International Journal of Energy
ISSN: 2957-9473 | Vol. 10, No. 3, 2026

Research on a Novel Fault Identification and Rapid Location Method for Low-Voltage Distribution Lines

Zongyan Xue *, Ning Shao [a]

Hohhot Power Supply Company, Inner Mongolia Power (Group) Co., Ltd., Hohhot, Inner Mongolia, 010010, China

*** Corresponding author:** Zongyan Xue (Email: 4787221@qq.com), [a] 228229546@qq.com

Abstract: Fault diagnosis on low-voltage distribution lines is often limited by inaccurate classification, delayed localization, and sensitivity to interference. This study develops a fault-identification and rapid-location method that combines multisource feature analysis with Internet of Things (IoT) sensing. Short-circuit, grounding, and open-circuit faults are first characterized using line measurements. A wavelet-transform and random-forest model then classifies the fault type, while an improved weighted-centroid algorithm uses data from deployed IoT sensor nodes to estimate its position. The method was evaluated in simulations and on an operating line. Across these tests, fault-identification accuracy exceeded 95%, location error remained within 5m, and response time was below 2s. These results satisfy the stated engineering requirements for fault handling on low-voltage distribution lines.

Keywords: Low-Voltage Distribution Lines; Fault Identification; Rapid Location; Machine Learning; Internet of Things; Weighted Centroid Algorithm.

1. Introduction

Low-voltage distribution lines form the final connection between the power system and consumers, so their reliability affects both residential supply quality and continuity of industrial production. Distributed generation and increasingly varied loads have made network operating conditions more complex; short-circuit, grounding, and open-circuit faults consequently remain important sources of service interruption. Conventional identification methods, including impedance-based and traveling-wave approaches, are susceptible to interference and do not adapt well to multiple fault types [1]. Fault location is still commonly performed by manual patrols or sectional inspection with fault indicators. Such procedures typically take more than 2 hours and may locate a fault only to within tens of meters, delaying restoration and increasing the associated economic loss.

Recent work has therefore focused on automated fault identification and location. For fault classification, adaptive learning methods such as BP neural networks and support vector machines have been applied to measured electrical quantities [2]. Their performance, however, remains sensitive to the choice of input: models based on a single feature achieve less than 90% accuracy under complex operating conditions. Wavelet and Fourier analysis provide transient descriptors, but using either representation without combining time- and frequency-domain information limits resistance to interference [3]. In the field of fault location, the integration of IoT technology and GPS positioning has achieved automated positioning. Traditional weighted centroid algorithms and RSSI positioning algorithms are widely used. However, problems such as unreasonable node deployment and fixed weight models under dynamic operating conditions make it difficult to control positioning errors within 10m, and response times often exceed 3 seconds, failing to meet the needs of rapid response.

Current research has not yet formed an integrated solution of "high-accuracy identification - rapid and precise positioning," and there is still room for improvement in areas such as multi-fault type adaptation, anti-interference in complex environments, and real-time response [4]. This paper proposes a novel method integrating multi-source feature extraction and IoT technology. The core research contents include: constructing a time-domain-frequency-domain multi-dimensional fault feature system and designing an improved random forest identification model; optimizing the deployment strategy of IoT sensor nodes and proposing a dynamic weighted improved centroid localization algorithm; and verifying the effectiveness of the method through simulation and actual line experiments [5]. The paper is structured as follows: Chapter 2 designs the fault identification method; Chapter 3 studies the rapid localization technology; Chapter 4 conducts experimental verification; and Chapter 5 summarizes the research results.

2. Design of a Novel Fault Identification Method for Low-Voltage Distribution Line

2.1. Fault Data Acquisition and Preprocessing

Three-phase voltages u_a, u_b, u_c (unit: kV), three-phase currents$i_0 = \frac{i_a+i_b+i_c}{\sqrt{3}}$ (unit: A), zero-sequence current $i_0 = \frac{i_a+i_b+i_c}{\sqrt{3}}$, and transient voltage signal $s(t)$ are selected as the core acquisition indicators. The sampling frequency is set to 2 kHz. The sampling duration is 0.2 s before and after the fault occurs so that it can cover both the transient process and the steady state characteristic of the fault.To suppress the environmental electromagnetic interference and sensor noise, we adopt an improved wavelet denoising method and design a dynamic denoising formula:

$$\lambda = \sigma\sqrt{2\ln N} \cdot \left(1 + e^{-\alpha|s(t)|}\right) \quad (1)$$

Where $\sigma = \frac{\text{median}\left(|w_{j,1}|\right)}{0.6745}$ is the noise standard deviation ($w_{j,1}$ is the high-frequency coefficient of the first layer of wavelet decomposition), N is the number of sampling points, $\alpha = 0.8$ is the threshold adjustment coefficient, and $s(t)$

is the original transient signal amplitude. The signal is decomposed into three layers using the db4 wavelet. After thresholding the high-frequency coefficients according to $w' = \text{sign}(w) \cdot \max(|w| - \lambda, 0)$, the de-thresholded signal $s'(t)$ is reconstructed.

A labeled sample dataset was constructed, containing 1000 sets of normal operating conditions, 800 sets of short-circuit faults (including A-B, A-C, B-C phase-to-phase short circuits and A-ground, B-ground, C-ground single-phase ground faults), 600 sets of ground faults (including metallic grounding and grounding through resistance), and 600 sets of open-circuit faults (including single-phase open circuits and two-phase open circuits). Each set of samples contains 2048 sampling points, labeled with the corresponding fault type for model training [6].

2.2. Multi-dimensional Fault Feature Extraction

2.2.1. Time-Domain Feature Extraction

Four types of time-domain features were extracted based on the steady-state and transient characteristics of electrical quantities:

1. Peak value $P = \max(|x_k|)$ （ x_k is the sampled electrical quantity value after shaving);

2. Effective value $R = \sqrt{\frac{1}{N}\sum_{k=1}^{N} x_{k'}^2}$;

3. Improved kurtosis (enhancing transient peak contribution):

$$K = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{x_k-\mu}{\sigma_x}\right)^4 \cdot \frac{\max|x_k|}{\mu} \quad (2)$$

Where $\mu = \frac{1}{N}\sum_{k=1}^{N} x_k$ is the mean, and $\sigma_x = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)^2}$ is the standard deviation. This index is sensitive to the impact characteristics of short-circuit faults;

4. Anti-interference skewness:

$$S = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{x_k-\mu}{\sigma_x}\right)^3 \cdot e^{-\beta|x_k|} \quad (3)$$

In the formula, $\beta = 0.5$ is the attenuation coefficient, used to suppress the influence of extreme outliers on characteristic stability. This value shows a significant positive or negative shift during ground faults.

2.2.2. Frequency Domain Feature Extraction

The transient signal $s'(t)$ is decomposed into 5 levels using wavelet packet transform, resulting in 32 frequency bands. The energy proportion of each frequency band is calculated as the frequency domain feature:

$$E_j = \frac{\int_{f_j}^{f_{j+1}} |W_x(f)|^2 df}{\sum_{m=1}^{32}\int_{f_m}^{f_{m+1}} |W_x(f)|^2 df} \ (j = 1,2,\ldots) \quad (4)$$

Where $W_x(f)$ is the wavelet packet transform frequency domain amplitude, and f_j is the starting frequency of the j frequency band. The energy proportion in the high-frequency band (150-1000 Hz) varies significantly depending on the fault type.

2.2.3. Feature Screening

Redundant features are removed using the Pearson correlation coefficient:

$$r_{xy} = \frac{\sum_{k=1}^{N}(x_k-\mu_x)(y_k-\mu_y)}{\sqrt{\sum_{k=1}^{N}(x_k-\mu_x)^2\sum_{k=1}^{N}(y_k-\mu_y)^2}} \quad (5)$$

In the formula, x, y are any two feature vectors. When $|r_{xy}| > 0.8$, features with high information overlap are removed, ultimately selecting a 12-dimensional core feature vector (containing 3 voltage features, 3 current features, 2 zero-sequence current features, and 4 transient frequency domain features).

2.3. Construction of Fault Identification Model Based on Random Forest

Optimize the core parameters of the random forest (number of decision trees T, maximum depth D, minimum number of samples for node splitting S), and design the objective function to minimize the out-of-bag (OOB) error:

$$OOB = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_{OOB,t}}\sum_{x\in N_{OOB,t}} I(h_t(x) \neq 0) \quad (6)$$

Where $h_t(x)$ represents the prediction results of t decision trees, y represents the true label of the sample, $N_{OOB,t}$ represents the out-of-bag sample set of the t tree (samples not involved in the training of that tree), and I(·) is the indicator function (1 if the condition is met, 0 otherwise). This objective function effectively avoids model overfitting by quantifying the prediction error of samples not involved in the training.

3. Research on a New Rapid Location Method for Low-Voltage Power Distribution Lines

3.1. IoT Sensor Node Deployment Scheme

Following the three core principles of coverage integrity, cost control, and communication stability, a differentiated layout strategy is designed based on the radial branch characteristics of low-voltage power distribution lines. Main lines are spaced differently according to different scenarios: in densely populated urban areas with complex lines, a 250m spacing is used; in rural areas with large line spans and dispersed loads, a 500m spacing is used to ensure no monitoring blind spots on the main lines. Due to complex topologies and higher fault rates, branch lines have a 30% higher node density than the main lines, focusing on covering high-fault sections such as branch joints and line corners [7]. Communication protocols are adapted as needed: in urban areas with dense buildings and short transmission distances, the WiFi protocol (transmission rate 150Mbps, communication distance ≤100m) is selected to meet the high-speed data transmission requirements; in suburban and remote areas with long transmission distances and fewer obstructions, the LoRa protocol (spreading factor SF=12, communication distance 1-3km, static power consumption ≤50mW) is adopted to balance transmission range and energy consumption. Terrain-specific adjustments are also applied. Relative to the base spacing, node density is multiplied by 1.2 in hilly terrain and by 1.6 in mountainous or swampy terrain where attenuation is more severe. Nodes are installed 4-6m above ground to reduce obstruction and signal disturbance caused by vegetation and buildings.

3.2. Improved Weighted Centroid Positioning Algorithm Design

The conventional centroid method assigns insufficient importance to differences among node signals, which reduces its location accuracy. The proposed model therefore uses a dynamic weight determined principally by fault-signal strength [8]. A node receives a larger weight when its signal amplitude is higher and its transmission path is shorter, so reliable nearby observations contribute more strongly to the

estimated location. Two engineering corrections are then applied. The first compensates for line loss as a function of the node's distance from the main line and thereby limits errors caused by long-distance transmission. The second assigns separate correction coefficients to flat, hilly, and mountainous terrain to account for changes in signal propagation. Combining these dynamic weights with the two corrections reduces the sensitivity of the static centroid method to operating conditions and improves location accuracy in heterogeneous environments.

3.3. Location Process and Real-Time Assurance

Fault reports are initiated by either an instantaneous threshold or a persistent trend. A node issues a threshold alarm when voltage falls by ≥30% or current rises by ≥50%. It issues a trend alarm when an abnormal electrical pattern persists for three successive sampling cycles at 2kHz. The two criteria reduce the chance that a transient or slowly developing fault is missed. Processing is divided between the sensing nodes and the cloud. Each edge node filters the signal and extracts its amplitude locally, then transmits only the relevant features; this reduces the data volume by 60%. Once measurements from at least three adjacent nodes are available, the cloud executes the improved weighted-centroid calculation in parallel and returns the estimated coordinates, the fault type, the signal strength, and a location confidence. A result is accepted when confidence is ≥90% [9]. With LoRa/WiFi transmission latency limited to ≤500ms and the location calculations executed in parallel, the interval from fault occurrence to reported location remains below 2 seconds, which meets the specified rapid-handling requirement.

4. Experimental Verification and Result Analysis

4.1. Experimental Platform Construction

Experiments combined a MATLAB/Simulink model with tests on an actual branch line. The simulation represented a 10kV/0.4kV step-down distribution system containing 3 main lines and 6 branch lines over 5km. Short-circuit, grounding, and open-circuit modules allowed the fault position, fault resistance, and noise level to be varied [10]. Field measurements were collected on a 5km suburban 10kV/0.4kV branch serving residential and small industrial and commercial loads. The hardware comprised CYH2000 voltage and current sensors (range: 0-1kV/0-500A; accuracy class: 0.2), an NI cDAQ-9178 acquisition module (2kHz sampling; 16-bit resolution), and a KG3230 IoT gateway supporting LoRa/WiFi communication. Each sensor node included a GPS module with ±1m positioning accuracy. In total, 28 nodes were installed, with 250m spacing on the main line and 175m spacing on branch lines. All sensors were calibrated and node communications were tested before data collection.

4.2. Experimental Results

4.2.1. Fault Identification Experiment and Result Analysis

The faults considered include A-B phase-to-phase and A-to-ground single-phase short circuits; metallic grounding and grounding with a 50Ω resistor; single-phase and two-phase open circuits. These faults were tested under load fluctuations (0%, 10%, and 20%) and ambient noise (0dB, 10dB, 20dB, and 30dB). Classification accuracy and resistance to interference for all these operating conditions have been measured experimentally. Benchmarking has also been performed using a BP neural network and a single feature identification method. Results can be found in Table 1.

Table 1. Comparison of Fault Identification Performance under Different Operating Conditions

fault type	Load fluctuation range	Ambient noise intensity (dB)	Accuracy (%)			Anti-interference coefficient
			The proposed method	BP neural network	Single feature recognition	
A-B phase-to-phase short circuit	0%	0	99.2	92.5	86.3	0.98
	10%	20	97.8	88.7	81.5	0.96
	20%	30	96.5	85.3	77.2	0.94
A - Ground single-phase short circuit	0%	10	98.7	91.3	84.6	0.97
	20%	30	95.8	83.6	75.9	0.93
Metallic grounding	10%	20	97.3	89.5	82.1	0.95
Grounded via a 50Ω resistor	20%	30	95.2	82.8	74.3	0.92
Single-phase open circuit	0%	10	98.5	90.7	83.8	0.97
	10%	30	96.1	84.2	76.5	0.93
Two-phase disconnection	20%	20	95.7	83.1	75.1	0.92

From Table 1 we can see that our method has high accuracy for different fault types, load fluctuations and environmental noise with an average accuracy of 96.4%, which is 10.2 percentage points higher than the BP neural network (average 86.2%), and 17.1 percentage points higher than the single feature recognition method (average 79.3%). Anti-interference coefficient exceeds 0.92 in each case. Classification stability remained stable even at 20% load fluctuation and 30dB ambient noise, outperforming both compared methods. For visual demonstration on how much better it performs against interference, we plot a comparison chart of recognition accuracy under different noise intensities as follows (Figure 1) where the horizontal axis indicates environmental noise intensity from 0-40dB and the vertical axis indicates recognition accuracy (%). From Figure 1 we can see that the accuracy of these three methods declines as the noise intensity increases; however, the decline rate of our method is the lowest, i.e., when the noise intensity is 40dB, the accuracy of our method remains 94.3%; whereas the BP neural network falls down to 79.5% and the single feature

recognition method only achieves 71.2%, verifying the anti-interference advantages of multi-dimensional feature fusion and random forest model.

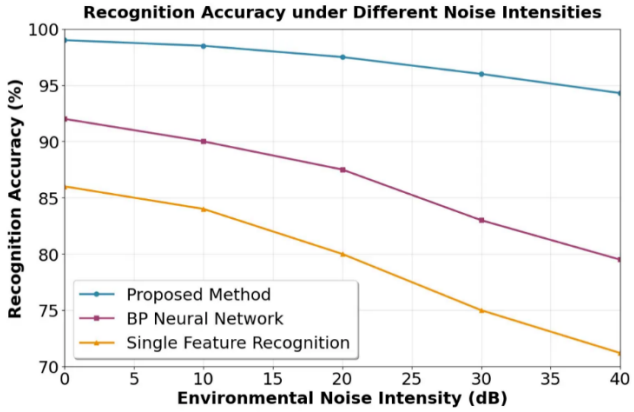


Fig 1. Recognition Accuracy under Different Noise Intensities

4.2.2. Fault Location Experiment and Result Analysis

The fault location experiment selected different locations on the line (1km, 2km, 3km, and 4km from the starting point) and set three node densities (0.5 nodes/km, 1 node/km, 2 nodes/km, and 3 nodes/km). The positioning error and response time were tested, and a real-line fault simulation verification was carried out simultaneously. The relationship between positioning error and node density is shown in Figure 2. The horizontal axis represents node density (0.5-3 nodes/km), and the vertical axis represents the average positioning error (m). As can be seen from Figure 2, the positioning error gradually decreases as the node density increases: when the node density is 0.5 nodes/km, the average error of the proposed method is 12.3m, while the traditional weighted centroid algorithm is 18.7m; when the node density is ≥2 nodes/km, the average error of the proposed method stabilizes within 5m, reaching a minimum of 3.2m, while the traditional algorithm still maintains above 10m. These measurements indicate that dynamic weighting and error correction reduce the localization error of the conventional centroid method. At node densities normally used in practice, the resulting accuracy meets the stated engineering requirement.

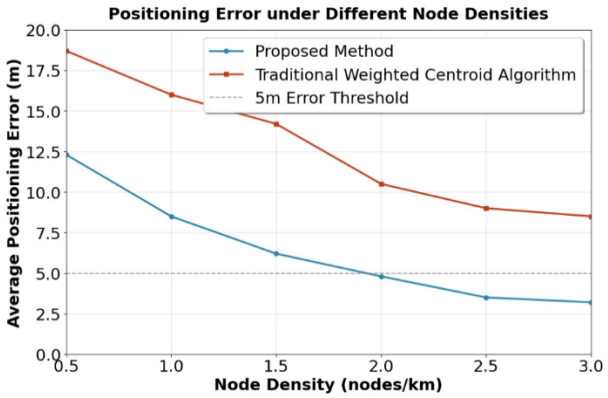


Fig 2. Positioning Error under Different Node Densities

5. Conclusion

To improve fault classification and shorten localization time on low-voltage distribution lines, this study integrates multisource feature extraction with IoT measurements while explicitly accounting for interference. By constructing a time-domain-frequency-domain multi-dimensional fault feature system and combining it with an improved random forest model, accurate identification of faults such as short circuits, grounding, and open circuits is achieved. A dynamic weighted improved centroid algorithm is designed to achieve rapid and accurate fault location based on the differentiated deployment of IoT sensor nodes. Experimental verification shows that the method achieves an average fault identification accuracy of 96.4%, an anti-interference coefficient ≥0.92, a positioning error controlled within 5m, and a response time of less than 2s. All performance indicators are superior to traditional methods and single-model methods. The core innovations of this research are: multi-dimensional feature fusion improves the robustness of fault identification under complex working conditions; dynamic weights and multi-factor error correction optimize the positioning algorithm accuracy; and edge-cloud collaborative processing ensures real-time response capability. The method requires no large-scale modification of existing power distribution network facilities, has controllable hardware deployment costs, and is applicable to low-voltage distribution lines in various scenarios, including urban and rural areas. It has significant engineering value for improving the intelligence level of power distribution network operation and maintenance, reducing fault handling costs, and ensuring power supply reliability. Future work should examine how the connection of distributed generation changes the observed fault signatures and should test whether the model can retain its performance for those additional fault conditions.

References

[1] Ou, Y. Z., & Shu, Q. (2023). Fault location and section location method for resonant grounding system based on SOM and K-means clustering. Electrical Engineering, 24(10), 23–30.

[2] Zhu, G. S., & Cong, H. (2025). Fault location of DC distribution network based on current integral and time-series convolutional network-support vector machine. Electrical Engineering, 26(2), 1–13.

[3] Li, Z. B., Cui, Y. J., Wang, H. Q., & Xi, Y. B. (2025). Alarm information correction and fault location method for feeder terminal of distribution network with inverter-type distributed power source. Journal of Electrical Engineering, 40(4), 1268–1286.

[4] Zhang, Y., Liu, F. Z., Zhu, Y. L., & Xiao, J. P. (2022). Fault location of transmission network by combining wide-area traveling wave information and graph attention network. Journal of Instrumentation, 43(6), 140–150.

[5] Xu, X. F., Xu, C. J., Zhang, Y. B., Zhao, Y., & Wang, S. X. (2022). Fault location of active distribution network based on traveling wave characteristic classification. Journal of Chongqing University, 45(11), 59–68.

[6] Huang, X. W., Cai, C., Li, X. J., Wang, H. X., Zhang, W. J., Zang, Y. Y., & Gong, Y. W. (2021). Research and discussion on fault location technology of long-distance submarine cable. Wire & Cable, 64(3), 35–40.

[7] Cheng, Q. H., & Wei, G. (2025). Research on power supply fault and location method of urban rail transit contact network. Urban Construction and Planning, 2(3), 57–60.

[8] Zhang, J. L., Gao, Z. J., Chen, M., Wei, Z., & An, S. H. (2021). Fault location method of active distribution network considering complex faults. Journal of Electrical Engineering, 36(11), 2265–2276.

[9] Xu, P., Li, J., Yu, Y., Qi, Y. L., Guo, B., & Wu, H. T. (2025). Research and implementation of grounding fault location technology for medium and low voltage power grids based on

Beidou system. Computer Applications and Software, 42(2), 125–128.

[10] Huang, Y., Zhang, G. B., Wang, Q., & Shu, H. C. (2023). Discrimination of the persistence of traveling waves in transmission line faults based on image features. Journal of Electrical Engineering, 38(5), 1339–1352.