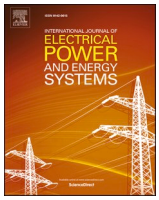
Electrical Power and Energy Systems 156 (2024) 109705

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



An incremental high impedance fault detection method under non-stationary environments in distribution networks

Mou-Fa Guo [a,d], Meitao Yao [a,d], Jian-Hong Gao [a,b,c,d,*], Wen-Li Liu [a,d], Shuyue Lin [b]

[a] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
[b] School of Engineering, University of Hull, Hull HU67RX, UK
[c] Department of Electrical Engineering, Yuan Ze University, Taoyuan 32003, Taiwan
[d] Engineering Research Center of Smart Distribution Grid Equipment, Fujian Province University, Fuzhou 350108, China

ARTICLE INFO

Keywords:
High impedance fault
Incremental learning
Data replay
Distribution network

ABSTRACT

In the non-stationary environments of distribution networks, where operating conditions continually evolve, maintaining reliable high impedance faults (HIF) detection is a significant challenge due to the frequent changes in data distribution caused by environmental variations. In this paper, we propose a novel HIF detection method based on incremental learning to handle non-stationary data stream with changing distributions. The proposed method utilizes stationary wavelet transform (SWT) to extract fault characteristics in different frequency domains from zero-sequence current data. Subsequently, a complex mapping from signal features to operational conditions is established using backpropagation neural network (BPNN) to achieve online detection of HIF. Additionally, signal features are analyzed using density-based spatial clustering of applications with noise (DBSCAN) to monitor the distribution of data. After encountering multiple distribution changes, an incremental learning process based on data replay is initiated to evolve the BPNN model for adapting to the changing data distribution. It is worth noting that the data replay mechanism ensures that the model retains previously acquired knowledge while learning from newly encountered data distributions. The proposed method was implemented in a prototype of a designed edge intelligent terminal and validated using a 10 kV testing system data. The experimental results indicate that the proposed method is capable of identifying and learning new distribution data information within non-stationary data stream. This enables the classifier model to maintain a high level of detection accuracy for the current cycle data, effectively enhancing the reliability of HIF detection.

1. Introduction

HIGH Impedance Faults (HIFs) are common events in distribution networks [1], typically caused by contact between line conductors and utility poles, the ground, or tree branches [2]. The transition resistance of the grounding medium in HIF often reaches several hundred or even thousands of ohms, resulting in weak fault signals that are challenging to trigger the activation threshold of traditional protection devices [3,4]. Moreover, HIF occurrences are frequently accompanied by intermittent arc discharges, random movements of grounding conductors [5], and nonlinear distortions, among other unstable characteristics. These factors significantly complicate the detection of HIFs [6]. Failure to promptly detect and eliminate HIF can lead to an increased likelihood of fire outbreaks and personal injury accidents due to the frequent occurrence of accompanying arc discharges [7,8]. Consequently, there is an urgent demand for robust and accurate technical solutions to detect HIFs [9].

With the advances in computing sciences and information technology, emerging data-driven artificial intelligence(AI) has been well-developed for fault detection, showing excellent application prospects [10]. Several studies have utilized AI techniques to address the issue of HIF detection. LALA H et al. in [11] employed variational mode decomposition (VMD) to acquire IMF1 of the fault phase current. Subsequently, the fault rising trend of IMF1 was extracted using singular value decomposition. Finally, classification was conducted utilizing support vector machines (SVM).Guo et al. in [12] introduced a variational autoencoder to extract features from zero-sequence currents and used the obtained features to train a decision tree for fault detection. Gomes et al. in [13] employed sparse coding to extract current and voltage features and trained a random forest for accurate detection of

* Corresponding author.
E-mail address: gaojianhong1994@foxmail.com (J.-H. Gao).

https://doi.org/10.1016/j.ijepes.2023.109705
Received 19 June 2023; Received in revised form 29 September 2023; Accepted 4 December 2023
Available online 21 December 2023

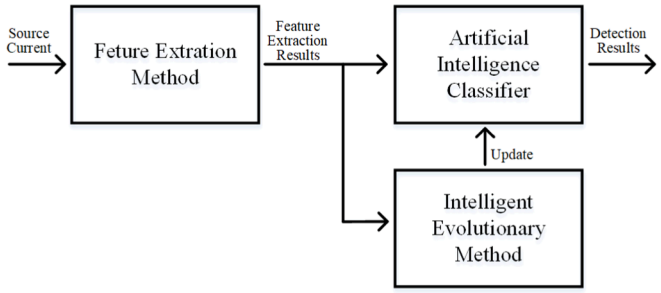


Fig. 1. Flowchart of the proposed HIF detection method.

HIF. Based on sufficient data learning, these methods are capable of extracting fault characteristics across various analysis domains, thereby establishing a nonlinear mapping from signal features to operational conditions. Consequently, they enable rapid identification of interfering factors and yield precise detection outcomes.

Existing AI-based algorithms typically utilize static intelligent classifiers, where the internal parameters are pre-constructed and non-adjustable. Such classifier models demand complete data during the training phase. However, in practical scenarios, data arrives in the form of data stream, making it unfeasible to obtain comprehensive training samples within a short timeframe. Additionally, due to environmental changes, the data distribution frequently undergoes alterations, resulting in non-stationary field data. When confronted with non-stationary data stream, static classifier models often experience performance degradation. Therefore, it is necessary to investigate an HIF detection method with continuous learning capability, which can both learn new information from non-stationary data stream [14] and adapt to changes in data distribution, thereby enhancing the reliability of HIF detection.

In this paper, an effective method based on incremental learning is proposed to tackle the performance degradation of AI-based HIF detection algorithms in non-stationary environments. In comparison to conventional approaches, this proposed method offers the capability for dynamic model adjustments through online incremental learning, identify and acquire non-stationary information from data stream, making it suited for complex distribution network environments.The method utilizes SWT to decompose zero-sequence current samples and extracts effective features through standard deviation calculation. A BPNN is used to classify these features. To facilitate the adaptive evolution of the classifier model, an evolutionary framework is utilized. This framework incorporates data evaluation, model updating, and exemplar dataset updating. This framework enables the classifier model to adapt to changes in field scenarios, ensuring high detection accuracy for the current cycle data.

The fault detection method proposed in this paper effectively monitors scenario changes within the field data stream and dynamically adapts to new scenarios by reorganizing the classifier model through incremental learning. This incremental detection method offers several contributions:

1) The method exhibits a strong data screening capability. Its data evaluation mechanism can perceive scenario changes in the data stream and accurately select new distributed data that differs from the representation information of the exemplar dataset.
2) The proposed method employs small-sample incremental training, enabling fast learning. It employs a lightweight AI classifier characterized by its simple structure and low computational requirements. The online incremental learning process utilizes a limited number of samples and is built upon a pre-trained classifier model, thereby leading to accelerated convergence.
3) The proposed method demonstrates strong resilience against catastrophic forgetting, which refers to the phenomenon where the model tends to forget the knowledge acquired from past data while learning new data. To address this challenge, the method constructs an exemplar dataset to store the information of previous scenario data and introduces a data replay mechanism. This ensures that the model can learn new distribution knowledge without forgetting the knowledge obtained from previous distributions.

The structure of this paper is as follows. Section II provides an in-depth introduction to the incremental HIF detection algorithm proposed in this study. In Section III, we describe the experimental test setup and present the results obtained from the application of the proposed method. Finally, Section IV summarizes the contributions of this paper and offers recommendations for further research.

2. Proposed method

Dynamic data distributions necessitate the continuous online evolution of model parameters [15–17]. In such cases, conventional machine learning models may fail to meet performance requirements due to their predetermined and non-adjustable internal parameters. The proposed detection method in this paper is characterized by its ability to adaptively adjust internal parameters based on the changing scenario characteristics, thus facilitating the machine learning model's adaptation to emerging fault patterns. Fig. 1 depicts the general flowchart of the detection method, which comprises three key elements:

1) Feature extraction method applied to zero-sequence current signals;
2) AI-based classifier models;
3) Intelligent evolutionary method for acquiring newly distributed knowledge in data stream.

The present study introduces an evolutionary approach for sustainable learning of the HIF detection model, as depicted in Fig. 3. The approach consists of four key components:

1) *Baseline model*: This component refers to the adoption of a pre-existing lightweight machine learning model for the HIF detection system. In this paper, the BPNN is selected as the underlying model due to its ease of system deployment and implementation.
2) *Data evaluation*: This component involves assessing the data stream to identify new distribution data exhibiting shifts and constructing a dataset for learning.
3) *Model update*: The model learns new distribution knowledge from the incoming data while revisiting the old distribution by replaying the data from the exemplar dataset. This prevents the loss of previously acquired knowledge.
4) *Exemplar dataset update*: This component merges the old and new data to construct an updated exemplar dataset based on the distribution.

Among these, the specific workflow for data evaluation, model update, and exemplar dataset update is illustrated in Fig. 2.

2.1. Baseline model

The core of the AI-based HIF detection method is an intelligent classifier model. This paper proposes a sustainable learning HIF detection approach intended for complex industrial environments, thus employing one of the lightweight classifier models, BPNN. The fault characteristics in different frequency domains from zero-sequence current data are extracted using SWT in this paper, followed by the establishment of a nonlinear mapping from signal features to operational conditions using BPNN.The input layer of the BPNN comprises six neurons that connect the output results of the feature extraction method to the hidden layer. Through extensive experimentation, it has been determined that configuring the hidden layer of the BPNN as $8 \times 18 \times 6$ yields superior detection performance on real-world HIF data. The

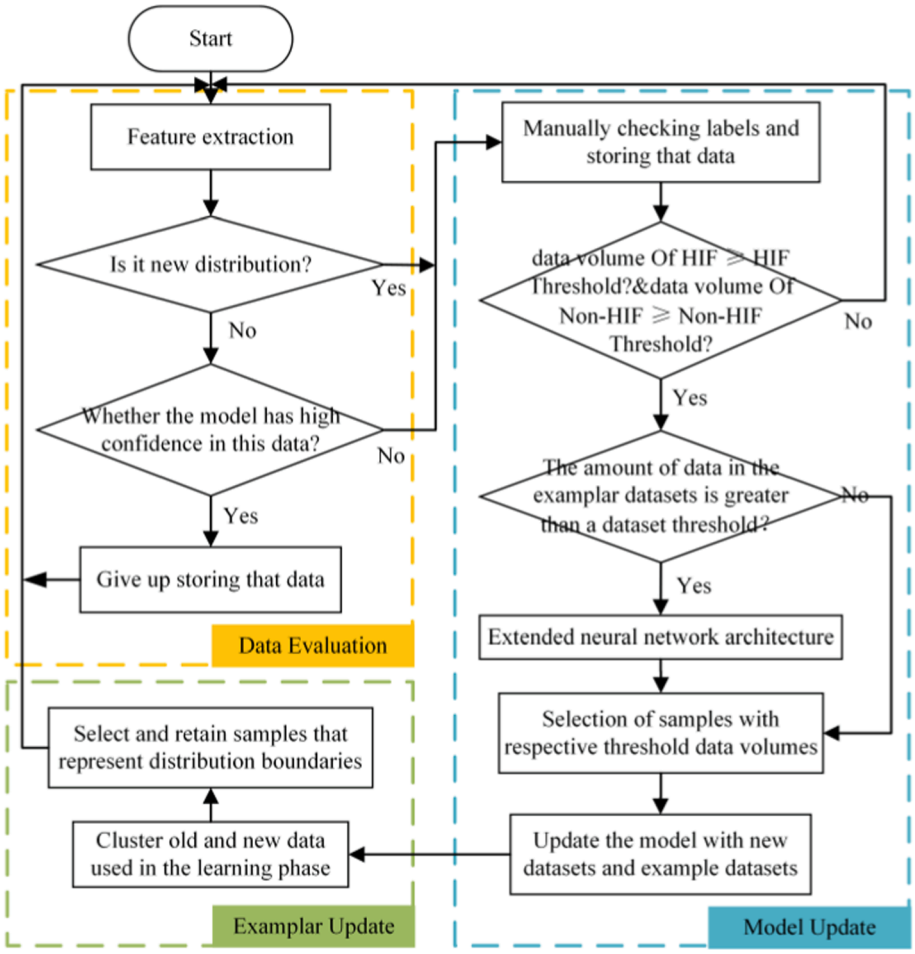


Fig. 2. Flowchart of the proposed HIF detection method.

output layer consists of two neurons, with each representing HIF and non-HIF, respectively.

The BPNN model is initially pre-trained using small-sample data to establish the initial classifier model with a structure of $6 \times 8 \times 18 \times 6 \times 2$. The training data undergoes data dimensionality reduction and clustering processing. Utilizing the results of the clustering analysis, a subset of data is selected to construct an exemplar dataset that accurately represents the information contained in the original training samples.

Training and testing were conducted using the PyTorch framework, with the specific hyperparameters detailed in Table 1. The selection of suitable hyperparameters is crucial for optimizing model performance. In our proposed method, the hyperparameters that deliver the best performance on the dataset for this method were determined using a trial-and-error approach. The hyperparameters presented in Table 1 were chosen through this process to achieve the optimal performance of the proposed method.

In practical distribution networks, data recorded by acquisition devices often exhibit noise. Noise interference can lead to significant waveform variations, thereby indirectly affecting the accuracy of HIF detection. Within the wavelet domain, noise typically manifests as small-amplitude high-frequency components. Therefore, in line with the wavelet denoising principle, this study employs soft thresholding after decomposing the signal using wavelet transformation to suppress these high-frequency noise wavelet coefficients. The denoised wavelet coefficients are then utilized to reveal fault characteristics in zero-sequence current data across different analysis domains. Taking into account the advantages of soft thresholding in balancing noise reduction and preserving signal details, the thresholding method mentioned above employs soft thresholding. The setting of the soft threshold aims to minimize noise while maximizing the preservation of useful signals. To assess the algorithm's robustness in the presence of noise interference, various levels of noise are introduced into the waveform data, and the test results are presented in Table 2, as illustrated.

As evident from Table 2, feature extraction employing the wavelet denoising algorithm enhances the detection stability of the approach. This enhancement enables the proposed algorithm to accurately extract fault characteristics and exhibit robust performance even in the presence of noise interference.

2.2. Data evaluation

Given the complex and dynamic nature of power scenarios in the field, the data presents a diverse and multi-level distribution. Attempting to learn all the data within the data stream would lead to substantial computational redundancy. Hence, it becomes crucial to employ a data evaluation mechanism that assesses the data stream, identifies data exhibiting distribution changes, and constructs a dataset for learning. This mechanism not only guarantees the detection system's performance but also minimizes computational costs and enhances overall system efficiency.

The data evaluation mechanism proposed in this paper comprises three components: feature extraction, data classification, and comprehensive evaluation.

1) feature extraction

In cases of HIFs and asymmetrical disturbances in distribution

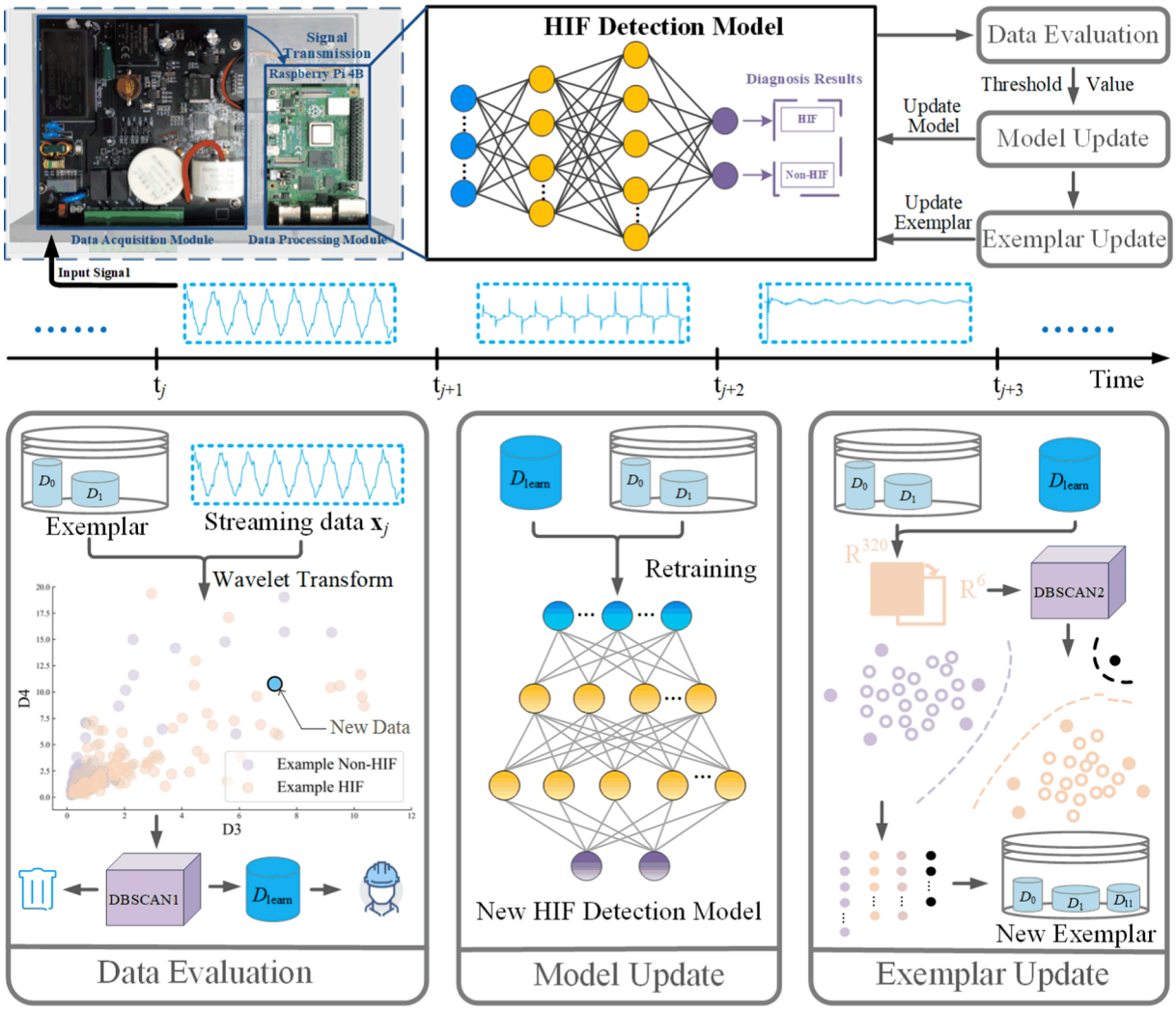


Fig. 3. Schematic diagram of the HIF detection method based on small-sample incremental learning.

Table 1
Noise immunity performance test results.

Learning rate	Training epoch	Optimizer	Batch Size	Eps
0.001	100	Adam	25	5.24
Denoising threshold	Non-HIF quantity threshold	HIF quantity threshold	Confidence level	MinPts
0.1	5	5	80 %	5
Model Update	Training epoch	Learning rate	Optimizer	Batch Size
	10	0.0001	Adam	10

Table 2
Noise immunity performance test results.

SNR(dB)	40	20	15	10	5	1
Accuracy(%)	98.35	99.21	98.50	98.23	98.63	97.84

networks, transient fluctuations arise in the line's zero-sequence current. By analyzing the time-varying, nonlinear transient signals, classifier models can enhance their capacity to detect HIF more effectively. The wavelet transform stands out as a potent tool in signal processing and analysis. Essentially, it concurrently provides temporal and frequency information for a given signal. Utilizing wavelets for signal

Table 3
Detection accuracy and computation time based on SWT feature extraction.

	Haar	**Db2**	**Db4**	**Coif1**	**Sym2**	**Sym4**
Accuracy of simulated datasets (%)	97.84	**98.43**	98.24	98.08	96.49	97.78
processing time (ms)	1.23	1.31	1.48	1.40	1.31	1.49
Accuracy of real-word datasets (%)	85.37	**86.79**	83.03	83.86	82.85	82.86
processing time (ms)	1.22	1.31	1.48	1.40	1.32	1.49

Table 4
Detection accuracy and computation time based on MALLAT feature extraction.

	Haar	**Db2**	**Db4**	**Coif1**	**Sym2**	**Sym4**
Accuracy of simulated datasets (%)	97.81	97.25	97.57	96.90	95.12	94.20
processing time (ms)	**0.89**	0.90	0.89	0.91	0.90	0.93
Accuracy of real-word datasets (%)	81.35	80.76	82.27	77.59	80.35	80.76
processing time (ms)	**0.89**	0.89	0.90	0.91	0.91	0.93

analysis is rooted in the principle of scaling and translating the signal through the mother wavelet. The mother wavelet undergoes expansion and compression via scaling operations to yield low and high-frequency signals, respectively.

Table 5
Detection accuracy and computation time based on Wavelet Packet Transform feature extraction.

	Haar	Db2	Db4	Coif1	Sym2	Sym4
Accuracy of simulated datasets (%)	90.15	91.37	92.96	91.91	89.19	88.74
processing time (ms)	1.77	1.78	1.75	1.79	1.78	1.81
Accuracy of real-word datasets (%)	81.43	78.34	75.08	68.89	73.32	73.49
processing time (ms)	1.77	1.78	1.75	1.79	1.78	1.80

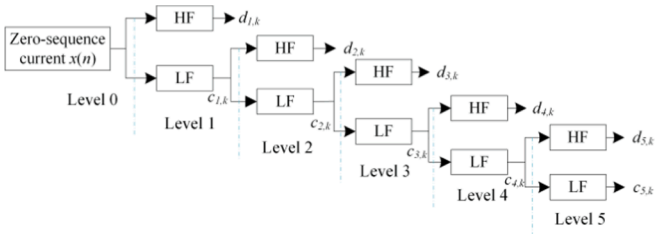


Fig. 4. Five level decomposition with SWT.

Regarding the detection algorithm embedded in the on-site microcomputer, the accuracy of feature extraction and the computational efficiency of the algorithm are both pivotal factors that demand attention. This study takes both factors into comprehensive consideration. Specifically, it employs the Stationary Wavelet Transform, Mallat transform, and Wavelet Packet Transform to analyze the zero-sequence currents. Subsequently, the analysis results are applied within the proposed framework for HIF detection. The outcomes of these detection efforts, along with the corresponding computation times, are detailed in Tables 3 through 5. The test microcomputer utilized is a Raspberry Pi 4B with 1.5 GHz BCM2711 (CPU) and 8.00 GB of processor and RAM. Due to the constraints posed by the Raspberry Pi's memory and computational resources, employing longer wavelet bases may surpass the processing capabilities of the Raspberry Pi. Consequently, this study opts for several commonly used compact wavelet bases during testing.

Based on the findings presented in Tables 3 to 5, it becomes evident that, irrespective of whether the data is simulated or real-world, the most effective signal processing method is the SWT employing the Db2 wavelet base. For computational efficiency, the Mallat transform using the Haar wavelet base demonstrates the highest performance. Conversely, the Wavelet Packet Transform exhibits both lower accuracy and computational efficiency compared to SWT and Mallat. For simulated data, the accuracy of feature extraction achieved by SWT surpasses Mallat by 1.18 %, whereas for real-world data, SWT outperforms Mallat by 6.03 % in terms of feature accuracy. It is worth noting that the precision of feature extraction directly impacts the algorithm's overall performance and utility, as it furnishes a dependable foundation for decision-making. While computational efficiency holds significance, especially on edge microcomputers, addressing this concern can be accomplished through algorithmic optimization and hardware acceleration. This can be achieved while ensuring the accuracy of feature extraction. Consequently, under the condition that computational efficiency requirements are met, this paper employs SWT as the preferred tool for signal processing and analysis.

As referenced in [18], it is established that longer wavelet bases introduce significant time delays when detecting HIF transients, whereas compact wavelet bases do not incur such delays. Compact wavelet bases, when applied to signal processing, offer a more dependable feature for HIF detection due to the increased distortion components at higher periodicized signal boundaries. Consequently, compact wavelet coefficients exhibit superior performance in fast transient detection when compared to their long-wavelet counterparts. However, it is worth noting that the Haar wavelet base presents undesirable low-frequency oscillations, which complicate transient detection. Conversely, other compact wavelets derived from db2 do not exhibit such problematic low-frequency oscillations [19]. This aligns with the previously discussed test results, thus, in this paper, the db2 wavelet base is employed as the mother wavelet for the SWT.

SWT is a multi-scale analysis method that decomposes signals into sub-signals at different scales, suitable for analyzing time-varying nonlinear transient signals [20]. In this paper, a high sampling rate of 4000 Hz is used to capture high-frequency features in the zero-sequence current waveform. The zero-sequence current undergoes 5-level decomposition using SWT, as schematically depicted in Fig. 4. Based on coefficients $c_{j,k}$, we can formulate the general transformation equation for SWT, as follows.

$$c_{j,k} = \langle f(x), \phi_{j,k}(x) \rangle \tag{1}$$

$$\phi_{j,k}(x) = 2^{-j}\phi\left(2^{-j}x - k\right) \tag{2}$$

The discrete detail coefficients at the resolution 2^j are given by.

$$d_{j,k} = \langle f(x), 2^{-j}\psi\left(2^{-j}x - k\right) \rangle \tag{3}$$

where $\psi(x)$ is called the wavelet function.

As the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ can be expressed as convolution with high pass and low pass filters as follows.

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_n h(n)\phi(x-n) \tag{4}$$

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \sum_n g(n)\phi(x-n) \tag{5}$$

The approximate coefficients at level $j+1$ $c_{j+1,k}$ can be directly computed from the previous step i.e., $c_{j,k}$ as follows.

$$c_{j+1,k} = \sum_n h(n-2k)c_{j,n} \tag{6}$$

Similarly, the detail coefficients at level $j+1$ also can be computed as follows.

$$d_{j+1,k} = \sum g(n-2k)c_{j,n} \tag{7}$$

Eqs. (4) and (7) are used for multi-resolution DWT analysis, in which the signal is down-sampled in each step and the length of the signal is halved in each step. However, in SWT, instead of down-sampling, up-sampling of the signal is carried out before filter convolution. Thus, in the case of SWT algorithm, the approximate and detailed coefficients are obtained as given below.

$$c_{j+1,k} = \sum_m h(m)c_{j,k+2^j m} \tag{8}$$

$$d_{j+1,k} = \sum_m g(m)c_{j,k+2^j m} \tag{9}$$

Here, $h(\cdot)$ and $g(\cdot)$ are the coefficients of low-pass and high-pass filters, respectively.

To minimize redundancy in the original waveform or its transformed formats, feature extraction is required. The standard deviation, serving as an indicator of signal distribution dispersion, offers insights into the degree of variation in signal frequency distribution. Therefore, standard deviation is computed for the detail coefficients from the first to the sixth level to achieve time-localized feature extraction and data dimensionality reduction for the zero-sequence current.

2) data classification

After the above feature extraction, the data undergoes dimensionality reduction from the original high-dimensional time series to six wavelet standard deviation features. These features include components

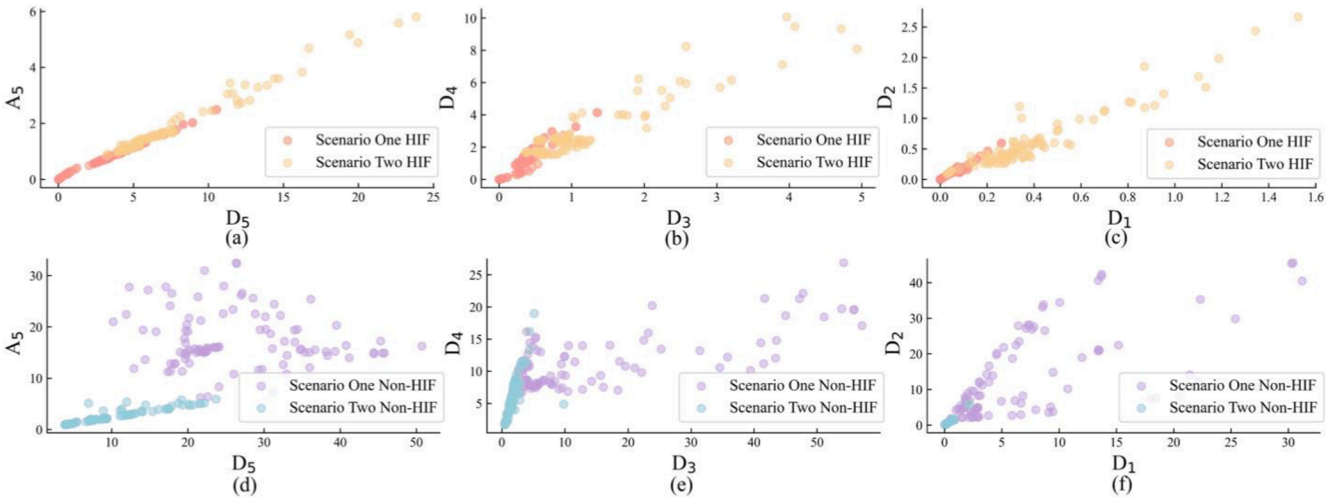


Fig. 5. Visualization for the features of the real-word dataset: (a) HIF low-frequency features (b) HIF mid-frequency features (c) HIF high-frequency features (d) non-HIF low-frequency features (e) non-HIF mid-frequency features (f) non-HIF high-frequency features.

A5 and D5, representing low-frequency characteristics, components D4 and D3, characterizing mid-frequency features, and classifications D2 and D1, signifying high-frequency attributes.

The conclusions drawn from Shapley additive explanation (SHAP) attribution theory in the referenced literature [21] concerning the principles of BPNN detection of HIF reveal that HIF waveforms contain abundant odd harmonics and high-frequency components. Consequently, there exists an overall positive correlation between the self-values of features D4, D3, D1, and D2, as well as their Shapley values. Conversely, the self-values of D5 and A5 exhibit an overall negative correlation with their Shapley values. Based on these findings, this paper defines the data distribution characterized by the six wavelet standard deviation features as follows:

$$C = [M_1 \bullet D_1, M_2 \bullet D_2, M_3 \bullet D_3, M_4 \bullet D_4, M_5 \bullet D_5, M_6 \bullet A_5] \tag{10}$$

$$M_i = 6 \times SHAP_{D_i} / \sum_{i=1}^{6} SHAP_{D_i} \tag{11}$$

Among these, $SHAP_{D_i}$ is determined based on experimental results from the literature.

To investigate the outcomes of feature extraction in diverse electrical scenarios, this study utilizes the 10 kV distribution network model constructed within the PSCAD/EMTDC software to acquire zero-sequence current data under various conditions. Scenario One is composed of capacitor switching (CS) events and HIF events simulated by the Mayr arc model, while Scenario Two comprises Low Impedance Fault (LIF) events and HIF events simulated by the Emanuel arc model. Fig. 5 illustrates the results of feature extraction and distribution-weighted visualization of the data for both scenarios. From the visual representation in Fig. 5, it becomes evident that certain data points in scenario one exhibit a notable distribution shift compared to the data distribution observed in scenario two.

Based on the data distribution characteristics of the zero-sequence current dataset following SWT feature extraction and SHAP distribution weighting, this study employs Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to assess the distribution of new data within the data stream. The specific process for data distribution evaluation based on DBSCAN is outlined as follows:

i) Based on extensive experimental results, set the neighborhood parameters (eps, minPts), input the example dataset data and the feature-weighted vector C of the new data samples in the data stream.
ii) Arbitrarily select a data object point p from the data set.
iii) If the chosen data object point p qualifies as a core point according to the parameters eps and minPts, proceed to identify and gather all data object points that are density-reachable from p, thereby forming a cluster.
iv) If the selected data object point p is an edge point, select the next data object point.
v) Repeat steps (ii) and (iii) until all points have been processed, and output the data point cluster labels.

The high-dimensional raw zero-sequence current data is projected onto a 6-dimensional space that characterizes information from each frequency band after undergoing SWT feature extraction and SHAP distribution weighting. Various distributions in this space form clusters of varying shapes. DBSCAN, utilizing density reachability, identifies data points within dense regions and groups them into clusters. Simultaneously, it detects noise or boundary points in sparse regions, free from the constraints of cluster shapes.Using DBSCAN to cluster new data samples with example set data, the focus lies on the cluster labels of the new data samples. When a cluster label is assigned as noise or a boundary point, it signifies that the new data samples' density cannot reach that of the data in the example dataset. In other words, these new data samples cannot form clusters with the data in the example dataset within the six-dimensional space following SWT feature extraction and SHAP distribution weighting. This implies a significant difference in their distributions from the old data, necessitating manual scrutiny and relearning of such data.Conversely, when a cluster label corresponds to a specific cluster, it indicates that the density of the new data samples matches that of the data in the example dataset. In this case, their distribution is considered less different from the old data, and there is no need for further learning.

4) comprehensive evaluation

After performing the aforementioned feature extraction and classification of the zero-sequence currents, the data distribution evaluation result is obtained, which indicates the dissimilarity between the data in the data stream and the data in the exemplar dataset. In this study, in addition to relying on the distribution evaluation results, we also utilize the probability output from the HIF detection model's output layer to determine the model's confidence level in classifying new data. The model's assessments for certain data points often fall near the boundary between faults and non-faults, resulting in low confidence in this subset of data and making detections more error-prone. To address this issue, the proposed method selects these data points by applying a suitable

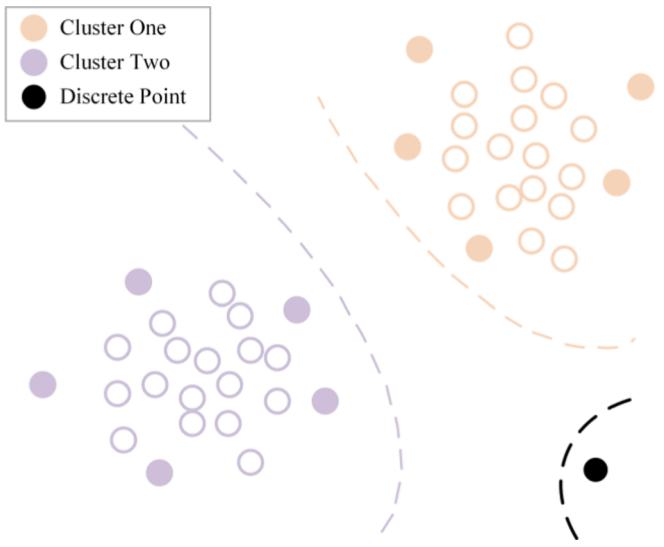


Fig. 6. Schematic diagram of Exemplar dataset update.

confidence threshold. When the confidence level exceeds the predefined threshold, it indicates that the model's detection of the current data is relatively reliable, eliminating the need for further learning. Conversely, if the confidence level falls below the set threshold, the model's detection of the current data is considered insufficiently reliable, requiring manual examination and subsequent re-learning.

In this paper, we recognize that new data with shifted data distribution can contribute to the discovery of new detection patterns for the detection model. Consequently, the proposed data evaluation system emphasizes the determination of data distribution as the primary criterion, with the evaluation of the model's confidence in the detection results serving as a supplementary criterion. The data evaluation mechanism categorizes the evaluation results of new data into three groups:

i) High confidence in the existing distribution: When the new data exhibits minor deviations from the distribution of the old data, and the model displays high confidence in the detection results of the new data, the information provided by the new data is redundant, and the model does not need to learn from this data.
ii) Low confidence in the existing distribution: When the new data shows slight deviations from the distribution of the old data, but the model demonstrates low confidence in the detection results, it indicates an insufficient understanding of this data by the model, necessitating further learning.
iii) New distribution: The distribution of the new data has shifted, indicating the presence of new feature information. Regardless of the confidence level in the model's detection results, the system stores this data and conducts manual verification, considering it as one of the samples for model learning in the next stage.

2.3. Model update

As illustrated in the Model Update process box in Fig. 2, when the stored data volume of pending-to-learn HIF and Non-HIF samples reaches the predefined thresholds for the number of HIFs and Non-HIFs, respectively, samples equal to these threshold data volumes are selected to constitute the pending-to-learn dataset, while the remaining data proceeds to the next incremental cycle. Building upon the inheritance of all the parameter weights of the original model, a data replay mechanism is introduced to update the classifier model using samples from the exemplar dataset and the pending-to-learn dataset. The settings of the two quantity thresholds are synchronized with the amount of replayed data to prevent the potential influence of dataset imbalance introduced by retrospective training. The pending-to-learn dataset comprises newly distributed data encountered by the model in the data stream and data with low confidence levels. The utilization of this data segment enables the model to acquire new knowledge, while the exemplar dataset represents the data distribution of previously trained samples, allowing the model to reinforce the old knowledge it has acquired and avoid "catastrophic forgetting" during its evolution.

During the learning of the nth distribution, the model can only access the current training dataset, $\boldsymbol{D}_n$. However, updating the model solely with $\boldsymbol{D}_n$ can lead to catastrophic forgetting. To address this issue, a method for preserving an additional exemplar dataset, denoted as $\mathcal{E}=\{(\mathrm{x}_j, \mathrm{y}_j)\}$, which contains a certain number of representative samples for each encountered distribution, has been proposed. Here, $\boldsymbol{x}_j$ represents the data of exemplar dataset samples, and y_j represents the labels of exemplar dataset samples. By retaining these samples, the model is better equipped to resist catastrophic forgetting during the process of learning new distributions. This approach mimics the way humans actively review previously acquired knowledge when acquiring new information, a concept referred to as data replay [22,23]. The aforementioned review process can be described as follows:

$$\mathscr{L} = \sum_{(\boldsymbol{x}_i, y_i)\in\left(\mathscr{D}^b \cup \mathscr{E}\right)} \ell\left(f(\mathbf{x}_i), y_i\right) \tag{12}$$

where $\mathscr{D}$ represents the model training process, $\ell()$ represents the loss calculation function, and $f(\boldsymbol{x}_i)$ represents the output of the model for sample $\boldsymbol{x}_i$. Equation (12) indicates that when the model learns the new task $\boldsymbol{D}_n$, it needs to consider the loss on the previous exemplar dataset of old distributions simultaneously, and optimize the model $f()$ to enable it to possess the detection capability for both the old and new distributions simultaneously.

To prevent overfitting of the BPNN model, this paper employs a structurally adaptive approach to align the model's complexity with the volume of training data. In scenarios with limited data, a simpler neural network structure is selected to reduce model complexity and mitigate the risk of overfitting. As the volume of data increases, the simplicity of the neural network hampers its learning capabilities, making it ineffective at acquiring complex feature representations, thereby constraining the model's performance. Consequently, when the training data volume reaches a certain level, this paper expands the model's structure by adding hidden layer neurons. To ensure that this structural expansion does not lead to performance degradation in the original data distribution task, the weights and biases associated with the added neurons are initially set to 0.01 % of those from the preceding neuron.

2.4. Exemplar dataset update

The exemplar dataset represents the knowledge that the model has already acquired, necessitating updates to the exemplar dataset following model evolution. As the model evolves, the amount of acquired data grows. However, due to storage constraints on edge devices, the exemplar dataset cannot expand indefinitely. Consequently, an under-sampling update approach is implemented for the exemplar dataset after model evolution to maintain a stable storage overhead.

As shown in Fig. 6, the proposed approach for updating the exemplar dataset aims to select and retain samples that effectively represent distribution boundaries. Algorithm 1 outlines the specific process of sample selection to update the exemplar dataset. Firstly, the old and new data used in the model's learning phase are extracted and clustered. The discrete points from these clusters form datasets $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Next, the center of each cluster is determined, and the Euclidean distance between each dimension-reduced data point and its corresponding cluster center is calculated. The desired number of datasets $\boldsymbol{X}_3$ is selected by sorting these distances in descending order. Subsequently, a random subset of data is chosen to form a supplementary dataset $\boldsymbol{X}_4$, based on the selected

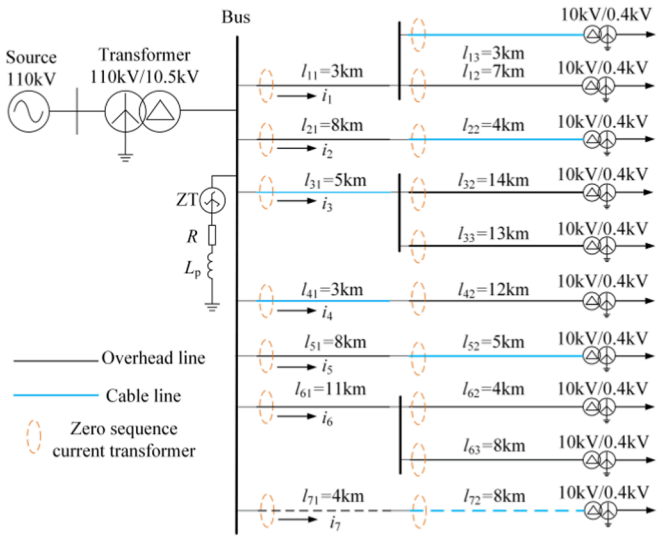


Fig. 7. Simulation model of 10 kV distribution system.

Table 6
Line parameters.

Line type	Sequential Component	Resistance Ω/km	Capacitance μF/km	Inductance mH/km
Cable lines	Zero	2.7000	0.2800	1.0190
	Positive	0.2700	0.3390	0.2550
Overhead lines	Zero	0.2300	0.0080	5.4780
	Positive	0.1700	0.0097	1.2100

data volume. Finally, datasets $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$ are combined to construct the updated exemplar dataset, denoted as $\boldsymbol{X}_{\text{new}}$.

Algorithm 1.

Input
Learn dataset $\boldsymbol{X}_{\text{learn}}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m\}$,Exemplar dataset $\boldsymbol{X}_{\text{old}}=\{\mathbf{x}_{m+1},\mathbf{x}_{m+2},\ldots,\mathbf{x}_{m+n}\}$
Minimum number of points to be included in the category min*Pts*
HIF scanning radius eps_1,Non-HIF scanning radius eps_2
Output
New exemplar dataset $\boldsymbol{X}_{\text{new}}$
Steps

(*continued on next column*)

(*continued*)

1. Stationary wavelet transform $\varphi(\bullet):\mathbb{R}^c\rightarrow\mathbb{R}^d$
$\mathbf{Y}_1,\mathbf{Y}_2\leftarrow\varphi(\mathbf{X}_{\text{learn}}+\mathbf{X}_{\text{old}})$
2.Density-Based Spatial Clustering of Applications with Noise DBSCAN
$\mathbf{L}_1\leftarrow\text{DBSCAN}(\mathbf{Y}_1),\mathbf{L}_2\leftarrow\text{DBSCAN}(\mathbf{Y}_2)$
3.Select all discrete points $\boldsymbol{X}_1,\boldsymbol{X}_2$
$\mathbf{X}_1\leftarrow\mathbf{L}_1=-1,\mathbf{X}_2\leftarrow\mathbf{L}_2=-1$
4.Select category boundary points $\boldsymbol{X}_3$
unique_labels = unique($\boldsymbol{L}_1$)
for i **in** unique_labels
$C^i=\frac{1}{\text{len}(L^i)}\sum\mathbf{Y}^i$//cluster center
end for
for $j=1,2,\ldots,m+n$ **do**
$d_j=|\mathbf{Y}_j^i-C^i|$//distance
end for
Arrange d_j in descending order and select $\boldsymbol{X}_3$ according to min*Pts*
5.Select supplemental points $\boldsymbol{X}_4$
if len($\boldsymbol{X}_1+\boldsymbol{X}_2+\boldsymbol{X}_3$)>=len($\boldsymbol{X}_{\text{old}}$)
$\boldsymbol{X}_4=\emptyset$
else
Randomly select len($\boldsymbol{X}_{\text{old}}$)-len($\boldsymbol{X}_1+\boldsymbol{X}_2+\boldsymbol{X}_3$) points from $\boldsymbol{X}_{\text{old}}+\boldsymbol{X}_{\text{old}}$-$\boldsymbol{X}_1$-$\boldsymbol{X}_2$-$\boldsymbol{X}_3$ to form $\boldsymbol{X}_4$
6.$\boldsymbol{X}_1$, $\boldsymbol{X}_2$, $\boldsymbol{X}_3$, and $\boldsymbol{X}_4$ constitute $\boldsymbol{X}_{\text{new}}$

3. Simulation and real-world data

3.1. Simulation model and real-world 10 kV distribution system

A distribution network model at 10 kV was created using the PSCAD/EMTDC software, as depicted in Fig. 7. The system consists of six lines, which are a combination of overhead lines and cable lines. The lines in the system are implemented using the Bergeron model, and their parameters are presented in Table 6. Each feeder line is equipped with zero-sequence current transformers placed at the starting point. The system operates in an overcompensated state with an overcompensation degree of 6 %.

A variety of fault conditions were simulated in a real-world 10 kV distribution system, considering different high impedance grounding media. The system consists of four feeders with simulated lengths of 9.7 km, 25 km, 17.35 km, and 4.9 km. The total capacitive current of the system is 41.4 A, with individual capacitive currents of 20 A, 3.4 A, 8 A, and 10 A for the four lines, respectively. The experiments focused on the 20 A line, and after arc extinguishing compensation, the residual current ranged from approximately 3 A to 4 A. The line has a total capacitance to

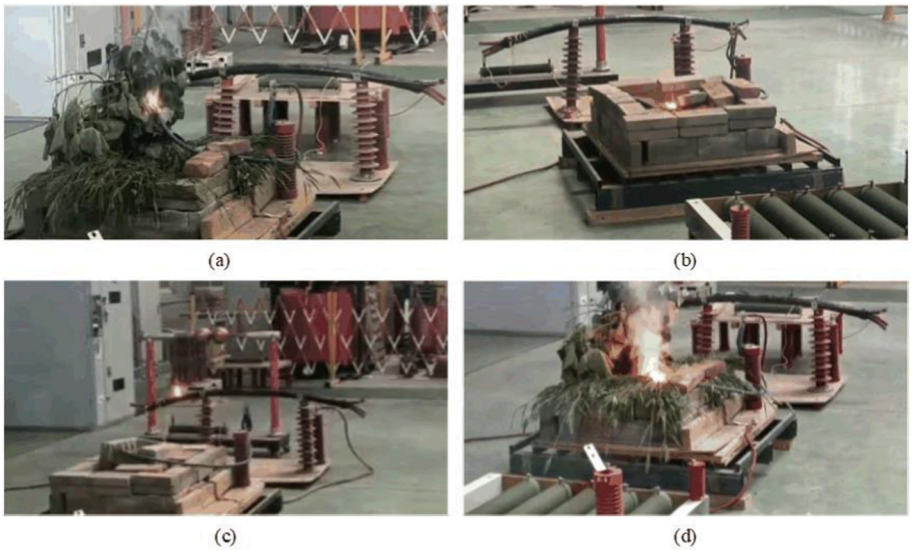


Fig. 8. Experimental scenarios:(a) branch grounding (b) masonry grounding (c) cable arc grounding (d) grassland grounding.

Table 7
Parameters of simulated events.

Event type	Topology	FL	Delay time (ms)	Simulated model	Number
HIF	*Add* *Init* *Delete*	l_{11} l_{12} l_{13} l_{21} l_{22} l_{31} l_{32} l_{33} l_{41} l_{42} $l_{51}$$l_{52}$	—	Mayr、Cassie、Emanuel and Control Theory Model	2560
CS		l_{61} l_{62} l_{63}	**0,1,2,3**	3-phase capacitor	520
IC			—	No-load transformer	520
LIF			—	Resistance	520
LS			—	3-phase asymmetric	520
Total					**4640**

Table 8
Parameters of simulated events.

R_{arc}					$R_c(\Omega)$
Mayr	P_{loss}(kW) 4.05–6.77/ 34.24–35.67		τ_m(us) 0.30/0.27–0.32		800–3000
Cassie	E_0(kV) 3.66–4.05/0.22–0.26		τ_c(us) 1.05–1.35/230–260		
Emanuel	$R_n(\Omega)$	V_n(kV)	$R_p(\Omega)$	V_p(kV)	
	250–350	3.91–4.63	250–350	3.91–4.63	
	450–550	4.91–5.63	450–550	4.05–4.77	
Control theory	L_k(cm)	I_k(kA)	β	V_k(V/cm)	
	10–100	4	2.85×10^{-5}	15	

ground of 10.8 μF and is connected to parallel resistors of 300 kΩ and 6 kΩ. Fig. 8 illustrates the experimental scenarios involving arc grounding through tree branches, masonry, cables, grass, and sand.

3.2. Acquisition of simulation and real-world data

In the resonant earthed system distribution network model depicted in Fig. 7, various HIF and HIF interference events occurring at different fault locations (Fls) and on different lines are implemented, as detailed in Table 7. HIF is simulated using a combination of variable resistance R_{arc} modeled by the Mayr model, Cassie model, Emanuel model, and cybernetic model, in conjunction with a constant resistance R_c. These models are utilized to simulate different grounding medium conditions by adjusting their parameters, as specified in Table 8. Specifically, the simulation encompasses CS employing shunt 3-phase capacitors, inrush current (IC) simulated with an no-loaded single-phase transformer, LIF represented by a Low Impedance Model (5 Ω-100 Ω), and load switching (LS) modeled using 3-phase asymmetrical loads. Additionally, the experiments incorporate asynchronous closing of the CS to simulate non-fault transient conditions, aligning with practical engineering scenarios. In this study, 3-phase asynchronous closing entails phase A being connected to the system first, followed by phases B and C with the same delay. The initial fault angle is set to 0°, 30°, 60°, 90°, and 120°. To broaden the distribution of samples, this research considers variations in the topology of the distribution system, which involve the addition of line l_7 and the removal of line l_3.

A total of 134 sets of HIF full-scale experiment data were collected from a real-world 10 kV distribution system. These experiments simulated various grounding scenarios, including pure resistance, tree branches, masonry, grass gravel, and cable arcing. In addition, 87 sets of HIF field data and 249 sets of non-HIF field data were collected from the field distribution system. Typically, HIF characteristics persist for approximately 8 to 10 cycles [12]. In this paper, the input sample length was set to 3 cycles, and each data group was divided into two non-overlapping samples representing the transient and steady-state periods. Fig. 9 displays the zero-sequence current waveforms and V-I arc characteristics of high impedance faults (HIFs) in both simulated and real-world scenarios. For HIFs based on various simulation models and those occurring in full-scale experiments within scenarios involving branches, grassland, and masonry, the V-I arc characteristics typically exhibit well-defined hysteresis loops, as illustrated in Fig. 9(a)–(d), and (f)–(h). These characteristics feature steep slopes near the origin and

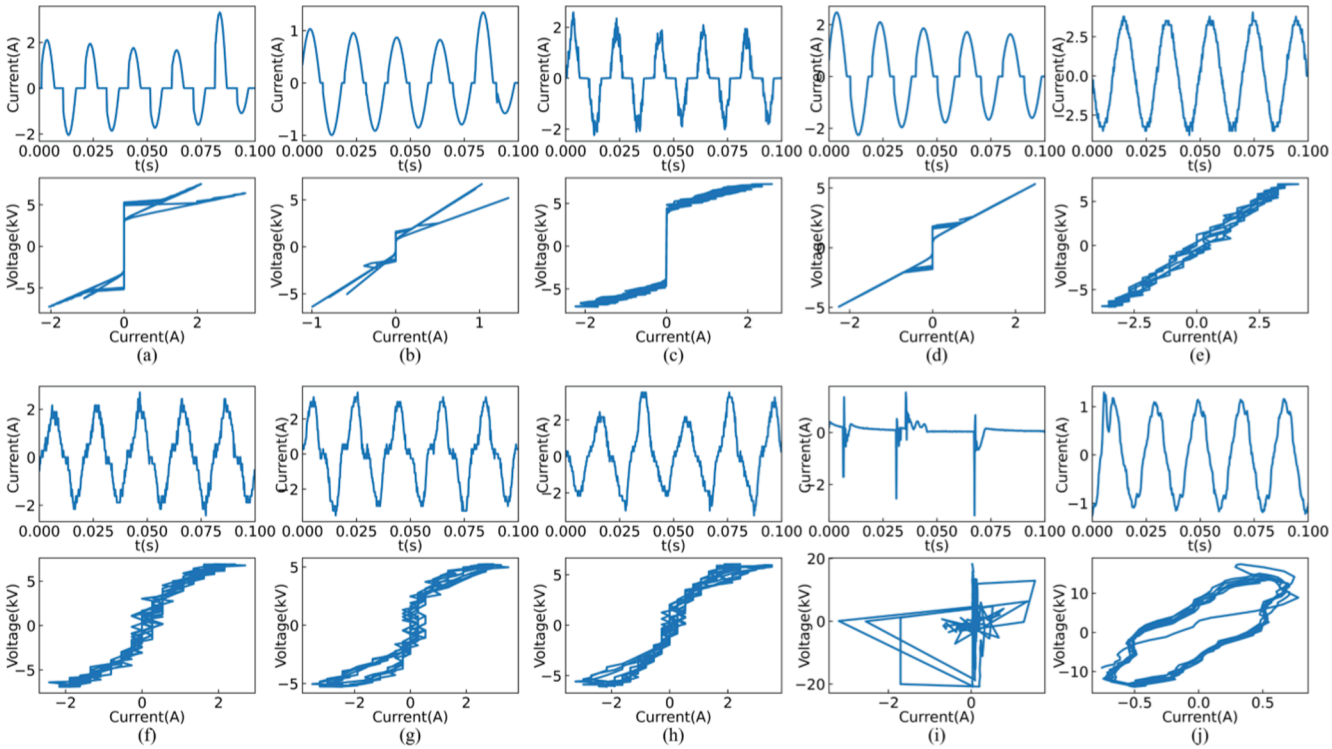


Fig. 9. zero-sequence current waveforms and V-I arc characteristics of HIF in various scenarios: (a)mayr (b) cassie (c) emanuel (d) cybernetic (e) resistance (f) branch (g) grassland (h) masonry (i) cable arc (j) field HIF.

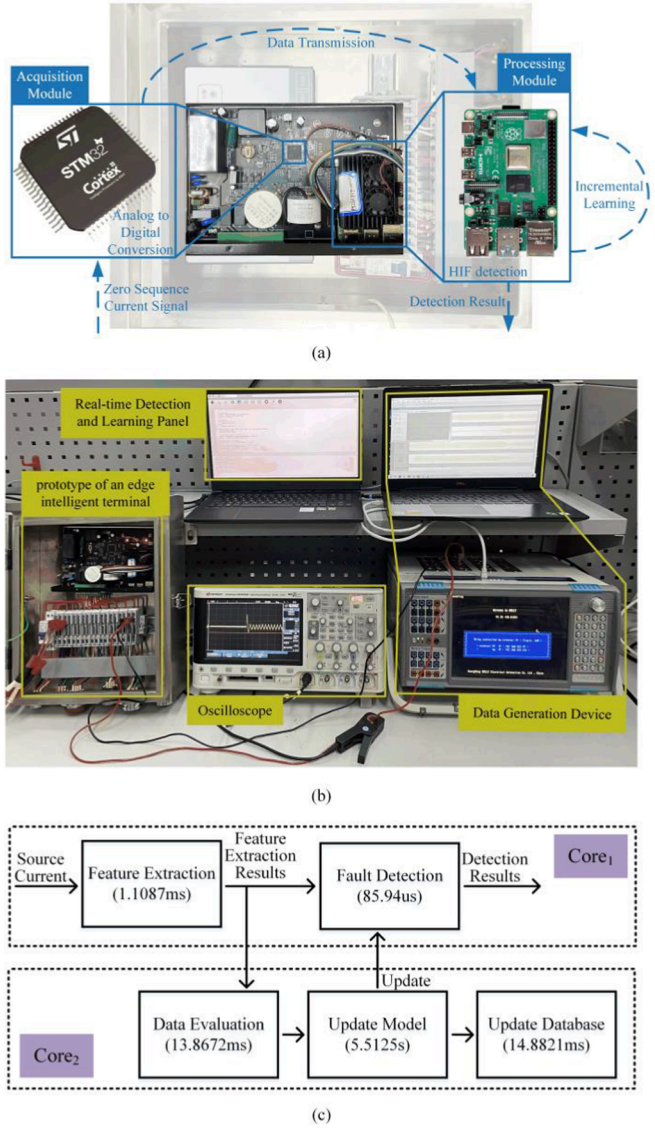


Fig. 10. A simulating scenario of real-time detection and learning: (a) prototype of an edge intelligent terminal (b) system composition (c) process and time used.

Table 9
Comparison of Time Consumption of Different Methods.

Comparison Literature	Feature extraction time	Detection time
(Lala H & Karmakar S, 2020)	234.8 ms	4.703us
(S. Kar & S. R. Samantaray, 2016)	1.773 ms	1.672us
(Gautam & Brahma, 2013)	3.281 ms	6.423us
(Gao et al., 2022)	158.8 ms	62.5 ms

Table 10
Simulation scenario composition.

	Initial Scenario	Incremental Scenario One	Incremental Scenario Two
Composition (samples)	Mayr(640) Cassie (640) LS(520) IC(520)	Emanuel(640) LIF(520)	Cybernetic(640) CS(520)

considerably smaller slopes near the extremes. However, in the cases of pure resistance, cable arcing, and field HIFs, the V-I arc characteristics do not demonstrate the aforementioned behavior but instead display entirely distinct trajectories.

4. Experimental validation

In this paper, to validate the feasibility of the proposed method in engineering applications, we developed a prototype of an edge intelligent terminal deployed in a 10 kV distribution network field. This prototype can be installed in both substations and distribution feeders. The structure of the prototype, as shown in Fig. 10(a), comprises a data acquisition module and a data processing module. The data acquisition module is based on the STM32 microcontroller and is responsible for analog-to-digital conversion and data transmission. The data processing module, based on the Raspberry Pi 4B, is responsible for real-time HIF detection and incremental learning.

Based on the prototype of the edge intelligent terminal described earlier, this paper establishes a simulated system for real-time detection and learning. The components of the system are depicted in Fig. 10(b). In this system, PC1 machine and relay tester form a data generator responsible for generating virtual scenarios of real-time data signals. PC2 machine is used as a real-time detection and learning panel to display the corresponding detection results and incremental learning status. Raspberry Pi is a common edge microcomputer based on ARM architecture. In this study, the Raspberry Pi utilizes a 1.5 GHz BCM2711 (CPU) processor, 8.00 GB of RAM, and the Linux operating system. It is noteworthy that the Raspberry Pi 4B employs the Broadcom BCM2711 SoC, which integrates four ARM Cortex-A72 processor cores. Consequently, the Raspberry Pi can concurrently execute multiple threads, with each core dedicated to processing an individual thread. To assess the algorithm's performance, we computed the approximate time required for the proposed method and several methods from the literature [11,24,25,26] to complete feature extraction, detection, and learning tasks on the developed prototype, as presented in Table 9 and Fig. 10(c).

From Table 9 and Fig. 10(c), it is evident that the proposed method in this paper exhibits the shortest combined feature extraction and detection time. However, the proposed incremental framework consumes more time. The proposed method addressed this issue by employing multi-threaded programming techniques, enabling the execution of HIF detection tasks on core one and incremental learning tasks on core two, thereby achieving a time frame suitable for engineering applications.

4.1. Adaptability analysis based on simulation data

To simulate the non-stationary data stream faced by the HIF detection system in engineering applications, the data samples in Table 10 are randomly partitioned into 8 groups, with each group's data samples further randomly partitioned into training, testing, and validation sets in a 7:2:1 ratio. The training set of the first dataset serves as the initial training set for training the primitive classifier model, while the remaining seven training sets are utilized as the simulated data stream dataset. In the initial training set, there is a partial distribution of the total samples, which, however, is insufficient to cover all distribution information. This is done to simulate scenarios where initial training samples may be incomplete in field applications. The simulated data stream dataset is employed to mimic the data stream encountered by the model at different time intervals. The test set for each dataset comprises 20 % of the data samples in that dataset and is employed to assess the model's performance on the current cycle of data. Each validation dataset from every group is used to assist the model in selecting the best hyperparameter configuration for the current data volume.

The process of the simulated data stream experiment is as follows. The model undergoes initial training in the initial training dataset, followed by incremental learning in batches using the simulated data

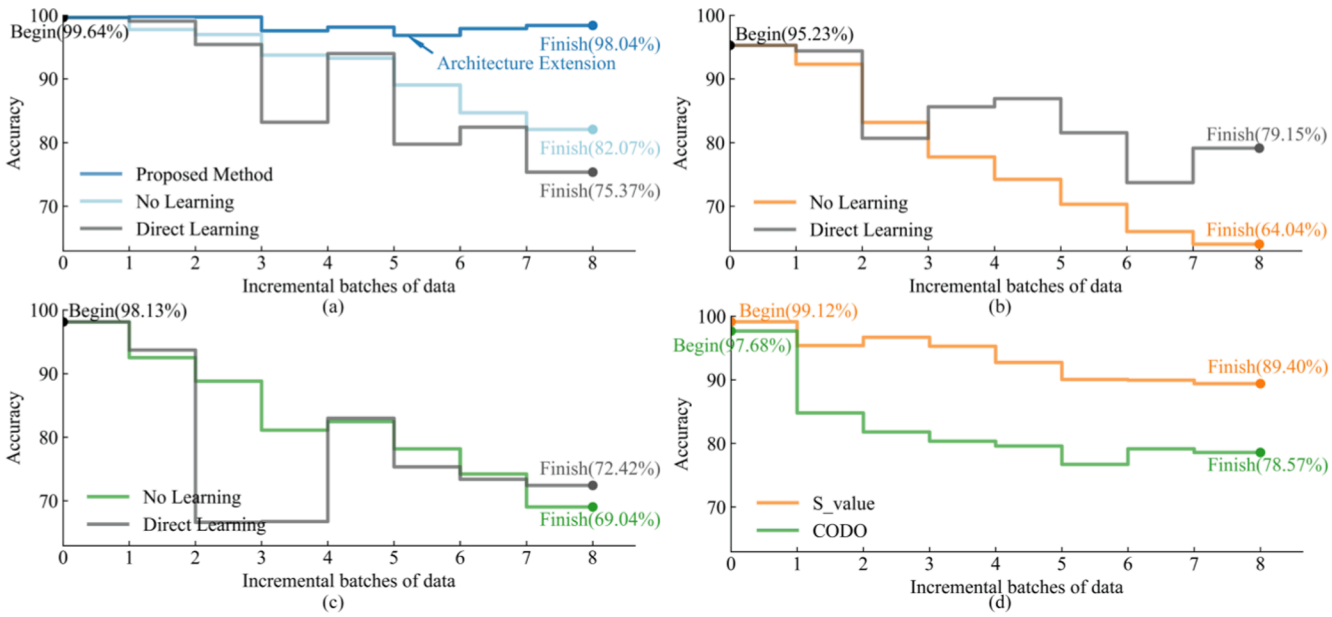


Fig. 11. Model performance in simulation data stream in different methods.

stream dataset. The comprehensive accuracy of the model in both the current and previous batched test datasets is recorded.

In this paper, we have employed several methods from literature [11,24,25,26] as benchmarks to conduct a comprehensive comparison between the proposed approach and existing techniques. The first two methods utilize Machine Learning (ML) models for HIF detection based on Support Vector Machine (SVM) and Decision Tree (DT), respectively, while the latter two employ threshold-based approaches, emphasizing signal analysis across time, frequency, and time–frequency domains. Additionally, in this paper, we have also applied a direct incremental learning mode to existing ML algorithms, which involves updating the ML model directly with new data. The results of the simulated data stream experiments are presented in Fig. 11.

Fig. 11, the detection accuracy of the proposed method remains consistently high for different batches of incremental data, significantly outperforming the other four methods in terms of sustaining performance. This illustrates the ability of the proposed method to adapt to new distribution data through the incremental learning framework, while existing methods gradually degrade in detection performance as new distribution data arrives, especially those relying on ML models. It is important to highlight that in this paper, we have fine-tuned the thresholds in [25] and [26] to achieve optimal performance. However, threshold-based approaches depend on the normal system operating state and tend to produce errors when confronted with unfamiliar data distributions. This is a common limitation of algorithms lacking continuous learning capabilities, and therefore, the method presented in this paper aims to assist ML models in achieving lifelong learning.

The proposed method exhibits strong detection performance in the initial three data batches, validating that the straightforward structure employed in this paper mitigates classifier model overfitting issues when dealing with limited data. However, a slight decline in detection accuracy is observed in the 3rd to 5th incremental data batches. Interestingly, this performance dip is followed by a recovery and improvement in performance after adjusting the network structure during the 5th batch of learning. This performance shift is attributed to the structural adaptation technique proposed for model structure adjustments. As the model's learning distribution widens and the dataset expands, the initial simplistic model proves inadequate for capturing the increasing complexity in data distribution. By expanding the model structure to accommodate this broader distribution range, the problem of decreased detection performance due to model underfitting is effectively circumvented.

Nevertheless, ML algorithms theoretically equipped with the ability for incremental learning when employing the direct incremental learning model yield subpar results. This is primarily attributed to the issue of "catastrophic forgetting," a challenge frequently encountered in the course of incremental learning. The proposed method effectively mitigates this problem through the utilization of data replay, thereby ensuring the stability of the ML model's detection performance throughout the learning process.

To investigate the mechanism of data replay technology in the above conclusions, this paper divides the data samples in Table 5 into three scenarios based on the different HIF models and Non-HIF events used. The composition of these scenarios is presented in Table 10, with numbers in parentheses indicating sample quantities. Each of the three scenario datasets is randomly partitioned into training and test sets at a ratio of 7:3. Evidently, there are significant variations in data distribution across different scenarios. The aim of this article is to magnify these distribution differences to observe the proposed method's learning process with new data and the review process with old data.

The flow of the scenario transformation experiments is shown below. The model is trained with the initial scenario training set to train the primitive classifier model in the proposed framework and direct learning mode, respectively, and then the training samples of incremental scenario one and incremental scenario two are learned, and the performance of the model in the three scenario test sets after each update is recorded. The results of the scenario transformation experiments are shown in Fig. 12.

As evident from Fig. 12(b), the ML algorithm utilizing the direct incremental learning model possesses the capability to learn new distribution data. However, it often forgets the characteristics of the old distribution data during the process of learning the new distribution data, resulting in a decline in the model's ability to recognize the data from the old distribution. In contrast, the proposed method constructs an example set that includes previously learned information about the old distribution and employs retrospective training of the model through data playback techniques. This approach ensures the model's effective recognition of the old scenario is maintained while learning the new scenario.

After learning the new distribution data, this portion of the data is transformed into the learned old distribution data through the example set updating technique. The proposed method retains its capability to

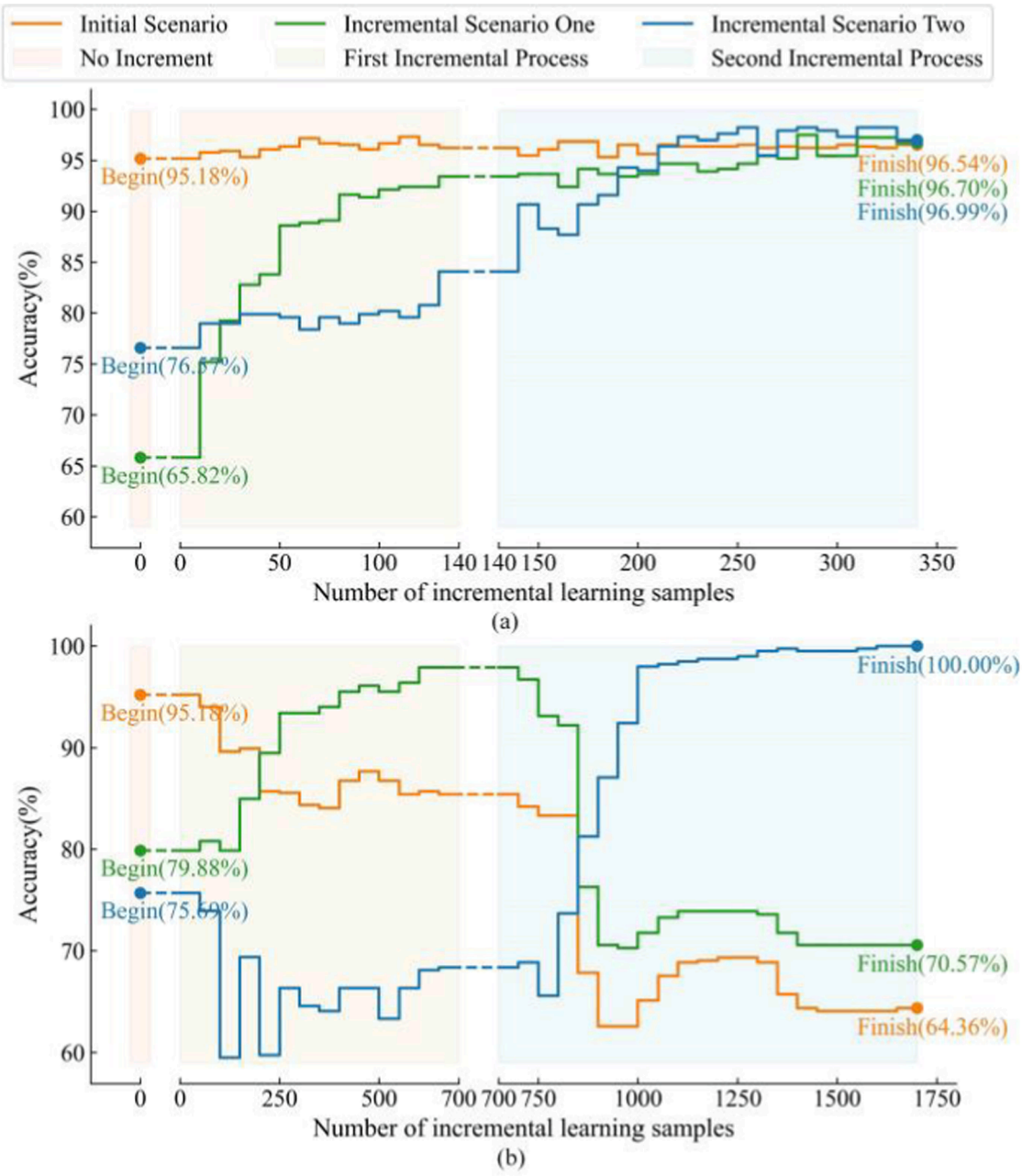


Fig. 12. Adaptation of models using different methods to adjust to simulation scenario changes.

Table 11
Real-world scenario composition.

	Initial Scenario	Incremental Scenario One	Incremental Scenario Two
Composition (samples)	Branch(50) Masonry(28) Resistance (106) Non-HIF_1(200)	Grassland (68) Non-HIF_2(146)	Practical HIF(178) Non-HIF_3(120)

detect the data from the first incremental scenario during the second incremental scenario learning (as shown by the latter part of the green dashed line in Fig. 12(a)). This demonstrates that the example set updating technique can integrate the acquired distributional features into the new example set, thereby enabling the preservation of newly acquired knowledge during the continuous learning process.

In the proposed method, samples from the data stream undergo data evaluation using data evaluation techniques. Only the data that deviates from the distribution observed in the exemplary set is subjected to further learning. In the scenario transformation experiments, the direct incremental learning model learns all the data samples, whereas the proposed method selectively learns 340 data samples, which represent 20 % of the total samples. As shown in Fig. 12(a), the model effectively captures the distribution characteristics of the specific scenario by learning the selected data, thus enabling accurate recognition of data within that scenario's distribution. This validates the capability of the proposed data evaluation technique to monitor distribution changes and identify new distribution data.

Combining the results of both experiments, it can be observed that within the proposed incremental learning framework, the BPNN is capable of learning new data through incremental learning even in scenarios where the initial training samples are incomplete.

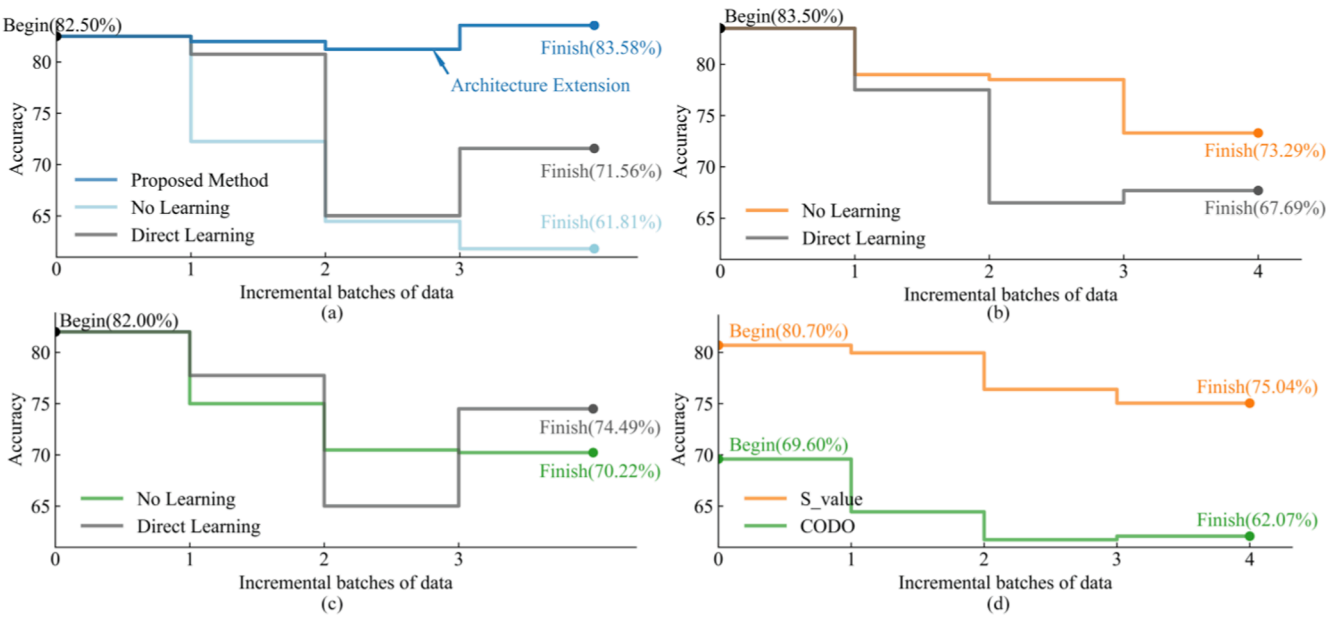


Fig. 13. Model performance in real-word data stream in different methods.

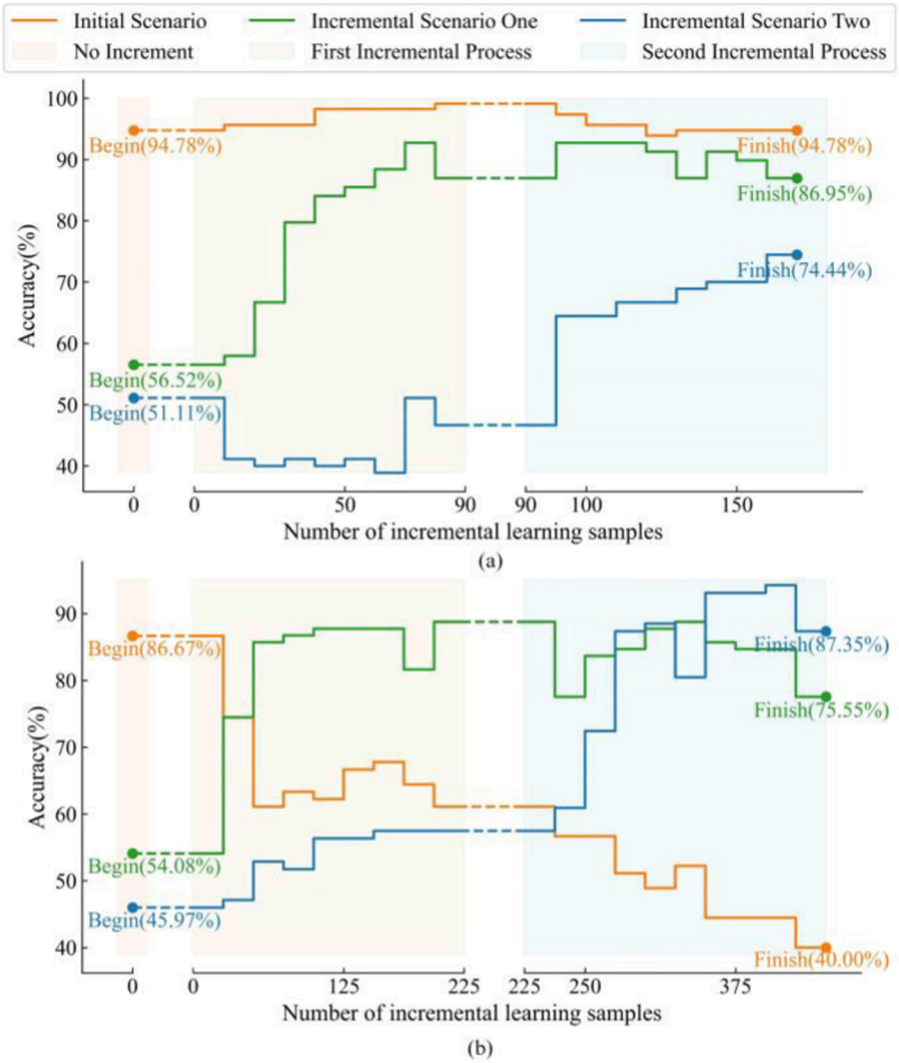


Fig. 14. Adaptation of models using different methods to adjust to real-world scenario changes.

Furthermore, it demonstrates a strong resistance to forgetting old data.

4.2. Adaptability analysis based on real-world data

In this section, 940 sets of existing real-world data have been chosen to assess the proposed method. Due to the scarcity of field data, this section randomly allocates the measured data samples into four groups for the simulated data stream test. Each group's data samples are then randomly divided into training, testing, and validation sets in a 7:2:1 ratio. In the scenario transformation test, these data are categorized into three different scenarios based on grounding medium and interference, as outlined in Table 11. These scenario-based data are further divided into real training and real testing data in a 7:3 ratio. The testing procedures for both sets of data are consistent with the previous section, and the results of these tests are presented in Fig. 13 and Fig. 14.

Considering the inherent time-varying and stochastic nature of real-world data, it significantly challenges the model's ability to recognize such data, resulting in a final recognition accuracy of only 83.58 % for the real-world data. However, when looking at the entire incremental learning process, it becomes evident that the stability-plasticity dilemma faced by the model is substantially mitigated within the proposed incremental learning framework. The model adeptly balances the interplay between new and old data, ensuring that recognition accuracy for field data remains stable under the proposed incremental learning framework.

In summary, when confronted with non-stationary data streams and scenarios characterized by significant variations in operating conditions, the model continues to demonstrate its incremental learning capabilities. It can acquire common knowledge from both old and new data, effectively mitigating forgetfulness regarding old data and assimilating new data from the waveform data streams. This highlights the promising engineering applications of the proposed method.

5. Conclusion

Traditional artificial intelligence approaches can achieve high accuracy in HIF detection when provided with complete training samples. However, in real-world scenarios, dynamic distribution networks often face continuously incoming data with new distributions, which are challenging to obtain as complete training samples during the early stages of training. In oeder to facilitate lifelong learning for the HIF detection model and consistently acquire the capability to detect new distributed data from non-stationary data streams, this paper presents an incremental learning-based framework for HIF detection in distribution grids. This framework employs a data evaluation mechanism, an incremental learning algorithm, and an exemplar dataset updating mechanism to handle real-time data streams. The effectiveness of this technique has been verified through both simulation and real-world data.

The results can be summarized as follows:

i) In non-stationary data streams with time-varying distributions, the proposed method, when compared to other traditional artificial intelligence approaches, exhibits the capability to learn new distributional data information, thereby enhancing its detection capabilities.
ii) The data evaluation mechanism allows for monitoring changes in data distribution within the data stream, facilitating the selection of data samples that require learning.
iii) The incremental learning algorithm based on data replay can effectively counteract the forgetting of previously acquired knowledge through retrospective training. Furthermore, the proposed exemplar dataset updating mechanism can incorporate learned distributional features into new exemplar dataset while significantly reducing data storage costs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Yuan J, Jiao Z. Faulty feeder detection for single phase-to-ground faults in distribution networks based on waveform encoding and waveform segmentation. IEEE Trans Smart Grid 2023:1.
[2] Zhang BL, Guo MF, Zheng ZY, Guo CH. A novel method for simultaneous power compensation and ground fault elimination in distribution networks, CSEE J Power Energy Syst, to be published. https://doi.org/10.17775/CSEEJPES.2022.03830.
[3] Deshmukh B, Lal DK, Biswal S. A reconstruction based adaptive fault detection scheme for distribution system containing AC microgrid. Int J Electrical Power Energy Syst 2023;147.
[4] Wang X, Liu W, Liang Z, et al. Faulty feeder detection based on the integrated inner product under high impedance fault for small resistance to ground systems. Int J Electr Power Energy Syst 2022;140:108078.
[5] Zhang BL, Guo MF, Zheng ZY, Hong Q. Fault current limitation with energy recovery based on power electronics in hybrid AC–DC active distribution networks. IEEE Trans Power Electron 2023;38(10):12593–606.
[6] Lopes GN, Menezes TS, Santos GG, et al. High Impedance Fault detection based on harmonic energy variation via S-transform. Int J Electr Power Energy Syst 2022; 136.
[7] Yuan J, Wu T, Hu Y, et al. Faulty feeder detection based on image recognition of voltage-current waveforms in non-effectively grounded distribution networks. Int J Electr Power Energy Syst 2022;143:108434.
[8] Gao J, Wang X, Wang X, et al. A high-impedance fault detection method for distribution systems based on empirical wavelet transform and differential faulty energy. IEEE Trans Smart Grid 2022;13(2):900–12.
[9] Gao J-H, Guo MF, Lin S, et al. Application of semantic segmentation in High-Impedance fault diagnosis combined signal envelope and Hilbert marginal spectrum for resonant distribution networks. Expert Syst Appl 2023;231:120631.
[10] Guo MF, Liu WL, Gao JH, et al. A data-enhanced high impedance fault detection method under imbalanced sample scenarios in distribution networks. IEEE Trans Ind Appl 2023:1–14.
[11] Lala H, Karmakar S. Detection and experimental validation of high impedance arc fault in distribution system using empirical mode decomposition. IEEE Syst J 2020; 14(3):3494–505.
[12] Xiao Q-M, Guo M-F, Chen D-Y. High-Impedance fault detection method based on one-dimensional variational prototyping-encoder for distribution networks. IEEE Syst J 2022;16(1):966–76.
[13] Gomes DPS, Ozansoy C, Ulhaq A. Vegetation high-impedance faults' high-frequency signatures via sparse coding. IEEE Trans Instrum Meas 2020;69(7): 5233–42.
[14] Van De Ven GM, Tuytelaars T, Tolias AS. Three types of incremental learning. Nat Mach Intell 2022;4(12):1185–97.
[15] Leite D, Costa P, Gomide F. Evolving granular neural network for fuzzy time series forecasting. In: The 2012 International Joint Conference on Neural Networks (IJCNN), 2012. p. 1-8.
[16] Angelov P, Filev DP, Kasabov N. Evolving intelligent systems: methodology and applications. John Wiley & Sons; 2010.
[17] Leite D, Costa P, Gomide F. Evolving granular neural network for semi-supervised data stream classification. In: The 2010 International Joint Conference on Neural Networks (IJCNN), 2010. p. 1-8.
[18] Costa FB, Souza BA, Brito NSD, et al. Real-Time detection of transients induced by high-impedance faults based on the boundary wavelet transform. IEEE Trans Ind Appl 2015;51(6):5312–23.
[19] Costa FB. Fault-Induced transient detection based on real-time analysis of the wavelet coefficient energy. IEEE Trans Power Delivery 2014;29(1):140–53.
[20] Yusuff AA, Jimoh AA, Munda JL. Fault location in transmission lines based on stationary wavelet transform, determinant function feature and support vector regression. Electr Pow Syst Res 2014;110:73–83.
[21] Gu S, Qiao J, Shi W, et al. Multi-task transient stability assessment of power system based on graph neural network with interpretable attribution analysis. Energy Rep 2023;9:930–42.
[22] Xu T, Zou P, Xu T, et al. Study on weight function of meshless method based on B-spline wavelet function. In: 3rd International Joint Conference on Computational Sciences and Optimization, CSO 2010: Theoretical Development and Engineering Practice, May 28, 2010 - May 31, 2010; 2010. p. 36-40.
[23] Robins A. Catastrophic forgetting in neural networks: the role of rehearsal mechanisms. In: Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, 1993. p. 65-68.

[24] Kar S, Samantaray SR. High impedance fault detection in microgrid using maximal overlapping discrete wavelet transform and decision tree. In: International Conference on Electrical Power & Energy Systems; 2016.
[25] Gautam Brahma S. Detection of high impedance fault in power distribution systems using mathematical morphology. IEEE Trans Power Syst 2013;28(2):1226–34.
[26] Wang X, Wei X, Gao J, et al. High-Impedance fault detection method based on stochastic resonance for a distribution network with strong background noise. IEEE Trans Power Delivery 2022;37(2):1004–16.