ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Using Convolutional Neural Network Algorithm to Improve Multi-Dimensional Information Identification Capabilities of Distribution Network Cable Faults

Wei Li[1,*], *Shaobin Liu*[1], *Tiao Li*[1], *Xuzhe Xu*[1] *and Hanxuan Guo*[1]

[1] Guangdong Electric Power Grid Co., Ltd. Chaozhou Power Supply Bureau, Chaozhou, Guangdong, 521000, China

SUMMARY: *The current cable fault detection is usually based on a single sensor and signal source, coupled with the constraints of the field environment, which greatly enhances the difficulty of personnel inspection. In this paper, by integrating five kinds of multi-dimensional cable fault signal feature information. The extracted multi-dimensional information features are downscaled using the KPCA downscaling algorithm, SVM is used as the multi-dimensional information fusion cable fault identification algorithm, and the KPCA-SVM multi-dimensional information fusion identification model is optimized with kernel parameters through the APSO algorithm, and finally the APSO-KPCA-SVM multi-dimensional information fusion identification cable fault model based on APSO-KPCA-SVM is proposed. The experiments are validated using simulated datasets, and the fusion and dimensionality reduction of five fault signal features is used as the input feature vector of the cable fault identification model, which reduces the accuracy of the model by only 0.22% compared with the model before dimensionality reduction, while the training time is reduced by 413.14 s. The tests are conducted under the condition of noise interference as well as data loss, and compared with the SVM and the DBN, the APSO-KPCA-SVM has stronger anti-interference performance, when the signal-to-noise ratio is reduced to 20dB, only the model cable fault identification accuracy of this paper is still maintained at more than 90%.*

KEYWORDS: *KPCA; SVM; APSO; information fusion; cable fault identification*

1 Introduction

Due to the fast development of the Chinese economy and the further progress of industrialization, electricity has become one of the indispensable aspects of everyday living and manufacturing. The distribution network cables form a significant part of distribution-network operation and the continuous operation of the distribution network plays a very important role [1, 2]. In reality, the cable faults are often observed when operating, which can lead to line outages and endanger the safety of the entire power grid [3, 4]. With regard to the risks of distribution-cable failure, literature [5] states that failures of power-distribution infrastructure including poles and cables represent an influential factor of wildfires, and it explains how power-distribution facilities could ignite fires and the studies on the prediction and detection of wildfires. Early fault detection of power-distribution infrastructure like cables is emphasized in literature [6], which indicates that the faults might severely compromise the stability and reliability of the power-distribution system. Thus, the process of cable-fault identification

*lldym0103@163.com
https://doi.org/10.65102/is2026387
Published: 30 April, 2026

should be enhanced to enable regular functioning of the cables.

In general, the cable generates failure problems are roughly divided into the following aspects, the first is mechanical damage, which is formed due to abnormal power outage of the circuit system, and is a relatively common type of failure [7]; The second is insulation moisture, mainly in the generation of leakage accidents, in the cable in the joints or in the terminal sealing is relatively poor part of the cable, due to the long-term impact of the humid environment, the insulation layer produces cracking, which in turn produces a leakage accident [8, 9]; The third is due to the cable long-term use in the overloaded state, and saturated with electricity and thermal effects of the double impact of the insulation layer resulting in aging or deterioration of the situation, which in turn produces leakage faults [10, 11]. In order to take effective countermeasures against cable faults, prerequisite work such as digging out the causes of faults, fault classification, and pre-detection are essential.Literature [12] analyzes the faults of distribution networks based on the fault database of the distribution network and identifies the causes of faults in order to estimate the fault rate and improve the quality of power supply. Literature [13] points out that underground cables play an important role in the transmission of electrical energy, but the difficulty of detecting their faults also poses a number of risks, by examining the different types of underground cable faults and their causes, and demonstrating the incidence of different cable fault types. Literature [14] addresses the shortcomings of the current existing fault classification techniques for distribution network cables, proposes a method to realize fault classification by detecting the attenuated DC component of the faulted phase current, and verifies that the method can improve the reliability of the distribution network. Literature [15] shows the important role played by underground cable facilities in the context of growing electricity demand, while pointing out the possible threat posed by ageing cable faults and emphasizing the importance of detecting and locating faults before they occur. Literature [16] describes the causes and consequences of faults occurring in transmission lines, stressing that the development and design of new methods for identifying fault types and locations is crucial for improving system reliability and optimal performance, and it also reviews existing methods for locating faults in underground cables.

For identifying the above cable faults, the main approaches include the bridge method and the pulse-detection method. The bridge method, as the most widely used and longest-applied cable-fault detection technology, has gradually become unable to meet the current needs of the power industry and is therefore being phased out [17-19]. Literature [20] describes the risk that cables may experience faults during operation and points out that there are major differences in accuracy and efficiency between traditional bridge methods and newer techniques such as time-domain reflection. The pulse-detection method is easy to operate, and its test results are intuitive and significant; however, it requires prior knowledge of the cable path direction, which also limits its applicability to current power-industry needs [21-23]. For this reason, literature [24] proposes a cable-fault detection and location method based on pulse technology, which determines the position of cable faults by using high-resolution reflectometry and is validated in the detection of various fault types at different distances. Literature [25] notes that the operating condition of power cables is closely related to the quality of the power system and that continuous improvement of cable equipment fault-analysis and detection technology can enhance overall power-system performance; it further summarizes relevant cable-fault problems and detection techniques. Literature [26] introduces cable-line diagnostic methods, including the pulse method and the return flame method, and shows that the pulse method has high measurement accuracy and non-destructive characteristics, making it suitable for locating defects in aging power cables.

Even though the benefits and drawbacks of classical detection approaches have been discussed, in the new age of artificial intelligence, deep learning has started playing a significant

role in the identification of distribution-network cables faults [27]. Deep learning is a machine-learning methodology that learns and builds abstract features using multi-layer neural networks, allowing automatic analysis and extraction of data and subsequent autonomous processing and learning of other tasks. In comparison with the traditional methods of cable-fault diagnosis, the deep-learning-based diagnosis has benefits such as automating the process of diagnosis, better extraction of complex features and high accuracy of diagnosis [28, 29]. With multidimensional information, it may also be combined to enhance greatly the precision and real-time behavior of fault identification. Literature [30] presented a cable-fault monitoring scheme by means of deep learning and established a fuzzy basis function network due to the fuzziness of cable faults in order to obtain the dynamic characteristics of the cable system and ensure the validity of the scheme. Literature [31] talked about the benefits of artificial intelligence in cable-fault detection and localization and examined the applications of the traditional machine-learning and deep-learning methods in this field, including but not limited to their usage conditions, strong points, weaknesses, and success stories. Literature [32] suggested a new wide-frequency-range transfer-function measurement-based and deep-learning-based detection and location method of aging cables, and the experimental outcomes proved its effectiveness and its generalization ability and indicated that it can perform well in different aging conditions. Literature [33] suggested an automated, efficient and precise cable-fault classification and localization method founded upon deep learning, which achieves data collection, feature extraction, dimension reduction, fault detection, fault classification and localization. Literature [34] came up with an online cable-fault diagnosis model relying on deep learning and, in comparison with the traditional backpropagation neural network, the method was found to be significantly superior. Literature [35] proposed a deep-learning-based localized-fault pattern-recognition method, which helps to increase considerably the accuracy of cable-fault detection as compared with traditional methods of detection.

In this study, we first analyze and extract the features of cable fault signals, including acoustic signals, electromagnetic wave pulse signals, low-voltage pulse test signals, path-finding signals, and electromagnetic field signals. In order to solve the problems of sparsity, multicollinearity and other problems of multidimensional information features in high dimensional space, the extracted multidimensional information features are downscaled by using the kernel principal component analysis algorithm, which improves the fault identification efficiency of the multidimensional information fusion algorithm. Then, based on the support vector machine, the KPCA-SVM multidimensional information fusion cable fault recognition model is optimized by the adaptive particle swarm algorithm for the kernel parameters to improve the accuracy and recognition rate of cable fault recognition. Through example analysis, multiple cable fault signal features are used as input feature vectors of the model to verify the effectiveness of the extracted features. The valid feature vectors are input into multiple identification models, and in the anti-jamming analysis test, it is proved that the model in this paper has stronger robustness.

2 Method

2.1 Multi-dimensional information feature extraction of fault sources in distribution cables

2.1.1 Interference and noise

In the process of effective impact discharge occurring in power cables, when acquiring the multimode signal of the fault source, the interfering signal that is mixed with the signal is the

main source of external interference. Complex field environment, should be on the sensor received signals to take measures for interference suppression, in order to accurately identify the fault point of the fault signal. This is particularly important for cable fault pinpointing when the anti-interference, interference signals can be divided into periodic interference signals and random interference noise. Usually there are two types of hardware and software methods that can be used to suppress interference. In conditions where the interference is particularly severe, a combination of the two methods is used so that the interference is suppressed to the level of the permissible range, and the characteristics of the fault signal can be accurately determined.

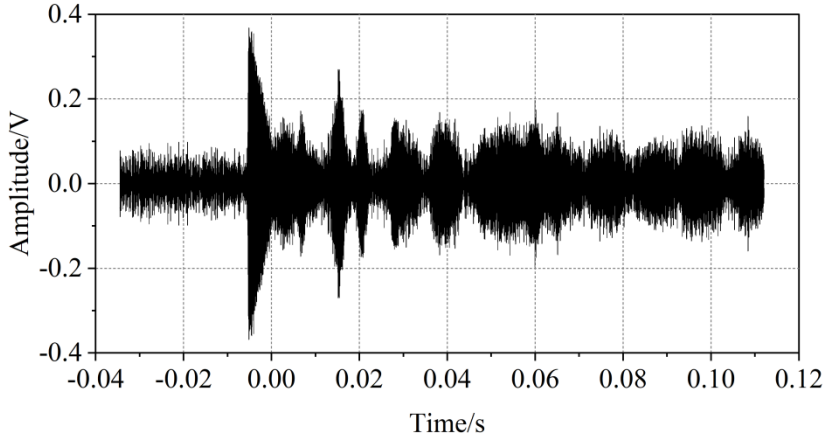
2.1.2 Characterization of fault signals

The precise positioning signals based on multi-mode information fusion are mainly acoustic signals, electromagnetic wave pulse signals, low-voltage pulse test signals, path-finding signals, and electromagnetic field signals. These signal characteristics can truly reflect the fault point information.

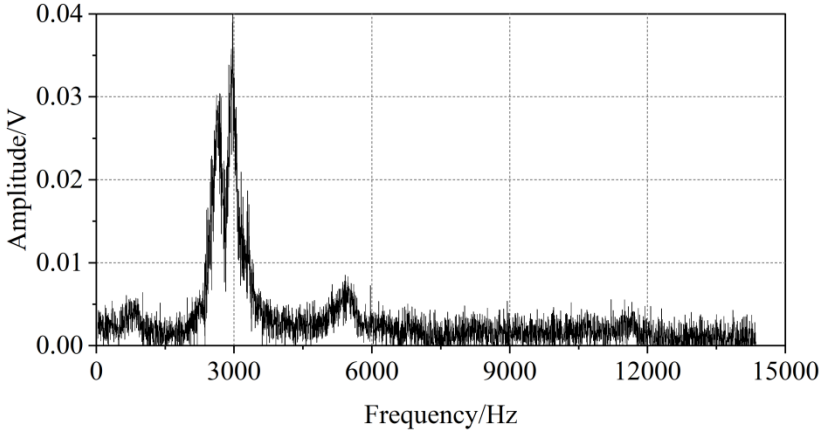
(1) Acoustic signal

Power cable inter-phase breaks or break faults, the impact of energy-controlled signal generator loaded at the beginning of the fault cable, injecting high-energy pulse signals, so that the cable faults between the two phases of the air breakthrough, the effective breakdown of the moment, the source of the fault will be released in a different direction, “pop, pop,” the sound of the discharge. In the laboratory, run the construction of the experimental environment, the sensor will receive the sound signal, the use of Tektronix oscilloscope recorded, and then input the recorded data into the MATLAB for spectral analysis, can be derived from the characteristics of the fault point sound signal. The laboratory measured acoustic signal and its spectrum analysis are shown in Fig. 1, Figs. (a) and (b) show the original model of sound and its spectrum, respectively.

Figure (a) from the characteristics of the waveform, the envelope of the waveform is a pulse waveform, in line with the phenomenon of the discharge sound “pop, pop”. Figure (b) can be seen in the cable effective impact discharge process, the fault clicks through the air issued by the sound frequency band distribution between 1.5kHz-4.5kHz. These characteristics for the judgment of the fault signal gives a clear signal characteristics, the difference between random interference noise. Of course, only from the above signal is unable to determine is the fault signal, but also need to signal denoising and signal estimation to determine whether the fault point signal.



(a)Original signal



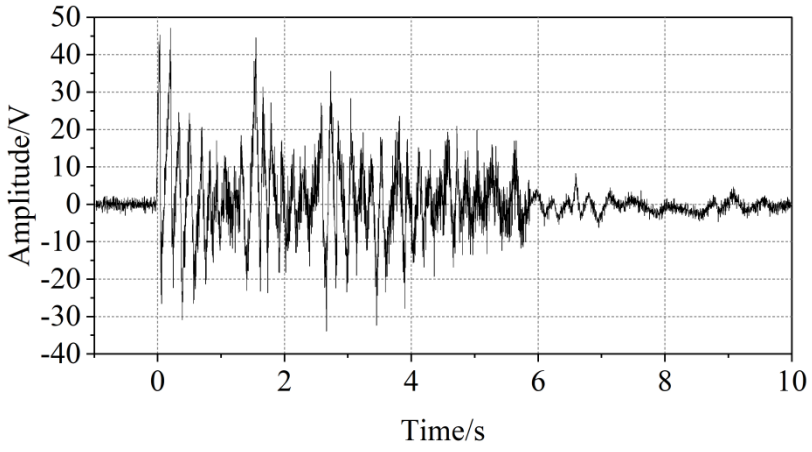
(b)The rapid Fourier transform of sound signals

Figure 1: Source sound signal and spectrum

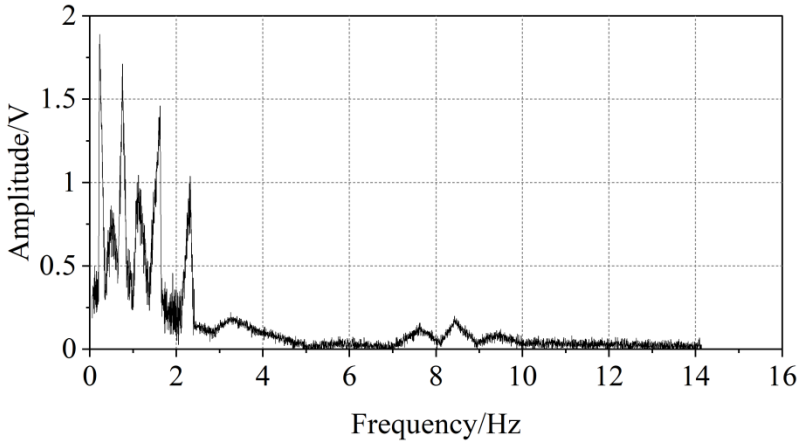
(2) Electromagnetic wave pulse signal

During the effective impact discharge of the cable-fault point, the electromagnetic wave signal is generated, and it can be detected by the antenna. With the help of the antenna, the oscilloscope has a direct connection to the receiver end such that the electromagnetic wave signal which was generated at the moment of fault discharge can be acquired. The time-domain waveform of this electromagnetic wave signal is depicted in Fig. 2, while Figs. (a) and (b) represent the initial waveform and its frequency spectrum respectively.

In this figure, the electromagnetic wave comes from two parts, the first part is the discharge electromagnetic wave at the fault point, and the second part is the electromagnetic wave generated by the discharge between the energy-gap balls of the high-energy shock pulse signal, and due to the limited space in the laboratory, the actual observed time-domain waveform is the superposition of the two electromagnetic waves. Of course, the frequency of these two electromagnetic waves is not the same, so there are multiple spikes inside the frequency, for the electromagnetic wave we do not care about its frequency characteristics, similar to the lightning on a rainy day, but rather concerned about the occurrence of this signal signal marking the presence of the fault point.



(a)Original signal



(b)Fast Fourier transform of the original signal

Figure 2: Signal characteristics of source electromagnetic wave

(3) Low-voltage pulse test signal

Figure 3 shows the outcome of the low-voltage pulse test signal used to test cables and the measured low-voltage pulse signals at two different characteristic conditions, where Figure (a) and Figure (b) represent the open-circuit fault point and the short-circuit fault point, respectively.

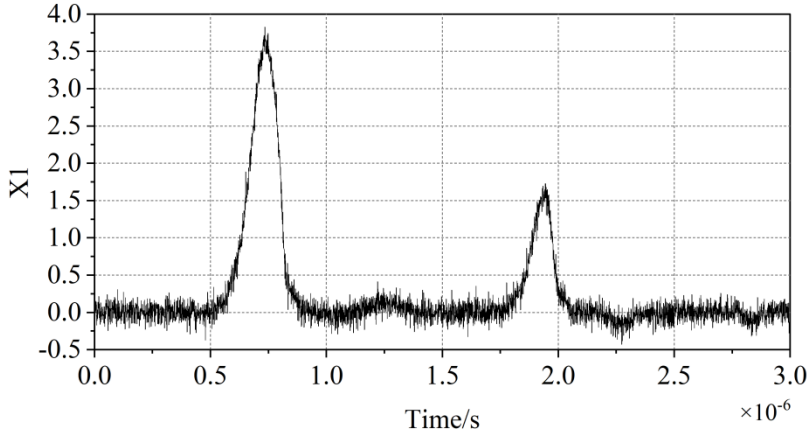
The features of the return signal during the low-voltage pulse test can be calculated based on the distance between the fault point and the beginning of the cable, along with the signal. This data serves as a crucial consideration in determining the exact location of cable faults. It is described as the distance influence factor of the transmission line:

$$\alpha = \frac{\ell - \Delta\ell}{\ell} \tag{1}$$

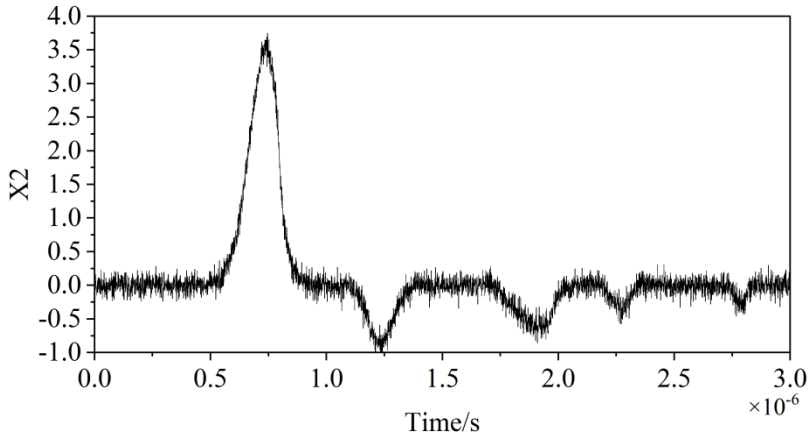
where ℓ is the distance to the fault point obtained from the low-voltage pulse-echo test signal, $\Delta\ell$ is the radius distance from the predicted fault point as the center of the circle, and the influence factor α calibrates the confidence of the precise location. A faulty cable in real life cannot have only one fault, which is manifested by multiple pulses on the low-voltage return signal, so there may be multiple influence factors α on a faulty cable, but each influence factor is independent. It can also reflect the attenuation of the signal by the ratio of the amplitude of the return signal to the amplitude of the incident wave, reflecting the size of the reflection coefficient. By defining the ratio of the amplitude of the return signal to the amplitude of the incident wave as the reflection coefficient impact factor as a reference factor for accurate fault location. The reflection coefficient influence factor is defined as:

$$\bar{\rho} = \frac{u_{-}}{u_{+}} \tag{2}$$

where u_{-} is the pulse amplitude of the return signal, u_{+} is the amplitude of the incident wave of the pulse signal, and $\bar{\rho}$ is the reflection coefficient influence factor. This reflection coefficient influence factor is an important reference for reflecting the characteristics of the fault point.



(a)When the fault point is open circuit



(b)When a fault point is short-circuited

Figure 3: Echo condition of the measured low voltage pulse test signal

(4) Path finding signal

The precise location of the fault is first of all to determine the precise direction of the faulty cable. Currently the most popular way is the audio induction method, by injecting audio signals into the faulty cable, and using sensors above the cable to receive the audio signals of electromagnetic induction, the closer the cable conductor, the stronger the electromagnetic induction, the stronger the audio signals, and the stronger the path direction of the faulty cable is known. The principle is shown in Figure 4.

Cable around the electromagnetic induction must be generated and the signal source constitutes a signal loop, in the figure you can see the signal current loop consists of three parts, the earth circuit, other cables and pipelines circuit, cable metal armor circuit.

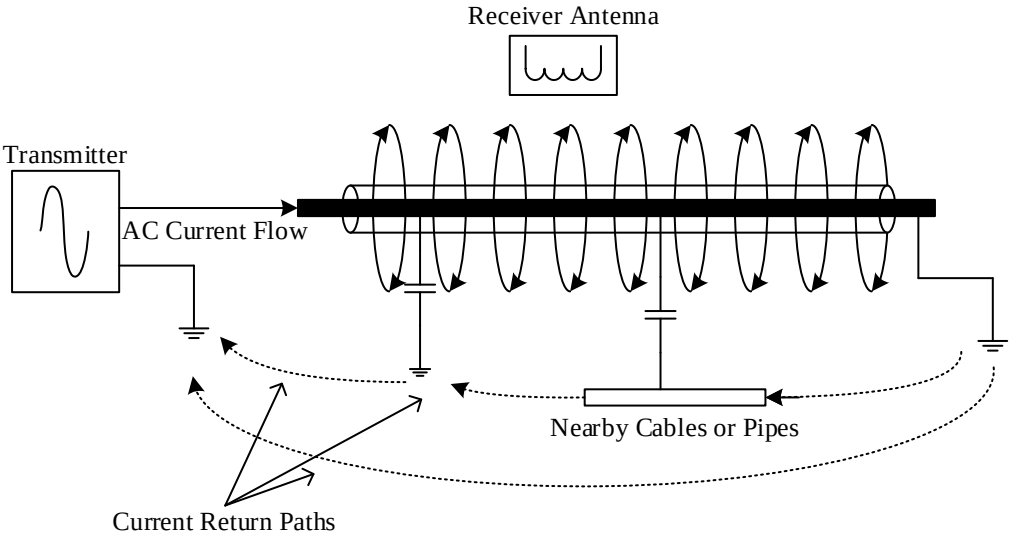


Figure 4: Path detection principle

(5) Electromagnetic field signal

When an effective impact discharge occurs at the cable fault point, the fault clicks through instantaneously, and there will be an inrush current flowing back through the fault point to the high-energy pulse signal source. Electromagnetic field signal is also an important characteristic signal in the process of fault pinpointing, through the pulse current sensor to measure this electromagnetic field signal. The principle of EMF signal generation at the point of fault is shown in Figure 5. The breakdown current at the fault point generates a pulsed EMF signal, which is received by the solenoid sensor, and the signal strength is strongest when the sensor is closest to the fault source.

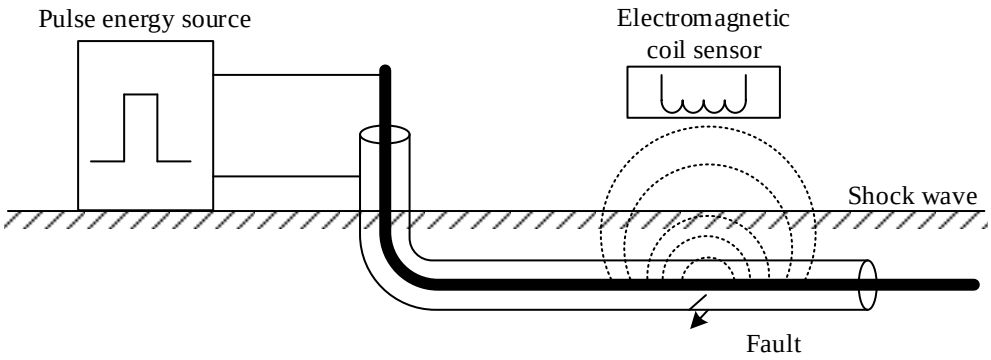


Figure 5: The principle of the electromagnetic signal of the fault point

2.2 Multidimensional data fusion methods

Multi-source information fusion technology, also known as multi-sensor data fusion, refers to the multi-level and multi-level information processing of one or more sources of data and information that have been obtained in accordance with a certain purpose or requirement, and then obtaining reliable and useful data or decision-making through relevant estimation and comprehensive analysis. Using the multi-source information fusion method, when there is some deviation in the response of some single source of information to the target object, the complementarity between multiple information is utilized to obtain more accurate and complete information about the target object through the comprehensive analysis of each data source.

The fusion of data-level information is based on the identification level of the obtained information, and there are mainly three levels: data level, feature level and decision level.

Cable fault identification in the existence of structured information differences, multiple data sources and other characteristics, multiple information fusion technology for cable fault identification, can be a greater degree of improvement in the one-sided reliance on data information sources for cable fault identification when the deficiencies that exist in the cable fault identification, is the trend of cable fault identification research.

Combined with the current power system operation can be obtained in the distribution network cable fault multi-source field information, including acoustic wave signals, electromagnetic wave pulse signals, low-voltage pulse test signals, path-finding signals, electromagnetic field signals and other multi-dimensional data, this paper adopts feature layer fusion multi-dimensional information fusion approach.

2.2.1 Support Vector Machines

(1) Linear support vector machine

Set the training data $\{x_i, y_i\}, i = 1, 2, \cdots, l$, where $x_i \in R^n, y_i \in \{-1, 1\}$. Set R^n to have a categorical hyperplane $w^{\mathrm{T}} x + b = 0$, where $w \in R^n$ is the normal to the hyperplane, i.e., the parameter vector, $b \in R$ is the intercept, and $|b|/\|w\|$ is the perpendicular distance of that hyperplane to the origin of the coordinates (and $\|\cdot\|$ is the Euclidean parameter). The so-called hyperplane is able to separate the pre-calibrated positive and negative samples. And suppose d_+ is the shortest distance from the positive class samples to the hyperplane, and d_- is the shortest distance from the negative class samples to the hyperplane. And $d_+ + d_-$ is defined as the residual of this hyperplane.

The so-called linear support vector machine means that the separating hyperplane with the largest margin can be found, that is, all samples of positive and negative categories are required to satisfy some constraints:

$$\begin{cases} x_i^{\mathrm{T}} w + b \geq +1, \\ x_i^{\mathrm{T}} w + b \leq -1, \end{cases} \begin{cases} y_i = +1, \\ y_i = -1, \end{cases} i = 1, 2, \cdots, l \tag{3}$$

Or:

$$y_i \left(x_i^{\mathrm{T}} w + b \right) - 1 \geq 0, \forall i \tag{4}$$

And the points satisfying equation (4) are called support vectors.

Consider equation (3), the hyperplane satisfying $x^{\mathrm{T}} w + b = +1$ is H_1, its normal is w and its perpendicular distance to the origin is $|1-b|/\|w\|$. And the hyperplane satisfying $x^{\mathrm{T}} w + b = -1$ is H_2, whose normal is also w, and whose perpendicular distance to the origin is $|-1-b|/\|w\|$: H_1 and H_2 These two hyperplanes are parallel to each other. And $d_+ = d_- = 1/\|w\|$, then cosine $d_+ + d_-$ is $2/\|w\|$. Remember that there are no training sample points between these two parallel hyperplanes.

So, a linear support vector machine is a problem that transforms the problem into a problem of finding the maximum margin, that is, the problem of minimizing $\|w\|^2$ on the basis of satisfying Eq. (4). The constraints are as follows:

$$\begin{cases} \min \dfrac{1}{2}\|w\|^2 \\ s.t. y_i\left(x_i^T w + b\right) - 1 \geq 0, \forall i \end{cases} \tag{5}$$

This transforms the problem into a convex quadratic programming problem. Figure 6 shows a typical linearly separated straight line in the two-dimensional case. And it happens to be satisfied by points on the hyperplanes H_1 and H_2:

$$\begin{cases} x_i^{\mathrm{T}} w + b = +1, \\ x_i^{\mathrm{T}} w + b = -1, \end{cases} \begin{cases} y_i = +1, \\ y_i = -1, \end{cases} i = 1, 2, \cdots, l \tag{6}$$

The change in these points affects the change in the residual, these points are support vectors.

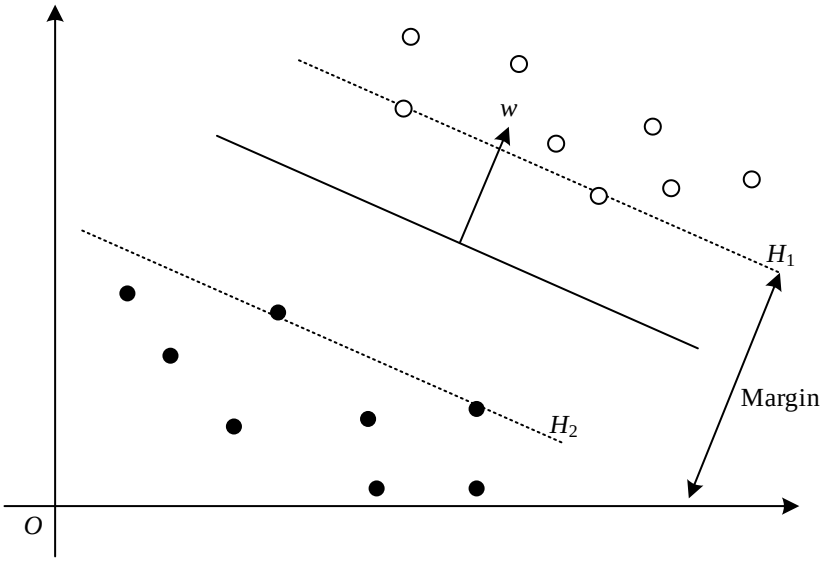


Figure 6: Linear separation hyperplane

(2) Lagrange method

The above problem is now transformed into a Lagrange formulation. This has the advantage that on the one hand the constraints in the optimization problem can be replaced using the constraints of the Lagrange multipliers, which is easier to handle. On the other hand it is easy to introduce the problem into a nonlinear one. The Lagrange multiplier $\alpha_i, i = 1, 2, \cdots, l$ is introduced to equation (4). Thus the Lagrange function can be introduced:

$$L_P = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i y_i \left(x_i^{\mathrm{T}} w + b\right) + \sum_{i=1}^{l} \alpha_i \tag{7}$$

And for the optimization problem of Eq. (5), its Wolfe dyadic problem can be solved according to the theorem, viz:

$$\begin{cases} \max\limits_{w,\alpha} L_P = \dfrac{1}{2}\|w\|^2 - \sum\limits_{i=1}^{l} \alpha_i y_i \left(x_i^{\mathrm{T}} w + b \right) + \sum\limits_{i=1}^{l} \alpha_i \\ s.t. \dfrac{\partial L_P}{\partial w} = w - \sum\limits_{i=1}^{l} \alpha_i y_i x_i = 0, \alpha_i \geq 0, i = 1, 2, \cdots, l \end{cases} \tag{8}$$

In Eq. $\alpha = \left(\alpha_1, \alpha_2, \cdots, \alpha_l \right)^T$, which is equivalent to Eq. (5) and has the same w value.

The constraints of the above equation are equal to the given equation:

$$w = \sum_{j=1}^{l} \alpha_j y_j x_j \tag{9}$$

$$\frac{\partial L_p}{\partial b} = \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{10}$$

Substituting Eq. (9) and Eq. (10) into Eq. (7) leads to:

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j \tag{11}$$

Thus, the training problem for a separable linear support vector machine can be transformed into the following planning problem:

$$\begin{cases} \max\limits_{\alpha} L_D = \sum\limits_{i=1}^{l} \alpha_i - \dfrac{1}{2} \sum\limits_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j x_i^{\mathrm{T}} x_j \\ s.t. \sum\limits_{i=1}^{l} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \cdots, l \end{cases} \tag{12}$$

The optimal α-value and hence the optimal w-value can then be obtained.

For the above constrained problem, the equivalent KKT condition can be written in the following form:

$$\begin{cases} \dfrac{\partial}{\partial w_j} L_p = w_j - \sum\limits_{i=1}^{l} \alpha_i y_i x_{ij} = 0, j = 1, 2, \cdots, n \\ \dfrac{\partial}{\partial b} L_p = -\sum\limits_{i=1}^{l} \alpha_i y_i = 0 \\ y_i \left(x_i^{\mathrm{T}} w + b \right) - 1 \geq 0, i = 1, 2, \cdots, l \\ \alpha_i \geq 0, \alpha_i \left(y_i \left(x_i^{\mathrm{T}} w + b \right) - 1 \right) = 0, \forall i \end{cases} \tag{13}$$

The optimal α and b values can be found from equation (13).

After training the SVM using l samples, it can be utilized to classify the test data. The functional expression of the classification hyperplane can be expressed as:

$$f(x) = \mathrm{sgn}\left(w^{\mathrm{T}} x + b\right) \tag{14}$$

where $\mathrm{sgn}(\bullet)$ is the sign function, and the sign of $f(x)$ determines exactly the class of the test data x.

(3) Nonlinear Support Vector Machine Classification

The previously described are linear classification hyperplane, while many problems encountered in practice are nonlinear. This requires us to generalize linear problems to nonlinear problems, which is also an important performance in SVM applications. At this point it is necessary to use nonlinear decision functions. Fig. 7 Schematic diagram of the conversion from nonlinear space to linear space.

The so-called nonlinear support vector machine, which passes through some pre-selected nonlinear mapping:

$$\Omega : \Gamma \to \Pi \tag{15}$$

Perform the transformation where $\Gamma = R^2$ is a low-dimensional Euclidean space and Π is a high-dimensional linear eigenspace, usually a Hilbert space. And make the following transformations:

$$x = \left(x_1, \cdots, x_n\right) \to \Omega(x) = \left(\Omega_1(x), \cdots, \Omega_N(x)\right) \tag{16}$$

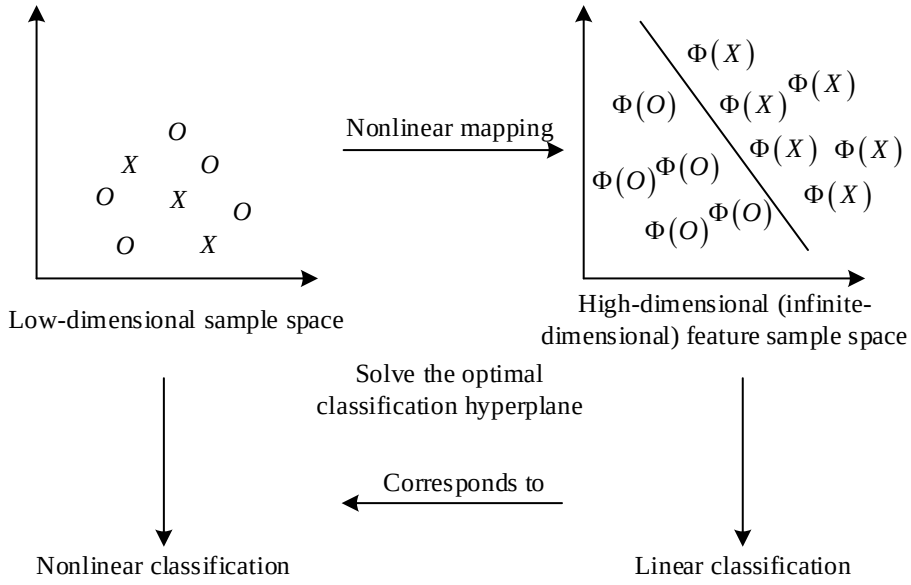


Figure 7: Transformation of nonlinear space to linear space

Here, a support vector machine is the conversion of a nonlinear problem in a low-dimensional space into a linear classification problem in a high-dimensional space. If the optimal hyperplane for classification is to be constructed in the high-dimensional feature space, the concept of dot product $\phi\left(x_i\right) \cdot \phi\left(x_j\right)$ in the feature space must be introduced in the training algorithm. Thereby, the relationship of the inner product operation between the training samples

will appear in the pairwise problem of the original problem, both in the objective function and the decision function. Therefore, it is sufficient to perform the inner product operation only in the high-dimensional space, and in a way, there is no increase in the complexity that the high-dimensional operation may bring. The appropriate inner product function (i.e., kernel function) $K\left(x_i, x_j\right)$ in the optimal classification hyperplane according to the Mercer condition such that $K\left(x_i, x_j\right)=\Omega\left(x_i\right)\cdot\Omega\left(x_j\right)$, the nonlinear problem transformed into a linearly differentiable problem can be achieved. At this point, the objective optimization function becomes:

$$L_D=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K\left(x_i, x_j\right) \tag{17}$$

The prediction function of the decision categorization surface becomes:

$$f(x)=\operatorname{sgn}\left\{\sum_{i=1}^{N_t} y_i\alpha_i K\left(x, x_i\right)+b\right\} \tag{18}$$

2.2.2 Selection of the kernel function

In the optimal classification surface described above, the concept of kernel function is introduced, and by adopting an appropriate kernel function $K\left(x_i, x_j\right)$, the linear classification of a certain nonlinear problem can be realized without any increase in computational complexity as a result. By replacing the objective function in Eq. (12) with Eq. (18), the principle of the optimization problem is the same. After the introduction of the kernel function, the problem of nonlinear classification in low-dimensional space can be mapped to a high-dimensional linearly divisible space for analysis, successfully avoiding the so-called "dimensional catastrophe" problem.

On the other hand, the introduction of the kernel function successfully avoids the so-called "dimensional catastrophe" by converting the samples that are not easy to be classified in the low-dimensional nonlinear space to be solved in the high-dimensional feature space. By introducing the kernel function, the ability of the learning machine to solve nonlinear problems has been improved to a large extent, and the internal linear characteristics of the learning machine in the high-dimensional feature space remain unchanged, so that the machine's learning can be easily controlled. The inner product relation in the original nonlinear space is replaced with a kernel function, which is equivalent to transforming the data into the corresponding high-dimensional vector space through some kind of mapping, which is called the mapping associated with the kernel, and the feature space is given a definition using the introduced kernel function. By introducing the kernel function, the inner product operation in the high-dimensional feature space is able to utilize this kernel function to carry out a series of operations, that is, after utilizing the kernel function, the data is implicitly described into the feature space and a linear learning machine can be obtained in its space, thus skipping the process that should have been computed originally.

Define a kernel function K assuming that there is a nonempty set A and that the inner product space E,Ω is a mapping of $\Gamma\rightarrow\Pi$. Suppose that the function $K:\Gamma\times\Gamma\rightarrow R$ satisfies: $\forall x_i, x_j\in\Gamma$ with $K\left(x_i, x_j\right)=\left\langle\Omega\left(x_i\right)\cdot\Omega\left(x_j\right)\right\rangle$.

The kernel function is an indispensable part of the support vector machine, the training sample set and the kernel function can be a complete and comprehensive description of the support vector machine. For specific problems encountered in practice, it is also very important

to choose which kernel function to use. Different kernel functions $K\left(x_i, x_j\right)$ are chosen, the final results of classification realized by the support vector machine are also different. By choosing different inner product kernel functions, various support vector machines can be combined, and thus many support vector machine methods can appear. Below is a list of some of the kernel functions that we often use:

Linear kernel function: $K\left(x_i, x_j\right) = x_i \cdot x_j$.

Polynomial kernel function: $K\left(x_i, x_j\right) = \left(x_i^{\mathrm{T}} x_j + 1\right)^p$. This is a polynomial data classifier of order p.

Radial Basis Kernel Function (RBF): $K\left(x_i, x_j\right) = \exp\left(-\left\|x_j - x_i\right\|^2 / \left(2\sigma^2\right)\right)$. This kernel function constitutes a classifier that can be trained by SVM to determine the support vector, the weight vector and the intercept automatically, and each center of the basis function corresponds to a support vector, obtaining better results than the classical classifier composed of radial basis kernel functions.

Sigmoid kernel function: $K\left(x_i, x\right) = \tanh\left(kx_i \cdot x_j + \delta\right)$. This kernel function is a neural network classifier with 2 layers of sigmoid.

In this paper, the radial basis kernel function (RBF) is chosen as the kernel function for multi-source data fusion to recognize the cable fault model.

2.3 Multi-source data feature dimensionality reduction

As can be seen from the analysis in Section 2.1, this paper comprehensively extracts the fault multidimensional information characteristics of faults such as acoustic wave signals, electromagnetic wave pulse signals, low-voltage pulse test signals, path-finding signals, electromagnetic field signals and other faults through the fault information collected by the sensors, with a view to obtaining more comprehensive fault information, while the calculation of the fault characteristics of the different data sources makes the fault data dimensional characteristics increase, although the multidimensional feature data can be more comprehensively Although the multidimensional feature data can be more comprehensive to obtain the feature information of different fault types, the feature set in high dimensional space has sparsity, multicollinearity and other problems, which will have an impact on the accuracy of the SVM model fault identification. Therefore, in this paper, multisource data features are downscaled by dimensionality reduction algorithm on the extracted multisource data while maintaining the comprehensive reflection of fault feature information.

PCA is a linear dimensionality reduction algorithm, which can effectively reduce the dimensionality of most of the linear data, but it cannot effectively reduce the dimensionality when dealing with nonlinearly related data. Kernel Principal Component Analysis (KPCA) is applied in many algorithms based on kernel learning methods, which is an improvement of PCA algorithm and retains the linear dimensionality reduction process of PCA algorithm.The core idea of KPCA algorithm is to map the original input data into a high-dimensional feature space through nonlinear kernel functions, calculate the kernel matrix eigenvalues and eigenvectors of the high-dimensional feature space, and then realize the efficient nonlinear data dimensionality reduction. Based on this, this paper selects the KPCA algorithm to downscale the extracted features, and the downsizing steps are as follows:

Let the input data matrix $X=\left[x_1, x_2, , x_m\right]^{\mathrm{T}} \in \mathbf{R}^{m\times n}$, where $\mathbf{x}_i(i=1,2,,m)\in\mathbf{R}^n, m$ and n are the number of samples, and the number of variables, respectively. Mapping the data of the initial matrix to the high-dimensional feature space by the nonlinear mapping function $\boldsymbol{\Phi}(x)$ yields:

$$X=\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \rightarrow \boldsymbol{\Phi}(X) \begin{bmatrix} \boldsymbol{\Phi}\left(x_{11}\right) & \cdots & \boldsymbol{\Phi}\left(x_{1n}\right) \\ \cdots & \cdots & \cdots \\ \boldsymbol{\Phi}\left(x_{m1}\right) & \cdots & \boldsymbol{\Phi}\left(x_{mn}\right) \end{bmatrix} \tag{19}$$

Centering the kernel matrix $\boldsymbol{\Phi}(x)$ of the above equation, with $\boldsymbol{\Phi}\left(\mathbf{x}_i\right)$ denoting the average value in each dimension, yields:

$$\boldsymbol{\Phi}'(X)=\begin{pmatrix} \boldsymbol{\Phi}\left(x_{11}\right)-\overline{\boldsymbol{\Phi}\left(x_1\right)} & \cdots & \boldsymbol{\Phi}\left(x_{1n}\right)-\overline{\boldsymbol{\Phi}\left(x_1\right)} \\ \cdots & \cdots & \cdots \\ \boldsymbol{\Phi}\left(x_{m1}\right)-\overline{\boldsymbol{\Phi}\left(x_m\right)} & \cdots & \boldsymbol{\Phi}\left(x_{mn}\right)-\overline{\boldsymbol{\Phi}\left(x_m\right)} \end{pmatrix} \tag{20}$$

Calculate the covariance matrix of the samples in the feature space as:

$$\mathbf{C}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\Phi}\left(x_i\right)\left[\boldsymbol{\Phi}\left(x_i\right)\right]^{\mathrm{T}} \tag{21}$$

Find the eigenvalues, eigenvectors of the matrix C, and then sort them from largest to smallest, with the eigenvalues denoted by $\lambda_1, \lambda_2, \lambda_m$, and the eigenvectors comprising the matrix notated as $\mathbf{V}_l=\left[\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_m\right]$. Define the matrix Q, and the diagonalized matrix QCQ^T respectively:

$$\mathrm{Q}=\left[\mathbf{V}_1, \mathbf{V}_2, , \mathbf{V}_n\right] \tag{22}$$

$$QCQ^T=\begin{pmatrix} \lambda_1 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \tag{23}$$

Calculate the cumulative contribution rate corresponding to the first k eigenvalues and its corresponding covariance cumulative contribution value p, select the first k eigenvalues whose cumulative contribution rate exceeds the desired target according to the requirement, find the corresponding eigenvectors, and finally form the eigenmatrix after dimensionality reduction process, take the first k columns in the matrix QCQ^T to form a new matrix Q':

$$\mathrm{Q}'=\left[\mathbf{V}_1', \mathbf{V}_2', , \mathbf{V}_k'\right] \tag{24}$$

The matrix after dimensionality reduction is computed:

$$Y'=Q'\boldsymbol{\Phi}(\mathrm{X}) \tag{25}$$

2.4 Based on APSO-KPCA - SVM multidimensional data fusion cable fault identification methods

In this paper, adaptive particle swarm (APSO) algorithm is used to optimize the kernel parameters of the KPCA-SVM model. APSO algorithm is able to jump out of the local optimum while maintaining better population diversity, effectively avoiding the defects of the traditional particle swarm optimization algorithm (PSO) such as premature convergence, poor local search ability at the late stage of particle search, and has a better ability to search for composite shape rays.

In order to avoid the problem of premature convergence of traditional particle swarm algorithm, APSO introduces the variation operator on the basis of PSO algorithm to extend the particle search range to improve the algorithm optimization accuracy. The improved velocity update formula and position update formula are:

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 \left(P_{id}^{k} - X_{id}^{k} \right) + c_2 r_2 \left(P_{gd}^{k} - X_{gd}^{k} \right) + flag \cdot c_3 \cdot v_{\max} \tag{26}$$

$$X_{id}^{k+1} = X_{id}^{k} + \alpha \cdot V_{id}^{k} \tag{27}$$

The above equation (26) introduces a mutation operator $flag \cdot c_3 \cdot v_{\max}$, which can be adjusted to make it jump out of the original solution space when the particle search falls into local convergence in the late stage, and to mutate the speed of the particles' flights. flag is generally taken as a flag value of 0 or 1, with 0 taken as the default value, and 1 when the particles fall into the early state. The c_3 is the acceleration constant.

$$c_3 = \min\left\{ \left| \frac{f_g(t) - f_t}{f} \right|, 1 \right\} \tag{28}$$

where: $f_g(t)$ - is the adaptation value of the current global optimal particle p_g. f_t - is the theoretical best quality or empirical optimal value.

In this paper, based on the feature quantity of each sensing line data source to extract the fault wave acoustic wave signal, electromagnetic wave pulse signal, low-voltage pulse test signal, path-seeking signal, electromagnetic field signal and other feature quantities, the use of KPCA - SVM model based on the optimization of the APSO to carry out multi-dimensional data fusion for cable fault identification, the specific distribution network cable fault identification method flow is shown in Figure 8.

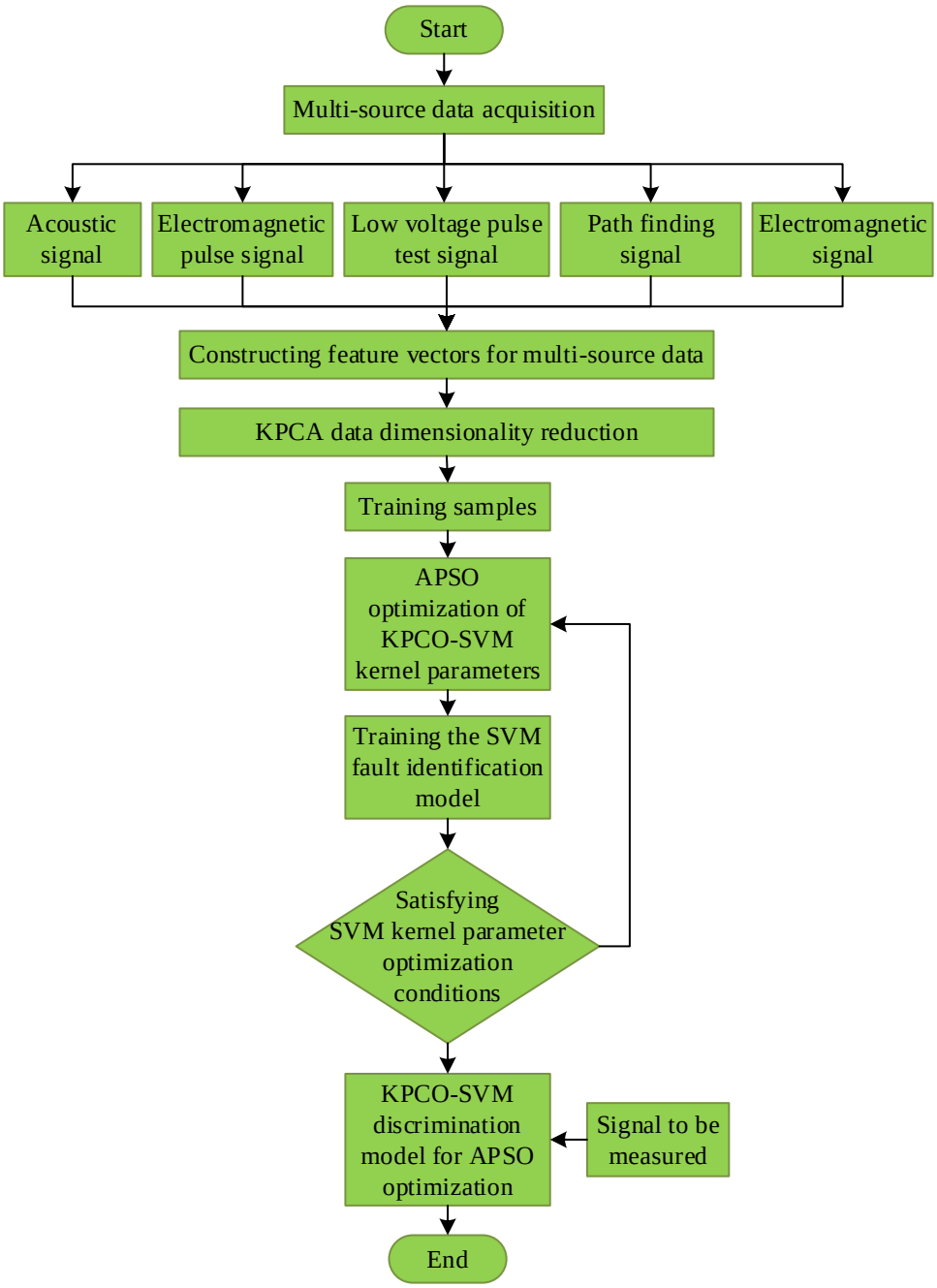


Figure 8: Cable fault identification flow chart

3 Results and discussion

3.1 Performance assessment indicators

In this paper, four evaluation metrics are introduced to evaluate the recognition performance of the fault recognition model, which are accuracy (P_{acc}), precision (P_{pre}), recall (P_{rec}), and F_1 values. The formulas of the four evaluation indexes are shown below.

Accuracy P_{acc} denotes the ratio of the number of positive class samples to the total number of samples, i.e:

$$P_{acc} = \frac{TP + TN}{TP + FP + TN + FN} \tag{29}$$

The precision rate P_{pre} denotes the ratio of the number of samples in the positive category to the total number of samples predicted to be in that category, i.e:

$$P_{pre} = \frac{TP}{TP + FN} \tag{30}$$

The recall rate P_{rec} denotes the ratio of the number of samples in the positive category to the total number of samples in that category, i.e:

$$P_{rec} = \frac{FP}{TP + FP} \tag{31}$$

Since the actual task of cable-fault identification may involve an imbalanced sample distribution, the F_1 score is introduced to further evaluate model-output performance. The F_1 score represents the harmonic mean of precision and recall:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{32}$$

In this case, *TP* means the number of samples that have both the actual and predicted values as positive; *FP* means the number of samples that are predicted negative but in reality have a positive value; *TN* means the number of samples that have both the actual and predicted values as negative; and *FN* means the number of samples predicted to be positive but actually have a negative value.

3.2 System configuration

The hardware equipment for this experiment is a personal computer with NVDIA GeForce GTX850M graphics card, Intel(R) Core(TM) i5-4200H CPU, Samsung SSD 860EVO 500G hard disk, and 12GB RAM. The original data sample set of the experiment is simulated in PSCAD/EMTDC, and the implementation framework is PyTorch 1.5.0 deep learning platform.

3.3 Experimental data acquisition

For engineering applications in which the time and location of cable faults or disturbances are uncertain, and at the same time the load of each line is different, in order to obtain a more realistic massive data set, the real-time digital simulation system RTDS is utilized to build a simulation model of the distribution network, and the following simulation conditions are set:

(1) All types set two ways of ungrounded and resonant grounding, and three kinds of loads: light load, heavy load and full load.

(2) Early faults set 5 cable fault lines, each line set 4 fault locations, the moment of fault occurrence randomly selected before and after the voltage peak value of 0.002s within 10 moments, set the arc resistance of 40 ohms, 120 ohms and 200 ohms in three cases, semi-circular early faults and multi-circular early faults, respectively, there are 2,174 groups.

(3) Motor casting and cutting selected 5 casting and cutting lines, set the input and removal of two kinds of state, casting and cutting 5 groups of different power size of motor, fault casting

and cutting moments for the random selection of 10, motor casting and cutting a total of 1,326 groups.

(4) Load casting and cutting has the same settings as motor casting and cutting except for different load values, and the total number of load casting and cutting is 1308 groups.

(5) For constant impedance grounding faults, five faulty lines are selected, and four fault locations are chosen for each line. The grounding resistance values are available in three scenarios: 1 ohm, 100 ohm, and 1000 ohm. The fault angles are set as 0°, 90°, 180°, 270°, and [0°, 90°], [90°, 180°], [180°, 270°], [270°]. There are a total of 1,934 groups of constant impedance grounding faults at any Angle within the 360° interval.

(6) Capacitor casting: 5 casting lines are selected, 2 states of input and removal are set, 8 groups of capacitors with different capacitance values are casted, and a total of 1128 groups of motors are casted.

The simulation is carried out according to the above settings, in order to reduce the simulation time, the Multi-Run module in PSCAD is used to realize multiple runs, and a total of 10044 groups of samples are obtained. According to the ratio of 4:1, the samples are divided into training set and test set, and the distribution of the data set is shown in Table 1.

Table 1: Sample data set distribution

Type	Sample size	Training concentration	Test set
Early failure of the semi-cycle	2174	1739	435
Early failure of multiple cycles	2174	1739	435
Fixed impedance ground fault	1934	1547	387
Motor cutting	1326	1061	265
Load cutting	1308	1046	262
Condenser cutting	1128	902	226

3.4 Experimental setup

The feature vectors of semi-circumferential early faults, fault-impedance grounding faults, and capacitor-casting faults are applied to the APSO-KPCA-SVM-based multidimensional data-fusion cable-fault identification model using the training set in order to train this model and then use the test set to test it thus achieving the accuracy that corresponds to the various features input. All the selected features are thoroughly assessed on the basis of the time spent on training the model and the accuracy of identification.

In order to confirm that the APSO-KPCA-SVM model is superior to other cable-fault recognition models such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and Deep Belief Network (DBN) employed in this paper, the legitimate variables after dimensionality reduction are also fed into various recognition models. The first three types of data considered in the experiments are semi circumferential wave early faults, fixed impedance grounding faults and the capacitor casting data, whereas motor casting, load casting, and multi-cycle early faults are incrementally increased to 4, 5 and 6 classes respectively.

3.5 Analysis of results

3.5.1 Feature validity verification

In order to verify the effectiveness of the features extracted in this paper, the acoustic wave signals, electromagnetic wave pulse signals, low-voltage pulse test signals, path finding signals and electromagnetic field signals of the early semi-circumferential faults, fixed impedance ground faults and capacitor casting are imported into the APSO-KPCA - SVM fault recognition

model after dimensionality reduction, and the comparison of recognition performance of various types of features is shown in Table 2 after training. When there are acoustic wave signal + electromagnetic wave pulse signal + low-voltage pulse test signal + path finding signal + electromagnetic field signal features as input feature vectors, although the recognition accuracy can reach the highest, the model training time consumes the most. While the accuracy of the reduced dimensionality is only 0.22% lower than the highest accuracy, the training time is reduced by 413.14 s. Considering the accuracy and the training time, the reduced dimensionality of the kernel principal component analysis method has a better discriminative performance, so the reduced dimensionality of acoustic wave signals + electromagnetic wave pulse signals + low-voltage impulse test signals + path-finding signals + electromagnetic field signals is chosen as the input feature vector of the APSO-KPCA **-**SVM. KPCA**-**SVM input feature vectors can effectively complete the cable fault identification.

Table 2: Comparison of diagnostic performance of various features

Signal characteristic	Accuracy rate (%)	Training time (s)
Sound signal	41.25	213.38
Electromagnetic pulse signal	71.94	89.34
Low voltage pulse test signal	78.06	99.41
Path finding signal	43.59	125.97
Electromagnetic signal	76.35	69.07
Sound signal + Electromagnetic pulse signal + Low voltage pulse test signal + Path finding signal + Electromagnetic signal	97.63	532.77
Sound signal + Electromagnetic pulse signal + Low voltage pulse test signal + Path finding signal + The electromagnetic field signal is descending dimensional	97.41	119.63

3.5.2 Validation of the superiority of the identification methods

Each recognition model learns the training set and then recognizes cable faults on the test set, and the results are shown in Table 3.The average recognition accuracy of APSO-KPCA **-** SVM is 98.17%, which is the highest among the five methods.The average recognition accuracy of DBN is 94.82%, which is second only to APSO-KPCA **-** SVM.According to the above analysis, the recognition accuracy of the multidimensional data fusion cable fault recognition model based on APSO-KPCA **-** SVM is better than the other four methods in terms of feature extraction and processing nonlinear data, and has better robustness and stability. KPCA**-**SVM multidimensional data fusion cable fault identification model has better identification accuracy than the other four methods, more robustness and stability in feature extraction and processing nonlinear data, and better generalization ability.

Table 3: Diagnostic performance comparison of various methods

Identification method		APSO-KPCA**-**SVM	SVM	RF	KNN	DBN
Class 3	Accuracy rate (%)	97.83	85.33	78.23	67.30	95.71
	Training time (s)	167.99	138.47	181.33	111.06	150.72
Class 4	Accuracy rate (%)	98.19	80.68	78.07	62.65	96.08
	Training time (s)	232.56	152.54	272.26	138.54	232.88
Class 5	Accuracy rate (%)	99.83	93.38	74.24	62.62	96.20
	Training time (s)	306.63	188.83	296.34	178.63	304.90
Class 6	Accuracy rate (%)	96.83	85.74	78.01	61.48	91.27
	Training time (s)	380.49	198.08	356.14	221.69	349.13

3.5.3 Discriminative method anti-interference analysis

In the actual field, the collected signal data is often disturbed by various factors in the environment. Various power equipment in the distribution network generates noise during its operation, which makes the signal collected by the data acquisition device also carry a large amount of noise. These noises can distort the signal waveforms, thus masking the key features of the signal. In order to verify the anti-interference performance of the cable fault identification model, it is necessary to consider the fault samples carrying noise. In addition, due to the backward level of power system intelligence in some areas, the signal data collected by the monitoring system may be lost. If the lost data happens to have key features, it will seriously affect the fault discrimination. Therefore, the validation of the interference resistance of the fault identification model also needs to consider the situation of signal data loss. In this section, SVM and DBN are selected as comparison models.

(1) Noise Resistance Analysis

In order to simulate the actual data carrying noise, Gaussian white noise with signal-to-noise ratios of 40dB, 30dB, 20dB, and 10dB is added to the original data of each type of fault in MATLAB. The noise-containing fault samples are input into the three fault identification models for testing after feature extraction. The four types of evaluation indexes of the three fault identification models under the influence of different signal-to-noise ratios of noise are calculated separately, and the results are shown in Table 4. As the signal-to-noise ratio decreases, the cable fault identification accuracy of the three models decreases. Among them, the accuracy of SVM decreases most severely, falling below 90% under the influence of noise up to 20 dB. The cable fault identification accuracy of DBN decreases at a relatively slower rate. Although the cable fault identification accuracy of APSO-KPCA - SVM also decreases, its cable fault identification performance is still better than the other two models. In addition, the other three evaluation indexes of APSO-KPCA-SVM are larger than those of SVM and DBN, and combining the four evaluation indexes, APSO-KPCA-SVM is still able to recognize early cable faults under the influence of signal with noise.

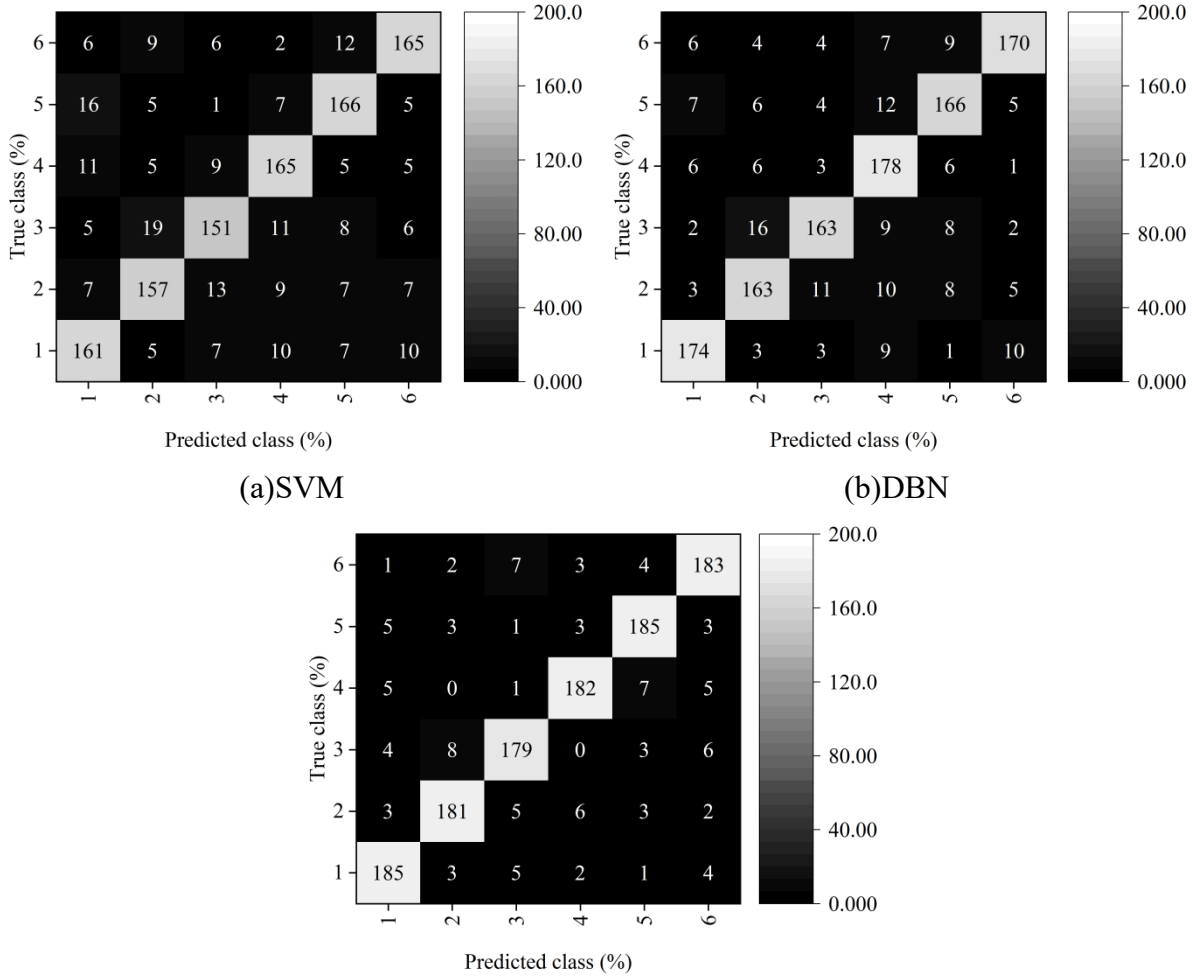
Table 4: Evalution results of network anti-noise

Noise(dB)	Model	P_{acc} (%)	P_{pre} (%)	P_{rec} (%)	F_1 (%)
40	SVM	90.59	89.44	90.75	88.75
	DBN	93.67	95.49	93.48	93.36
	APSO-KPCA-SVM	98.45	98.02	96.93	98.57
30	SVM	91.57	90.58	89.57	88.26
	DBN	92.13	93.14	92.39	91.97
	APSO-KPCA-SVM	96.67	97.29	94.95	95.51
20	SVM	87.97	86.41	86.21	85.12
	DBN	90.86	88.03	88.46	88.83
	APSO-KPCA-SVM	95.01	95.42	94.14	93.75
10	SVM	85.08	83.60	83.22	82.50
	DBN	86.08	85.99	85.40	83.87
	APSO-KPCA-SVM	91.69	89.75	88.03	87.90

(2) Anti-data loss analysis

In order to simulate the data loss situation that may exist in practice, 200 groups of data are selected from each of the six types of cable fault test sets, and data loss is performed on 2000 sampling points including fault signals. Due to the short duration of half-cycle early faults and capacitive switching disturbance fault signals, a large number of data loss will result in the loss

of all the characteristic information of these two types of faults. Therefore only a small percentage of data loss is considered. The proportion of lost data is chosen to be 8%, i.e., 160 sampling points are lost. In addition, in order to consider the randomness of data loss, as well as to retain the original length of the sample data, the random function is used to randomly select 100 sample points to be replaced by 0 values. The processed sample data are input into the three models for testing after feature extraction. Fig. 9 shows the confusion matrix of fault identification results of individual models after data loss processing, and Figs. (a)~(c) show the results of SVM, DBN, and APSO-KPCA-SVM, respectively. Figures 1~6 are half-cycle early fault, multi-cycle early fault, fixed impedance ground fault, motor casting, load casting, and capacitor casting, respectively. When data loss occurs in the original samples, the characteristic information of the original samples is changed greatly, which in turn makes the constructed feature samples change accordingly. From the misclassification situation, it can be seen that the number of misidentifications of the six cable faults by the APSO-KPCA-SVM model proposed in this paper is generally lower than that of the SVM and DBN models.



(a)SVM (b)DBN

(c)APSO-KPCA-SVM

Figure 9: Model test results

According to the confusion matrix results, the evaluation indexes of the three models are calculated separately, and the results are shown in Table 5. As can be seen from the table, compared with the samples without data loss processing, the cable fault identification accuracy of the three models has undergone a significant decrease. SVM had the lowest cable fault

identification accuracy, and DBN cable fault identification accuracy improved compared to SVM, but still did not exceed 87%. APSO-KPCA-SVM has the highest accuracy rate of 91.04%. Combining the four evaluation metrics, APSO-KPCA-SVM still has good cable fault identification performance in response to sample data loss.

Table 5: Evaluation results of network anti data loss

Model	P_{acc} (%)	P_{pre} (%)	P_{rec} (%)	F_1 (%)
SVM	81.35	80.04	81.90	83.34
DBN	85.96	84.29	86.11	85.19
APSO-KPCA-SVM	91.04	90.91	91.37	90.68

4 Conclusion

In order to overcome the restrictions of conventional cable fault diagnosis models which consider only one dimension of the fault, this paper presents a way of identifying cable faults using multidimensional information features and APSO-KPCA-SVM, and the proposed approach offers the following:

(1) The acoustic signal, electromagnetic wave pulse signal, low-voltage pulse test signal, path-finding signal, and electromagnetic field signal of cable faults can be used as the input feature vectors of the APSO-KPCA-SVM fault identification model after dimensional reduction, which can make the model's comprehensive ability of identification better. Compared with the signal features without dimensionality reduction, the accuracy of the model after dimensionality reduction is not much different, but the training time is saved by 77.55%.

(2) Compared with SVM, RF, KNN and DBN, the model in this paper performs better in recognizing six kinds of cable fault information at the same time, namely, half-circumferential early faults, fixed impedance ground faults and capacitor casting, motor casting, load casting and multi-circumferential early faults, and the accuracy of the model in recognizing them is 98.17%.

(3) In the anti-interference performance analysis, in the signal-to-noise ratio of 40dB, 30dB, 20dB, 10dB, this paper's model is always higher than the recognition accuracy of SVM and DBM. The cable fault identification accuracy of this paper's model is the highest in the anti-data loss analysis, reaching 91.04%, followed by the accuracy of DBN.

Funding

This work was supported by the Science and Technology Project of China Southern Power Grid Company Limited under Grant 035100KC23070001 (GDKJXM20230758).

References

[1] Bandara, S., Rajeev, P., & Gad, E. (2024). A review on condition assessment technologies for power distribution network infrastructure. Structure and Infrastructure Engineering, 20(11), 1834-1851.

[2] Li, Y., Chen, Q., Strbac, G., Hur, K., & Kang, C. (2023). Active distribution network expansion planning with dynamic thermal rating of underground cables and transformers. IEEE Transactions on Smart Grid, 15(1), 218-232.

[3] Ali, K. H., Aboushady, A. A., Bradley, S., Farrag, M. E., & Maksoud, S. A. A. (2022). An industry practice guide for underground cable fault-finding in the low voltage distribution network. IEEE Access, 10, 69472-69489.

[4] Kim, J. H., Kim, J. Y., Cho, J. T., Song, I. K., Kweon, B. M., Chung, I. Y., & Choi, J. H. (2014). Comparison between underground cable and overhead line for a low-voltage direct current distribution network serving communication repeater. Energies, 7(3), 1656-1672.

[5] Bandara, S., Rajeev, P., & Gad, E. (2023). Power distribution system faults and wildfires: Mechanisms and prevention. Forests, 14(6), 1146.

[6] Ibrahim, A. H. M., Sadanandan, S. K., Ghaoud, T., Rajkumar, V. S., & Sharma, M. (2024). Incipient fault detection in power distribution networks: review, analysis, challenges and future directions. IEEE Access.

[7] Su, J., Wei, L., Zheng, J., Liu, J., Zhang, P., Pang, X., & Xing, Y. (2022). Effects of mechanical stress on insulation structure and performance of HV cable. Polymers, 14(14), 2927.

[8] Li, G., Liang, X., Zhang, J., Li, X., Wei, Y., Hao, C., ... & Li, S. (2022). Insulation properties and interface defect simulation of distribution network cable accessories under moisture condition. IEEE Transactions on Dielectrics and Electrical Insulation, 29(2), 403-411.

[9] Li, X., Liu, T., Sun, W., Liang, X., Wei, Y., Hao, C., ... & Li, G. (2023). Influence and mechanism analysis of acid or alkali damp environment on insulation performance of distribution cable accessories. Engineering Failure Analysis, 152, 107469.

[10] Buhari, M., Levi, V., & Awadallah, S. K. (2015). Modelling of ageing distribution cable for replacement planning. IEEE Transactions on Power Systems, 31(5), 3996-4004.

[11] Zapf, M., Blenk, T., Müller, A. C., Pengg, H., Mladenovic, I., & Weindl, C. (2021). Lifetime assessment of PILC cables with regard to thermal aging based on a medium voltage distribution network benchmark and representative load scenarios in the course of the expansion of distributed energy resources. Energies, 14(2), 494

[12] Clavijo-Blanco, J. A., & Rosendo-Macías, J. A. (2020). Failure rates in distribution networks: Estimation methodology and application. Electric Power Systems Research, 185, 106398.

[13] Shaalan, E. M., Ward, S. A., & Youssef, A. (2021, December). Analysis of a practical study for under-ground cable faults causes. In 2021 22nd International Middle East Power Systems Conference (MEPCON) (pp. 208-215). IEEE.

[14] Zhu, K., & Pong, P. W. (2019). Fault classification of power distribution cables by detecting decaying DC components with magnetic sensing. IEEE Transactions on Instrumentation and Measurement, 69(5), 2016-2027.

[15] Carmo, E. C., da Silva, L. A., & Maia, T. A. (2025). Survey on incipient fault localization methods in underground cables. Computers and Electrical Engineering, 123, 109961.

[16] Rengaraj, R., Venkatakrishnan, G. R., Shalini, S., Subitsha, R., Suganthi, S., & Carolyn, S. S. (2021, July). Identification and classification of faults in underground cables–A review. In IOP Conference Series: Materials Science and Engineering (Vol. 1166, No. 1, p. 012018). IOP Publishing.

[17] Ni, Y., Zhang, Q., & Xin, R. (2021). Magnetic flux detection and identification of bridge cable metal area loss damage. Measurement, 167, 108443.

[18] Mo, S., Zhang, D., Li, Z., & Wan, Z. (2021). The possibility of fault location in cross-bonded cables by broadband impedance spectroscopy. IEEE Transactions on Dielectrics and Electrical Insulation, 28(4), 1416-1423.

[19] Tariq, R., Alhamrouni, I., Rehman, A. U., Tag Eldin, E., Shafiq, M., Ghamry, N. A., & Hamam, H. (2022). An optimized solution for fault detection and location in underground cables based on traveling waves. Energies, 15(17), 6468.

[20] Bascom, E. C. R., von Herrmann, M. J., & Zhao, T. T. (2014, April). Power cable failure modes and fault location methods, practices and strategies. In 2014 IEEE PES T&D Conference and Exposition (pp. 1-5). IEEE.

[21] Wu, Q., & Liu, Z. (2013). Design of a high-frequency high-voltage pulse power supply for cable fault detection. Australian Journal of Electrical and Electronics Engineering, 10(2), 199-208.

[22] Sheng, B., Zhou, C., Hepburn, D. M., Dong, X., Peers, G., Zhou, W., & Tang, Z. (2015). A novel on-line cable pd localisation method based on cable transfer function and detected pd pulse rise-time. IEEE Transactions on Dielectrics and Electrical Insulation, 22(4), 2087-2096.

[23] Li, G., Zeng, S., Wang, Q., & Zhang, Z. (2024). Distribution network cable detection based on terahertz pulse and imaging. Russian Journal of Nondestructive Testing, 60(3), 318-325.

[24] Liu, Y., Shi, Y., Guo, J., & Wang, Y. (2018). Application of pulse compression technique in fault detection and localization of leaky coaxial cable. IEEE Access, 6, 66709-66714.

[25] Wang, Y., Ma, X., Zhao, L., Li, H., & Liu, J. (2020, June). Analysis of power cable fault diagnosis and electric field detection technology based on computer control system. In Journal of Physics: Conference Series (Vol. 1574, No. 1, p. 012080). IOP Publishing.

[26] Preduş, M., Spunei, E., & Piroi, I. (2016, October). A study on failure diagnosis methods in power cables and their applications. In 2016 International Conference on Applied and Theoretical Electricity (ICATE) (pp. 1-6). IEEE.

[27] Fu, H., Qiu, L., Ai, Y., Tu, J., & Yan, Y. (2025). Deep learning-based fault detection and location in underground power cables using resonance frequency analysis. Electrical Engineering, 107(4), 4051-4062.

[28] Maduako, I., Igwe, C. F., Abah, J. E., Onwuasaanya, O. E., Chukwu, G. A., Ezeji, F., & Okeke, F. I. (2022). Deep learning for component fault detection in electricity transmission lines. Journal of Big Data, 9(1), 81.

[29] Guomin, L., Yingjie, T., Changyuan, Y., Yinglin, L., & Jinghan, H. (2019). Deep learning-based fault location of DC distribution networks. The Journal of Engineering, 2019(16), 3301-3305.

[30] Zhou, Q. (2024). Research on Fault Diagnosis Algorithm of Power Cable Based on Deep Learning. J. Electrical Systems, 20(3), 333-343.

[31] Shao, Q., Fan, S., Zhang, Z., Liu, F., Fu, Z., Lv, P., & Mu, Z. (2025). Artificial intelligence in cable fault detection and localization: Recent advances and research challenges. Energies, 18(14), 3662.

[32] Wu, Y., Zhang, P., & Lu, G. (2021). Detection and location of aged cable segment in underground power distribution system using deep learning approach. IEEE Transactions on Industrial Informatics, 17(11), 7379-7389.

[33] Said, A., Hashima, S., Fouda, M. M., & Saad, M. H. (2022). Deep learning-based fault classification and location for underground power cable of nuclear facilities. Ieee Access, 10, 70126-70142.

[34] Zhang, Y., Mei, W., Dong, G., Gao, J., Wang, P., Deng, J., & Pan, H. (2018). A cable fault recognition method based on a deep belief network. Computers & Electrical Engineering, 71, 452-464.

[35] Saad, M. H., Hashima, S., Omar, A. I., Fouda, M. M., & Said, A. (2025). Deep learning approach for cable partial discharge pattern identification. Electrical Engineering, 107(2), 1525-1540.