani, M. E., & Sheta, O. E. (2019). Fault Detection in Wireless Sensor Networks through the Random Forest Classifier. *Sensors*, 19, 7, 68.**https://doi.org/10.3390/s19071568**

Statement and Declarations

Competing Interests

The authors declared no competing interests

Ethical Consideration

Not applicable

Funding

The authors declared no external source of funding

Availability of data and material

All data used in this manuscript are original except those duly acknowledged and cited by the author.

Author Contribution

Monsurat O. Balogun conceived the study, designed the methodology, performed data analysis, and drafted the manuscript. Tosin O. Akomolafe and Adesina M. Lambe contributed to model development, simulation, and validation. Musa Abdul-Waheed, Olamilekan Ogunbiyi, and Bilkisu Jimada-Ojuolape reviewed the manuscript, interpreted results, supervised the research, and approved the final version for publication.