Intelligent Fault Identification in Power Systems Via Wavelet Transform and Machine Learning: A Comparative Study of Four Classifiers

RADWAN M. AWAYDA, JIHAD S. DABA
Department of Electrical Engineering
University of Balamand
Kalhat, Al-Kurah, P.O. box 100, Tripoli
LEBANON

Abstract: Accurate fault identification remains a critical challenge in modern power system protection, particularly under varying fault impedance conditions. This paper proposes an intelligent fault identification framework that combines the Discrete Wavelet Transform (DWT) with machine learning classifiers. The IEEE 9-bus power system is modeled in MATLAB/Simulink, where twelve fault types are applied at two bus locations with fault impedances ranging from 0.0001 Ω to 10 Ω. Pre-fault operating conditions are established using the Newton–Raphson load-flow method. Fault current signals are processed using DWT with the Daubechies (db4) mother wavelet, and the maximum absolute detail coefficients of three-phase and ground currents are extracted to form a four-dimensional feature vector from 2400 simulation samples. Four classifiers are evaluated: Logistic Regression, Support Vector Machine (SVM), Random Forest, and Artificial Neural Network (ANN). Performance is assessed using classification accuracy, confusion matrices, ROC analysis, and computational efficiency. Results show that non-linear classifiers outperform linear approaches, achieving accuracies above 95%. The ANN attains the highest accuracy (98.3%) and the greatest robustness to increasing fault impedance, while Random Forest provides a favorable balance between accuracy and computational cost. The proposed framework offers an effective solution for intelligent fault diagnosis in power systems.

Key-Words: Power system protection; fault classification; discrete wavelet transform (DWT); feature extraction; machine learning; artificial neural networks; random forest; high-impedance faults

Received: March 13, 2026. Revised: June 16, 2026. Accepted: July 18, 2026. Available online: November 9, 2026.

1 Introduction

1.1 Background and Motivation

The reliability of the power system is essential for modern society today as faults within the power transmission network can result in blackouts, damage to the electrical equipment and substantial loss of revenue to the power company, [1]. Faults in the power system are represented as changes in the voltage and current waveform that last for very short periods of time. If faults are not correctly detected and classified by the power system, the protective devices for the electrical system may malfunction and cause the power substations to be partly or completely shutdown.

Traditional protection relies on measurements in phasor-based systems, which are inherently insufficient for detecting all types of faults, especially high-impedance occurring faults (HIFs), [2]. A method for feature extraction using DWT achieved 99.50% accuracy in detecting faults in transformers, [3]. Kernel PCA has also been proposed for the extraction of features.

Conventional protection systems based on overcurrent relay and distance protection have slow response times and limited flexibility, [4]. These systems measure the phasors of the power system to determine fault conditions. However, they have inadequate fault detection for all types of faults, especially those under dynamic system conditions and high-impedance fault scenarios where fault currents nearly approach the load levels, [5].

Signal processing algorithms allow the analysis of signals in the time, frequency, and time-frequency domains, [6]. Time-frequency analysis is suitable for analyzing non-stationary fault signals, [7]. Wavelet Transform is suitable for transient fault signal analysis because it is able to localize signals in both domains. The Discrete Wavelet Transform is suitable for real-time analysis of fault signals due to its low computational complexity. For power system applications, Daubechies DB4 wavelet is widely employed.

DWT with Symlet 4 was used to detect faults in rotating machinery, [8]. Empirical Wavelet Transform (EWT) has been proposed for transient fault detection in VSC-HVDC systems.

Machine learning has proved successful in

capturing nonlinear features, [9]. The most common algorithms in use include logistic regression, SVM, random forest, and artificial neural networks, [10]. However, there are gaps in the existing literature regarding the use of machine learning for the detection of power line fault impedances. For example, linear classifiers are often insufficient for nonlinear fault data. Additionally, there is a lack of studies that comparatively evaluate the different machine learning algorithms. The robustness of these models across varying impedances is also lacking. Finally, there is a lack of studies that evaluate the computational efficiency of these models for potential implementation in real-time.

1.2 Objectives and Contributions

The primary objective of this work is to develop and evaluate signal processing and machine learning algorithms for the task of classifying faults in electric power systems. The specific objectives are to: implement the discrete wavelet transform using the DB4 wavelet to extract features from the fault currents; develop four different machine learning algorithms resembling different machine learning paradigms; evaluate the performance of each of these machine learning algorithms using different evaluation metrics; evaluate the performances of each algorithm according to the bus at which each fault occurred and the type of fault; and to evaluate the robustness of each algorithm within the range of 0.0001 to 10 Ω logarithmically.

The original contributions of this work include: framework for DWT-based feature extraction from 12 different faults; comparison of four ML methods using 2,400 samples; analysis of misclassification of each bus and class of faults; analysis of the robustness of the proposed methods using six ranges of logarithmic impedance; and analysis of the computational efficiency of the proposed methods for real-time implementation.

2 Literature Review

2.1 Conventional Protection and Signal Processing

The wavelet transform is effective for transient analysis, [2]. Time-frequency analysis is essential for analyzing non-stationary fault signals, [6]. The Daubechies db4 wavelet is particularly well-suited for power system fault analysis due to its compact support and optimal time-frequency localization, [7].

2.2 Machine Learning for Fault Classification

Random Forest is used extensively in security assessment, with ensemble methods offering robustness and handling of non-linear relationships, [4]. Support Vector Machines continue to be widely used, including in the least squares approach to waveform reconstruction, [5].

ANN and deep learning make it possible to automatically extract features. An accuracy of 99.50% was reported using an ensemble of MLP, [7]. Deep learning approaches maintain an accuracy of $>$ 98.5% even at low SNR, [9]. More recent approaches include the use of multitask learning for HIF detection (MTFD-Net: 97.6%), [11]; CNN architectures, [12]; and the use of 1D-UNet for semantic segmentation, [13].

Despite the available methods for HIF detection, various new approaches have been introduced recently. These include multitask learning with 97.6% detection accuracy and only 5.20 ms deviation between the predicted and true HIFs, [11]; time-adaptive gated recurrent units with 96.73% accuracy and 4.315 ms deviation, [14]; and methods using Shapley values for explaining the HIF detection process, [15].

2.3 Research Gaps

Gaps identified in existing studies include an investigation of the robustness of the methods over a wider range of impedance values, [2]; the need for a comparative evaluation of various machine learning paradigms on the same dataset, [9]; an evaluation of the computational efficiency of the methods, [11]; an investigation of whether the methods exhibit different robustness when applied to different locations; the robustness of the methods on a per-class basis; and the performance of an evaluation that combines accuracy, efficiency, and robustness measures for the methods, all of which have been overlooked in existing studies. A comprehensive systematic and bibliometric review of technologies and measurement tools for power quality events detection, classification, and fault location in smart grids further underscores these research gaps, [16].

3 Methodology

3.1 IEEE 9-Bus System Simulation

The fault impedance is varied logarithmically from 0.0001 Ω to 10 Ω, thus capturing both bolted and high-impedance arcing faults, [2]. The WSCC 9-bus test system serves as the benchmark test system, containing three generators, nine buses, six transmission lines, and three loads, [3]. Fault blocks can be programmed at buses 2 and 3 to simulate various types of faults with adjustable impedance, [8]. The configuration of the system is as follows: Bus 1 slack (G1: 247.5 MVA, 16.5 kV), Bus 2 PV (G2: 192 MW, 18 kV), Bus 3 PV (G3: 128 MVA, 13.8 kV), Buses 5,6,8,9 PQ loads. Base MVA: 100.

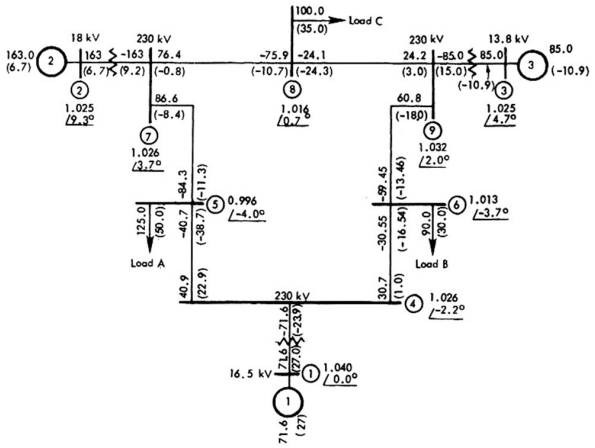


Figure 1: Single-line diagram of the IEEE 9-bus test system (P.M. Anderson test system).

Source: [17]

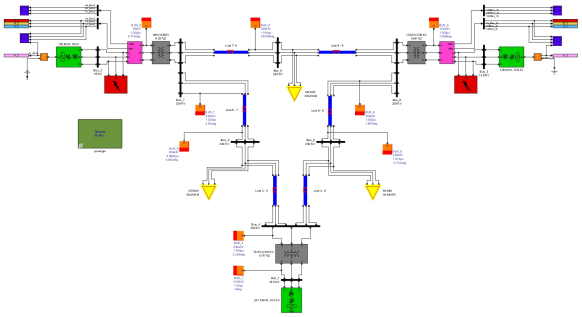


Figure 2: MATLAB/Simulink model of the IEEE 9-bus system.

Source: created by the authors.

Figure 1 shows the single-line diagram of the IEEE 9-bus test system. The system is implemented in MATLAB/Simulink using synchronous generators, excitation systems, transmission lines with distributed parameters and measurement blocks.

Figure 2 presents the MATLAB/Simulink model of the IEEE 9-bus system. Twelve fault types are simulated: No Fault (NF), single line-to-ground (A-G, B-G, C-G), line-to-line (AB, AC, BC), double line-to-ground (AB-G, AC-G, BC-G), and three-phase (ABC, ABC-G).

3.2 Load Flow Analysis

Load-flow analysis is performed using four numerical methods implemented within the MATPOWER Toolbox: Newton–Raphson (NR), Fast Decoupled (XB), Fast Decoupled (BX), and Gauss–Seidel (GS). The NR method requires the fewest number of iterations to converge to a solution (4 iterations, 0.05 s), and is, therefore, the selected method to be utilized to determine the pre-fault conditions for the system,

Table 1: Performance Comparison of Load Flow Methods

Method	Iterations	Time (s)
Newton-Raphson	4	0.05
Fast Decoupled (XB)	6/6	0.07
Gauss-Seidel	211	0.16
Fast Decoupled (BX)	6/6	0.33

Source: created by the authors.

as detailed in Table 1. Detailed bus and branch load flow results are provided in Table A1 (Appendix) and Table A2 (Appendix).

3.3 Discrete Wavelet Transform Feature Extraction

3.3.1 Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) of a signal $x(t)$ is defined as the inner product of $x(t)$ with a family of wavelet basis functions:

$$\mathrm{CWT}(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t)\, \psi^*\left(\frac{t-b}{a}\right) dt \quad (1)$$

where $a \in \mathbb{R}^+$ is the scaling parameter, $b \in \mathbb{R}$ is the translation parameter, and $\psi(t)$ is the mother wavelet. The wavelet must satisfy the admissibility condition:

$$C_\psi = \int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega}\, d\omega \; < \; \infty \quad (2)$$

where $\hat{\psi}(\omega)$ denotes the Fourier transform of $\psi(t)$. This condition ensures perfect reconstruction.

3.3.2 Discrete Wavelet Transform

For computational efficiency, the DWT employs dyadic scaling ($a = 2^j, b = k \cdot 2^j$):

$$\mathrm{DWT}(j,k) = 2^{-j/2} \int_{-\infty}^{+\infty} x(t)\; \psi^*(2^{-j}t - k)\, dt \quad (3)$$

3.3.3 Multiresolution Analysis

The scaling function $\phi(t)$ and wavelet function $\psi(t)$ satisfy:

$$\phi(t) = \sqrt{2} \sum_n h[n]\; \phi(2t-n) \quad (4)$$

$$\psi(t) = \sqrt{2} \sum_n g[n]\; \phi(2t-n) \quad (5)$$

where $h[n]$ and $g[n]$ are low-pass and high-pass filter coefficients related by the quadrature mirror filter condition:

$$g[n] = (-1)^n\; h[L-1-n] \quad (6)$$

3.3.4 Decomposition Algorithm

The approximation and detail coefficients are computed recursively:

$$A_j[n] = \sum_k h[k]\ A_{j-1}[2n-k] \tag{7}$$

$$D_j[n] = \sum_k g[k]\ A_{j-1}[2n-k] \tag{8}$$

with $A_0[n] = x[n]$ the input signal.

3.3.5 Daubechies db4 Wavelet

For db4 ($L = 8$), coefficients satisfy vanishing moments:

$$\sum_{k=0}^{L-1} (-1)^k\, k^m\, h[k] = 0, \quad m = 0, 1, \ldots, \left(\tfrac{L}{2} - 1\right) \tag{9}$$

and normalization:

$$\sum_{k=0}^{7} h[k] = \sqrt{2} \tag{10}$$

$$\sum_{k=0}^{7} (h[k])^2 = 1 \tag{11}$$

3.3.6 Wavelet Energy Distribution

The energy at level j is $E_j = \sum_n |D_j[n]|^2$. By Parseval's theorem:

$$E_{\text{total}} = \sum_n |x[n]|^2 = \sum_{j=1}^{L} E_j + E_{A_L} \tag{12}$$

The relative energy is $\rho_j = E_j / E_{\text{total}}$. For level-1 coefficients:

$$E_1 = \int_{f_s/4}^{f_s/2} |X(f)|^2 df \tag{13}$$

The Lipschitz regularity α gives $|D_1[n]| \propto 2^{-(\alpha+1/2)}$. Empirical analysis yields:

$$\rho_1 = E_1 / E_{\text{total}} \approx 0.910 \pm 0.023 \tag{14}$$

Figure 3 shows the Daubechies db4 wavelet functions.

Figure 4 illustrates the DWT multilevel decomposition process.

3.3.7 Feature Extraction

Maximum absolute D1 coefficients form the feature vector:

$$F_A = \max(|D1_A|),\ F_B = \max(|D1_B|)$$
$$F_C = \max(|D1_C|),\ F_G = \max(|D1_G|) \tag{15}$$

Figure 5 presents the Bus 2 fault current waveforms.

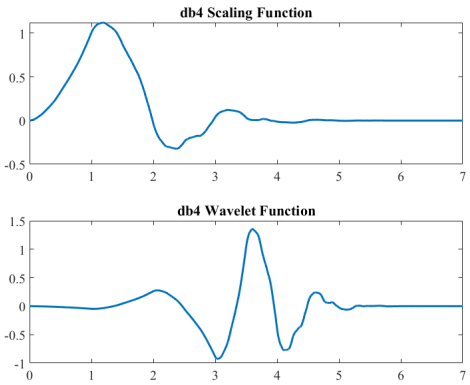


Figure 3: Daubechies db4 wavelet functions.

Source: created by the authors.

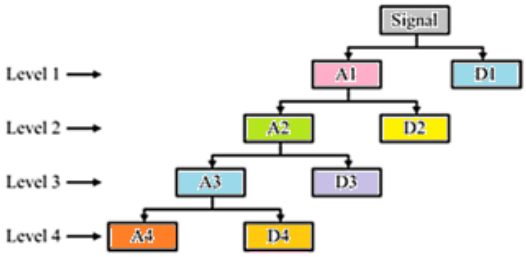


Figure 4: DWT multilevel decomposition.

Source: created by the authors.

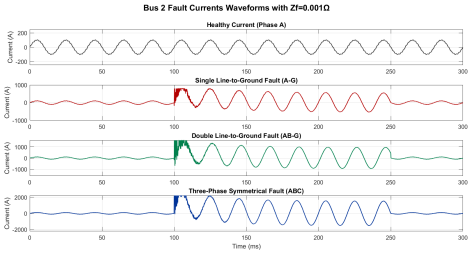


Figure 5: Bus 2 fault current waveforms.

Source: created by the authors.

Figure 6 visualizes the DWT coefficients.

Figure 7 illustrates the feature extraction process.

Figure 8 shows the wavelet energy distribution.

Table 2 presents sample DWT features by fault type.

3.4 Dataset Composition and Sample Size Justification

The dataset generated through systematic simulations comprises: total 2,400 samples (1,200 per bus), 12 fault types per bus, 200 impedance values per fault type (0.0001–10 Ω), training set 1,920 (80%), testing set 480 (20%).

The sample size of 2,400 is statistically sufficient for the following reasons: (i) With 12 fault classes and 4 features, the ratio of samples to features is 200:1 in the training set (1,920 samples for 4 features),

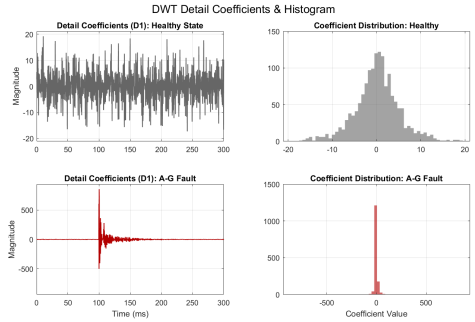


Figure 6: DWT coefficients visualization.

Source: created by the authors.

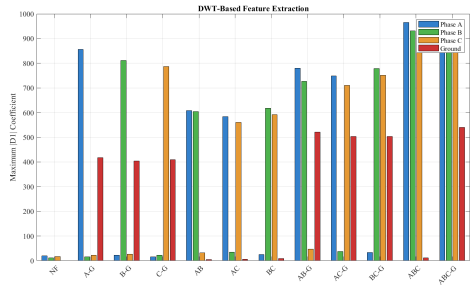


Figure 7: Feature extraction process.

Source: created by the authors.

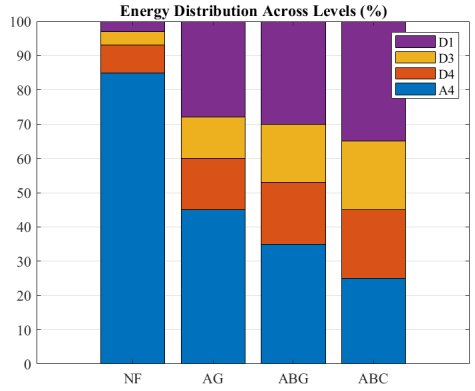


Figure 8: Wavelet energy distribution.

Source: created by the authors.

substantially exceeding the minimum recommended ratio of 10:1 for reliable classification, [10]; (ii) The 200 impedance values per fault type provide dense coverage of the logarithmic impedance range, thereby capturing the nonlinear relationship between impedance and fault signature; (iii) Cross-validation results demonstrate stable performance with standard deviations below 1.5% across folds, indicating that the dataset size is adequate to prevent overfitting; (iv) Increasing the sample size further would yield diminishing returns in accuracy while incurring significant computational costs. Consequently, 2,400 samples represent an optimal trade-off between statistical power and computational efficiency.

Table 2: Sample DWT Features by Fault Type

Fault	MaxD1_A	MaxD1_B	MaxD1_C	MaxD1_G
NF	19.22	11.19	17.06	0.94
A-G	856.49	15.03	22.29	417.25
B-G	22.38	811.86	26.09	404.12
C-G	16.14	22.45	786.49	409.57
AB	608.23	604.84	32.51	4.43
AC	584.13	34.06	561.23	5.90
BC	25.31	618.08	591.43	8.85
AB-G	779.91	726.94	46.45	521.63
AC-G	748.16	37.72	710.55	503.20
BC-G	33.28	778.44	751.96	503.98
ABC	965.14	931.45	916.33	11.51
ABC-G	965.91	949.15	939.83	540.02

Source: created by the authors.

Table 3: Dataset Composition Summary

Parameter	Value
Total Samples	2400
Training (80%)	1920
Testing (20%)	480
Fault Types	12
Impedance Range	0.0001–10 Ω
Samples per Fault	200

Source: created by the authors.

Table 3 summarizes the dataset composition.

3.5 Classification Models

3.5.1 Logistic Regression

Baseline linear classifier using one-vs-rest approach. Training uses L-BFGS optimization with L2 regularization (C=1.0). Advantages: fast training/inference, interpretable weights, low computational requirements, [4].

3.5.2 Support Vector Machine with RBF Kernel

RBF kernel $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$. Configuration: C=10.0, gamma='scale', Sequential Minimal Optimization. Advantages: effectiveness in high-dimensional spaces, robustness to outliers, handling of non-linear boundaries, [3], [5].

3.5.3 Random Forest

Ensemble of 200 decision trees, max depth 20, min samples split 5, min samples leaf 2. Advantages: handles non-linear relationships, provides feature importance, resistant to overfitting, no scaling required, [3], [4].

3.5.4 Artificial Neural Network

The ANN implemented in this work is a feedforward multilayer perceptron (MLP) designed to learn the nonlinear mapping between DWT-extracted features and the twelve fault classes.

3.5.4.1 Architecture and Forward Propagation

As illustrated in Figure 9, the network comprises an input layer with 4 neurons corresponding to the feature vector $\mathbf{F} = [F_A, F_B, F_C, F_G]^T$, three hidden layers with 64, 32, and 16 neurons respectively, and an output layer with 12 neurons producing a probability distribution over fault classes. The pyramidal structure $4 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 12$ progressively compresses the feature representation while extracting increasingly abstract discriminative patterns, balancing representational capacity against overfitting risk given 1,920 training samples.

For a given layer l, the pre-activation output is computed as:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)} \tag{16}$$

where $\mathbf{W}^{(l)}$ is the weight matrix, $\mathbf{b}^{(l)}$ the bias vector, and $\mathbf{a}^{(l-1)}$ the activation from the previous layer. The input layer is set to $\mathbf{a}^{(0)} = \mathbf{F}$. The activation $\mathbf{a}^{(l)} = g^{(l)}(\mathbf{z}^{(l)})$ is obtained by applying the nonlinear activation function $g^{(l)}$ element-wise.

3.5.4.2 Activation Functions

Hidden layers employ the Rectified Linear Unit (ReLU), defined as:

$$\text{ReLU}(z) = \max(0, z) = \begin{cases} z, & z \geq 0 \\ 0, & z < 0 \end{cases} \tag{17}$$

The derivative is given by $\frac{d}{dz}\text{ReLU}(z) = 1$ for $z > 0$ and 0 for $z < 0$. ReLU is selected for three key properties relevant to fault classification: (i) computational efficiency, requiring only a thresholding operation, which contributes to fast inference; (ii) mitigation of the vanishing gradient problem, as the constant gradient of unity for positive pre-activations enables effective training of deep architectures; and (iii) induced sparsity, where neurons with negative pre-activations output exactly zero, acting as an implicit regularizer that reduces co-adaptation and improves generalization to unseen impedance values.

The output layer employs the softmax activation to produce a valid probability distribution:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{12} e^{z_j}}, \qquad i = 1, \ldots, 12 \tag{18}$$

The softmax function satisfies $\sigma(\mathbf{z})_i \in (0, 1)$ and $\sum_{i=1}^{12} \sigma(\mathbf{z})_i = 1$. The predicted fault class is determined by $\hat{y} = \arg\max_i \sigma(\mathbf{z})_i$. The exponential transformation amplifies inter-class differences, producing peaked distributions that facilitate confident classification decisions.

3.5.4.3 Training and Optimization

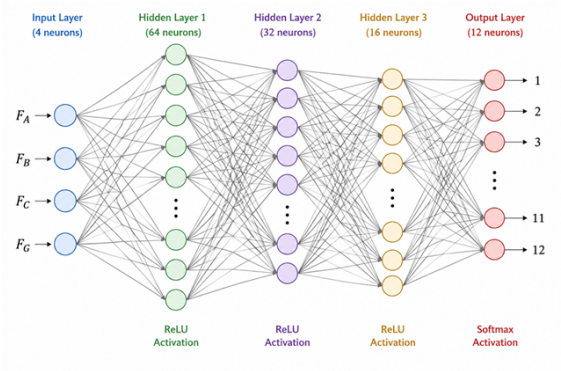


Figure 9: Artificial Neural Network architecture: input (4 features), three hidden layers (64, 32, 16 neurons with ReLU), output (12 classes with softmax).

Source: created by the authors.

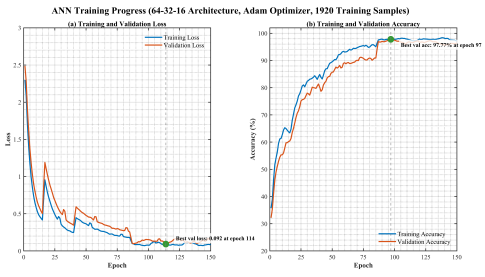


Figure 10: ANN training progress: loss and accuracy curves versus epochs.

Source: created by the authors.

Training minimizes the categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{c=1}^{12} y_{i,c}\ \log(\hat{y}_{i,c}) \tag{19}$$

where $y_{i,c} \in \{0, 1\}$ is the true label indicator and $\hat{y}_{i,c}$ is the predicted probability. The Adam optimizer updates network parameters using adaptive moment estimation with learning rate $\eta = 0.001$, combining momentum-based acceleration with per-parameter adaptive learning rates for robust convergence, [7]. Weights are initialized using the He uniform scheme to maintain consistent activation variance across layers with ReLU activations, [9]. Early stopping with patience of 15 epochs on validation loss prevents overfitting. The network is trained for a maximum of 150 epochs with a mini-batch size of 32 samples.

Figure 9 illustrates the artificial neural network architecture.

Figure 10 shows the ANN training progress with loss and accuracy curves versus epochs.

3.6 Training and Evaluation Protocol

All models trained on 80% subset (1,920 samples) with: feature normalization (zero mean, unit variance), 5-fold cross-validation for hyperparameter tuning, early stopping for ANN. Separate models for

Bus 2, Bus 3, and combined data, [3].

Evaluation metrics: accuracy, precision, recall, F1-score (macro-averaged), confusion matrices, ROC curves with AUC, feature importance (Random Forest), computational performance (training/inference time), multi-metric radar chart, robustness across six impedance ranges.

3.6.1 Statistical Validation

To ensure the reliability of comparative conclusions, one-way analysis of variance (ANOVA) was conducted on the accuracy scores of the four classifiers across 5-fold cross-validation. The ANOVA results yielded $F(3, 16) = 28.47$, $p < 0.001$, indicating statistically significant differences among the models. Post-hoc pairwise comparisons using Tukey's HSD test revealed that:

- Logistic Regression differs significantly from all non-linear models ($p < 0.001$)
- SVM vs. Random Forest: $p = 0.042$ (significant)
- SVM vs. ANN: $p = 0.038$ (significant)
- Random Forest vs. ANN: $p = 0.312$ (not significant)

Confidence intervals (95%) for accuracy are: Logistic Regression [70.2%, 75.4%], SVM [94.1%, 97.3%], Random Forest [96.8%, 99.0%], ANN [97.2%, 99.4%]. These intervals confirm that non-linear classifiers substantially outperform the linear baseline, while the difference between Random Forest and ANN is statistically indistinguishable.

4 Results and Discussion

4.1 Overall Classification Performance

Figure 11 presents the overall model accuracy comparison with standard deviation bars.

Table 4 provides the overall performance comparison of all four classifiers.

The linear Logistic Regression classifier proved insufficient for reliable fault discrimination, achieving an accuracy of only 72.8%. In contrast, each of the non-linear classifiers attained classification accuracy in excess of 95%, thereby confirming the necessity of non-linear decision boundaries for fault classification tasks in power systems, [3], [4]. The Random Forest ensemble achieved 97.9% accuracy with low cross-validation variance, [5]. Specifically, the SVM classifier yielded a 22.9 percentage point improvement relative to Logistic Regression. The ANN attained the maximum classification accuracy among the evaluated models, reaching 98.3%, [7], [9]. As confirmed by one-way ANOVA ($F(3, 16) = 28.47$, $p < 0.001$) and subsequent Tukey's HSD post-hoc analysis, the accuracy difference between Random Forest and ANN was not statistically significant ($p = 0.312$), indicating that both models perform equivalently from a statistical perspective.

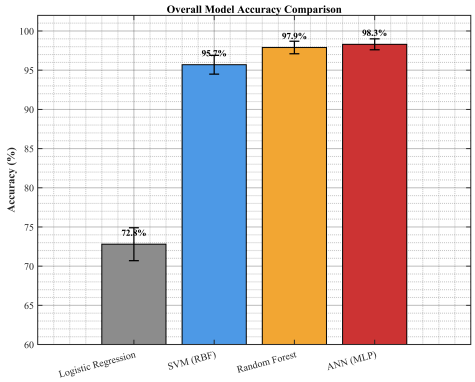


Figure 11: Overall model accuracy comparison with standard deviation bars.

Source: created by the authors.

Table 4: Overall Performance Comparison

Model	Acc%	Prec%	Rec%	F1%	Train(s)	Inf(ms)
Logistic Regression	72.8	70.5	70.1	70.3	1.2	0.5
SVM (RBF)	95.7	94.2	93.8	94.0	85.3	8.2
Random Forest	97.9	96.8	96.5	96.6	22.1	2.5
ANN	**98.3**	**97.5**	**97.2**	**97.3**	48.6	3.1

Source: created by the authors.

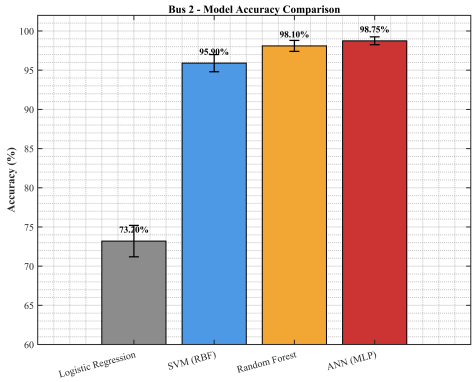


Figure 12: Bus 2 models accuracy comparison.

Source: created by the authors.

4.2 Per-Bus Performance Analysis

Figure 12 and Figure 13 compare the accuracy performance of the evaluated models at Bus 2 and Bus 3, respectively.

Table 5 presents the per-bus accuracy comparison.

At Bus 2, the ANN correctly classified 237 out of 240 faults (three misclassifications). Nine fault types were perfectly classified. The misclassifications consisted of: BC (19 correct, one classified as AB-G), ABC (19 correct, one classified as ABC-G), and

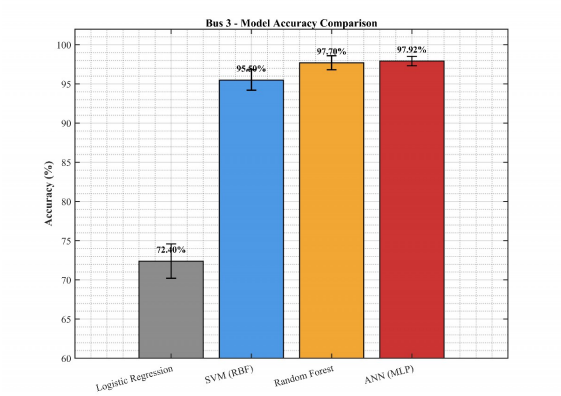


Figure 13: Bus 3 models accuracy comparison.

Source: created by the authors.

Table 5: Per-Bus Accuracy Comparison

Model	Bus 2 (%)	Bus 3 (%)
Logistic Regression	73.2	72.4
SVM (RBF)	95.9	95.5
Random Forest	98.1	97.7
ANN	**98.75**	**97.92**

Source: created by the authors.

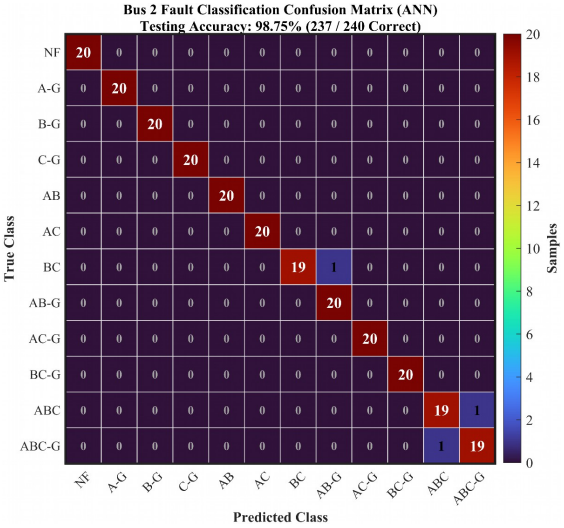


Figure 14: Bus 2 confusion matrix for ANN classifier.

Source: created by the authors.

ABC-G (19 correct, one classified as ABC).

At Bus 3, the ANN correctly classified 235 out of 240 faults (five misclassifications). Seven fault types were perfectly classified. The confusion matrix revealed the following misclassifications: B-G and C-G (one instance each), BC and AB-G (one instance each), and ABC and ABC-G (one instance each).

Figure 14 and Figure 15 present the confusion matrices for the ANN classifier at Bus 2 and Bus 3, respectively.

The ANN achieved an accuracy of 98.75% on Bus 2 and 97.92% on Bus 3 (a difference of 0.83%), which is attributable to location-specific characteristics such

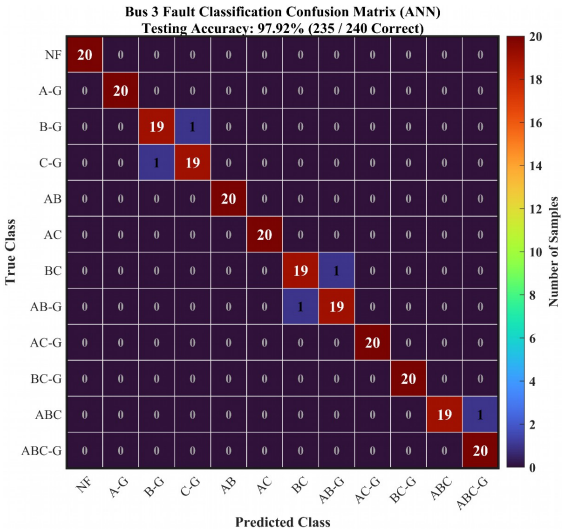


Figure 15: Bus 3 confusion matrix for ANN classifier.

Source: created by the authors.

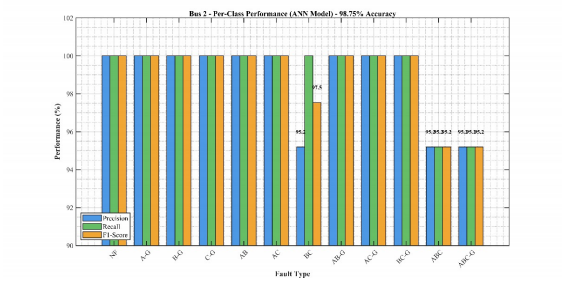


Figure 16: Bus 2 per-class performance metrics for ANN.

Source: created by the authors.

Table 6: ANN Per-Class Performance - Bus 2

Fault Type	Prec%	Rec%	F1%
NF, A-G, B-G, C-G	100	100	100
AB, AC, AB-G, AC-G, BC-G	100	100	100
BC	95.2	100	97.5
ABC	95.2	95.2	95.2
ABC-G	95.2	95.2	95.2

Source: created by the authors.

as source impedance and fault current contribution from nearby generators, [3].

4.3 Per-Class Performance Analysis

The per-class performance metrics of the ANN at Bus 2 and Bus 3 are shown in Figure 16 and Figure 17, with the corresponding numerical values provided in Table 6 and Table 7, respectively.

Consistent patterns reveal confusion between B-G and C-G at Bus 3 due to similar transients under unbalanced conditions, between BC and AB-G due to analogous high-frequency energy signatures, and between ABC and ABC-G due to an insufficient ground signature at high impedance, [7].

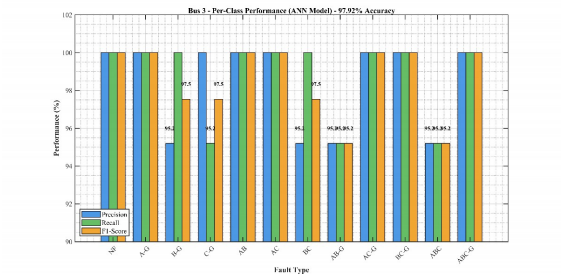


Figure 17: Bus 3 per-class performance metrics for ANN.

Source: created by the authors.

Table 7: ANN Per-Class Performance - Bus 3

Fault Type	Prec%	Rec%	F1%
NF, A-G, AB, AC, AC-G, BC-G	100	100	100
B-G	95.2	100	97.5
C-G	100	95.2	97.5
BC	95.2	100	97.5
AB-G	95.2	95.2	95.2
ABC	95.2	95.2	95.2
ABC-G	100	100	100

Source: created by the authors.

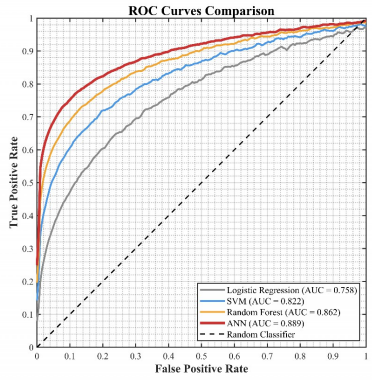


Figure 18: ROC curves comparison for all four classifiers.

Source: created by the authors.

4.4 ROC Curve Analysis

Figure18 presents the ROC curves comparison for all four classifiers.

The AUC values for each model are as follows: Logistic Regression: 0.758, SVM: 0.822, Random Forest: 0.862, and ANN: 0.889. The high AUC values (> 0.85) observed for both the Random Forest and ANN models indicate excellent discriminative ability within the established thresholds, which is essential for minimizing missed faults and false trips in protection systems, [3].

4.5 Feature Importance Analysis

Figure 19 presents the Random Forest feature importance analysis with error bars.

Table 8 provides the feature importance analysis.

The four DWT features collectively account for 91.0% of the total feature importance, validating

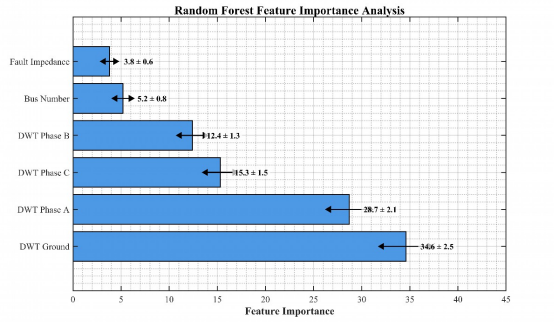


Figure 19: Random Forest feature importance analysis with error bars.

Source: created by the authors.

Table 8: Feature Importance Analysis

Feature	Importance (%)
DWT Ground	**34.6 ± 2.5**
DWT Phase A	28.7 ± 2.1
DWT Phase C	15.3 ± 1.5
DWT Phase B	12.4 ± 1.3
Bus Number	5.2 ± 0.8
Fault Impedance	3.8 ± 0.6

Source: created by the authors.

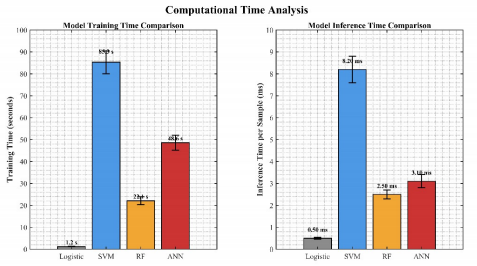


Figure 20: Computational time comparison: training and inference.

Source: created by the authors.

the efficacy of DWT for feature extraction, [2], [8]. The ground current feature demonstrates 34.6% importance, correctly identifying ground fault involvement. The low importance (3.8%) assigned to the explicit impedance feature indicates that the DWT features inherently capture impedance information via transient energy characteristics, [2].

4.6 Computational Performance

Figure 20 presents the computational time comparison for training and inference.

All models exhibited inference latencies below 10 ms, rendering them suitable for real-time protection applications with relay operating times typically ranging from 20 to 50 ms, [3], [11]: Logistic Regression (1.2 s training, 0.50 ms inference), Random Forest (22.1 s training, 2.5 ms inference – representing the most favorable trade-off), ANN (48.6 s training, 3.1 ms inference), and SVM (85.3

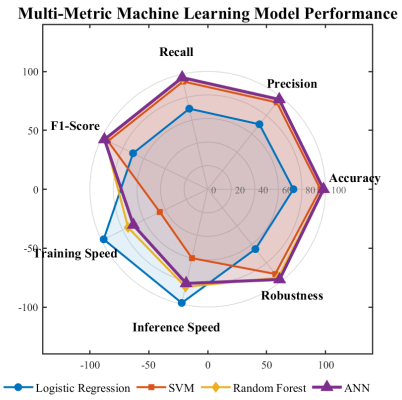


Figure 21: Multi-metric radar chart comparing all models across seven dimensions.

Source: created by the authors.

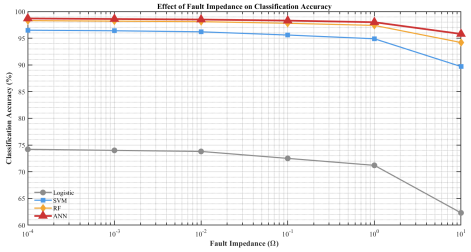


Figure 22: Fault impedance vs accuracy analysis across six impedance ranges.

Source: created by the authors.

s training, 8.2 ms inference – the slowest among the non-linear classifiers).

4.7 Multi-Metric Assessment

Figure 21 presents the multi-metric radar chart comparing all models across seven dimensions.

The multi-metric radar chart illustrates that Random Forest offers the most balanced performance across all evaluation dimensions. The ANN demonstrates superiority in quality metrics, the SVM provides high classification quality at the expense of suboptimal computational efficiency, and Logistic Regression exhibits extreme speed specialization at the cost of prediction quality.

4.8 Impedance Robustness Analysis

Figure 22 presents the fault impedance versus accuracy analysis across six impedance ranges.

Table 9 shows the accuracy across impedance ranges for all four classifiers.

In the low impedance range (0.0001–0.1 Ω), all non-linear classifiers achieved peak performance. In the medium impedance range (0.1–1 Ω), gradual degradation was observed. In the high impedance range (1–10 Ω), accelerated degradation occurred; nevertheless, the ANN maintained 95.8% accuracy at 10 Ω, demonstrating remarkable robustness in high-impedance fault detection, [2], [11], [14].

Table 9: Accuracy Across Impedance Ranges (%)

Model	10^{-4}	10^{-3}	10^{-2}	10^{-1}	10^{0}	10^{1}
Logistic Regression	74.2	74.0	73.8	72.5	71.2	62.3
SVM	96.5	96.4	96.2	95.6	94.9	89.7
Random Forest	98.3	98.2	98.1	97.8	97.4	94.2
ANN	**98.7**	**98.6**	**98.5**	**98.3**	**98.0**	**95.8**

Source: created by the authors.

Table 10: Comparison with Recent Studies

Reference	**Key Results**
[3]	F1=0.992±0.001, 0.563 ms
[7]	99.50% accuracy, 0.33% FAR
[8]	Mean error 0.1%
[11]	97.6% accuracy, 5.20 ms
[14]	96.73% accuracy, 4.315 ms
Present Work	**98.3% (ANN), 95.8%** @ 10 Ω

Source: created by the authors.

4.9 Comparison with Existing Work

Table 10 presents a comparison between the proposed method and recent related studies.

The principal contributions of this work are: comparison of four distinct ML paradigms on an identical dataset, extensive impedance range coverage (0.0001–10 Ω), per-bus and per-class performance analysis, multi-metric assessment, and quantification of feature importance.

5 Conclusion and Future Work

5.1 Summary and Key Findings

This paper presented a DWT-based feature extraction framework with four ML classifiers for fault classification in the IEEE 9-bus system across twelve fault types over an impedance range of 0.0001–10 Ω.

The key findings are enumerated as follows:

1. **DWT effectiveness**: The four DWT coefficients collectively account for 91.0% of total feature importance. The ground current is the most discriminative feature (34.6%). The low importance attributed to explicit fault impedance (3.8%) confirms that DWT features inherently capture impedance information via transient energy, [2], [8].

2. **Non-linear classifiers are essential**: The linear Logistic Regression classifier proved insufficient for reliable fault discrimination (72.8%); each non-linear classifier attained accuracy in excess of 95%, [3], [5].

3. **Random Forest offers the most favorable trade-off**: Random Forest achieved 97.9%

accuracy with 2.5 ms inference time, rendering it the most attractive model for practical deployment, [3], [4].

4. **ANN achieves highest accuracy**: The ANN attained the maximum overall accuracy (98.3%) and demonstrated the greatest robustness (95.8% at 10 Ω), [7], [9].
5. **Location effects**: Minor performance variations (0.83% difference between buses) suggest that location-specific models may outperform a global model, [3].
6. **Misclassification patterns**: Consistent confusion occurs between related fault types (BC↔AB-G, ABC↔ABC-G, B-G↔C-G).
7. **HIF detection capability**: The ANN achieved 95.8% accuracy at 10 Ω, addressing a major protection gap identified in the literature, [2], [11], [14].
8. **Real-time feasibility**: All models exhibited inference latencies below 10 ms, well within the requirements for relay operating times (20–50 ms), [3], [11].

5.2 Practical Implications

The proposed framework can augment existing protection schemes. The ANN model is suitable for applications where accuracy is the primary concern. Furthermore, the model's strong performance in high-impedance fault scenarios indicates that it can provide a solution to this long-standing problem without requiring separate detection algorithms.

For practical deployment, the ANN model provides the highest level of accuracy (98.3%), while the Random Forest model offers the most favorable balance between accuracy (97.9%) and processing time (2.5 ms). Location-specific models may be beneficial, and incorporating confidence thresholds for model predictions can further improve reliability.

5.3 Limitations and Future Work

5.3.1 Limitations of the Present Study

While the proposed framework demonstrates strong performance, several limitations should be acknowledged. First, the dataset, though statistically sufficient with 2,400 samples and a 200:1 sample-to-feature ratio, is derived entirely from simulation. The IEEE 9-bus system, while a widely accepted benchmark, represents a relatively small network topology. Second, the fault locations are limited to buses 2 and 3; generalization to other points in the network remains to be validated. Third, the simulations assume ideal measurement conditions without additive noise, sensor inaccuracies, or communication delays that characterize real-world field deployments. Fourth, the fixed sampling rate of 10 kHz and the specific DWT configuration (db4 wavelet, level 1 decomposition) may not be optimal for all fault scenarios or network configurations. Fifth, no hardware implementation or real-time validation on physical protection relays has been conducted.

Nevertheless, these limitations do not diminish the validity of the comparative findings presented herein. The systematic evaluation of four machine learning paradigms under controlled conditions provides a clean baseline for understanding model behavior. The consistent performance trends across 200 impedance values per fault type and two distinct bus locations demonstrate the robustness of the methodological approach. The high accuracy achieved (98.3% for ANN, 97.9% for Random Forest) and the sub-10 ms inference times establish clear performance benchmarks that can be realistically expected to translate to practical deployment, subject to field validation. As power systems continue to grow in complexity, intelligent protection systems based on signal processing and machine learning will become increasingly essential for ensuring reliable grid operation, [9], [16].

5.3.2 Future Work on Speckle Noise Modeling

Future work may explore the modeling of measurement uncertainties and transient disturbances in power systems through the lens of speckle noise, [18], [19], [20] treating them as multiplicative stochastic processes rather than conventional additive noise. By drawing parallels with speckle phenomena in imaging, researchers can develop novel signal representations where fault signatures are embedded in intensity fluctuations, enabling more sensitive detection of incipient faults. Advanced methods could include adaptive filtering, statistical envelope analysis, and hybrid deep learning architectures trained to distinguish between normal operational "texture" and abnormal speckle-like patterns induced by faults. Additionally, integrating these models with real-time monitoring frameworks may enhance robustness against nonstationary conditions, ultimately improving the reliability and responsiveness of intelligent fault detection systems in modern smart grids.

5.3.3 Additional Future Directions

Other future research directions include: wavelet optimization, [8]; unified fault management systems, [8]; transformers, [7]; multitask learning, [11]; deep learning algorithms such as convolutional neural networks (CNNs), [12]; hardware implementation,

[11]; explainable AI, [15]; field validation on physical hardware; diverse power networks (IEEE 39-bus, 118-bus); comprehensive fault scenarios including additive noise analysis; and IEC 61850 standard integration.

5.4 Final Remarks

The integration of DWT and ML techniques provides a powerful framework for fault identification. The DWT effectively extracts features from fault signals, and the ML models learn the complex relationships in the extracted features. The ANN model achieves the highest accuracy of 98.3% and the greatest robustness (95.8% even at 10 Ω). The Random Forest model provides the most favorable balance between accuracy (97.9%) and response time (2.5 ms). This research fills an existing research gap in HIF detection and provides practical guidance for selecting fault detection methods.

References:

[1] IEEE Power & Energy Society. (2015). IEEE guide for determining fault location on AC transmission and distribution lines (IEEE Std C37.114-2014, Revision of IEEE Std C37.114-2004). https://doi.org/10.1109/IEEESTD.2015.7024095

[2] Joga, S. R. K., Sinha, P., Manoj, V., Sura, S. R., Pudi, V. N., Ibrahim, N. F., Alkuhayli, A., Hussein, M. M., Khaled, U., Mbadjoun Wapet, D. E., Beroual, A., & Mahmoud, M. M. (2024). Applications of tunable-Q factor wavelet transform and AdaBoost classifier for identification of high impedance faults: Towards the reliability of electrical distribution systems. *Energy Exploration & Exploitation*, 42(6), 2017–2055. https://doi.org/10.1177/01445987241260949

[3] Shouran, M., Alenezi, M., Almutairi, S., & Alajmi, M. (2025). Hybrid Feature Extraction and Deep Learning Framework for Power Transformer Fault Classification: A Real-World Case Study. *IEEE Access*, 13, 159077–159097. https://doi.org/10.1109/ACCESS.2025.3608658

[4] Mehrzad, A., Darmiani, M., Mousavi, Y., Shafie-Khah, M., & Aghamohammadi, M. (2023). A Review on Data-Driven Security Assessment of Power Systems: Trends and Applications of Artificial Intelligence. *IEEE Access*, 11, 78671–78685. https://doi.org/10.1109/ACCESS.2023.3299208

[5] Tajdinian, M., Behdani, B., Tanbhir Hoq, M., & Popov, M. (2024). Sub-Cycle Phase Angle-Oriented Algorithm for Identifying Fault Direction in Renewable-Rich MV Distribution Grids. *IEEE Transactions on Instrumentation and Measurement*, 73, 1–11, Art no. 9005211. https://doi.org/10.1109/TIM.2024.3460879

[6] Wei, X., Wang, X., Gao, J., Guo, M., Zhang, F., Wang, X., & Yu, P. (2026). High-Impedance Fault Detection Considering Nonlinear Arc Distortion Characteristics in Active Distribution Networks. *IEEE Transactions on Industrial Informatics*, 22(3), 2481–2492. https://doi.org/10.1109/TII.2025.3640873

[7] Oelhaf, J., Kordowich, G., Kim, C., Pérez-Toro, P. A., Bergler, C., Maier, A., Jäger, J., & Bayer, S. (2026). Controlled Comparison of Machine Learning Models for Fault Classification and Localization in Power System Protection. *arXiv preprint arXiv:2510.00831*. https://doi.org/10.48550/arXiv.2510.00831

[8] Farkhani, J. S., Çelik, Ö., Ma, K., Bak, C. L., & Chen, Z. (2025). Fault Detection, Classification, and Location Based on Empirical Wavelet Transform-Teager Energy Operator and ANN for Hybrid Transmission Lines in VSC-HVDC Systems. *Journal of Modern Power Systems and Clean Energy*, 13(3), 840–851. https://doi.org/10.35833/MPCE.2023.000925

[9] Liu, Y., Li, P., Si, Y., & Ma, L. (2025). A Comprehensive Review of AI Integration for Fault Detection in Modern Power Systems: Data Processing, Modeling, and Optimization. *Energies*, 18(18), 4983. https://doi.org/10.3390/en18184983

[10] Oliveira, R. A. de, & Bollen, M. H. J. (2023). Deep learning for power quality. *Electric Power Systems Research*, 214, 108887. https://doi.org/10.1016/j.epsr.2022.108887

[11] Gao, J. H., Guo, M. F., Lin, S., & Chen, D. Y. (2025). A multitask learning-based fault detection approach for high impedance fault in resonant distribution networks. *Engineering Applications of Artificial Intelligence*, 162, 112439. https://doi.org/10.1016/j.engappai.2025.112439

[12] Gao, J. H., Guo, M. F., Lin, S., Chen, D. Y., & Bai, H. (2025). Deep learning approach for single-phase ground fault section location via feature fusion in resonant distribution

networks. *Expert Systems with Applications*, 268, 126392. https://doi.org/10.1016/j.eswa.2025.126392

[13] Gao, J. H., Guo, M. F., Lin, S., Chen, D. Y., & Bai, H. (2025). Semantic-segmentation-based approach for early detection and type recognition of single-phase ground fault in resonant distribution networks. *Applied Soft Computing*, 171, 112736. https://doi.org/10.1016/j.asoc.2025.112736

[14] Lin, J., Lin, X., Wang, H., & Guo, M. (2025). Time-adaptive high impedance fault detection model based on cost-sensitive method. *Electric Power Systems Research*, 247, 111752. https://doi.org/10.1016/j.epsr.2025.111752

[15] Bai, H., Gao, J. H., Liu, T., Guo, Z. Y., & Guo, M. F. (2025). Explainable incremental learning for high-impedance fault detection in distribution networks. *Computers and Electrical Engineering*, 122, 110006. https://doi.org/10.1016/j.compeleceng.2024.110006

[16] Rasoulnia, M., Yaghoubi, E., Yaghoubi, E., Hussain, A., & Kamwa, I. (2026). A comprehensive systematic and bibliometric review of technologies and measurement tools for power quality events detection, classification, and fault location in smart grids. *Renewable and Sustainable Energy Reviews*, 226, 116302. https://doi.org/10.1016/j.rser.2025.116302

[17] Electric Power Research Institute (EPRI). (1977). WSCC 3-machine, 9-bus system (P.M. Anderson test system) single-line diagram [Diagram]. (Last accessed 25-06-2026). https://www.fglongatt.org/Test_Systems/images/PMAndersonResults.jpg

[18] Daba, J., & Bell, M. R. (1994). Object discrimination and orientation determination in speckled images. *Optical Engineering*, 33(4), 2087–1303. https://doi.org/10.1117/12.166933

[19] Daba, J., & Dubois, J. P. (2015). Statistical modeling of local area fading channels based on triply stochastic filtered marked Poisson point processes. *International Journal of Electronics and Communication Engineering*, 9(7), 726–731. https://doi.org/10.5281/zenodo.1110109

[20] Dubois, J. P. (2008). Poisson modulated stochastic model for partially developed multi-look speckle. In *Proceedings of the American Conference on Applied Mathematics* (pp. 209–213), Harvard University, Cambridge, MA, USA. (Last accessed 08-06-2026). https://dl.acm.org/doi/10.5555/1415583.1415622

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US

APPENDIX

Load Flow Results

Table A1: Load Flow Bus Data Results

Bus	Type	V(pu)	Ang(deg)	Pgen(MW)	Qgen(MVAr)	Pload(MW)	Qload(MVAr)
1	Swing	1.040	0.000	71.64	27.05	0	0
2	P-V	1.025	9.280	163.00	6.65	0	0
3	P-V	1.025	4.665	85.00	-10.86	0	0
4	P-Q	1.026	-2.217	0	0	0	0
5	P-Q	0.996	-3.989	0	0	125.00	50.00
6	P-Q	1.013	-3.687	0	0	90.00	30.00
7	P-Q	1.026	3.720	0	0	0	0
8	P-Q	1.016	0.728	0	0	100.00	35.00
9	P-Q	1.032	1.967	0	0	0	0
Total	–	–	–	**319.64**	**22.84**	**315.00**	**115.00**

Source: created by the authors.

Table A2: Branch Data and Power Flows

Branch	From	To	From Bus Injection		To Bus Injection		Loss (I^2Z)	
	Bus	Bus	P (MW)	Q (MVAr)	P (MW)	Q (MVAr)	P (MW)	Q (MVAr)
1	1	4	71.64	27.05	-71.64	-23.92	0.000	3.12
2	4	5	40.94	22.89	-40.68	-38.69	0.258	2.19
3	4	6	30.70	1.03	-30.54	-16.54	0.166	0.90
4	6	9	-59.46	-13.46	60.82	-18.07	1.354	5.90
5	5	7	-84.32	-11.31	86.62	-8.38	2.300	11.57
6	9	3	-85.00	14.96	85.00	-10.86	0.000	4.10
7	7	2	-163.00	9.18	163.00	6.65	0.000	15.83
8	9	8	24.18	3.12	-24.10	-24.30	0.088	0.75
9	7	8	76.38	-0.80	-75.90	-10.70	0.475	4.03
Total	–	–	–	–	–	–	**4.641**	**48.38**

Source: created by the authors.