

Transform Waveforms into Signature Vectors for General-purpose Incipient Fault Detection

Yang Weng, *Senior Member, IEEE*, Qiushi Cui, *Member, IEEE*, and Muhao Guo, *Student Member, IEEE*

Abstract—Power system equipment presents special signatures at the incipient stage of faults. As more renewables are integrated into the systems, these signatures are harder to detect. If faults are detected at an early stage, economical losses and power outages can be avoided in modern power grids. Many researchers and power engineers have proposed a series of signature-specific methods for one type of equipment's waveform abnormality. However, conventional methods are not designed to identify multiple types of incipient faults (IFs) signatures at the same time. Therefore, we develop a general-purpose IF detection method that detects waveform abnormality stemming from multiple types of devices. To avoid the computational burden of the general-purpose IF detection method, we embed the abnormality signatures into a vector and develop a pre-training model (PTM) for machine understanding. In the PTM, signal “words,” “sentences,” and “dictionaries” are designed and proposed. Through the comparison with a machine learning classifier and a simple probabilistic language model, the results show a superior detection performance and reveal that the training radius is highly related to the size of abnormal waveforms.

Index Terms—Incipient fault, distribution system protection, representation learning, pre-training model.

I. INTRODUCTION

FAULT analysis and prediction are with significant importance to distribution network operation and protection. Establishing effective fault analysis and prediction theory can help distribution network operators monitor distribution network systems and equipment health status. The advanced monitoring is useful for effective overhaul and maintenance decisions, which reduces the risk of large-scale power outages. Many factors affect distribution network fault characteristics, such as equipment type, operating conditions, and noise, etc. Therefore, the research on fault analysis and prediction methods [1]–[4] has been extensively studied.

Distribution network fault analysis needs to detect anomalies in voltage and current waveforms [5]. Once anomalies are detected, the waveforms and root-mean-square values associated with the anomalous cycles can be extracted for detailed analysis. Ultimately, the state of the device can be determined from the results. There are a wide variety of equipment fault characteristics, many of which are not yet well understood. Furthermore, it is hard to find a general method that incorporates the equipment signature together. Therefore, if a general method that can detect all types of equipment anomalies is established, it will be helpful in understanding equipment anomalies coherently and comparably. However,

most of the present research proposes one algorithm for a specific device [6]–[9]. For example, various fault detection algorithms are developed specifically for cables [6], [9], transformers [1], [10], [11], and lines [7].

In the fault detection field, IF detection is beneficial to distribution network operations due to its preventive capability. Prior to an equipment failure, there are repeated predictive anomaly signals [12]–[15]. Effective IF detection helps avoid catastrophic failures stemming from different devices. Faulty equipment can be replaced in advance to effectively improve power supply reliability. In addition, it transforms reactive processing into predictive maintenance, which greatly improves the traditional thinking pattern of power protection [16]. According to the IEEE Power and Energy Society report [5], there are five classes of methods for detecting waveform anomalies [5], [8], [17]–[22]: current-based methods, voltage-based methods, methods based on integrated current and voltage, methods based on hypothesis testing, and the interpretive anomaly detection method. Characterization of the IF waveform generally exists in the time domain, frequency domain, and time-frequency domain. In the time domain, the magnitude of the fault current and the fault duration are usually recorded. In the frequency domain, the harmonic components associated with the fault are generally monitored, and the total harmonic distortion of the voltage at the fault point is used as a criterion, which should exceed the threshold value when the initial fault occurs. In the time-frequency domain, the transient behavior of the fault is generally analyzed using wavelet transform; then the initial fault is classified by certain detection rules. In the work of [23], a simple algorithm based on five main characteristics of voltage and current waveforms in early fault conditions is proposed. To identify the system parameters and characterize the observed initial behavior, field data recorded from underground distribution feeders were evaluated in the literature [15], which is based on time and frequency domain analysis. In the literature [24], the authors proposed a pattern analysis technique for classifying load variation transients and early anomalies in underground distribution cables using a KNN classifier. The scheme proposed in the literature [23] detects self-clearing transient faults based on the magnitude and rate of change of a neutral current. An intelligent system-based approach is proposed in the literature [25]. In the frequency domain-based approach introduced in the literature [26], the S-transform takes into account the harmonic components of the arc current and/or voltage. In the literature [27], a rule-based and SVM-based pattern classifier is used to classify the transient patterns of underground cables.

Conventional IF detection methods are not designed to

Y. Weng and M. Guo are with the Department of Electrical and Computer Engineering, Arizona State University, Tempe, AZ, 85281 USA. Q. Cui is with the State Key Laboratory of Power Equipment & System and New Tech., the School of Electrical Engineering, Chongqing University, China. Corresponding author: Q. Cui. Email: qiushicui@gmail.com.

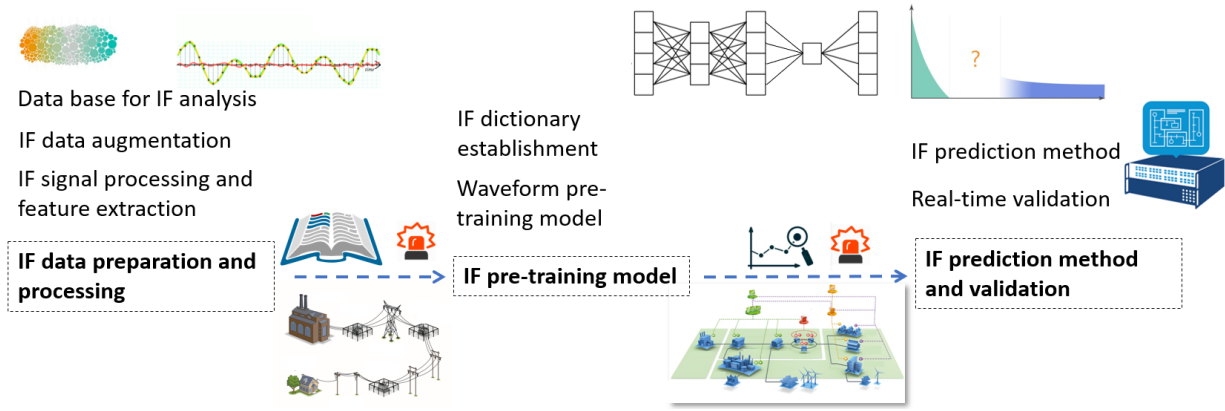


Figure 1. The overview of the proposed method.

comprehend the waveform signatures of different IF faults in a coherent way, therefore, they usually focus on the signature detection of one equipment. However, incipient faults are hard to detect since different equipment exhibits different behaviors in various operating conditions. There were efforts in the past to combine the nonlinear systems theory with language theory [28], but the natural language processing technique was not very successful before 2006 [29]. In this paper, we formulate signature vectors and their dictionary through the embedding of the physical characteristics of the power grid. This enables the machine's understanding of distribution network fault analysis and prediction. Then, we construct signature vectors and their dictionaries to help build a general quantitative model of faults and disturbances as well as a probabilistic analysis model, which provides theoretical basis, methods, and analysis tools for power system IF detection. In the end, a signature vector pre-training model is introduced into the field of fault analysis and prediction, which provides a new solution for initial fault detection. The idea of introducing dictionaries and words may not be completely new, but the new philosophy that is evolving brings significant new ideas and applicability to incipient faults. For example, the dictionary coupled with the automatic feature extraction and construction capability of the dictionary is brand new and necessary for cyber-physical systems such as power systems with incipient faults. Inspired by the momentum of new research ideas in the field of natural language processing, this paper not only transforms the way word2vec is used, but also shows how to embed physical knowledge from power protection to improve the knowledge database. This leads to constructing fault models by making full use of fault databases and signal processing techniques. Technically, we use Pytorch to implement our own word2vec algorithm with libraries like Sklearn, Nltk, NumPy, etc. Then, we improve the proposed algorithm through numerical testing and redesign it by introducing new concepts and functions that traditional word2vec that does not have.

This paper is organized as follows: Section II describes the distribution system anomaly characteristics and anomaly dictionary concept. Section III illustrates the formation of the distribution network pre-training model for IF detection. Section IV shows the numerical results, followed by Section V the conclusions.

II. DISTRIBUTION SYSTEM ANOMALY CHARACTERISTICS AND WAVEFORM DICTIONARY

Features extracted from time/frequency domains vary when used to detect faults of multiple types of equipment. It could be a mathematical burden when dealing with multidimensional data, especially when different types of power systems equipment are involved. To address this issue, we develop a distributed representation of the waveforms in a low-dimension vector. There is no specific meaning in this vector, but it contains rich information when viewed as one vector. The vector can be realized through a designed IF pre-training model (IF-PTM). The proposed IF-PTM has three advantages:

- The extensive training over large-volume power measurement data ensures the generalized representation of the power grid anomalies. This is helpful for the subsequent prediction task.
- Has a ready-to-use IF detection model with excellent parameter initialization. This enhances the IF-PTM generality and convergence speed.
- Provides a way of normalization when viewed on a small dataset, avoiding model overfitting.

This paper proposes to construct a signature vector pre-training theoretical system, as shown in Fig. 1. It is inspired by research results in the field of natural language processing. The paper is carried out by involving cross-application and the innovation of knowledge in computer science and electric power fields, to construct fault models by making full use of fault databases and signal processing techniques. At the same time, a fault word embedding model focusing on learning context waveform is used to obtain advanced machine understanding of distribution network faults. Finally, fault analysis, identification and prediction theories, and techniques are developed based on the aforementioned research procedures.

A. Build a fault database.

The training and testing of the proposed PTM requires a high volume database to become effective. The database to be studied includes three parts: the realistic data from open resources, the simulation data that is complementary to the realistic data, and the artificial data that is based on realistic data. First, we extract data from technical reports, top journal articles, and shared data. The data is in image format at the beginning, shown as the original signal in Fig.

2. We then digitalize the waveform using specialized software. Furthermore, we add Gaussian noises to mimic more realistic scenarios, which expands the datasets.

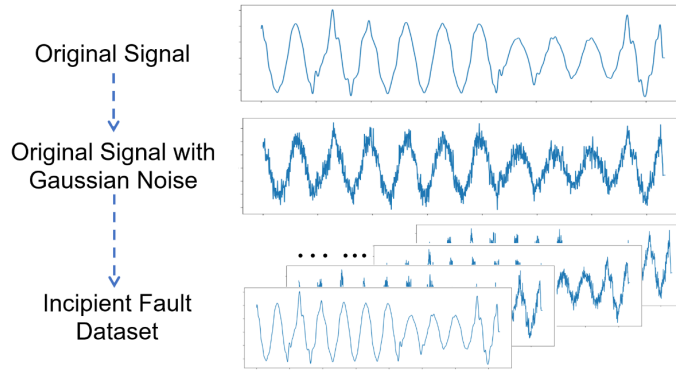


Figure 2. Build the incipient fault dataset.

Second, we use the distribution network edge equipment (including power quality meter, relay protection equipment, PMU, etc.) as data collection devices and intelligent computing terminals. First, we simulate various system operating conditions. At the system level, we change the system topology, line parameters, load type, normal operating conditions, etc. to obtain fault data of different equipment and fault data from early to late-stage. At the equipment level, we change the equipment model, operation mode, and capacity size to obtain normal operation and fault operation data from equipment. Based on this, the relationship between operating conditions and faults is used to populate the database. Fig. 3 shows some of the disturbance data. It includes permanent fault data since they are disturbances to the IF identification tasks.

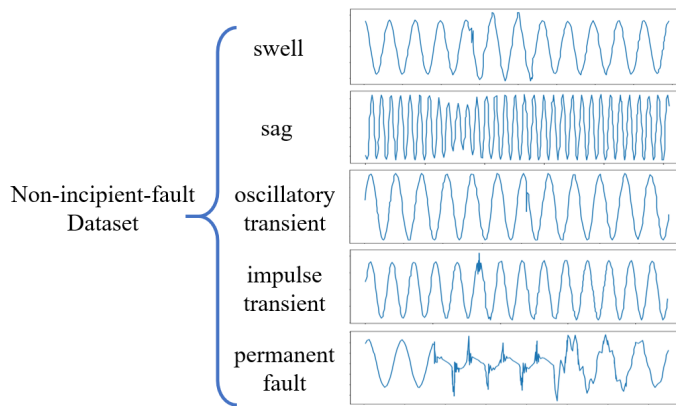


Figure 3. Build a non-incipient-fault dataset.

Third, we generate artificial data based on the realistic data we obtain. We construct it by assigning different amplitudes and simply repeating the initial fault waveform two or three times (see figure 4). The purpose is to simulate the IF events that possess repetitive characteristics, instead of looking at only one IF event in the literature.

B. Waveform processing and signal sentences.

To convert the waveform into digital values, the link between signal analysis and processing is essential. Signal features are extracted by Fast Fourier Transform (FFT), which

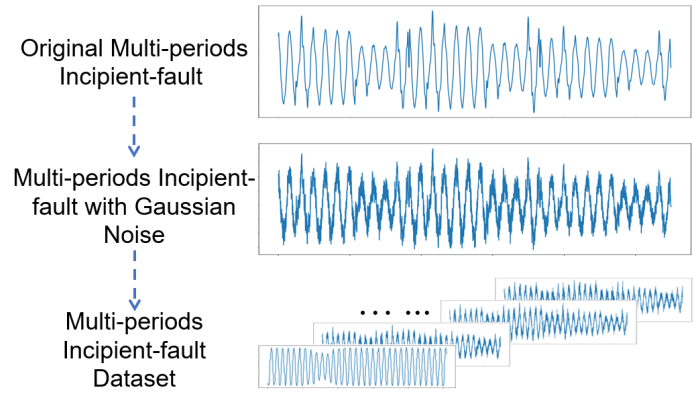


Figure 4. Build multi-periods IF datasets.

is simple, reliable, and can be implemented in many engineering fields. FFT is used to capture most physical quantities and is widely implemented in microcomputer-based relays. Since the open-source data is obtained at varying sample frequencies, we re-sample it according to the sample rate of power system protection or monitoring devices once we get a signal waveform. After the re-sample, we split the entire signal waveform by cycle. Then we apply the FFT to each cycle and acquire the amplitude single-sided spectrogram as shown in Fig. 5.

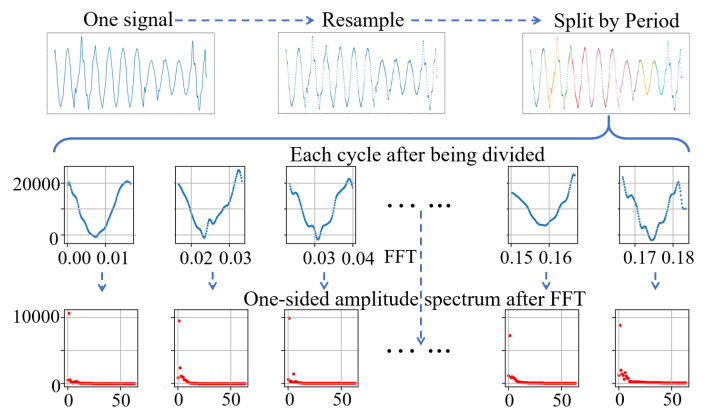
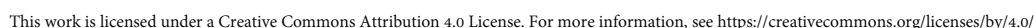


Figure 5. Signal processing and feature extraction.

Based on the selected features above and our designed *Value-Letter Table*, we translate the cycle information into “letters” and then combine them into one “word.” The example in Fig. 6 shows the continuous values are discretized into multiple intervals. The first 10 intervals are equally split and the last interval has a larger range since the contributions from the high order harmonics are subtle. These intervals can be designed according to specific cases. Once we obtain the “words” for signals, many cycles of waveforms form a “signal sentence” in the same order as the cycles in signal waveforms.

In particular, when the effective features are determined after much data analysis and research, an “alphabet” of signature vectors is obtained. Unlike English words, each fault waveform unit ($segment, s_i$) corresponds to a fixed alphabet $[F_1, F_2, \dots, F_n]$, in which the numerical difference of each feature represents a different waveform unit s_i . For example, taking $n = 3$, F_1 , F_2 and F_3 represent the FFT transformed direct current (DC) component A_0 , fundamental amplitude



“word”) in the dictionary corresponds to $1 \sim L$ one by one. Given the total length T of the waveform dictionary provided in the waveform recorder and the contextual waveform window size of m , the probability of generating a target waveform unit, given any surrounding waveform unit, is required to maximize the function $J = \prod_{t=m}^T p(s_t | s_{t-m}, \dots, s_{t-1}; \theta)$, which is equivalent to

$$\min_{\theta} J = \min_{\theta} \left(-\frac{1}{T-m} \sum_{t=m}^T \log p(s_t | s_{t-m}, \dots, s_{t-1}; \theta) \right), \quad (1)$$

where the conditional probability $p(s_t | s_{t-m}, \dots, s_{t-1}) = \text{softmax}(g_j^T q_i) = \frac{\exp(g_j^T q_i)}{\sum_{k \in L} \exp(g_k^T q_i)}$; q_i is a vector when the faulty unit is the target unit, whose vector dimension is $d \times 1$. g_j is a vector when the faulty unit is the prediction unit, whose vector dimension is $d \times 1$. Softmax is introduced so that it can convert the output value of a multi-classification into a probability distribution in the range $[0, 1]$ and the sum is 1. The disadvantage of softmax is its high computational cost. The complexity of the softmax equation is $O(V + V) \approx O(V)$. This is computationally very expensive, as the signal vocabulary and their weights can reach millions or more as previously discussed. To solve this problem, we utilize the negative sampling technique [30]. This technique reformulates the problem into a set of independent binary classification tasks of algorithm complexity $= O(K + 1)$, where K typically has a range below 20. When implemented in a real distribution network system, the computational burden can be approximated in such a way.

Eq. (1) is based on a simple idea: minimizing the loss function so that the probability of predicting the target waveform using the contextual waveform is maximized. However, in power systems, there are much more normal waveforms than abnormal waveforms. Since it is not necessary to learn those normal waveforms all the time, we introduce the concept of “jumping level” to control the probability in (1). The “jumping level” is defined as $JL(s_i, s_{i+1})$, which means the “waveform difference” between two consecutive signal waveforms reflected by their signal words, as shown in Fig. 7.

It is assumed that the “jumping level” from letter “a” to “b” or from “d” to “e” is 1, and that from letter “a” to “c” or from “d” to “f” is 2. This assumption does not depend on where the letter is located in the “word.” For example, $JL('aad', 'aab') = 1$, $JL('aad', 'acd') = 2$, and $JL('aad', 'abc') = 3$. After defining the “jumping level,” we show the proposed loss function as follows:

$$J = -\frac{1}{T-m} \sum_{t=m}^T \log \lambda(s_{t-m}, \dots, s_t) p(s_t | s_{t-m}, \dots, s_{t-1}; \theta), \quad (2)$$

where $\lambda(s_{t-m}, \dots, s_t) = 1$, when $\frac{1}{m} \sum_{i=t-m}^{t-1} JL(s_i, s_{i+1}) < n$; $\lambda(s_{t-m}, \dots, s_t) = 10^{-6}$, when $\frac{1}{m} \sum_{i=t-m}^{t-1} JL(s_i, s_{i+1}) \geq n$. n is an empirical parameter that indicates the “level” of signal word change.

In the above training framework, the high dimensionality of the signature vector is represented by the one-hot and can cause a dimensionality catastrophe when solving certain tasks (e.g., building fault synthesis models). Thus, we use the signature vectors generated by the distributed representation. This is usually multidimensional and each dimension represents its feature. The distance between two signature vectors is the Euclidean distance between them.

B. Constructing pre-training models for before and after waveform correlation.

The trained signature vectors of a single waveform pre-training model can also capture the meaning of simple waveforms. Although these pre-trained signature vectors can also capture these types of faults implied by the waveforms, they are not constrained by context and simply learn the “co-occurrence probability.” Such an approach is clearly unable to understand higher-level waveform information, such as initial faults, grid disturbances, colored noise, etc. Therefore, this project will also focus on designing signature vector embedding algorithms that learn the context. In the field of natural language processing, word embedding algorithms, such as CoVe, ELMo, OpenAI GPT, and BERT, which learn more reasonable word representations that encapsulate contextual information about words, can be used for subsequent tasks such as question and answer systems, machine translation, etc.

C. The idea behind general-purpose IF detection

For each type of IF, it is one “word” in the created “dictionary”. In the dictionary, each signal word is treated equally. This is the main difference between the proposed general-purpose IF-PTM method and other special-purpose methods. We can imagine that each signal vector is one point in the high dimensional space. A specific type of IF will formulate a high-dimensional shape, which can be a high-dimensional sphere or most probably anything else. The distance between every two shapes indicates the similarity of two IFs. The shorter the distance is, the more similar two IFs are to each other. It is pointless to visualize individual IF type, but we can infer the comparative operating space of devices using signature vectors in the three dimension rectangular coordinate.

IV. NUMERICAL RESULTS

This section first focuses on the demonstration of general-purpose detection among multiple types of equipment. Then, the prediction advantages of the proposed method are compared with other methods under two datasets. Next, the proposed method is tested under complicated fault conditions based on one dataset. Last, the effect of vector size and loss function selection of IF-PTM test scores are presented.

As previously mentioned, the detection alphabet depends on the IF tasks and fault types. The waveforms in [5] are collected and studied. After harmonic analysis, we utilize only the first seven FFT amplitudes as the selected features, including the DC offset, and the amplitudes from the fundamental frequency to the sixth order harmonics. Since we realize 7 amplitudes are good enough for the incipient fault detection problem based on the waveform data in [5]. Based on our observation, the magnitude of the high frequency that is higher than the 7th order is low. Therefore, we choose the first seven orders of harmonics in the numerical example. It is necessary to emphasize and not to mislead the readers that the selection of the “alphabet” of signature vectors is flexible based on the incipient fault types. When the IF under study contains, for example, high frequency components, it is necessary to include related signal “alphabet” for better performance.

The proposed IF-PTM method edges towards a generic solution of machine understanding of all of the IFs through learning the meaning in the signal “corpus”, which is the measured signals over a long time. Therefore, this idea is different from the past solutions that focus on one application for one device. What this method provides is the machine understanding of the waveform abnormality. This is hard to visualize since the output is just the signal vectors of abnormal waveforms.

A. General-purpose detection with equipment IF signatures embedded.

In this subsection, we first study IFs of three types of equipment, including cables (IF on one phase of a 27 kV underground feeder), lines (tree branch to burn and fall to the ground), and transformers (load tap changer failure). Then, we study 16 types of IFs to show the general purpose nature of the proposed method.

The data source of the first study comes from [5] that provides rich information on a variety of the IF scenarios. With the IF datasets, we train a PTM for the detection of the three types of power equipment IF. One important output of the IF-PTM model is the signal signature vector, therefore, we can easily get the waveform dictionary based on the historical waveform data. Consequently, with a focus on three important pieces of equipment in power systems, we visualize the learned faulty waveform of lines, transformers, and cables from the waveform dictionary, as shown in Fig. 8. It is noteworthy that Fig. 8 is not for classification purposes, but for visualization only. In the visualization below, we choose the best training window size (9 cycles in this case) based on the performance evaluation index (to be discussed in the following subsections). Each signature vector has 20 dimensions in our IF-PTM, so the representative waveform has 180 dimensions in this example. Different equipment fault signature vectors are differentiable, as seen by the computer but not human beings. Therefore, we visualize them in Fig. 8. We plotted three self-contained application cases in 3-D in Fig. 8. We can see that the IF of lines is relatively far from cables and transformers. Interestingly, cables and transformers are close to each other and are distributed in a larger space. It indicates IFs associated with cables and transformers have a larger operating space compared to lines.

The data source of the second study also comes from [5]. The main advantage of the proposed method is its general-purpose nature, therefore, examples with additional failure types are provided for the benefit of the reader. We included 16 types of IF according to the field data in [5]. They are summarized in Table I, including some typical and common IFs like failures of cable joints, failures of tap changers, etc. To further demonstrate the efficacy of the proposed method, we plot the signal words in Fig. 9. It is hard to show the signal word separability in high dimensional space, therefore, in this figure, we visualize each signal vector against the other vectors in the upper triangle. For example, after dimension reduction, the sub-figure at row 1 and column 2 is the visualization of the first and second fault types that are corresponding to Table I. For the lower triangle, waveforms are associated

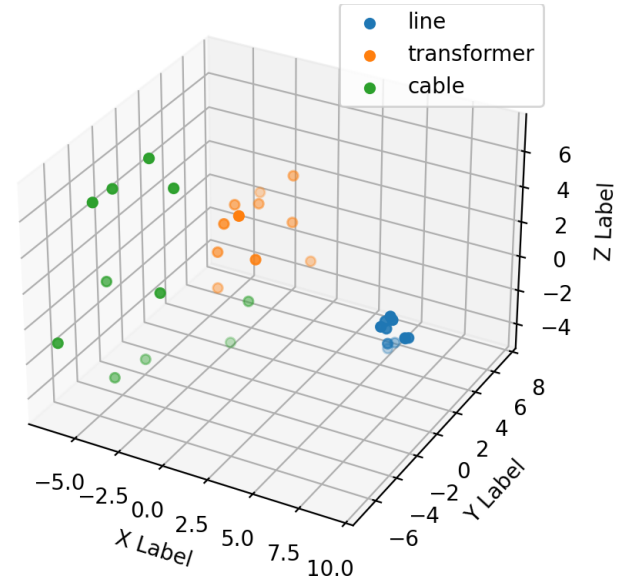


Figure 8. Visualization of three devices' signature vectors. Data source: Fig. 10 (cable), Fig. 22 (line), and Fig. 28 (transformer) in [5].

with their counterparts in the upper triangle to demonstrate their difference in the time domain. Additionally, we randomly select four sub-figures from the upper triangle in Fig. 10 for the readers to have a close-up view of the IF separability. The sub-figures are first delimited by the method of support vector classifier with Gaussian kernel and then plotted with contour lines that are the boundaries of the red and blue dots. A darker colored area means this area is further from the classification boundary.

Table I
INCIPIENT FAULT TYPES IN [5]. THE FIGURE NUMBER AT THE END OF EACH IF TYPE REFERS TO THE FIGURE IN [5].

Devices	IF types
Cables	1. Failures of Cables (Fig.4(a))
	2. Failures of Cable Joints (Fig.13)
	3. Experimental Results on Cable Insulation Failure (Fig.20)
Overhead Lines	4. Failure of Overhead Lines (Fig.22)
Transformers	5. Failures of Tap Changers (Fig.28)
	6. Failures of Transformer Bushing(Fig.36)
Switches	7. Failure of Line Switch (Fig.42)
	8. Arcing of Capacitor Bank Switch (Fig.43)
	9. Restrikes during Capacitor De-energization (Fig.46)
Capacitors	10. Capacitor Failure Caused by Misoperation of Controller (Fig.51)
	11. Unsuccessful Synchronous Closing Control (Fig.53)
	12. Capacitor Energization Triggering Resonance (Fig.54)
	13. Other types of failures in a capacitor bank (Fig.58)
Lightning and Surge Arresters	14. Failure of Lightning Arresters (Fig.60)
	15. Failure of Surge Arresters (Fig.61)
Potential Transformers	16. Failure of Potential Transformers (Fig.62)

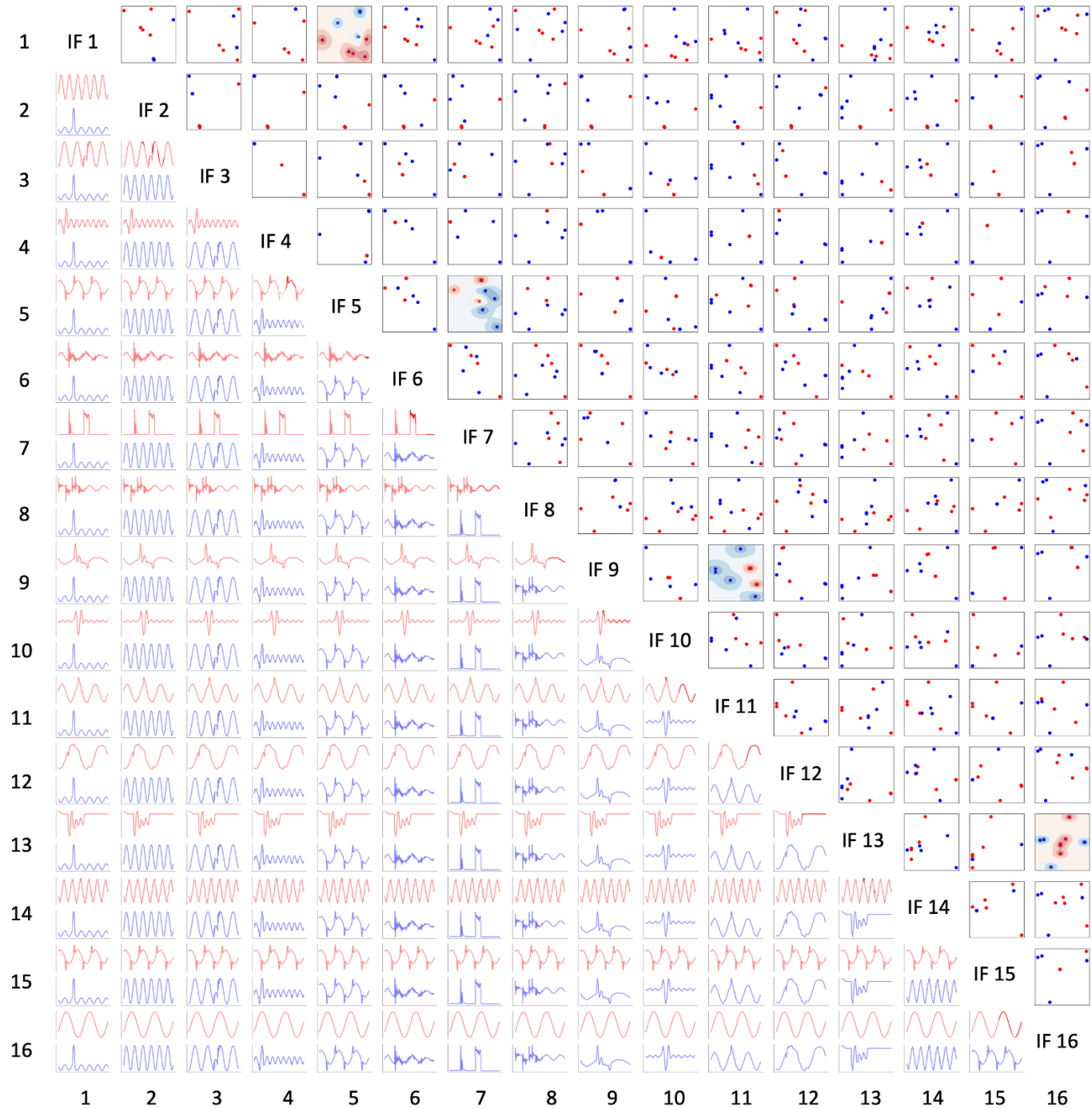


Figure 9. Visualization of the general-purpose IF detection method with 16 types of IF in 2-D space. The lower triangle is in soft white shadow to balance the contrast with the upper triangle. Each element indicates the separability of its row and column element number related to the IF type number in Table I.

B. The prediction advantages over machine learning and simple natural language processing methods.

To show the comparison performance, two datasets are created. Dataset A includes two parts. The first part consists of the initial failure waveforms in Fig. 10 of [5], which illustrate the occurrence of an IF on phase-A of a 27 kV underground feeder, and 10 waveforms with random Gaussian noise with an Signal-To-Noise Ratio (SNR) of 8 added to this original waveform. The second part comes from known disturbance waveforms that include swell, sag, oscillatory transient waveform, impulse transient waveform, and permanent

fault waveform. Dataset B is similar to dataset A. The only difference is the addition of a third part, i.e., the incipient fault waveform from the first part is repeated multiple times. For example, two waveforms with amplitudes of 1.1 and 0.8 times the amplitude of the original waveform are inserted into the original waveforms respectively. The purpose is to take into account the repetitive nature of the incipient fault waveform of signals at a longer time horizon.

The proposed PTM is compared with other methods at two different angles. The first angle is a simple machine learning classifier – logistic regression. The logistic regression method

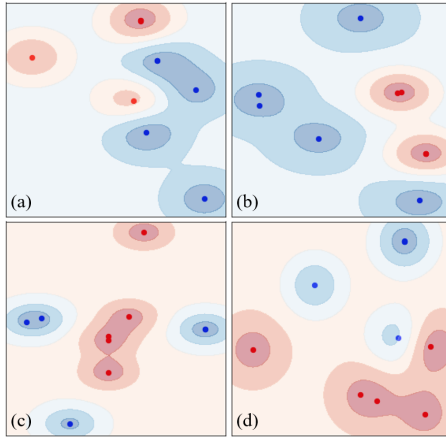


Figure 10. Closeup view of the general-purpose IF separability. Sub-figures (a), (b), (c), (d) presents the classification between IF 5 and 7, IF 9 and 11, IF 13 and 16, IF 1 and 5 related to the IF type number in Table I.

utilizes a sliding window to classify IF data. The second angle is a comparison with a simple model that assigns probabilities to waveforms and sequences of waveform units. Each method was trained using our randomly sliced training set and tested in the corresponding test set. Ten experiments were conducted for each method, and the average of the ten experiments was used as our final result.

We use different methods to predict whether a permanent failure will occur. If a permanent failure occurs after a known waveform, we record the predicted true value as 1 (a positive sample), otherwise as 0 (a negative sample). Next, we compare the proposed PTM with Logistic Regression and N-gram methods. Based on data A, we compare the precision, recall, and F1 scores of different methods for positive samples, i.e., with permanent faults, as shown in Fig. 11. We have chosen ten different sizes of windows from 1 to 10. In the IF-PTM, we specify the length of the signature vector to be 20. We note that in dataset A, N-gram is unable to predict whether a permanent failure will occur, as indicated by (*Recall_1, F1_1, Precision_1*). This is because dataset A is small and the N-gram algorithm works poorly when the data

volume is not sufficient. Meanwhile, the proposed IF-PTM has a better performance than Logistic Regression. Besides, the IF-PTM method and N-gram method each have advantages and disadvantages in most of the evaluation scenarios, except for the overall accuracy with a window size of less than 8. However, selection of the IF-PTM window size is flexible and we can choose the window size of 9 in our experiments.

Then, we tabulate the accuracy, dependability, security, F1 score, etc. in Table II and Table III. In these tables, 1 represents the positive samples and 0 represents the negative samples. Bold font indicates the highest score on a specific evaluation. The value in brackets indicates the size of the window. Since the F1 score of positive samples is our primary concern, we have identified it in red.

From both tables, it is noticed that N-Gram and IF-PTM in general perform better than Logistic Regression. Moreover, N-Gram performs no worse than IF-PTM. However, the N-Gram method is not performing stably especially in the positive samples under the index of Precision, Recall, and F1 score. In sum, the proposed IF-PTM has the best performance among the three methods.

C. High IF Detection Performance Under Complicated Fault Conditions.

The IF-PTM performance is further tested under complicated fault conditions with the dataset of B. In this dataset, the IF scenarios demonstrate a long and varying complexity, which imposes difficulty on the IF detection task. The results are illustrated in Fig. 12. In dataset B, N-Gram performs better comparing with its performance in dataset A. Additionally, N-Gram is as good as the proposed IF-PTM in this dataset. For example, the precision, recall, and F1 score of the N-Gram method are almost equally good as the IF-PTM method in negative sample tests; while both methods have their pros and cons in positive sample tests. However, Logistic Regression is not as good as the other two methods. By evaluating both Fig. 11 and Fig. 12, the IF-PTM exhibits a stable performance in different datasets. Table IV and V present the performance comparison of Logistic Regression, N-Gram, and IF-PTM.

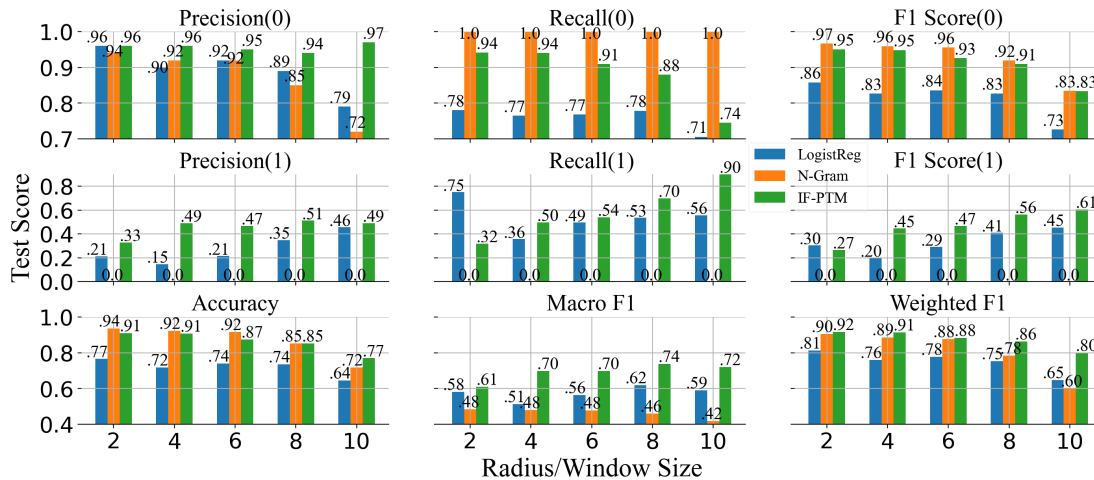


Figure 11. Nine evaluation criteria for different methods under different window/radius values in dataset A. Namely, negative samples precision, negative samples recall, negative samples F1 score, positive samples precision, positive samples recall, positive samples F1 score, overall accuracy, overall macro average F1 score and overall weighted average F1 score. In the IF-PTM, we specified the length of the signature vector to be 20.

Table II
THE ACCURACY, MACRO AVERAGE F1 SCORE AND WEIGHTED AVERAGE F1 SCORE FOR ALL SAMPLES BASED ON DATASET A.

Window Size		1	2	3	4	5	6	7	8	9	10
Methods	accuracy	0.7444	0.7655	0.7717	0.7167	0.7116	0.7395	0.7030	0.7357	0.6696	0.6444
	macro average F1 score	0.5868	0.5806	0.5476	0.5126	0.5450	0.5628	0.5569	0.6182	0.5985	0.5903
	weighted average F1 score	0.7984	0.8125	0.8046	0.7586	0.7585	0.7772	0.7345	0.7539	0.6824	0.6479
N-Gram	accuracy	0.9297	0.9356	0.9358	0.9229	0.9209	0.9162	0.8812	0.8519	0.7955	0.7176
	macro average F1 score	0.4817	0.4832	0.4834	0.4799	0.4791	0.4778	0.4683	0.4596	0.4424	0.4171
	weighted average F1 score	0.9050	0.9047	0.9050	0.8861	0.8836	0.8768	0.8260	0.7844	0.7062	0.6012
IF-PTM	accuracy	0.8484	0.9085	0.8962	0.9062	0.8698	0.8730	0.8562	0.8519	0.8318	0.7706
	macro average F1 score	0.4649	0.6085	0.6200	0.6970	0.6351	0.6968	0.7140	0.7365	0.7784	0.7200
	weighted average F1 score	0.8648	0.9168	0.8929	0.9133	0.8688	0.8823	0.8655	0.8633	0.8392	0.7982

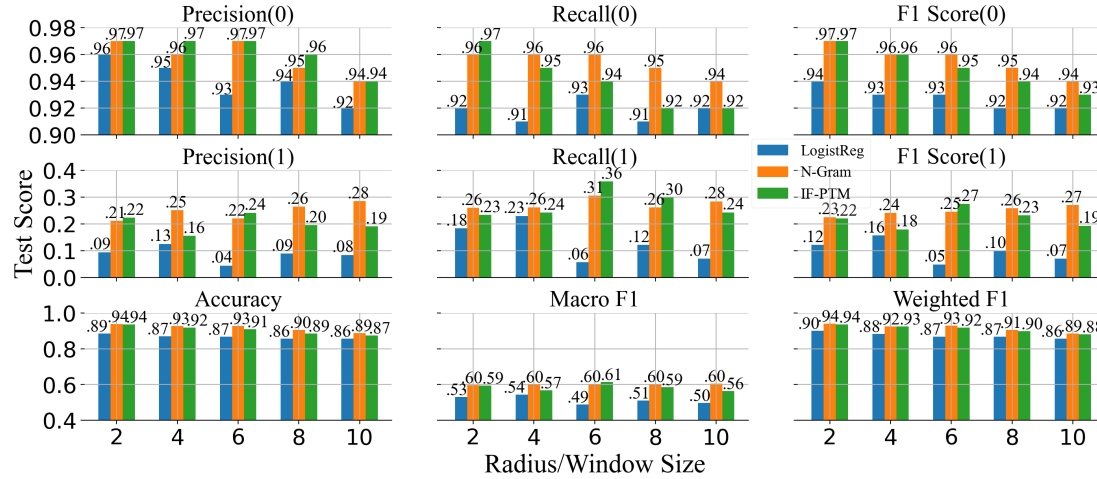


Figure 12. Nine evaluation criteria for different methods under different window/radius values in dataset B. In the IF-PTM, we specified the length of the signature vector to be 20.

Table III
EVALUATION INDEX FOR DIFFERENT METHODS IN IF DETECTION ON DATASET A. THE VALUE IN BRACKETS INDICATES THE SIZE OF THE WINDOW.

Methods	LogistReg	N-Gram	IF-PTM
Precision 0	0.9665 (1)	0.9385 (1)	0.9669 (10)
Recall 0	0.8170 (3)	1.0000 (2)	0.9477 (3)
F1 score 0	0.8636 (3)	0.9667 (3)	0.9509 (2)
Precision 1	0.4551 (10)	0.0000 (1)	0.6102 (9)
Recall 1	0.7520 (2)	0.0000 (1)	0.9000 (10)
F1 score 1	0.4545 (10)	0.0000 (1)	0.6710 (9)
Accuracy	0.7717 (3)	0.9358 (3)	0.9085 (2)
Macro average F1 score	0.6182 (2)	0.4834 (3)	0.7784 (9)
Weighted average F1 score	0.8125 (2)	0.9050 (1)	0.9168 (2)

These tables show that IF-PTM has a better comprehensive performance.

D. The effect of vector size and loss function selection on IF-PTM test scores.

We first investigate the vector size of the proposed waveform representation in the IF-PTM and its impact on prediction performance. As shown in Fig. 13, a larger vector size tends to perform better according to the evaluation index. However, it is a trade-off among the accuracy, precision, F1 score, and the computational burden. In our study, we use the vector size of 20, which does not compromise the performance and computational speed. It is an optimization problem to determine the optimal length of vector size. Through the experiments indicated in Fig. 11, we notice the score of

precision 1 is low for most of the vector sizes. Therefore, we choose the highest precision score that is corresponding to a vector size of 20. However, when the vector size is 20, its Recall and F1 score are not the highest among all sizes. If we choose the vector size with the highest Recall and F1 score, its accuracy and precision are compromised. Therefore, based on the general performance in Fig. 11, we choose the vector size of 20.

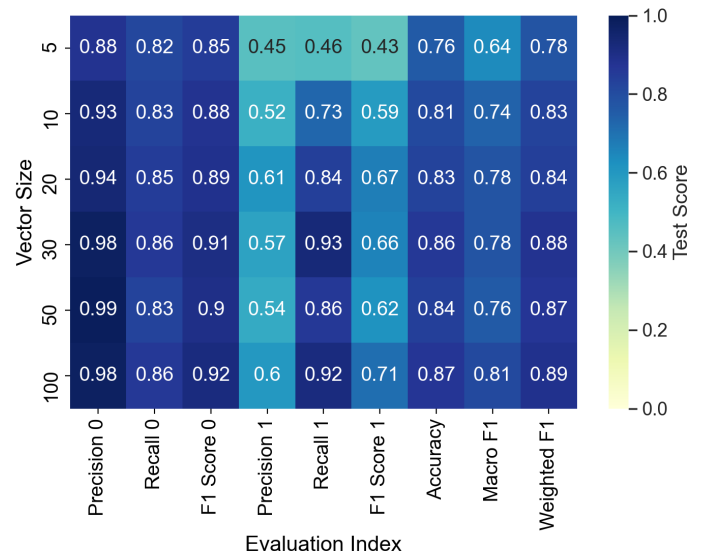


Figure 13. The effect of vector size on IF-PTM test scores.

Table IV
THE ACCURACY, MACRO AVERAGE F1 SCORE AND WEIGHTED AVERAGE F1 SCORE FOR ALL SAMPLES BASED ON DATASET B.

Window Size		1	2	3	4	5	6	7	8	9	10
Methods	accuracy	0.8736	0.8851	0.8776	0.8694	0.8743	0.8676	0.8664	0.8566	0.8623	0.8571
	macro average F1 score	0.5108	0.5301	0.5369	0.5429	0.5212	0.4890	0.5060	0.5111	0.5125	0.4964
	weighted average F1 score	0.8875	0.9002	0.8904	0.8824	0.8849	0.8678	0.8729	0.8678	0.8619	0.8571
N-Gram	accuracy	0.9321	0.9383	0.9217	0.9261	0.9216	0.9273	0.9100	0.9041	0.8991	0.8865
	macro average F1 score	0.5639	0.5966	0.5909	0.6010	0.6050	0.6036	0.5871	0.6035	0.5778	0.6042
	weighted average F1 score	0.9364	0.9400	0.9203	0.9240	0.9246	0.9300	0.9085	0.9057	0.8948	0.8859
IF-PTM	accuracy	0.9272	0.9354	0.9283	0.9191	0.9088	0.9094	0.8977	0.8860	0.8839	0.8740
	macro average F1 score	0.5705	0.5936	0.6051	0.5686	0.5514	0.6129	0.5538	0.5852	0.5428	0.5622
	weighted average F1 score	0.9300	0.9361	0.9334	0.9251	0.9127	0.9189	0.9023	0.8982	0.8812	0.8809

Table V
EVALUATION INDEX FOR DIFFERENT METHODS IN IF DETECTION ON DATASET B. THE VALUE IN BRACKETS INDICATES THE SIZE OF THE WINDOW.

Methods	LogistReg	N-Gram	IF-PTM
Evaluations			
Precision 0	0.9580 (2)	0.9708 (2)	0.9706 (3)
Recall 0	0.9283 (6)	0.9650 (2)	0.9655 (2)
F1 score 0	0.9384 (2)	0.9678 (2)	0.9663 (2)
Precision 1	0.1256 (4)	0.2850 (10)	0.2416 (6)
Recall 1	0.2299 (4)	0.3058 (6)	0.3588 (6)
F1 score 1	0.1567 (4)	0.2701 (10)	0.2744 (6)
Accuracy	0.8851 (2)	0.9383 (2)	0.9354 (2)
Macro average F1 score	0.5429 (4)	0.6050 (5)	0.6129 (6)
Weighted average F1 score	0.9002 (2)	0.9400 (2)	0.9361 (2)

To highlight the advantages of the proposed loss function, we compare the original loss function in (1) (loss1), the proposed loss function in (2) (loss2), and a variation of the proposed function (loss3) that is shown as follows:

$$\min_{\theta} J = \min_{\theta} \left(-\frac{1}{T-m} \sum_{t=m}^T \lambda(s_{t-m}, \dots, s_t) \log p(s_t | s_{t-m}, \dots, s_{t-1}; \theta) \right) \quad (3)$$

Loss3 function is different from loss2 by having the $\lambda(\cdot)$ term outside of the logarithmic function, while the proposed loss2 has the $\lambda(\cdot)$ term inside. Fig. 14 shows the comparison among the three designed loss functions. It is observed that each loss function outperforms the other two in some of the evaluation indices. However, loss3 does not work well in positive samples. As for the performance of loss1 and loss2, loss2 performs no worse than loss1 in 8 evaluation indices out of 9. This builds confidence and justifies the adoption of the proposed loss function.



Figure 14. The comparison among three loss functions.

V. CONCLUSIONS

This paper focuses on IF detection in power distribution systems. The signature vector is constructed to realize a

general fault analysis and prediction method to protect the electric power equipment in the distribution network. It can be effectively used for IF analysis and prediction, as well as later equipment maintenance and overhaul. This paper introduces the “signature vector” model, “signature vector dictionary” and pre-training model for waveform correlation in the field of power system protection. These can realize the machine understanding of fault waveform, meet the protection needs of various devices in the complex state of intelligent power distribution network protection, and effectively improve the reliability and economy of power distribution networks. In the future work, it will be meaningful to consider concurrent faults and investigate the impact of power electronics equipment on the proposed method.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (Grant No. 1810537, 2048288), the Key Research and Development Program of China (2018YFB2100100), and the 111 Project (BP0820005).

REFERENCES

- [1] Q. Su, C. Mi, L. Lai, and P. Austin, “A fuzzy dissolved gas analysis method for the diagnosis of multiple incipient faults in a transformer,” *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 593–598, 2000.
- [2] Q. Cui and Y. Weng, “Enhance high impedance fault detection and location accuracy via μ -PMUs,” *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 797–809, 2020.
- [3] Q. Cui, K. El-Arroudi, and Y. Weng, “A feature selection method for high impedance fault detection,” *IEEE Transactions on Power Delivery*, vol. 34, no. 3, pp. 1203–1215, 2019.
- [4] J. Poon, P. Jain, C. Spanos, S. K. Panda, and S. R. Sanders, “Fault prognosis for power electronics systems using adaptive parameter identification,” *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 2862–2870, 2017.
- [5] W. Xu *et al.*, “Electric signatures of power equipment failures,” in *Transmission & Distribution Committee Power Quality Subcommittee*, 2018.
- [6] W. Zhang, X. Xiao, K. Zhou, W. Xu, and Y. Jing, “Multicycle incipient fault detection and location for medium voltage underground cable,” *IEEE Transactions on Power Delivery*, vol. 32, no. 3, pp. 1450–1459, 2017.
- [7] S. Xiong, Y. Liu, J. Fang, J. Dai, L. Luo, and X. Jiang, “Incipient fault identification in power distribution systems via human-level concept learning,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5239–5248, 2020.
- [8] T. S. Sidhu and Z. Xu, “Detection of incipient faults in distribution underground cables,” *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1363–1371, 2010.
- [9] M. Jannati, B. Vahidi, and S. H. Hosseini, “Incipient faults monitoring in underground medium voltage cables of distribution systems based on a two-step strategy,” *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1647–1655, 2019.

- [10] D. Morais and J. Rolim, "A hybrid tool for detection of incipient faults in transformers based on the dissolved gas analysis of insulating oil," *IEEE Transactions on Power Delivery*, vol. 21, no. 2, pp. 673–680, 2006.
- [11] T. Kari, W. Gao, D. Zhao, Z. Zhang, W. Mo, Y. Wang, and L. Luan, "An integrated method of anfis and dempster-shafer theory for fault diagnosis of power transformer," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 1, pp. 360–371, 2018.
- [12] K. Manivinnan, C. L. Benner, B. D. Russell, and J. A. Wischkaemper, "Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders," in *IEEE International Conference on Intelligent System Application to Power Systems (ISAP)*, 2017, pp. 1–6.
- [13] M. J. Mousavi and K. L. Butler-Purry, "Detecting incipient faults via numerical modeling and statistical change detection," *IEEE transactions on power delivery*, vol. 25, no. 3, pp. 1275–1283, 2010.
- [14] C. Kim, T. Bialek, and J. Awylik, "An initial investigation for locating self-clearing faults in distribution systems," *IEEE transactions on smart grid*, vol. 4, no. 2, pp. 1105–1112, 2013.
- [15] R. Moghe, M. J. Mousavi, J. Stoupis, and J. McGowan, "Field investigation and analysis of incipient faults leading to a catastrophic failure in an underground distribution feeder," in *IEEE/PES Power Systems Conference and Exposition*, 2009, pp. 1–6.
- [16] S. Xiong, Y. Liu, J. Fang, J. Dai, L. Luo, and X. Jiang, "Incipient fault identification in power distribution systems via human-level concept learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5239–5248, 2020.
- [17] B. Kasztenny, I. Voloh, C. G. Jones, and G. Baroudi, "Detection of incipient faults in underground medium voltage cables," in *IEEE Annual Conference for Protective Relay Engineers*, 2008, pp. 349–366.
- [18] M. J. Mousavi, J. J. McGowan, J. Stoupis, and V. D. Donde, "Apparatus and method for adaptive fault detection in mv distribution circuits," Mar. 5 2013, US Patent 8,390,302.
- [19] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 51–59, 2013.
- [20] O. Poisson, P. Rioual, and M. Meunier, "Detection and measurement of power quality disturbances using wavelet transform," *IEEE transactions on Power Delivery*, vol. 15, no. 3, pp. 1039–1044, 2000.
- [21] C. Benner, K. Purry, and B. Russell, "Distribution fault anticipator," *EPRI, Palo Alto, CA*, vol. 1001879, 2001.
- [22] B. Li, Y. Jing, and W. Xu, "A generic waveform abnormality detection method for utility equipment condition monitoring," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 162–171, 2016.
- [23] N. Stringer and L. A. Kojovic, "Prevention of underground cable splice failures," in *IEEE Industrial and Commercial Power Systems Technical Conference. Conference Record (Cat. No. 00CH37053)*. IEEE, 2000, pp. 83–93.
- [24] M. J. Mousavi, K. L. Butler-Purry, R. Gutierrez-Osuna, and M. Najafi, "Classification of load change transients and incipient abnormalities in underground cable using pattern analysis techniques," in *IEEE PES Transmission and Distribution Conference and Exposition*, 2003, pp. 175–180.
- [25] S. Ebron, D. L. Lubkeman, and M. White, "A neural network approach to the detection of incipient faults on power distribution feeders," *IEEE Transactions on Power Delivery*, vol. 5, no. 2, pp. 905–914, 1990.
- [26] M. Faisal, A. Mohamed, and H. Shareef, "Prediction of incipient faults in underground power cables utilizing s-transform and support vector regression," *International Journal on Electrical Engineering and Informatics*, vol. 4, no. 2, p. 186, 2012.
- [27] M. J. Mousavi and K. L. Butler-Purry, "A novel condition assessment system for underground distribution applications," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1115–1125, 2009.
- [28] D. Friedlander, I. Chattopadhyay, A. Ray, S. Phoha, and N. Jacobson, "Anomaly prediction in mechanical systems using symbolic dynamics," in *IEEE Proceedings of the American Control Conference*, vol. 5, 2003, pp. 4275–4280.
- [29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [30] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, 2013, pp. 3111–3119.