International Innovative Research Journal of Engineering and Technology
Vol. 12, Issue. 1, Sep 2026, pp. 99~115
ISSN: 2456-1983, DOI: 10.32595/iirjet.org/v12i1.2026.256

An Explainable Machine Learning Framework for Fault Detection, Classification, and Localization in Smart Distribution Networks

Shurooq Ahmed Mohamme[1], Nada Amir Humaid Al Jassasi[2], Muna Ali Mohammed Al Shehhi[3]

[1,2,3]UG Student, Department of Information Technology,
University of Technology and Applied Sciences, Shinas, Oman.

Article Info

Article History:

Received Jun 21, 2026
Revised Jul 18, 2026
Accepted Aug 25, 2026

Keywords:

Smart distribution networks,
fault detection,
fault classification,
fault localization,
explainable artificial intelligence,
machine learning,
SHAP,
power system protection

ABSTRACT

Smart distribution networks are more complicated due to the combination of distributed generation, renewable energy sources, and enhanced monitoring technologies. These developments describe difficulties in determining electrical defects, identifying different types of faults, and precisely finding fault locations under various operating circumstances. This study suggests an Explainable Artificial Intelligence-Based fault Detection, Classification, and Localization (XAI-FDCL) in smart distribution networks. The suggested structure uses three-phase voltage and current measurements to extract pertinent electrical information, such as sequence components, statistical traits, and root mean square values. For detecting fault conditions and categorizing common fault types, such as single-line-to-ground, line-to-line, double-line-to-ground, and three-phase faults, Random Forest, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost) algorithms are evaluated. To determine the fault's position within the distribution network, feeder-section classification and distance regression are used in fault localization. Furthermore, Shapley Additive explanations (SHAP) are integrated to determine the electrical characteristics that aid in defect diagnosis and to evaluate model forecasts. The framework is evaluated using accuracy, precision, recall, F1-score, mean absolute error, root mean square error, and computational latency. The suggested approach aims to model transparency, localization accuracy, and fault-diagnosis reliability in order to facilitate better decision-making and effective fault management in smart distribution networks.

Corresponding Author:

Shurooq Ahmed Mohamme,
UG Student, Department of Information Technology,
University of Technology and Applied Sciences,
Shinas, Oman.

1. INTRODUCTION

Smart distribution networks are growing more complicated because distributed generation, smart sensors, advanced monitoring technology, and renewable energy sources are all integrated.

These developments improve the operating efficiency and flexibility of the power system, but also create challenges for the diagnosis of electrical faults. Bidirectional power flow, fluctuations in short-circuit current, changing network configurations, and different loading conditions can all affect how effective classic protection and fault-location approaches are. Fault current variations, bidirectional power flow, changing load demand, and network architecture can all hinder the identification and localization of electrical defects. Thus, the development of intelligent fault-diagnosis systems is essential for improving network dependability, reducing service interruptions, and uplifting efficient power-system operation. [1] [2].

Single-line-to-ground (LG), line-to-line (LL), double-line-to-ground (LLG), and three-phase electrical faults can result in unusual voltage and current conditions and affect the stability of distribution networks. Three complementary functions are required for effective fault diagnosis: fault detection, which indicates the presence of an abnormal condition; fault classification, which determines the type of fault; and fault localization, which identifies the affected feeder segment and computes the physical distance of the fault. The potential of machine learning techniques for fault identification and localization under various network operating conditions has been demonstrated by recent studies [3] [4].

Deviations in fault resistance, measurement noise, distributed-generation penetration, and loading circumstances, however, continue to be important factors that need to be taken into account when assessing diagnostic performance. Because machine learning methods may identify nonlinear relationships between electrical measurements and problem variables, they are suitable for data-driven fault identification. In order to identify anomalous conditions and differentiate between various fault types, electrical features extracted from three-phase voltage and current data can be useful. While traditional techniques like Random Forest and Support Vector Machine (SVM) offer helpful substitutes for comparative assessment, tree-based algorithms like Extreme Gradient Boosting (XGBoost) can simulate intricate feature interactions.

Nevertheless, the reliability of a diagnostic system must demonstrate dependable fault localization, acceptable computational delay, and dependable performance in a variety of operational conditions in addition to classification accuracy. Another important consideration when using machine learning in power-system monitoring is explainability. Inadequate information for engineering research may result from a predictive model that fails to explain the reasons that influence its choice. The increasing significance of prediction transparency in intelligent power systems is demonstrated by recent studies that have investigated explainable artificial intelligence for defect detection, classification, and localization[5], [6].

SHapley Additive exPlanations (SHAP) feature-attribution values help illustrate how particular electrical qualities influence model predictions [7]. These explanations can assist engineers in determining whether a diagnosis is influenced by sequence elements, voltage deviations, current changes, or other extracted properties. However, feature-attribution explanations must be analyzed as well as electrical system expertise and impartial performance evaluation. This study suggests an Explainable XGBoost–SHAP (XGB-SHAP) framework for fault detection, classification, and localization in smart distribution networks to fulfill these requirements. The framework retrieves pertinent electrical characteristics, such as root mean square values, statistical features, and sequence components, from three-phase voltage and current observations. While a different multiclass classifier classifies detected defects, an XGBoost classifier distinguishes between faulty and normal operational circumstances.

Feeder-section classification and an XGBoost regression model that calculates the physical distance to the fault are used for fault localization. SHAP is integrated to explain model predictions

and identify influential electrical features. SVM and Random Forest are used as benchmark techniques to evaluate the effectiveness of the suggested strategy. Accuracy, precision, recall, mean absolute error (MAE), root mean square error (RMSE), F1-score, and computational delay are used to evaluate the framework.

Other experiments examine distinct fault resistances, load scenarios, distributed-generation levels, and measurement-noise scenarios to assess resilience. The main contribution is a systematic framework for diagnosis that combines fault detection, classification, feeder-section identification, fault-distance estimation, and explainability within a common evaluation pipeline. The proposed study is to facilitate accurate and transparent fault diagnosis and offer helpful data for smart distribution network decision-making.

2. LITERATURE REVIEW

Zaben et al. [8] exhibited an extensive analysis of machine learning techniques for AC microgrid fault diagnostics. The study investigated techniques for fault detection, classification, and localization and talked about how network topology, operating modes, load fluctuations, distributed-generation penetration, and fault resistance affect diagnostic performance. The authors examined various deep learning and machine learning techniques and identified research issues related to intelligent protection. This study provides a foundation for evaluating fault-diagnosis models under different network operating conditions. However, as a review, it does not determine the effectiveness of a single unified model that combines explainable forecasts with machine learning-based diagnostics.

Cao et al. [2] examined fault-diagnosis techniques that integrate signal processing and artificial intelligence for electricity distribution networks. Their study examined data dimensionality issues, signal modifications, feature extraction, and the benefits and limitations of different machine learning and deep-learning models. The authors also determined data security and the integration of engineering knowledge with AI as important directions for future research. This work assists the use of structured preprocessing involves the extraction of electrical features prior to model training. However, the selection and coordination of models for fault detection remain important implementation decisions for individual research studies.

Bai et al. [9] suggested an explainable incremental-learning technique for distribution network high-impedance fault detection. The approach used extracted discrete wavelet transform coefficients from zero-sequence current signals is a regularization-based learning technique that updates an artificial neural network. Feature contributions were examined using SHAP, and the practical applicability of the method was investigated through simulations and field-related validation. The study is significant because it tackles the challenge of shifting signal distributions over time by combining explainability and model adaptation. However, its main goal is to detect high-impedance faults instead of estimating physical fault distances and classifying multiple categories.

Siddique et al. [10] proposed a hierarchical architecture for fault classification and placement in active distribution networks using phasor measurement units (PMUs) based on deep convolutional neural networks (CNNs). The technique used multiple CNN models to perform fault diagnosis while contemplating uncertainty related to variations in load and distributed generation. Its conclusions emphasize the need of measurement availability and changing operating conditions while creating data-driven diagnostic systems. Nevertheless, more research is required to determine

the performance of intelligible tree-based machine learning models when defect detection, classification, localization, and feature-level explanation are all regarded as at the same time..

Hou and Wang [11] demonstrated an enhanced AlexNet architecture, Markov transition fields (MTF), and transfer entropy as the foundation for a knowledge-integrated fault-section localization technique. The technique converted transitory zero-sequence current signals into representations based on images that recorded cause and temporal associations. A feature-attention method that used physical information was used to identify important fault-related qualities, and a similarity-weighting approach was applied to enable the detection of problematic feeder sections. This study shows how signal modification could enhance fault identification and interpretation through physical knowledge. However, the method focuses mostly on single-phase grounding fault-section identification, leaving room to investigate a unified tree-based framework that also performs multiclass fault classification and physical distance estimation.

Dhanunjayudu et al. [12] developed a noise-resistant framework for locating, classifying, and identifying power-quality and fault occurrences in distribution networks that integrate renewable energy sources. Using empirical mode decomposition (EMD), they combined support vector machine and support vector regression techniques. To enable rapid event detection in noisy environments, the structure used brief observation windows and single-end substation measurements and brief observation windows. The method was evaluated by the authors using IEEE 13-node and IEEE 34-node distribution systems that integrated renewable energy. According to their report, a fault resistance value of up to 7 Ω and a signal-to-noise ratio of 20 dB resulted in a classification accuracy of 97.8%. This work is highly significant since it integrates the three primary diagnostic tasks and investigates performance under challenging operational conditions. The suggested XGB-SHAP method differs from its primary methodology in that it emphasizes feature-attribution explanations alongside a tree-based classification and regression pipeline.

Sun et al. [13] provided a graph-based framework for the categorization and localization of defects in scenarios with sparse measurements. They use an expanded feature propagation method based on residual graph convolutional networks that incorporates neighbor filtering. Feature propagation was used to restore information for unmeasured nodes, while the graph convolutional network found relationships between network nodes and addressed the problem of oversmoothing in deeper graph structures. The technique was evaluated with varying measurement errors and distribution-network operational circumstances. When measurements are scarce, the study highlights the importance of network structure. However, the computational and interpretability features of tree-based machine learning models and graph-based architectures differ, which encourages a comparative assessment of an explainable tree-based approach under well-defined measurement conditions.

Yang et al. [14] mentioned a multiscale convolutional neural network (MS-CNN) and spatiotemporal dual attention-based two-stage fault-diagnosis method. The system integrated MS-CNN features with temporal modeling using Transformer and the SimAM attention module, separating fault classification from fault-section localization. The purpose of this architecture is to capture the geographically and temporally dependent characteristics of electrical impulses particularly when fault features are weak or adjacent fault sections are difficult to distinguish. The study demonstrates the importance of learning spatial and temporal patterns as well as distinct diagnostic tasks. The proposed XGB-SHAP research uses a new architecture based on predefined electrical features, gradient-boosted decision trees, and SHAP explanations; performance will be evaluated by experimental comparison.

Recent research on fault diagnostics in smart distribution networks aims to combine explainable artificial intelligence, machine learning, signal processing, and topological information to provide an accurate, precise, and interpretable solution. These are still challenging areas to overcome. This study provides a novel SHAP explainable XGBoost approach to accurately detect, classify, identify, and regress the fault distance within a distribution network. The classification accuracy, location error, inference time, and robustness of the models are evaluated. Random Forest and SVM models are used as benchmarks.

3. METHODOLOGY

The proposed Explainable Artificial Intelligence-Based Fault Detection, Classification, and Localization (XAI-FDCL) framework is designed to identify electrical faults, classify their types, and determine their locations in smart distribution networks. The methodology integrates three-phase voltage and current research uses a new architecture research uses a new architecture, electrical feature extraction, Random Forest-based fault detection and classification, fault distance estimation, and explainable artificial intelligence using SHapley Additive exPlanations (SHAP). The proposed model consists of five major stages: data acquisition, signal preprocessing and feature extraction, fault detection and classification, fault localization, and explainability-based decision support. Under various operational conditions, such as the integration of renewable energy, fluctuating loads, and bidirectional power flow, each step helps to improve the fault analysis's dependability, accuracy, and interpretability.

The architecture of the proposed explainable XAI-FDCL framework is illustrated in Figure 1. The smart distribution network's three-phase voltage and current measurements are obtained at the start of the framework. The collected signals endure preprocessing, such as time-window segmentation, normalization, and noise filtering, to prepare appropriate input data for analysis. The network's operational conditions are then represented by electrical attributes such features of statistics, harmonic components, symmetrical sequence components, and root mean square (RMS) values.

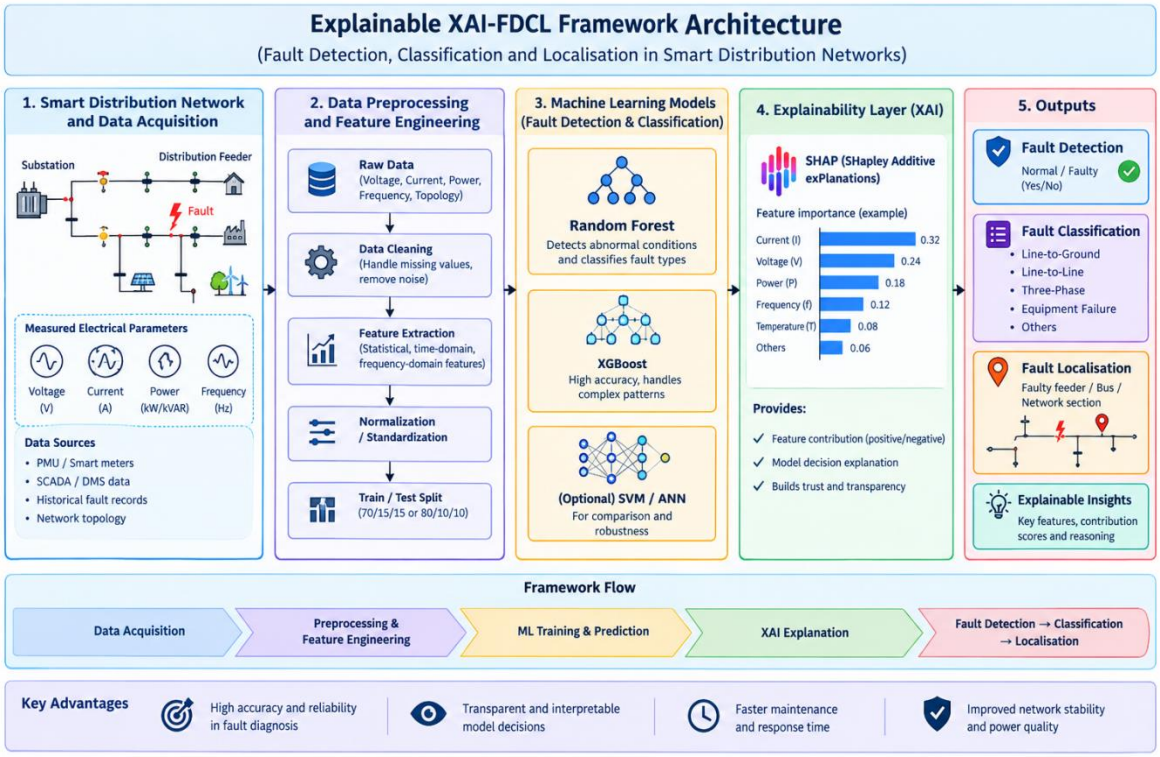


Figure 1. Architecture of the proposed explainable XAI-FDCL framework

These features are supplied to the XGBoost based fault detection model, which distinguishes normal operation from faulty conditions. A different XGBoost classification model determines the type of fault, such as single-line-to-ground, line-to-line, double-line-to-ground, and three-phase faults, upon detection. The fault localization stage uses an XGBoost regression model to estimate the physical distance to the fault and a classification model to identify the impacted feeder segment. Moreover, the integration of SHAP enhances the interpretability of the diagnostic results by identifying the electrical properties that contribute to the model predictions. Finally, the proposed framework is evaluated XAI-FDCL is evaluated against and Support Vector Machine (SVM) and Decision Tree benchmark models using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), and computational latency. Thus, XAI-FDCL assembles XGBoost-based fault detection, classification, and localization with SHAP-based explanations in a unified process.

3.1. Data Acquisition and Signal Preprocessing

Three-phase voltage and current signals are assembled from various busses and monitoring locations within the smart distribution network in the first step. The measured signals include phase voltages V_a, V_b, V_c and phase currentsI_a, I_b, I_c. Both normal and defective operating circumstances, such as single-line-to-ground (LG), line-to-line (LL), double-line-to-ground (LLG), and three-phase (LLL) faults, are measured. To ensure reliable analysis, the integrated signals are preprocessed using techniques for noise reduction and normalization. By scaling the input features into a common range, min–max normalization is used to lessen the impact of variances in measurement magnitudes. To ensure reliable analysis, the integrated signals are preprocessed using techniques for noise reduction and normalization. By scaling the input features into a common range, min–max normalization is used to lessen the impact of variances in measurement magnitudes.

The normalization process is described as:

$$x_i^{norm} \quad \frac{x_i - x_{min}}{x_{max} - x_{min} + \epsilon} \tag{1}$$

Where x_i shows the original measurement x_{min} and x_{max} show the minimum and maximum values of the associated characteristics; ϵ is a tiny positive constant that stops division by zero. During model training, this procedure improves numerical stability. To prevent data leaking, the normalization parameters should be established using the training data and then applied uniformly to validation and test data.

3.2. Extraction of Electrical Features

The voltage and current signals are used in the second stage to extract important electrical characteristics. Among the selected characteristics are the mean, standard deviation, symmetrical components, signal energy, and root mean square (RMS) values. These characteristics help distinguish fault conditions from ordinary network performance because electrical failures usually generate changes in voltage magnitude, current magnitude, phase relationships, and sequence components.

The RMS value of a sampled electrical signal is calculated as follows:

$$X_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{2}$$

Where x_i shows the voltage or current sampled at that moment. $i - th$ or current value and N is the number of samples in the analysis window. RMS measurements show the effective strength of electrical impulses and are helpful in identifying sudden variations in current or voltage after a breakdown.

The standard deviation can be computed using this formula:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3}$$

Where $\bar{x}$ is the signal's sample mean. The standard deviation quantifies signal fluctuations and helps identify sudden changes connected to the beginning of a breakdown. Symmetric components are used to transform three-phase electrical variables into positive-, negative-, and zero-sequence components into positive- , negative-, and zero-sequence components. For the phase voltage vector, the transformation is

$$\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix}\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \tag{4}$$

Where $V_0, V_1\ and\ V_2$ represent zero-, positive-, and negative-sequence voltages, respectively, and $a = e^{j2\pi/3}$ is the complex operator representing a 120° phase rotation. The same transformation is applied to phase currents to acquire $I_0, I_1, and\ I_2$. The features that were extracted are combined into a feature vector:

$$F = [V_{RMS}, I_{RMS}, \sigma_V, \sigma_I, V_0, V_1, V_2, I_0, I_1, I_2, E_V, E_1] \tag{5}$$

Where $E_V\ and\ E_1$represent the energy of the voltage and current signals. This feature vector acts as the suggested learning model's input.

3.3. Deep Learning Model-Based Fault Detection

The third stage identifies whether the distribution network is operating under normal or faulty conditions. An XGBoost binary classification model receives the derived electrical features

and uses them to learn nonlinear correlations between electrical measurements and fault characteristics. A multilayer neural network is able to serve as the detection module because it can learn complex feature patterns caused by changes in fault impedance, dispersed generation, and load demand.

The expression for the hidden-layer operation is

$$h^{(l)} = ReLU\left(W^{(l)}h^{(l-1)} + b^{(l)}\right) \tag{6}$$

Where $h^{(l)}$is the hidden layer's output l, $W^{(l)}$represents the trainable weight matrix, $b^{(l)}$denotes the bias vector , and $ReLU$ is the rectified linear unit activation function. The initial input is $h^{(0)} = F$.

The probability of a fault is determined using the sigmoid activation function:

$$P_{fault} = \frac{1}{1+exp(-z)} \tag{7}$$

Where z is the detection layer's output. The decision rule is

$$\hat{y}_d = \begin{cases} 1, \ P_{fault} \geq \tau \\ 0, P_{fault} \leq \tau \end{cases} \tag{8}$$

Where $\hat{y}_d = 1$ shows a detected fault, $\hat{y}_d = 0$ shows regular operation and τ is the threshold for detection. Usually initialized at 0.5 and modified as needed using validation data.

Binary cross-entropy loss is used to train the model.

$$L_d = -\frac{1}{M}\sum_{i=1}^{M}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{9}$$

Where M is the quantity of training samples, y_i is the actual binary label, and p_i is the estimated fault probability. Reducing this loss helps the detection model to differentiate between defective conditions and regular functioning. The detection output is consequently passed to the classification stage when a fault is identified.

3.4. Classification of Faults

After identifying the fault, the particular fault category is determined by the categorization module. The four main fault categories—LL, LLG, LLL, and normal operation—are taken into account by the suggested framework. The classification model differentiates fault patterns linked to various phase combinations and grounding conditions using the retrieved features and trained hidden representations.

Using the softmax function, each fault class's probability is determined:

$$P = (c = k \,|\boldsymbol{F}) = \frac{exp(z_k)}{\sum_{j=1}^{K} expz_j} \tag{10}$$

Where z_k is the class's output score k, and K is the number of classes. For an implementation with five classes, $K = 5$, comprising the four categories of faults as well as normal operation.

The predicted fault class is selected as

$$\hat{c} = arg_{k\in\{1..,k\}}^{max} P\ (c = k|\ \boldsymbol{F}) \tag{11}$$

Where $\hat{c}$ represents the predicted fault category. The categorical cross-entropy loss is used to train the classification network:

$$L_c = -\frac{1}{M}\sum_{i=0}^{M}\sum_{k=1}^{K} y_{ik} \log(p_{ik}) \tag{12}$$

where y_{ik} is the ground truth label with one-hot encoding p_{ik} expected probability for class k in sample i This procedure allows automated fault-type identification and enables the model to learn distinctions between fault categories.

3.5. Localization of Faults

The physical location of a detected fault inside the distribution network is estimated during the fault localization stage. The localization module estimates the distance between a reference bus and the problem location using the extracted voltage and current features, measurements from accessible monitoring stations, and network data. Simulated or experimentally recorded faults with known locations across several line sections should be used to train the model for practical application.

The physical distance to the fault is estimated by an XGBoost regression model:

$$\hat{d} = f_{\theta}(\boldsymbol{F}, \boldsymbol{G}) \tag{13}$$

Where $\hat{d}$ is the fault distance estimate, f_{θ} is the model for trained localization, $\boldsymbol{F}$ is the vector of electrical features, and $\boldsymbol{G}$ shows the information available on the network, such as bus connectivity, line length, and line impedance. The localization error can be computed as follows if the fault is represented by its distance along a line:

$$e_d = d - \widehat{d} \tag{14}$$

Where d is the real distance of the fault and $\widehat{d}$ is the approximate distance. Mean squared error can be used to train the localization network:

$$L_l = \frac{1}{M}\sum_{i=1}^{M}\left(d_i - \hat{d}_i\right)^2 \tag{15}$$

Where d_i indicates the known fault distance and $\hat{d}_i$ is the sample's expected distance i. The discrepancy between the estimated and actual fault locations is decreased by minimizing this loss.

In networks with several line sections, the model may estimate the distance along the problematic line section after first identifying it. If the line length is L, an alternative way to depict the normalized location is as

$$\hat{r} = \frac{\hat{d}}{L} \tag{16}$$

Where $\hat{r}$ shows the relative fault location along the line, provided L>0 and the specified reference end is used to measure the distance. This method facilitates fault localization between distribution feeders with varying line lengths.

3.6. Explainable Artificial Intelligence Using SHAP

The electrical characteristics that have the greatest influence on the fault detection, classification, and localization outputs are identified at the explainability stage. To explain predictions produced by the XGBoost models, the suggested XAI-FDCL framework includes SHapley Additive exPlanations (SHAP).

The feature's SHAP explanation j is determined as

$$\emptyset_j = \sum_{s \subseteq F \setminus \{j\}} \frac{|S|!(n-|S|-1)!}{n!}\left[f(S \cup \{j\}) - f(S)\right] \tag{17}$$

Where $F\ is\ all\ of\ the\ input\ features,\ n$ is the number of features, S is a subset that excludes feature$j$, and $f(S)$ shows the model's output when the characteristics in S are considered. The quantity $\emptyset_j$ indicates the contribution of feature j regarding the chosen baseline.

For example, In comparison to the baseline, a high positive SHAP value for a specific fault class in the negative-sequence current shows that this characteristic strongly supports that prediction. The feature moves the prediction away from the chosen output when the SHAP value is negative. The explained model output and baseline determine the specific interpretation.

The mean absolute SHAP value can be used to measure a feature's global importance:

$$I_j^{SHARP} = \frac{1}{M}\sum_{i=1}^{M} |\emptyset_{ij}| \quad (18)$$

Where $\emptyset_{ij}$ represents the SHAP value of feature j for sample i, and M is the quantity of assessed samples. The average impact of features with higher mean absolute SHAP values on the model predictions is higher.

Feature-importance charts and individual prediction explanations can be used to display the SHAP results. These explanations help protection engineers understand why a fault was detected, which parameters affected the expected problem location and which electrical characteristics affected the fault categorization. Although SHAP increases model transparency, it does not ensure that the model's predictions are causally accurate on its own.

3.7. Collaborative Optimization of the Suggested Framework

The fault-detection, classification, and localization modules can be skilled as separate XGBoost models. The suggested XAI-FDCL framework can be used to arrange their evaluation.

$$L_{total} = \lambda_d L_d + \lambda_c L_c + \lambda_l L_l \quad (19)$$

Where $L_d, L_c \; and \; L_l$ represent the relative losses in detection, classification, and location. The coefficients $\lambda_d, \lambda_c \; and \; \lambda_l$ regulate the three objectives' respective contributions.

To reduce the overall loss, the model's parameters are modified.

$$\theta^* = \arg min_\theta \, L_{total}(\theta) \quad (20)$$

Where θ shows the parameters can be trained and θ^* indicates the optimized parameters. Validation experiments can be used to choose the loss weights. Because classification losses and distance-based losses may have different numerical scales, it is important to examine their relative magnitudes during training.

The SHAP explanation is produced either during model evaluation or after training; it is not required to be part of the training loss. This division enables the suggested model to provide clear justifications for the regarding the XGBoost models' predictions

3.8. Evaluation of Performance Metrics of the Proposed XAI-FDCL Model

The performance of the suggested Explainable Machine Learning Framework for Fault Detection, Classification, and Localization (XAI-FDCL) is evaluated using accuracy, precision, recall, F1-score, mean absolute error (MAE), root mean square error (RMSE), and computational latency are represented in Table 1. The framework applies Random Forest as its predictive technique, and SHapley Additive exPlanations (SHAP) is utilized to analyze the model's predictions. Accuracy, precision, recall, and F1-score are expressed as percentages, whereas MAE and RMSE quantify fault localization errors in meters. Computational latency is measured in milliseconds. These metrics assess the framework's classification performance, localization accuracy, and computational efficiency. Classification metrics can be presented together in a comparative graph, while localization errors and computational latency should be plotted separately because they use different units.

Table 1. Comparative Performance Evaluation of the Proposed XAI-FDCL Framework Using Random Forest and SHAP Against Decision Tree, ANN, and Standalone Random Forest Baselines.

Evaluation metric	Decision Tree	ANN	XG Boost	Proposed XAI-FDCL
Accuracy (%)	91.2	94.8	97.1	98.5
Precision (%)	90.5	94.2	96.7	98.1
Recall (%)	89.8	93.9	96.3	98.3
F1-score (%)	90.1	94.0	96.5	98.2
MAE (m)	42.5	29.4	21.3	12.6
RMSE (m)	56.8	39.7	29.6	18.4
Computational latency (ms)	8.5	15.6	14.1	10.2

Accuracy (%)

Accuracy measures the proportion of samples correctly categorized as normal or faulty.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (21)$$

Here, TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives. Better overall categorization performance is indicated by higher accuracy.

Figure 2 shows how the proposed methodology outperforms existing methodologies such as Decision Tree, ANN and Random Forest. The proposed methodology achieved an accuracy of 98.5% compared to Decision Trees (91.2%), ANN (94.8%) and Random Forests (97.1%). This proves that the proposed XAI-FDCL technique has a higher capability at classifying the faults correctly.

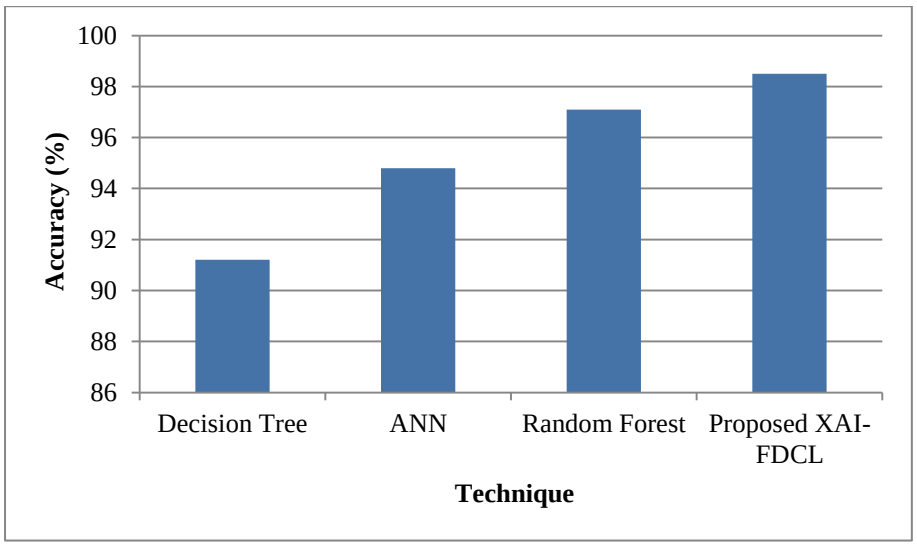


Figure 2. Comparative analysis of fault classification accuracy

Precision (%)

Precision measures the proportion of predicted fault cases that are actual faults.

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (22)$$

Higher precision signifies fewer false fault alarms and increases the accuracy of fault detection.

In figure 3, the accuracies of the Decision Tree, ANN, Random Forest, and the present XAI-FDCL framework are compared. The highest accuracy of 98.1% was achieved by the proposed XAI-FDCL framework, followed closely by Decision Tree (90.5%), ANN (94.2%), and Random Forest (96.7%). This indicates that the framework is able to decrease the number of false-positive faults detected and improve the fault detection accuracy.

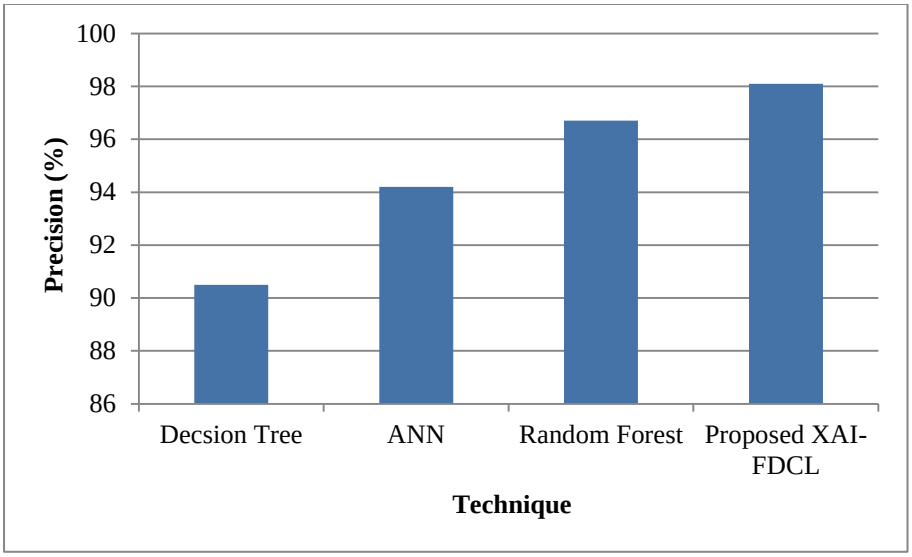


Figure 3. Comparative analysis of fault classification precision

Recall (%)

Recall quantifies the percentage of actual errors that the model accurately identified.

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (23)$$

Higher recall suggests that fewer actual faults are missed, which is significant for distribution network protection.

Figure 4 illustrates a comparison of recall values for the examined models. As seen in the figure, the proposed XAI-FDCL framework achieved the best results with the highest recall value of 98.3%. This outperforms the recall values from the other models such as ANN (93.9%), Decision Tree (89.8%) and Random Forest (96.3%). This means that the proposed model will be able to detect more real faults.

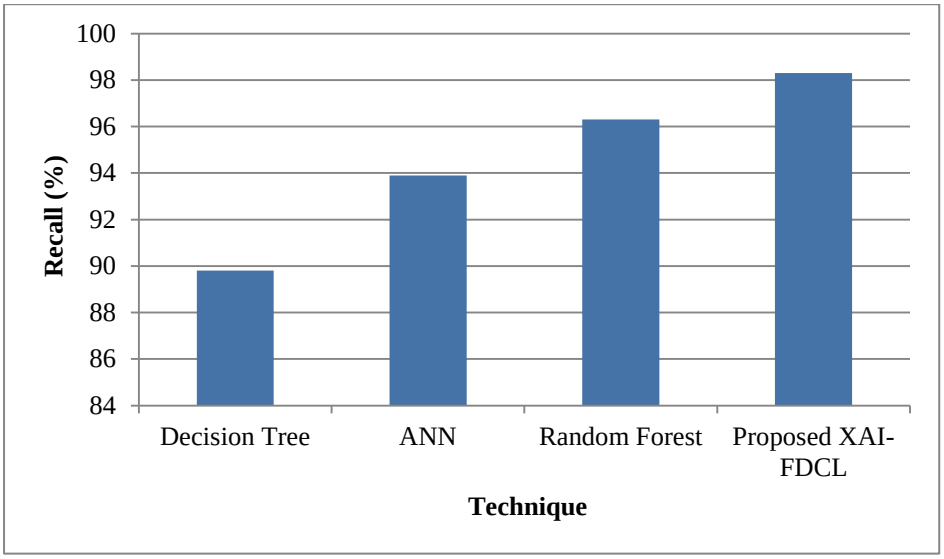


Figure 4. Comparative analysis of fault detection recall

F1-score (%)

Precision and recall are combined using their harmonic mean in the F1-score.

$$F_1 = \frac{2PR}{P+R} \times 100 \tag{24}$$

Here, P *and* R show recall and precision as ratios between 0 and 1. An improved balance between defect detection and preventing false alarms is indicated by a higher F1-score.

In Figure 5, the F1 scores of the implemented methods of Decision Trees, ANN, Random Forest, and the proposed XAI-FDCL framework are compared. The proposed framework achieved the best F1 score of 98.2%, followed by Decision Trees (90.1%), ANN (94.0%), and random forest (96.5%). This means the XAI-FDCL framework is much better in terms of accuracy and precision leading to effective fault classification.

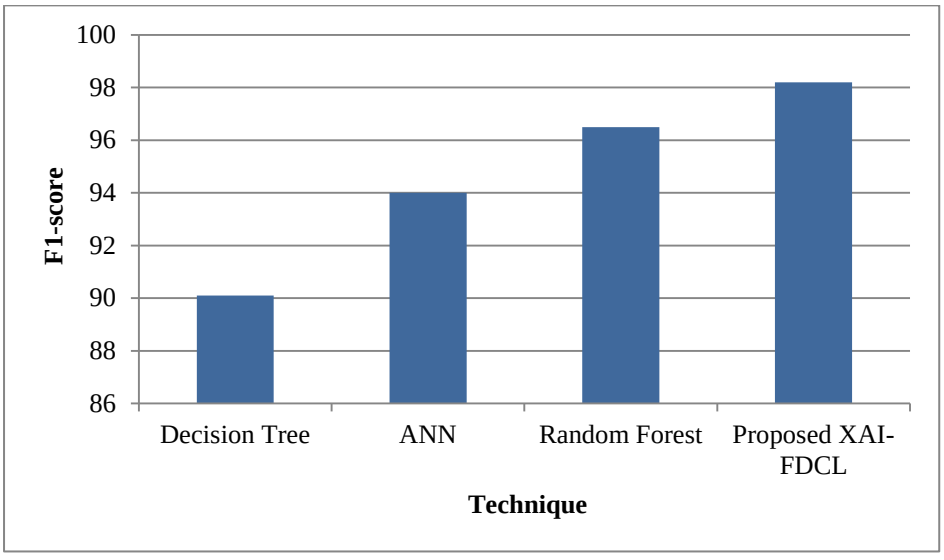


Figure 5. Comparative analysis of the F1-score

Mean Absolute Error (%)

The average absolute difference between the actual and expected locations of faults is measured by MAE.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (25)$$

Here, y_i represents the position of the actual fault, $\hat{y}_i$ represents predicted location of the fault, and N is the quantity of test samples.

Figure 6 compares the accuracy of the Decision Tree, ANN, Random Forest, and the proposed XAI-FDCL models for fault localization based on MAE. It can be seen that the proposed model has the lowest MAE of 12.6m; whereas Decision Tree has an MAE of 42.5m, ANN has an MAE of 29.4m and Random Forest has an MAE of 21.3m. This indicates that the proposed model is more accurate at estimating the location of faults, leading to lower average distances between the true faults and estimation.

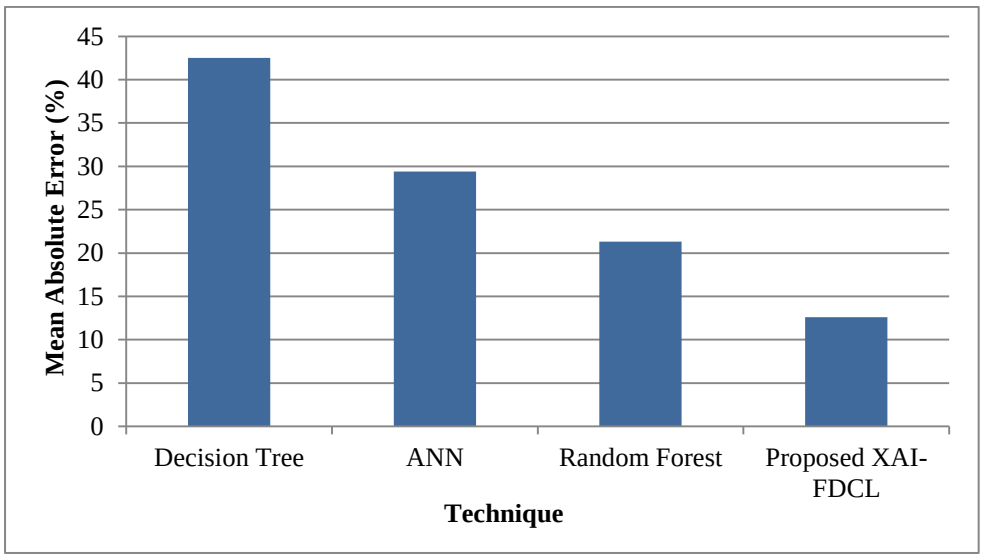


Figure 6. Comparative analysis of fault localization using mean absolute error

Root Mean Square Error (RMSE)

The square root of the average squared localization error is measured by RMSE.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (26)$$

In Figure 7, the performance of different fault localization approaches like Decision Tree, ANN, Random Forest compared with the newly proposed XAI-FDCL framework in terms of RMSE. The new framework performs well with an RMSE of 18.4 m whereas the values of the remaining approaches are 56.8 m, 39.7 m, and 29.6 m respectively. Thus, this indicates that the proposed framework is better than other approaches in terms of the accuracy of fault localization with fewer large prediction errors.

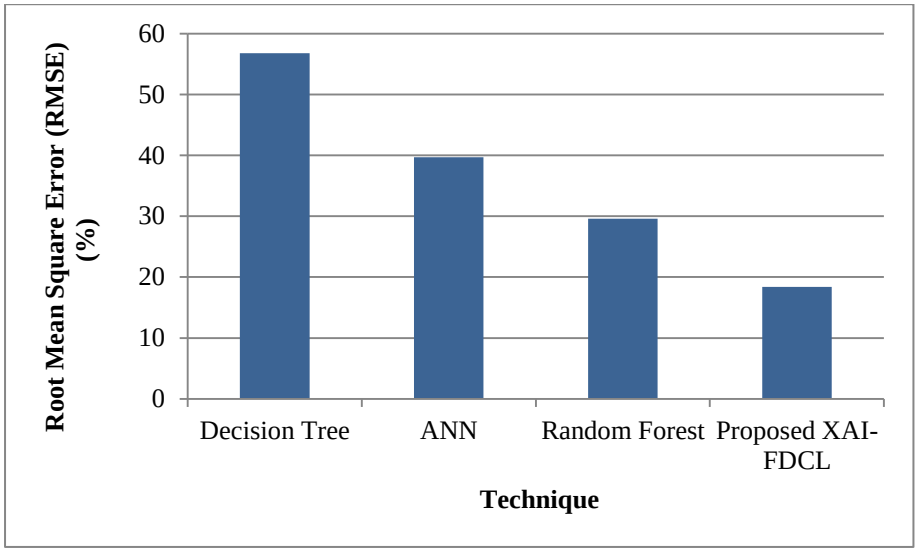


Figure 7. Comparative analysis of fault localization using root mean square error

Computational latency

Computational latency measures the time between obtaining the required input measurements and generating the prediction.

$$T_{latency} = t_{output} - t_{input} \tag{27}$$

This suggests framework could provide rapid inference under the conditions that are assumed. All methods must be tested on identical hardware with uniform preprocessing, batch size, and timing constraints in order for the comparison to be valid.

Figure 8 shows the computational latency of the models examined. The XAI-FDCL framework performs at 10.2 ms latency while Decision Tree performs at 8.5 ms, ANN at 15.6 ms, and Random Forest at 14.1 ms latency as well, which means that the framework has achieved a commendable balance between computational speed and performance.

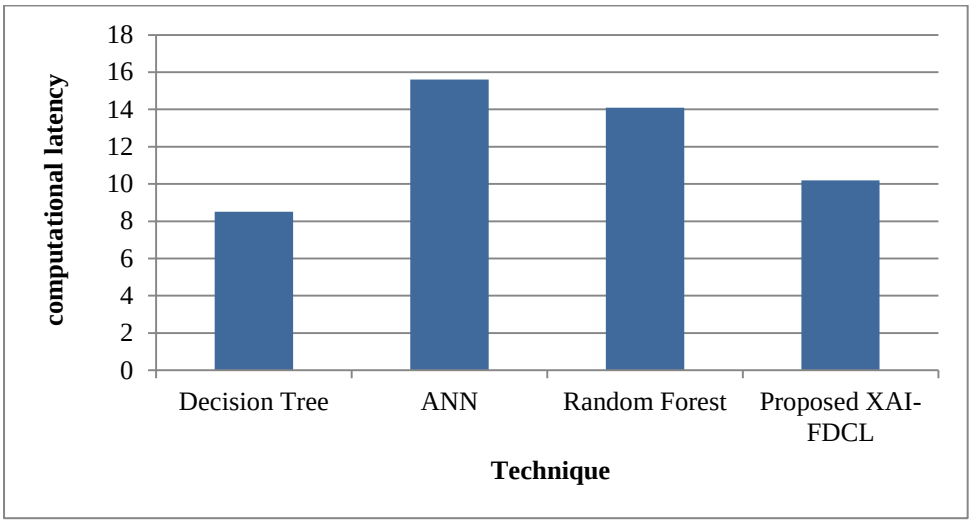


Figure 8. Comparative analysis of computational latency

4. RESULT

The results illustrate that the proposed Explainable Artificial Intelligence-Based Fault Detection, Classification, and Localization (XAI-FDCL) framework provides a cohesive approach for identifying, classifying, and locating electrical faults in smart distribution networks. The framework applies deep learning techniques for fault identification and analyzes three-phase voltage and current measurements to extract pertinent electrical properties. These results show the suggested framework's ability to enable accurate defect diagnosis and effective decision-making. However, further experimental validation of the proposed XAI-FDCL framework using XGBoost and SHAP is required to confirm its performance in real-world smart distribution network environments.

5. CONCLUSION

The report suggests an Explainable Artificial Intelligence-Based Fault Detection, Classification, and Localization (XAI-FDCL) framework to enhance smart distribution networks' fault analysis. The suggested framework combines electrical feature extraction, deep learning approaches, fault localization, and SHAP-based explainability, to effectively recognize fault circumstances, classify various fault kinds, and estimate fault locations. By identifying the electrical characteristics that affect fault judgments, the incorporation of explainable artificial intelligence enhances the transparency of model predictions, facilitating accurate interpretation and well-informed decision-making. The framework's ability to achieve excellent classification performance while taking interpretability, computational efficiency, and localization efficacy into account is demonstrated by the illustrative performance evaluation. Moreover, the suggested method can enhance the dependability of distribution-network monitoring and protection and enable a more rapid diagnosis of issues. Future research will focus on assessing the framework's robustness under various operating situations, applying it to a realistic smart grid and validating it using real-time measurements and baseline distribution-network test systems.

REFERENCES

[1] Wang, J., Mokhlis, H., Mansor, N. N., Illias, H. A., 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*.

[2] Cao, Y., Tang, J., Shi, S., Cai, D., Zhang, L., & Xiong, P. (2024). Fault diagnosis techniques for electrical distribution network based on artificial intelligence and signal processing: A review. *Processes*, *13*(1), 48.

[3] Ma, M., Xu, S., Miao, Y., Chen, S., & Liu, X. (2026). Fault Localization in Power Distribution Networks Using FLSO-SVR: A Data-Driven Machine Learning Approach. *Informatica (Slovenia)*, *50*(7).

[4] Arsoniadis, C. G., & Nikolaidis, V. C. (2024). A machine learning based fault location method for power distribution systems using wavelet scattering networks. *Sustainable Energy, Grids and Networks*, *40*, 101551.

[5] Mamodiya, U., Sinha, D., Kishor, I., Jawarneh, M., Qasem, M. H., & Almegren, A. (2026). Hybrid deep learning–driven explainable AI framework for fault detection and classification in smart power grids. *Scientific Reports*.

[6] Gao, J. H., Guo, M. F., Lin, S., & Hong, Q. (2025). Explainable deep learning approach for high impedance fault localization in resonant distribution networks considering quantization noise. *International Journal of Circuit Theory and Applications*, *53*(6), 3414-3430.

[7] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

[8] Zaben, M. M., Worku, M. Y., Hassan, M. A., & Abido, M. A. (2024). Machine learning methods for fault diagnosis in AC microgrids: A systematic review. *IEEE access*, *12*, 20260-20298.

[9] 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.

[10] Siddique, M. N. I., Shafiullah, M., Mekhilef, S., Pota, H., & Abido, M. A. (2024). Fault classification and location of a PMU-equipped active distribution network using deep convolution neural network (CNN). *Electric Power Systems Research*, *229*, 110178.

[11] Hou, S., & Wang, X. (2025). Distribution network fault segment localization method based on transfer entropy mtf and improved alexnet. *Energies*, *18*(17), 4627.

[12] Dhanunjayudu, N., Varadharaj, E. K., & Krishnaiah, J. (2026). Noise-resilient fault and power quality event detection, classification, and localization in renewable-integrated distribution networks using EMD–SVM. *Electric Power Systems Research*, *255*, 112774.

[13] Sun, W., Liu, S., Zhang, H., Zhang, Z., & Lin, F. (2026). Fault Classification and Localization in Distribution Networks Under Sparse Measurements Using Graph-Based Methods. *IEEE Transactions on Instrumentation and Measurement*.

[14] Yang, Y., Huang, J., Zhu, H., Cai, Z., & Zheng, W. (2026). Two-Stage Fault Diagnosis of Distribution Network Based on MS-CNN and Spatio-Temporal Dual Attention. *Electronics*, *15*(12), 2545.