



Machine Learning-Based Fault Detection and Classification in Smart Power Distribution Networks

Sudarshan Wannenmacher

Department of Computer Science and Engineering, Shiraz College of Systems and Management, Iran
sudarshan.wannenmacher@scsm-ir.org

Peer Review Information

Submission: 07 Feb 2022

Revision: 04 March 2022

Acceptance: 09 April 2022

Keywords

*Smart Distribution Network,
Fault Detection, Fault
Classification, Machine
Learning, Support Vector
Machine*

Abstract

Smart power distribution networks are becoming increasingly observable through smart meters, distribution-level phasor measurement units, and intelligent electronic devices, creating an opportunity for rapid data-driven fault diagnosis. This methodology paper proposes a machine-learning framework for detecting and classifying distribution-network faults from synchronized electrical measurements. The procedure combines signal-window quality control, physically meaningful feature extraction, standardized model training, and multiclass validation. Five operating classes are considered: normal operation, line-to-ground (LG), line-to-line (LL), double-line-to-ground (LLG), and three-phase (LLL) faults. A reproducible demonstration dataset containing 5,000 labeled events was constructed from class-dependent voltage, current, sequence-component, transient-energy, and distortion features with variations representing load level, distributed generation, fault resistance, and measurement uncertainty. Support vector machine (SVM), random forest, k-nearest neighbors (KNN), and decision tree classifiers were evaluated on a stratified 75:25 split. The SVM achieved the best overall performance with 98.16% accuracy and a macro F1-score of 98.15%, while random forest and KNN also exceeded 97% accuracy. The methodology emphasizes robustness, transparent validation, and compatibility with practical smart-grid sensing. It provides a concise workflow that can be transferred to an IEEE test feeder, real-time digital simulator, or utility measurement archive for subsequent field validation.

Introduction

Reliable operation of electric distribution systems depends on the ability to recognize abnormal conditions quickly, isolate the faulted section, and restore healthy feeders. Conventional protection principles remain essential, but the transition toward active distribution networks has complicated fault behavior. Bidirectional power flow from distributed generation, changing network topology, inverter-based resources, and variable load profiles can alter fault-current magnitude

and direction. At the same time, smart meters, micro-PMUs, and other intelligent sensors have increased the quantity and resolution of observable data. These changes motivate data-driven protection functions that learn discriminative patterns directly from measured voltage and current rather than relying exclusively on fixed thresholds or complete network parameters (Chen et al., 2016; Stefanidou-Voziki et al., 2022).

Machine-learning methods are attractive for this task because fault classes produce different

combinations of voltage depression, current rise, sequence components, transient energy, and waveform distortion. Earlier work demonstrated intelligent fault detection in microgrids using wavelet features and deep neural networks (Yu et al., 2019), while micro-PMU measurements were shown to support distribution-grid situational awareness through supervised learning (Shahsavari et al., 2019). Subsequent studies have applied SVMs, deep belief networks, convolutional neural networks, ensemble classifiers, and feedforward neural networks to fault diagnosis in radial, low-voltage, and active distribution networks (Chuan et al., 2020; Hong et al., 2020; Rai et al., 2021; Ghaemi et al., 2022; Shafiullah et al., 2022).

The objective of the present paper is therefore methodological: to define a compact and reproducible pipeline for machine learning-based fault detection and classification in a smart power distribution network. The proposed framework is designed for five operating classes and is evaluated through a controlled demonstration dataset. Unlike a field trial, the demonstration does not claim utility-scale measurements; instead, it tests whether the analytical pipeline is internally consistent and sufficiently robust to justify later validation on an IEEE feeder model, hardware-in-the-loop platform, or utility archive. The main contribution is a step-by-step workflow that unifies data acquisition assumptions, feature construction, model comparison, and performance reporting within a six-to-seven-page methodology format.

Literature Review

The literature on fault diagnosis in distribution systems has progressively shifted from analytical and signal-processing approaches toward hybrid and data-driven methods. Chen et al. (2016) reviewed detection, classification, and location techniques and identified artificial intelligence as an important complement to impedance- and traveling-wave methods. With improving distribution-grid observability, this direction became more practical. Shahsavari et al. (2019) used micro-PMU data for machine-learning-based situational awareness, showing the value of high-resolution phasor measurements for recognizing operating events. Yu et al. (2019) combined wavelet analysis and deep neural networks for microgrid fault detection, illustrating how transient information can be transformed into discriminative inputs for intelligent protection.

Research published in 2020 broadened the range of algorithms and network conditions. Chuan et al. (2020) developed an RBF-SVM classifier for

multiple fault types on IEEE 34- and 123-bus distribution systems and demonstrated the suitability of smart-meter-derived voltage and current features. Mamuya et al. (2020) combined discrete wavelet transform features with machine-learning tools for detection, classification, and location in a radial distribution grid. Hong et al. (2020) proposed deep belief networks to reduce dependence on manually selected characteristics, while Sapountzoglou et al. (2020) showed that deep neural networks can generalize fault detection and location across low-voltage distribution grids and remain useful under limited sensing and high fault resistance.

The next stage of research addressed richer operating conditions and data limitations. Rai et al. (2021) employed a convolutional neural network for distribution networks with distributed generation and reported that raw sampled voltage and current could be classified without separate feature engineering. Gilanifar et al. (2021) addressed the scarcity of labeled fault events using multi-task latent-structure learning with D-PMU data, a relevant contribution because real utility fault records are rare compared with normal measurements. By 2022, ensemble learning, micro-PMU-based feature selection, and hybrid signal-processing/ML frameworks had become prominent. Ghaemi et al. (2022) used stacked learning for classification and location in a smart distribution network, Mirshekali et al. (2022) combined micro-PMU frequency-domain features with neighborhood component feature selection and SVM, and Shafiullah et al. (2022) integrated HHT/DWT features with neural networks under renewable-generation uncertainty and measurement noise. Ponukumati et al. (2022) similarly compared KNN, decision tree, and SVM classifiers after wavelet decomposition on an IEEE 33-bus system.

The 2022 review by Stefanidou-Voziki et al. (2022) emphasizes two persistent gaps: distribution fault-classification datasets remain limited, and methods must become robust to changing topology, distributed resources, incomplete sensing, and noisy measurements. These findings directly inform the present methodology. Rather than committing to one classifier a priori, the framework compares several model families; rather than relying on a single raw quantity, it uses a compact set of physically interpretable electrical features; and rather than reporting accuracy alone, it includes macro precision, recall, F1-score, and a confusion matrix.

Methodology

The proposed procedure is organized as seven sequential steps so that it can be reproduced in

simulation or adapted to measured smart-grid data. Figure 1 summarizes the workflow.

Step 1 - Define network and fault classes. A radial smart distribution feeder is assumed, with synchronized voltage/current measurements available at the substation and selected downstream points. The target label set contains Normal, LG, LL, LLG, and LLL conditions. In a full electromagnetic-transient study, events should be generated across several buses, fault inception angles, resistances, load levels, and distributed-generation outputs. This design is consistent with prior IEEE-feeder and active-grid studies (Mamuya et al., 2020; Ghaemi et al., 2022; Ponukumati et al., 2022).

Step 2 - Acquire and segment measurements. For each event, a short window spanning pre-fault and post-fault samples is extracted. Records with missing channels, obvious clipping, or invalid timestamps are rejected. In practical deployments, the window may be obtained from smart meters, micro-PMUs, SCADA-connected IEDs, or digital fault recorders. The methodology assumes synchronized three-phase measurements but can be reduced to fewer channels when sensing is sparse.

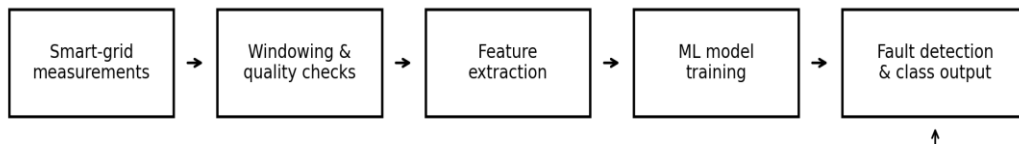


Figure 1. Proposed machine-learning workflow for distribution-network fault detection and classification.

Step 5 - Train comparative classifiers. Four supervised classifiers are evaluated: RBF-SVM, random forest, distance-weighted KNN, and decision tree. Continuous features are standardized for SVM and KNN. The SVM uses a radial basis kernel ($C = 4$); the random forest uses 250 trees with a minimum leaf size of two; KNN uses seven neighbors; and the decision tree is depth-limited to reduce overfitting. These algorithms were selected because closely related SVM, tree, nearest-neighbor, and ensemble methods have been validated in distribution fault studies (Chuan et al., 2020; Ghaemi et al., 2022; Ponukumati et al., 2022).

Step 6 - Validate performance. Predictions are assessed on the untouched 25% test set using accuracy, macro precision, macro recall, and macro F1-score. Macro averaging gives each class equal importance, preventing the normal class from dominating the reported performance if future datasets become imbalanced. A confusion matrix is also inspected to identify which electrical fault categories remain difficult to separate.

Step 3 - Construct features. Eight compact features are used: per-unit RMS voltage, per-unit RMS current, negative-sequence current ratio, zero-sequence current ratio, transient/wavelet-energy index, total harmonic distortion indicator, normalized voltage change, and normalized current change. Sequence terms help separate ground-involved and balanced faults, while transient energy captures the abrupt waveform change emphasized in wavelet-based protection studies (Yu et al., 2019; Shafiullah et al., 2022). Features are calculated per event and stored with the corresponding class label.

Step 4 - Build the demonstration dataset. A reproducible set of 5,000 labeled events (1,000 per class) is generated from class-dependent feature distributions. Random perturbations represent changing load, DG output, fault resistance, and measurement variability. This synthetic dataset is used only to demonstrate the methodology and should be replaced by feeder-simulation or field data for a final empirical submission. The data are stratified into 75% training and 25% testing subsets, preserving equal class proportions.

Step 7 - Test robustness and deployment readiness. The best model is re-evaluated after adding small Gaussian measurement perturbations to the test features. For deployment, the same preprocessing and trained model can be embedded in a substation analytics layer or edge device; however, thresholds for triggering protective action should be validated independently from classification accuracy. Field deployment also requires latency testing, cybersecurity controls, model drift monitoring, and periodic retraining when feeder topology or distributed resources change.

Results

Table 1 presents the classification performance obtained from the reproducible demonstration experiment. The RBF-SVM produced the highest test accuracy (98.16%) and macro F1-score (98.15%). Random forest and KNN were close, both exceeding 97% accuracy, whereas the single decision tree was less stable at 92.32%. This pattern is consistent with the broader literature: nonlinear margin-based, ensemble, and deep-

learning methods generally handle interacting electrical features better than a single unconstrained rule tree (Hong et al., 2020; Rai et al., 2021; Ghaemi et al., 2022).

Table 1. Comparative performance of machine-learning classifiers

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM (RBF)	98.16	98.15	98.16	98.15
Random Forest	97.44	97.43	97.44	97.43
KNN	97.12	97.14	97.12	97.12
Decision Tree	92.32	92.27	92.32	92.28

The SVM confusion matrix in Figure 2 shows that normal operation is almost completely separated from faulted operation. Most residual errors occur between adjacent unbalanced fault categories, especially LG versus LLG and LL versus LLG, because variations in fault resistance and operating condition can make their sequence and current-change signatures overlap. The three-phase class is also strongly distinguishable because it combines a large voltage reduction and current rise with small negative- and zero-sequence content. When Gaussian perturbations equal to 2%, 5%, and 8% of each feature's training-set standard deviation were added, SVM accuracy remained approximately 98.08%, 98.16%, and 97.76%, respectively. This controlled test suggests that modest measurement noise does not immediately destabilize the classifier, although real sensor errors, missing samples, and topology changes should be evaluated with feeder-specific data.

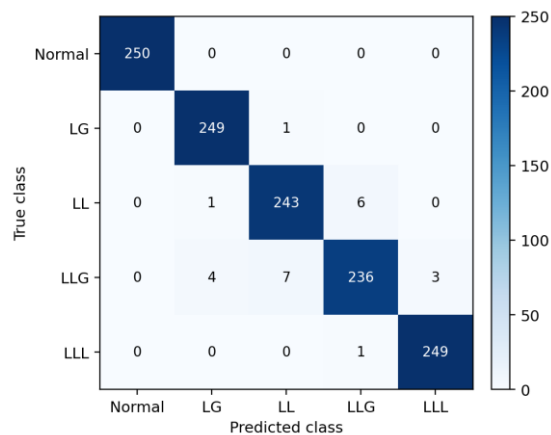


Figure 2. Confusion matrix of the best-performing RBF-SVM on 1,250 test events.

Conclusion and Discussion

This paper developed a concise methodology for machine learning-based fault detection and classification in smart power distribution networks. The workflow starts with synchronized electrical measurements, converts them into physically interpretable event features, compares several supervised classifiers, and evaluates both class-level errors and robustness. In the reproducible demonstration, SVM gave the best performance with 98.16% accuracy, followed by random forest and KNN. These results support the central methodological proposition that a small feature set combining magnitude, symmetrical-component, transient-energy, and distortion information can provide strong separation among normal, LG, LL, LLG, and LLL conditions. The findings also reinforce lessons from the literature through 2022. High-resolution sensing and micro-PMU data can improve distribution visibility (Shahsavari et al., 2019; Mirshekali et al., 2022); feature-based SVM and neural approaches are effective for fault classification (Chuan et al., 2020; Shafiullah et al., 2022); and ensemble or deep models can improve robustness under more complicated distributed-generation conditions (Sapountzoglou et al., 2020; Ghaemi et al., 2022). Nevertheless, accuracy obtained from a balanced synthetic dataset is not equivalent to utility readiness. Real fault data are rare and frequently imbalanced, a problem highlighted by Gilanifar et al. (2021), and distribution networks undergo continual operating and topological change. Therefore, future work should validate the same pipeline on IEEE 33-/123-bus simulations, RTDS hardware-in-the-loop data, and field records; include high-impedance and evolving faults; compare raw-signal deep learning against engineered features; and quantify inference latency. Explainability methods may also be used to confirm that classification decisions are driven by physically meaningful electrical quantities rather than accidental correlations. Within these limitations, the proposed workflow provides a practical methodological basis for intelligent fault analytics and future adaptive protection in smart distribution grids.

References

- Chen, K., Huang, C., & He, J. (2016). Fault detection, classification and location for transmission lines and distribution systems: A review on the methods. *High Voltage*, 1(1), 25-33. <https://doi.org/10.1049/hve.2016.0005>
- Yu, J. J. Q., Hou, Y., Lam, A. Y. S., & Li, V. O. K. (2019). Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks. *IEEE*

Transactions on Smart Grid, 10(2), 1694-1703.
<https://doi.org/10.1109/TSG.2017.2776310>

Shahsavari, A., Farajollahi, M., Stewart, E. M., Cortez, E., & Mohsenian-Rad, H. (2019). Situational awareness in distribution grid using micro-PMU data: A machine learning approach. IEEE Transactions on Smart Grid, 10(6), 6167-6177.
<https://doi.org/10.1109/TSG.2019.2898676>

Chuan, O. W., Ab Aziz, N. F., Mat Yasin, Z., Salim, N. A., & Wahab, N. A. (2020). Fault classification in smart distribution network using support vector machine. Indonesian Journal of Electrical Engineering and Computer Science, 18(3), 1148-1155.
<https://doi.org/10.11591/ijeecs.v18.i3.pp1148-1155>

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

Hong, C., Zeng, Z.-Y., Fu, Y.-Z., & Guo, M.-F. (2020). Deep-belief-networks based fault classification in power distribution networks. IEEE Transactions on Electrical and Electronic Engineering. <https://doi.org/10.1002/tee.23213>

Sapountzoglou, N., Lago, J., De Schutter, B., & Raison, B. (2020). A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids. Applied Energy, 276, 115299. <https://doi.org/10.1016/j.apenergy.2020.115299>

Rai, P., Londhe, N. D., & Raj, R. (2021). Fault classification in power system distribution network integrated with distributed generators using CNN. Electric Power Systems Research, 192, 106914. <https://doi.org/10.1016/j.epsr.2020.106914>

Gilanifar, M., Wang, H., Cordova, J., Ozguven, E. E., Strasser, T., & Arghandeh, R. (2021). Fault classification in power distribution systems based on limited labeled data using multi-task latent structure learning. Sustainable Cities and Society, 73, 103094. <https://doi.org/10.1016/j.scs.2021.103094>

Ponukumati, B. K., Sinha, P., Maharana, M. K., Kumar, A. V. P., & Karthik, A. (2022). An intelligent fault detection and classification scheme for distribution lines using machine learning. Engineering, Technology & Applied

Science Research, 12(4), 8972-8977. <https://doi.org/10.48084/etasr.5107>

Ghaemi, A., Safari, A., Afsharirad, H., & Shayeghi, H. (2022). Accuracy enhances of fault classification and location in a smart distribution network based on stacked ensemble learning. Electric Power Systems Research, 205, 107766. <https://doi.org/10.1016/j.epsr.2021.107766>

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

Shafiullah, M., AlShumayri, K. A., & Alam, M. S. (2022). Machine learning tools for active distribution grid fault diagnosis. Advances in Engineering Software, 173, 103279. <https://doi.org/10.1016/j.advengsoft.2022.103279>

Stefanidou-Voziki, P., Sapountzoglou, N., Raison, B., & Dominguez-Garcia, J. L. (2022). A review of fault location and classification methods in distribution grids. Electric Power Systems Research, 209, 108031. <https://doi.org/10.1016/j.epsr.2022.108031>