International Journal of Power and Energy Systems, Vol. 46, No. 10, 2026
An Open Access Paper
Special Issue
Next Generation Energy Systems

HIGH VOLTAGE CABLE FAULT IDENTIFICATION METHOD BASED ON BAGGING SAMPLING AND IMPROVED K-NEAREST NEIGHBOR ALGORITHM

Yingchun Xu,* Xiangmao Cheng,* Peifeng Huang,* Runjie Lin,* Yangrui Lin,* and Jia Weng*

Abstract

As a key transmission channel in the power grid, fault diagnosis of high-voltage cables directly affects the safety and reliable operation of the power system. However, traditional methods are susceptible to noise interference and have limited generalisation ability when processing high-dimensional cable monitoring data. A K-nearest neighbour cable fault diagnosis method based on a Bagging sampling strategy and adaptive sample weighting is proposed to address this issue. This method improves the ability to discriminate low-dimensional signals by embedding a learning process that preserves global and local structural features. It also uses a diverse ensemble of subclassifiers to improve adaptability to complex fault modes. Experimental verification showed that the proposed method achieved a recognition accuracy of 95.7% on the training set and 93.7% on the test set, both of which were superior to the other two comparison methods. In terms of computational resources, this method was computationally efficient while maintaining diagnostic accuracy. It had an average memory usage of 24 MB and a single inference time of 2.3 s. This effectively reduced the risk of misjudging faults and made online deployment feasible. The research results can provide reliable technical support for the intelligent operation and maintenance of power systems.

Key Words

High voltage cables; Fault identification; Bagging; K-nearest neighbour algorithm; K-means

* Chaozhou Power Supply Bureau, Middle Section of Xinyang Road, Chaozhou City, Guangdong Province, China; e-mail: xuyc8090@outlook.com, 13827337619@139.com, 13632014054@163.com, linrunjie0088@163.com, linyangrui@163.com, 13828328292@163.com
Corresponding Author: Yingchun Xu

Recommended by Zhenling Liu
(DOI:10.2316/J.2026.203-0659)

1. Introduction

Rapid and precise fault diagnosis is especially important as the power system's scale continues to grow and power supply reliability standards continue to rise. High-voltage cables are the main piece of equipment used for power transmission [1]. The high-voltage cable laying environment is complex, and the operational state is hidden. Once a fault occurs, not only is it difficult to locate, but it may also cause a wide range of power supply interruptions. This will cause serious economic losses and security risks [2]. Traditional signal processing methods, such as the analysis means based on the Fourier or wavelet transform, often suffer from insufficient recognition accuracy and fuzzy classification boundaries when facing complex cable fault waveforms in non-smooth and strong noise environments. Therefore, establishing a set of sophisticated and effective problem identification mechanisms is crucial to ensuring the electricity system operates steadily. Numerous academics have conducted studies on this. Zhang et al., for instance, suggested a defect detection technique based on a residual network and an enhanced Graham angle field for cable terminals installed on vehicles. The method converted the one-dimensional localized discharge signal into a two-dimensional image by converting it into a two-dimensional image. It also constructed a ResNet101 network to integrate anti-aliasing downsampling and attention mechanism, and combined the centre loss with the Softmax loss function to achieve 97.3% defect recognition accuracy [3]. Lacerda V A et al. proposed a novel distance protection algorithm suitable for modular multilevel converters in high-voltage DC power grids. This algorithm identified faulty cables by detecting the resonant frequency of DC capacitors. It was validated in a four-terminal, high-voltage DC power grid containing five cables. Results showed that the algorithm could quickly and selectively locate cable faults without terminal communication and perform reliably under strong noise conditions [4]. Peng N et al. proposed a hybrid line fault identification method to elim-

inate cross-node interference through modal domain analysis. The method first established a unified model of a hybrid line, proposed the concept of decoupled transient energy, and realised fault localisation based on the energy ratio feature. Simulation experiments verified the effectiveness of the method [5]. Zaky M S et al. proposed a fault section location method for non-uniform high-voltage direct current lines with hybrid cable overhead lines. This method identified the type of fault in cable or overhead line sections by analysing the distribution pattern of voltage signals at both ends of the line. This method did not require calculation of the specific fault distance and could effectively deal with various fault types and resistances [6].

In recent years, the combination of integrated learning and pattern recognition technology has provided new ideas for intelligent diagnosis of power systems [7]. Among them, Bagging uses self-sampling to construct multiple differentiated training subsets. Parallel training of independent classifiers followed by integrated decision-making effectively improves the stability and generalisation performance of the model [8]. An intelligent protection approach was presented by Raju G V et al. that combined regression-based Bagging integrated learning for defect localisation with a fuzzy logic system for problem detection and classification. The scheme extracted fault features by applying the discrete Fourier transform to renewable energy bus voltage signals. The features were also analysed using a Mamdani fuzzy inference system to determine the fault state. Experiments indicated that the scheme could accurately identify the fault type in one cycle under dynamic generation conditions with a detection accuracy of 99.56%, which verified its reliability [9]. Solimun et al. used the Bagging integration method to compare and analyse the performance of Bagging discriminant and Logistic regression Bagging based on bank mortgage data and simulated data. The experimental results indicated that both integration methods outperformed traditional discriminant analysis and Logistic regression in terms of classification accuracy, sensitivity, and specificity. Especially, it demonstrated better identification ability in credit recoverability assessment [10].

The K-nearest neighbour (KNN) algorithm is a typical nonparametric classification method. It is widely used in various kinds of fault diagnosis and state identification tasks because of its simple implementation and no need for a training process [11]. The method accomplishes classification mainly by measuring the distance between test samples and training samples, selecting the nearest K neighbours, and according to the majority voting principle. Abdul Z K et al. proposed a fuzzy monotone KNN method, which innovatively utilises the fuzzy dominance relationship between instances to construct a monotone classifier. By dynamically adjusting the fuzzy dominance threshold, the method effectively reduced the influence of noise on KNN selection. Experimental data revealed that the method achieved an average improvement of 28%, 11%, and 29% in ACCU, MAE, and NMI metrics, respectively. This significantly outperformed the existing monotonic classifiers [12]. Sukshitha R proposed a KNN regression algorithm based on empirical likelihood ratio (ELR). The algorithm was more advantageous in dealing with skewed distributions and suppressing outlier interference than the traditional distance metrics. Simulation and practical application validation indicated that the ELR-KNN model performed significantly better than the traditional KNN method [13].

In summary, existing research has introduced a variety of deep feature extraction methods and integrated classification frameworks in high-voltage cable fault identification, and achieved certain results. However, most of them still have the problems of complex model structure, low inference efficiency or insufficient consideration of data distribution structure. Especially, the recognition accuracy under high-dimensional redundant feature processing and sample heterogeneous distribution still needs to be improved. In view of this, the study innovatively proposes an improved KNN fault identification method that combines Bagging sampling with the introduction of a weighting mechanism. Moreover, the feature extraction strategy of global and local structure preservation is introduced in the dimensionality reduction stage to enhance the generalisation ability (GA) and discriminative performance of the model in high-dimensional space. The construction of sample subsets is guided by clustering to enhance the diversity of subclassifiers, aiming to comprehensively improve the intelligence and practicality of the high-voltage cable fault diagnosis system.

2. Methods and Materials

2.1 Bagging Integrated Classification Model Combined with K-means

The study builds a classification model that combines K-means clustering with the Bagging integration technique in an attempt to enhance the fault detection model's generalisation capability and classification stability. Bagging constructs multiple classifiers and integrates the prediction results through subsampling. Its working mechanism is shown in Figure 1 [14].

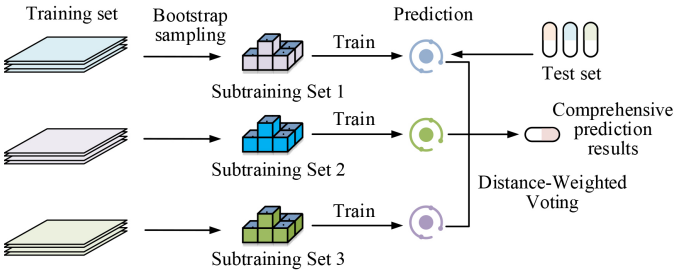

Figure 1. Bagging Working Mechanism (Icon source: https://iconpark.oceanengine.com/official)

In Figure 1, the mechanism first divides the original training set into feature sets and label sets. Then, it performs self-sampling on the original training set. Through a sampling method with replacement, it randomly selects

samples from the overall data, forming several sub-training sets. There is partial overlap between different subsets, but overall differences are maintained to increase sample diversity. Each sub-training set is used to train a base classification model. This improves robustness to noise and sample imbalance by enabling the model to independently learn the feature distribution of each subsample. Subsequently, each base classifier independently predicts the test set samples and outputs the classification results and their confidence levels. These prediction results are eventually integrated into the final classification decision through voting or weighted fusion. This improves the GA and the robustness of the overall model. However, the classification performance of every single classifier has certain limitations. To improve the classification accuracy, the study adopts an integrated learning approach to construct a strong classifier by comparing multiple weak classifiers two by two and fusing the results, as shown in Figure 2 [15].

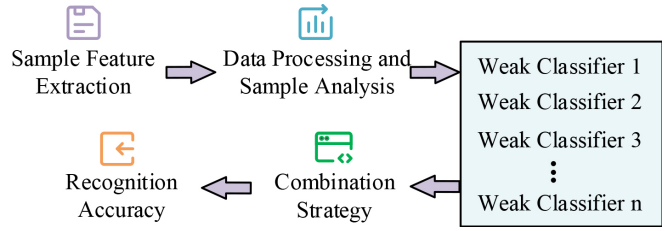

Figure 2. Bagging Integration Process (Icon source: https://iconpark.oceanengine.com/official)

Figure 2 shows the process of first extracting working condition features from the original cable monitoring data and then processing the data and analysing the samples. This includes normalisation, dimensionality reduction, and noise suppression to enhance sample separability. Then input the processed data into multiple weak classifiers (Weak Classifiers 1∼n) with differential learning ability for independent training. To enhance classification diversity and stability, there may be differences between weak classifiers in sampling distribution, model structure, or neighbourhood feature space. Finally, a combination strategy is used to comprehensively output the prediction results of all weak classifiers as recognition results. This reduces the risk of errors from single classifiers by using distance-weighted voting or confidence fusion mechanisms. To further reduce the information loss, the study builds a balanced subset of training base learners based on the Bagging integration method combined with K-means clustering sampling. First, clustering is performed on the training data, dividing samples with similar spatial distributions into the same cluster. Then, a random sample is taken from each cluster according to the sampling ratio. This ensures that each sub-classifier covers different fault modes, improving overall robustness. The K value is optimised and selected based on the intra-class contour coefficient, which can balance the ability of local feature mining with the cost of clustering calculation. The pseudocode is as follows:

K-means searches for the optimal cluster division through iterative optimisation and divides the samples into a number of clusters based on the intrinsic relationship of the data. The sum of squares of the errors of each sample from the centre of the cluster it belongs to is shown in Equation (1) [16].

Algorithm 1 K-means + Bagging

Input: Training dataset D, number of clusters K, sampling ratio ρ, number of base learners B
Output: Ensemble classifier H

Perform K-means clustering on $D \rightarrow \{C_1, C_2, \ldots, C_K\}$
for $b = 1$ to B **do**
  Initialize subset $D_b = \emptyset$
  for each cluster C_i **do**
    Randomly sample $\rho \times |C_i|$ samples $\rightarrow S_i$
    $D_b = D_b \cup S_i$
  end for
  Train a base KNN classifier h_b based on GLSP embedding of D_b
end for
Combine predictions of all h_b via weighted voting $\rightarrow H$
return H

$$E = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \bar{x}_i\|_2^2 \tag{1}$$

In Equation (1), E denotes the error squared. C_i denotes the set of sample points contained in the ith cluster. x denotes the sample points of the ith cluster. $\bar{x}_i$ is the mean vector of the ith cluster and K denotes the quantity of clusters clustered. The K-means clustering-based stratified sampling integration strategy divides the dataset into multiple strata with internal homogeneity through clustering. Subsequently, random sampling is performed within each stratum, which maintains the data distribution characteristics and enhances the inter-stratum variability, as shown in Figure 3 [17].

In Figure 3, the sample distribution of the clusters generated by K-means clustering is different, and an adaptive sampling strategy needs to be adopted. The degree of data dispersion reflected in the intra-cluster variance determines how many samples need to be extracted. Dispersed clusters with larger variance require more samples, while dense clusters with smaller variance require fewer. This effectively maintains the structural characteristics of the original data. The formula for calculating the number of samples taken from cluster C_i is shown in Equation (2).

$$\begin{cases} S_i = N_{\max} \cdot \dfrac{n_i \cdot \sigma_i^2}{\sum_{i=1}^{K} n_i \cdot \sigma_i^2} \\ \hat{\sigma}_i^2 = \dfrac{\sigma_i^2}{\max\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_i^2\right)} \end{cases} \tag{2}$$

In Equation (2), S_i denotes the sampling volume of the ith class in the subsample set. N denotes the total quantity of samples. n_i denotes the sample base. σ_i^2 denotes the sample variance of the ith cluster or class. $\hat{\sigma}_i^2$ represents the normalized variance. The normalised variance is used

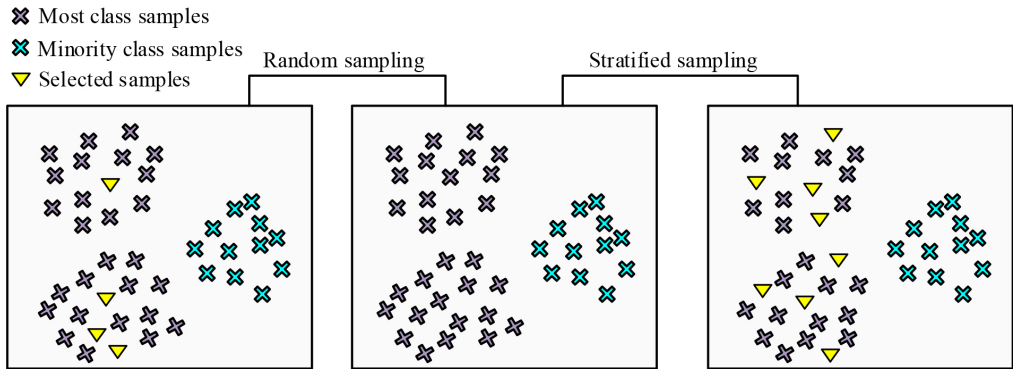

Figure 3. Random Sampling and Stratified Sampling

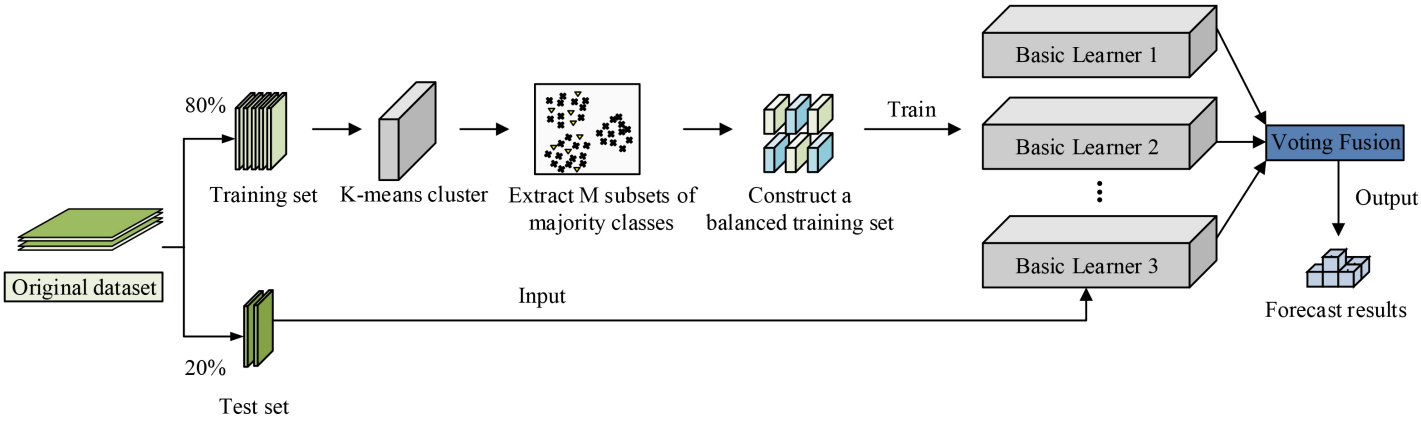

Figure 4. K-Bagging Framework

to calculate the number of sample selections, thereby limiting the abnormal amplification effect of variance caused by sample scarcity. To optimise data utilisation, the study uses the Bagging integration framework to construct diversity training subsets through multiple rounds of cluster sampling. Each subset is combined with a few class samples to form a balanced training set. It also trains M base classifiers and fuses the prediction results through a voting mechanism. The decision output of the final classifier is shown in Equation (3).

$$Q(x) = sign(\sum_{j=1}^{M} q_j(x)) \tag{3}$$

In Equation (3), $Q(x)$ denotes the final integrated classifier output result. $q_j(x)$ denotes the prediction result of the jth base classifier for x. $sign(\cdot)$ is the sign function. We propose a K-means enhanced Bagging method (K-Bagging). The framework is shown in Figure 4.

In Figure 4, K-Bagging randomly divides the original dataset into training and testing sets in the ratio of 8:2, keeping the distribution of classes consistent. Subsequently, K-means clustering is performed on the majority class samples to obtain K clusters. Samples are drawn from each cluster using a putative back sampling strategy to generate M majority class subsets. Each subset is combined with the minority class samples to form a balanced training set for training M base classifiers. Finally, the test set is predicted and output by voting integration.

2.2 Fault Identification Model Based on Improved KNN Algorithm

The degree of data dispersion, as reflected by the intra-cluster variance, determines the number of samples that need to be extracted. Clusters that are more dispersed with larger variance require more samples, while dense clusters with smaller variance require fewer. Kernel principal component analysis (KPCA), as a classical nonlinear feature extraction method, preserves global structural features by maximising the data variance. Kernel locality preserving projection (KLPP) preserves the local structure of the data by constructing sample nearest neighbour relationships [18]. Combining the advantages of KPCA and KLPP, the study proposes a global and local structure-preserving algorithm (GLSP). The objective function (OF) of this algorithm integrates global and local structure information. After introducing the kernel function, the KLPP OF is shown in Equation (4).

$$J_l(\alpha) = \min_{\alpha} \alpha^T L' \alpha \tag{4}$$

In Equation (4), $J_l(\alpha)$ denotes the loss function for the local preservation objective in KLPP. α denotes the vector of projection coefficients of the sample in the kernel space. L' is not a graph Laplace matrix. The KLPP method disre-

gards global characteristics while attempting to maintain the data's nearest-neighbour structure in high-dimensional space. Because the local OF does not explicitly incorporate the global attributes of the samples, it relies solely on local neighbourhood relationships to estimate the global structure. This results in a bias in the global features after dimensionality reduction. The global structure OF is shown in Equation (5).

$$J_g(\alpha) = \underset{\alpha}{max} \alpha^T C \alpha \tag{5}$$

In Equation (5), $J_g(\alpha)$ denotes the global structure preserving loss function in GLSP. C denotes the global covariance matrix in kernel space. KPCA achieves global feature extraction by maximising the data variance. However, it ignores local structural features, which may lead to the loss of important geometric information in low-dimensional representations. Therefore, the joint optimisation of local and global structure preservation objectives is investigated to construct a new OF that also preserves the local neighbourhood relationship. The OF is defined as shown in Equation (6).

$$J(\alpha) = \underset{\alpha}{max} (\alpha^T (C - L') \alpha) \tag{6}$$

In Equation (6), $C - L'$ is the combined matrix of global and local information. Faults such as breakdown and local discharge may occur in the operation of high voltage cables. Accurate identification of fault types is crucial for power system security. Therefore, the study adopts the KNN algorithm as the core classifier for fault diagnosis and identification. The KNN decision process is shown in Figure 5 [19].

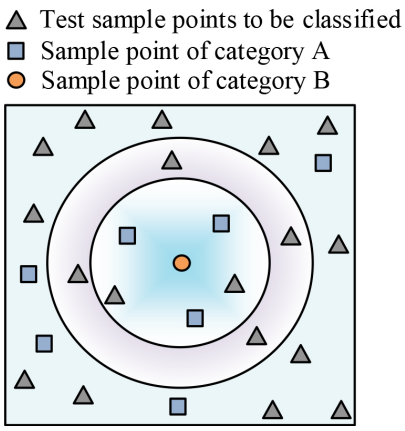

Figure 5. KNN Decision Process

In Figure 5, KNN first calculates the distance between the sample to be tested and the known sample, and selects the nearest k neighbours. Then, the category of the sample to be tested is judged according to the categories of these nearest neighbours. If the k nearest neighbours belong to the same category, they are directly categorised. Otherwise, the final category is determined according to specific rules. However, the traditional KNN algorithm is based on the Euclidean distance measure of sample similarity for categorisation. Although simple and efficient, it is prone to classification ambiguity and error when dealing with edge data. To address the computational intensity and sensitivity to category imbalance of the algorithm, the KNN distance metric is adopted for boundary samples, while the Euclidean distance is calculated for most samples. Meanwhile, the data weighting mechanism is introduced to optimise the classification effect. The confirmation of boundary samples is shown in Equation (7).

$$\left| \frac{D_1 - D_2}{D_1} \right| \leq \varepsilon \tag{7}$$

In Equation (7), D_1 and D_2 denote the Euclidean distance of the sample to be tested from other samples, respectively. ε is the boundary threshold used to determine whether the distance difference after projecting fault features belongs to a distinguishable interval. First, the threshold is selected based on the feature projection results of the validation set samples. Then, the distribution differences between intra-class and inter-class distances are calculated. The intersection point of these two distributions is then used as an initial candidate threshold. Additionally, grid search is used to optimise the threshold within a preset range to further enhance the discriminative ability. The final threshold maximises the interval between the minimum inter-class distance and the maximum intraspecies distance, thereby improving the robustness of fault boundary discrimination. In addition, the KNN algorithm has the disadvantage of high computational complexity. To optimise the efficiency, the study first identifies the edge data samples. Then, only the t nearest neighbours of these edge points are calculated. The study introduces category weight factors to consider both distance and category prior probability, thereby improving the ability to identify minority class faults. The calculation of sample point weights is shown in equation (8).

$$\begin{cases} w(D_i) = \frac{\lambda_c}{D_i} \\ \lambda_c = \frac{1/n_c}{\sum_{j=1}^{C} 1/n_j} \end{cases} \tag{8}$$

In Equation (8), w denotes the weight, λ_c denotes the category weight coefficient of the c-th class sample, n_c denotes the number of c-th class samples, and $\sum_{j=1}^{C} \frac{1}{n_j}$ is the normalization factor for the reciprocal of the number of all categories. Next, a distance weighting strategy is used, where the weights of the nearest neighbour points are inversely proportional to their distance to the point to be measured. The OF to be satisfied is shown in Equation (9).

$$F = argmax \sum_{i=1}^{k} w(D_i) \tag{9}$$

In Equation (9), F represents the final classification result. Aiming at the problems of high dimensionality of HV cable fault data, many redundant features, and complex structural distribution, the study proposes a fault recognition framework that integrates K-Bagging and an improved KNN classification strategy (KB-IKNN). The process is shown in Figure 6.

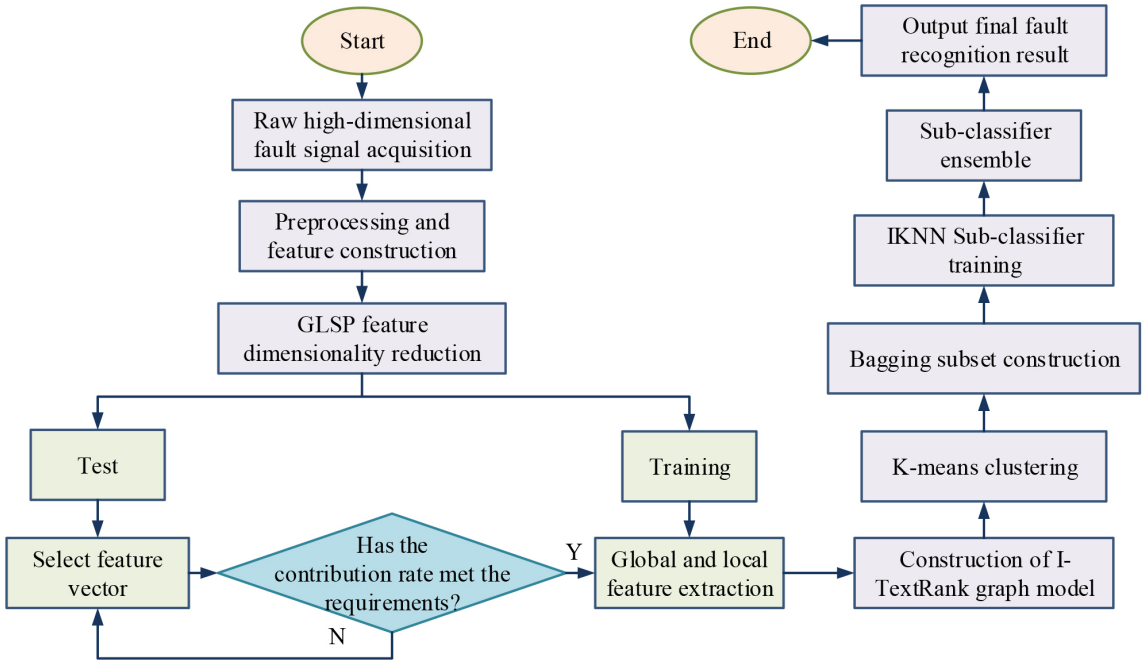

Figure 6. KB-IKNN Fault Identification Process

In Figure 6, the original high-dimensional cable fault signals collected are first preprocessed and features constructed. A low-dimensional feature subspace with greater discriminative power is extracted through GLSP dimensionality reduction, which reduces data dimensionality and model complexity while retaining key information. Second, the K-means clustering algorithm is introduced in the feature space to pre-group the samples. This clustering result is used to guide Bagging's sample sampling strategy. That is, subset partitioning is performed based on the intra-class structure. This can enhance the complementarity of subclassifiers by maintaining the sample diversity and avoiding the model redundancy problem brought by traditional random sampling. The above subsets are utilised to train the improved KNN that introduces the sample weighting mechanism, respectively. Finally, the final recognition results are output by integrating the prediction results of each subclassifier and using the weighted voting mechanism. This process fully utilises the advantages of Bagging in model robustness improvement, while overcoming the limitations of KNN in high-dimensional space. This provides an effective solution for fault identification under complex working conditions of high-voltage cables.

3. Results

3.1 Performance Test of Fault Identification Model Based on KB-IKNN

To validate the performance of the proposed KB-IKNN fault identification model for the study, the experimental platform uses an Intel Core i9-12900K processor, 32 GB DDR5 memory, and NVIDIA GeForce RTX 4090 GPU, with a Windows 11 64 -it operating system. The software environment includes Python 3.9, Scikit learn 1.3, NumPy 1.24, and Matplotlib 3.8. All experiments are conducted in a unified hardware environment to ensure comparability of performance indicators. Based on this experimental environment, the study selected the IEEE PHM dataset. The data originates from simulated and measured signals of common faults, including partial discharge, insulation breakdown, grounding short circuits, and conductor fractures, in high-voltage cables. The raw data contains approximately 4800 sets of samples, with the training and testing sets divided in an 8:2 ratio. After feature construction, each set of samples contains a 64-dimensional time-frequency feature vector, covering indicators such as amplitude, energy spectrum, envelope entropy, and phase statistics. The sample distribution of each fault type is relatively balanced: 1210 groups for partial discharge, 1175 groups for insulation breakdown, 1215 groups for grounding short circuit, and 1200 groups for conductor fracture. All algorithms run under the same random seed and parameter settings, ensuring consistency and reference value in comparisons between different methods. The performance of GLSP is compared with three dimensionality reduction methods: principal component analysis (PCA), locality preserving projection (LPP), and t-distributed stochastic neighbour embedding (t-SNE). The results are shown in Table 1.

In Table 1, GLSP has advantages in all core indicators: Precision and Recall reach 93.5% ± 0.5% and 94.2% ± 0.5%, respectively. AUC is 96.8% ± 0.3%, and F1's CI95% is only ± 0.4%, showing the smallest fluctuation range and most stable generalisation performance while maintaining high accuracy. Due to the fact that PCA only performs global linear projection, the recall rate significantly decreases to 86.9% ± 0.9%, resulting in F1 and AUC being lower than GLSP. LPP outperforms PCA overall by pre-

Table 1
Performance Comparison of Different Dimensionality Reduction Methods

Method	Precision (%)	Recall (%)	AUC (%)	F1 CI95%	p-value vs GLSP
GLSP	93.5 ± 0.5	94.2 ± 0.5	96.8 ± 0.3	93.7% ± 0.4%	/
PCA	90.1 ± 0.7	86.9 ± 0.9	93.1 ± 0.6	88.4% ± 0.7%	< 0.01
LPP	91.7 ± 0.6	89.6 ± 0.7	94.2 ± 0.5	90.6% ± 0.6%	< 0.01
t-SNE	87.4 ± 1.2	84.5 ± 1.3	91.0 ± 0.9	85.9% ± 1.1%	< 0.001

serving local adjacency relationships, with a Precision of 91.7% ± 0.6%, Recall of 89.6% ± 0.7%, an AUC of 94.2% ± 0.5%, and an F1 of 90.6% ± 0.6%. However, it still cannot balance global separability and local structural stability simultaneously, and its performance and robustness lag behind those of GLSP. Although t-SNE excels in visualisation, it performs the worst overall in classification tasks. Its precision and recall are only 87.4% ± 1.2% and 84.5% ± 1.3%, respectively. This is accompanied by the largest variance, indicating insufficient probability prediction and generalisation stability. The t-test results with GLSP confirm that the differences between the three comparison methods are statistically significant (PCA/LP: $p<0.01$; t-SNE: $p<0.001$), indicating that the advantage of GLSP is not coincidental. Overall, GLSP achieves synchronous improvements in precision and recall for identifying high-voltage cable faults. It also demonstrates more reliable engineering availability with a narrower confidence interval.

The KB-IKNN model is tested against the fault identification methods of a multi-layer neural network with multi-valued neurons (MLMVN) [20] and total least squares estimation of signal parameters via rotational invariance technique (TLS-ESPRIT) [21]. The loss functions of the different methods on the dataset are shown in Figure 7.

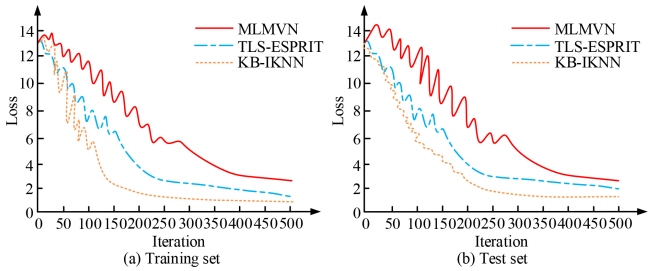

Figure 7. Different Methods of Loss Function Variation

Figures 7(a) and 7(b) show the comparison of the loss values with the number of iterations for the three methods on the training and test sets, respectively. The overall trend shows that all three methods show a decreasing trend in the loss values during the iterations. However, the ultimate performance and convergence speed vary significantly. In Figure 7(a), KB-IKNN converges significantly faster than TLS-ESPRIT and MLMVN, and the loss value decreases rapidly at the 100th iteration. Whereas TLS-ESPRIT starts to decline rapidly only after the 150th iteration, MLMVN declines more slowly overall, with the convergence tail remaining at 2.5. KB-IKNN has a loss value of 0.8 at the end of the iteration, which is better than TLS-ESPRIT and MLMVN. In Figure 7(b), KB-IKNN shows the same advantage in generalisation performance. The three methods fluctuate within the first 150 iterations. However, KB-IKNN still maintains the minimum loss value and finally converges to 1.7. Whereas TLS-ESPRIT is 2.3 and MLMVN is stabilised at 3.7. It indicates that KB-IKNN is not only fast in training, but also has a better fitting ability and robustness on unseen data. Figure 8 displays the recognition accuracy outcomes of several techniques.

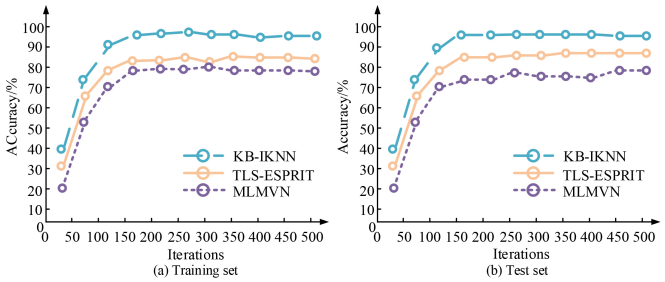

Figure 8. Recognition Accuracy Results of Different Methods

Figure 8(a) and Figure 8(b) show the trend of classification accuracy with the number of iterations for the three algorithms on the training and test sets, respectively. In Figure 8(a), KB-IKNN improves rapidly in the early iterations. Its accuracy is close to 90% at 50 iterations and eventually stabilises at 98.4% after the 200th, showing a very high fitting ability. In contrast, the accuracy of TLS-ESPRIT grows faster at the beginning, but stays at 81.8% after stabilisation. The MLMVN, on the other hand, converges even more slowly and ends up with an accuracy of only 77.5% on the training set. This indicates its relatively weak learning ability. In Figure 8(b), KB-IKNN also shows excellent generalisation performance. Its accuracy is basically stabilised at 97.6% after the 150th iteration with almost no significant oscillation. It indicates that the model's prediction ability on unseen samples is very strong. The accuracy of TLS-ESPRIT on the test set is 84.1%, which is slightly lower than its performance on the training set but still stable. MLMVN's accuracy on the test set is stable at 78.6%, with some tendency of overfitting. To verify the statistical reliability of performance differences, multiple independent experiments (N=10) are conducted on the classification indicators of each model, and a two-tailed t-test is used to compare the differences between KB-IKNN and the comparison method. The study takes the precision rate, recall rate, F1 value, and running time of classification as indicators. Table 2 displays the test results.

Table 2
Multiple Indicator Test Results

Data set	Algorithm	Precision/%	Recall/%	F1/CI95%	Running time/s	p-value (vs KB-IKNN)
Training set	MLMVN	75.3	74.1	74.7 ± 0.7	12.6	< 0.001
	TLS-ESPRIT	84.7	86.2	85.4 ± 0.5	18.3	< 0.01
	KB-IKNN	95.1	96.4	95.1 ± 0.3	20.9	/
Test set	MLMVN	72.8	71.2	72.0 ± 0.9	13.1	< 0.001
	TLS-ESPRIT	82.3	80.9	81.6 ± 0.6	18.7	< 0.01
	KB-IKNN	93.5	94.2	93.7 ± 0.4	21.4	/

In Table 2, there are significant differences in the performance of the three algorithms on the training and testing sets, especially in key classification indicators such as accuracy, recall, and F1 score. KB-IKNN performs the best. On the training set, KB-IKNN achieved an accuracy of 95.1% and a recall of 96.4%. The 95% confidence interval of the F1 score was narrow, at 95.7% ± 0.3%. This indicates that the model has a strong discriminative ability and stable convergence performance for training samples. On the test set, KB-IKNN still maintains the strongest generalisation performance, with an F1 value of 93.7% ± 0.4%, significantly better than TLS-ESPRIT and MLMVN. The differences between the two and KB-IKNN are significantly verified by a two-tailed t-test (p<0.01), indicating that its performance improvement under unknown conditions is statistically reliable. Especially, the accuracy and recall of KB-IKNN maintain a good balance, with 93.5% and 94.2%, respectively. On the one hand, it can effectively reduce the rate of false alarms and prevent unnecessary economic losses caused by erroneous shutdowns or the false triggering of maintenance. On the other hand, it can lower the probability of missing serious faults and reduce safety risks caused by insulation breakdown or short-circuit tripping. Therefore, the significant advantage of the F1 value reflects the comprehensive guarantee ability of this method for safety and economy in power system protection scenarios. In summary, KB-IKNN strikes a better balance between recognition accuracy, prediction credibility, and operational efficiency. This balance aligns more closely with the comprehensive requirements of "high security, high reliability, and deployability" for power grid operation and maintenance.

3.2 Effect of KB-IKNN-Based Fault Type Recognition Application

To verify the practical application value of the KB-IKNN model in the task of identifying fault types in high-voltage cables, the study is based on the cable simulation test platform of an electric power research to conduct experimental comparison tests. The platform simulates typical faults under actual working conditions, including partial discharges, insulation breakdowns, grounded short circuits, and conductor breaks by building 220 kV high-voltage AC cable experimental lines. To facilitate a comparative analysis of the distribution ability of different models in the feature space, the t-distributed stochastic neighbour embedding (t-SNE) algorithm is used to visualise the low-dimensional feature space output by the three methods. Among them, the dimensionality reduction target dimension of t-SNE is set to 2, the perplexity is set to 30, the learning rate is 200, and the maximum iteration step is 1000 to ensure the stability and interpretability of the projection results. The study introduces the boundary class separation ratio (BCSR). The BCSR is a quantitative indicator of interpretability. It can objectively evaluate the separation effect of each fault category in the decision space. The larger the BCSR value, the more obvious the inter-class separation and the more reliable the predicted boundary. The fault signal classification test results of the three methods are shown in Figure 9.

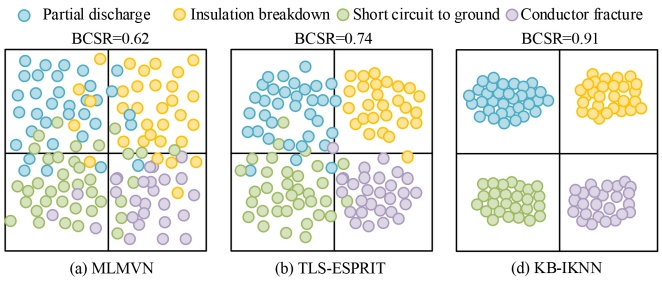

Figure 9. Fault Signal Classification Test Results

Figure 9(a) and Figure 9(b) show the visualised distributions of the results of dimensionality reduction processing and clustering distributions of the three methods in different fault identifications, respectively. In Figure 9(a), the feature distributions of the MLMVN method after dimensionality reduction have obvious category overlap, especially the insulation breakdown and conductor breakage are less distinguished. In Figure 9(b), the TLS-ESPRIT method shows a better clustering effect. Although the clustering effect of some categories is enhanced, there is still overlap between the partial discharge and grounded short circuit categories. It indicates that there is still inter-class interference after its dimensionality reduction. In Figure 9(c), the KB-IKNN method has the most superior results. The four classes of fault samples are compactly distributed with clear boundaries and almost no cross-region, and the clustering structure is obviously better than other methods. It shows that the proposed KB-IKNN model has stronger expressive ability and robustness in feature extraction and classification discrimination. The method can effectively maintain the structural information of high-dimensional fault data and realise efficient fault type recognition. In addition, the BCSR of MLMVN, TLS-ESPRIT,

Table 3
Mean Confidence Per Class With Standard Deviation

Fault type	Method	Mean confidence	Std confidence
Partial discharge	MLMVN	0.553	0.170
	TLS-ESPRIT	0.590	0.164
	KB-IKNN	0.596	0.161
Insulation breakdown	MLMVN	0.542	0.173
	TLS-ESPRIT	0.556	0.162
	KB-IKNN	0.617	0.166
Short circuit to ground	MLMVN	0.552	0.178
	TLS-ESPRIT	0.542	0.173
	KB-IKNN	0.625	0.171
Conductor fracture	MLMVN	0.571	0.175
	TLS-ESPRIT	0.584	0.181
	KB-IKNN	0.595	0.168

and KB-IKNN are 0.62, 0.74, and 0.91, respectively. The high value of KB-IKNN suggests that it possesses superior discriminative boundaries and a reduced risk of aliasing in low-dimensional decision spaces. Based on the visualisation results, it is evident that KB-IKNN maintains the structural information of the original high-dimensional features more effectively and achieves greater confidence in fault differentiation. To further quantify the reliability of the model's prediction results, a statistical evaluation is conducted on the first confidence level of each fault category using different classification methods. Table 3 summarises the average confidence and standard deviation values of the test set prediction results. These values reflect the certainty of the model's decisions and the stability of the confidence distribution under various fault scenarios.

As shown in Table 3, KB-IKNN has a higher average confidence level than MLMVN and TLS-ESPRIT for all four types of faults. The most significant improvement is observed for insulation breakdown and ground short circuit faults, with average confidence levels of 0.617 and 0.625, respectively. This indicates that the proposed method can still maintain strong decision certainty in high identification difficulty fault scenarios. Meanwhile, the difference in the standard deviation of confidence among the three methods is relatively small. This indicates that improving KB-IKNN confidence does not lead to increased instability and that its prediction process is controllable under different fault categories. As shown in Figure 9, KB-IKNN has better feature discrimination ability, a more reliable interclass separation structure, and a more accurate output of prediction probabilities. These features significantly enhance the interpretability and applicability of the model in practical engineering applications. Further, long short-term memory gated recurrent unit (LSTM-GRU) [22] and 1-dimensional convolutional neural network (1D-CNN) [23] are introduced, and comparative testing is conducted. Four fault diagnosis models are compared and tested using mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) as indicators. Table 4 displays the test results.

Table 4 includes the mean ± standard deviation error index of 10 independent experiments to systematically evaluate the stability of the model under different types of fault diagnosis. For example, taking partial discharge into account, the MSE and MAE of KB-IKNN are 0.008 ± 0.001 and 0.059 ± 0.002, respectively. These values are reduced by approximately 42.9% and 29.8%, respectively, compared to those of TLS-ESPRIT. Compared with 1D-CNN and LSTM-GRU, its MAE also decreases by about 36.6% and 38.5%, respectively. In insulation breakdown faults, the MSE and MAE of KB-IKNN are 0.009 ± 0.001 and 0.067 ± 0.002, respectively, which are lower than the 0.095 and 0.097 of 1D-CNN and LSTM-GRU. Compared with MLMVN, the error is reduced by more than 40%, indicating that its identification of high-risk faults is more stable and accurate. In the ground short circuit scenario, the RMSE and MAE of KB-IKNN are 0.116 ± 0.002 and 0.074 ± 0.002, respectively. These values are approximately 22.1% and 24.5% lower than the respective MAE values of 0.095 and 0.098 for 1D-CNN and LSTM-GRU. This indicates that it has a more stable discriminative ability on the confusing boundary of "partial discharge ground short circuit". In the extreme fault scenario of conductor fracture, KB-IKNN performs equally well in terms of error, with MSE, RMSE, and MAE of 0.007 ± 0.001, 0.084 ± 0.002, and 0.054 ± 0.001, respectively. This is beneficial for accurately isolating the fault point, shortening the repair time, and reducing the scope of power outages. Overall, the standard deviation of KB-IKNN across all indicators is maintained between 0.001 and 0.002. This range is generally lower than the 0.002 to 0.005 range observed for TLS-ESPRIT, 1D-CNN, and LSTM-GRU. This reflects not only its better repeatability, but also that the method has more controllable output performance under multiple random disturbances and in different operating states. KB-IKNN has the lowest error and standard deviation and achieves full coverage and stable diagnosis of everything from minor hidden dangers to serious fault scenarios. This makes it more in line with the practical needs of power systems, which require high safety, reliability, and a fast response. The computational overheads of the different methods in the identification work for the four fault types are shown in Figure 10.

Figure 10(a) and Figure 10(b) show the average memory

Table 4
Diagnostic Error Results of Different Models

Fault type	Method	MSE	RMSE	MAE
Partial discharge	MLMVN	0.021 ± 0.002	0.145 ± 0.004	0.111 ± 0.003
	TLS-ESPRIT	0.014 ± 0.001	0.118 ± 0.003	0.084 ± 0.003
	LSTM-GRU	0.017 ± 0.002	0.130 ± 0.003	0.096 ± 0.003
	1D-CNN	0.016 ± 0.002	0.126 ± 0.003	0.093 ± 0.003
	KB-IKNN	0.008 ± 0.001	0.089 ± 0.002	0.059 ± 0.002
Insulation breakdown	MLMVN	0.027 ± 0.003	0.164 ± 0.005	0.129 ± 0.004
	TLS-ESPRIT	0.016 ± 0.002	0.126 ± 0.004	0.091 ± 0.003
	LSTM-GRU	0.018 ± 0.002	0.134 ± 0.003	0.097 ± 0.003
	1D-CNN	0.017 ± 0.002	0.130 ± 0.003	0.095 ± 0.003
	KB-IKNN	0.009 ± 0.001	0.095 ± 0.002	0.067 ± 0.002
Short circuit to ground	MLMVN	0.025 ± 0.002	0.158 ± 0.004	0.122 ± 0.003
	TLS-ESPRIT	0.015 ± 0.002	0.122 ± 0.003	0.088 ± 0.002
	LSTM-GRU	0.018 ± 0.002	0.134 ± 0.003	0.098 ± 0.003
	1D-CNN	0.017 ± 0.002	0.130 ± 0.003	0.095 ± 0.003
	KB-IKNN	0.012 ± 0.001	0.116 ± 0.002	0.074 ± 0.002
Conductor fracture	MLMVN	0.019 ± 0.001	0.138 ± 0.003	0.101 ± 0.002
	TLS-ESPRIT	0.012 ± 0.001	0.119 ± 0.002	0.077 ± 0.002
	LSTM-GRU	0.015 ± 0.002	0.122 ± 0.003	0.094 ± 0.002
	1D-CNN	0.014 ± 0.002	0.118 ± 0.003	0.092 ± 0.002
	KB-IKNN	0.007 ± 0.001	0.084 ± 0.002	0.054 ± 0.001

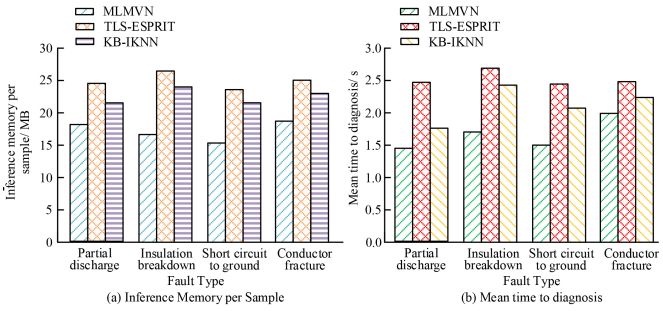

Figure 10. Different Methods for Calculating Cost Results

overhead versus diagnostic elapsed time of the three algorithms in the inference phase under different fault types, respectively. In Figure 10 (a), among the four typical cable fault identification methods, the inference memory usage of three of the methods fluctuates slightly with the complexity of the fault. However, the differences mainly stem from variations in the feature processing and online matching methods. MLMVN has a low feature dimension, so memory consumption is always kept to a minimum. TLS-ESPRIT relies on high-dimensional spectral estimation and matrix decomposition. Its resource overhead is the highest in all scenarios, especially during insulation breakdown, when it occupies 26 MB, which can easily strain edge device deployment. KB-IKNN performs moderately in comparison but is closer to a lightweight level, with an average inference memory of 24MB per inference. Thanks to the offline feature library and neighbourhood index, the online stage only requires the retrieval of necessary features and the calculation of similarities. This maintains stable and controllable memory requirements under different fault types, making it more suitable for online monitoring devices with limited resources. In Figure 10 (b), the difference in diagnostic time among the three methods can further reflect the source of time overhead and real-time adaptability. Due to the requirement of spectral feature analysis and parameter extraction for each inference, TLS-ESPRIT has the highest processing delay. This is true for all fault types. It reaches 2.5 seconds in identifying conductor fracture faults. This poses a potential risk of not meeting the requirements of rapid cutting. The MLMVN structure is relatively simple and therefore has the fastest response speed, but its accuracy is insufficient to handle high-risk scenarios. Although KB-IKNN uses distance weighting and integrated voting mechanisms, most of the computational workload during the online phase comes from local feature matching and the nearest neighbour search. The average inference time is controlled within the range of 2.1-2.3 s, meeting the general real-time requirements of power system fault warning at the second level. Overall, KB-IKNN achieves a better balance between resource cost and response speed while ensuring high-precision diagnosis. This enables it to more effectively support online rapid diagnosis and protection action triggering of high-voltage cables, which is beneficial for reducing fault propagation and power outage losses.

4. Discussion

The study aimed to address the issues of high feature dimensionality, weak classification robustness, and inadequate model generalisation in HV cable fault identification. To this end, it proposed an improved KNN-based identification method that incorporated Bagging sampling and a K-means auxiliary strategy. The study adopted the

GLSP algorithm to extract the low-dimensional discriminative features from high-dimensional fault signals. Combined with K-means clustering to optimise the Bagging sampling process, the weighted KNN classifier integration was constructed. Finally, fault identification was achieved through the voting mechanism. The experimental results indicated that KB-IKNN showed good generalisation and fitting ability on both the training and test sets. The recognition accuracy was 98.4% and 97.6%, respectively, which was better than the comparison methods. The KB-IKNN model exhibited excellent performance under different fault types. Compared to the comparative approaches, the accuracy of the training and test sets was substantially higher, at 95.7% and 93.7%, respectively. In terms of error metrics, the model maintained a low MSE of 0.008 $\pm$ 0.001 and 0.007 $\pm$ 0.001 under both partial discharges and conductor breakage faults, respectively. Meanwhile, a good balance between the average diagnosis time and inference memory consumption was also achieved. This method has three main advantages. First, GLSP maintains the structural relationship of fault signals in high-dimensional space, improves interclass separability, and solves the limitations of traditional methods, which are susceptible to noise interference. Second, introducing a sampling mechanism based on a clustering structure effectively solves the problem of missed detections of minority classes caused by class imbalance. Third, by using a distance weighting strategy to enhance the security of nearest neighbour voting, samples with similar features can obtain higher confidence and improve diagnostic reliability. Despite achieving good results, the research still has certain limitations. In extreme noise environments and when fault signals are too sparse, GLSP embedding may experience a decrease in stability. At the same time, the model still relies on CPU/GPU computing power. Further acceleration optimisation is needed on low-power edge devices. Additionally, the parameters selected and thresholds set in the study are based on optimisation of the validation set. These can be combined with incremental learning and model compression strategies to improve engineering deployment in the future. The introduction of additional time-frequency fusion features enhances the detection sensitivity for sudden faults. A cross-regional and cross-device migration testing and validation system will be established to enhance the model's generalisation capabilities across diverse operational scenarios. Overall, KB-IKNN provides a data-driven and secure solution for the rapid and accurate identification of high-voltage cable faults, with good engineering application prospects.

5. Conclusion

Aiming at the difficulties in identifying high-dimensional fault features of high-voltage cables and the insufficient generalisation ability of the model, an improved KNN fault recognition model (KB-IKNN) combining Bagging sampling, K-means structure optimisation, and GLSP dimensionality reduction algorithm is proposed, which achieves collaborative improvement of feature extraction, sample distribution optimisation, and classification decision-making mechanism. The experimental results show that the model achieves recognition accuracy of 98.4% and 97.6% on the training and testing sets, respectively, with a BCSR of 0.91 and low diagnostic error. The minimum MSE under typical fault types can reach 0.007 $\pm$ 0.001. Meanwhile, the inference phase takes an average of 2.1-2.3 seconds and occupies approximately 24 MB of memory, achieving a balance between diagnostic accuracy and online deployment efficiency. This method outperforms the comparative methods in terms of accuracy, stability, and real-time performance. The research results indicate that KB-IKNN not only maintains stable recognition ability under complex working conditions but also exhibits significant advantages in boundary sample discrimination, class imbalance processing, and model robustness, providing a feasible technical solution for rapid diagnosis and online warning of high-voltage cable faults. This method has good engineering applicability and potential for promotion, providing effective technical support for the construction of intelligent operation and maintenance systems in the future. Subsequent research will further consider model acceleration and migration deployment under extreme noise environments and edge computing conditions to improve its applicability in more complex power system scenarios.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by China Southern Power Grid Company Limited Science and Technology Project "Research on the Mechanism of Sheath Grounding Faults and Live-Line Detection Technology in High-Voltage Cables" (No. GDKJXM20230312).

References

[1] J. Liu, S. Wang, S. Yan, *et al.*, "Fast detection method on water tree aging of MV cable based on nonsinusoidal response measurement," *IEEE Transactions on Power Delivery*, vol. 38, no. 1, pp. 146–153, 2022.

[2] G. K. Rao and P. Jena, "A novel fault identification and localization scheme for bipolar DC microgrid," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 12, pp. 11752–11764, 2023.

[3] T. Zhang, S. Dai, and Z. Cai, "Planning and fault control of urban distribution lines through optimal design," *International Journal of Power and Energy Systems*, vol. 45, no. 1, pp. 19–25, 2025.

[4] V. A. Lacerda, R. M. Monaro, R. Peña-Alzola, D. Campos-Gaona, and D. V. Coury, "Nonunit distance protection algorithm for multiterminal MMC-HVDC systems using DC capacitor resonance frequency," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 12, pp. 12924–12933, 2022.

[5] N. Peng, Y. Li, R. Liang, C. Jiang, H. Xu, W. Jin, Y. Wang, Y. Guan, and R. Yang, "Fault section identification of hybrid transmission lines by the transients in modal domain free from

the refractions and reflections at cross-bonded nodes," *IEEE Transactions on Power Delivery*, vol. 38, no. 4, pp. 2864–2878, 2023.

[6] M. S. Zaky, H. E. Ahmed, M. Elsadd, and M. Elgamasy, "Protection of HVDC transmission systems for integrating renewable energy resources," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12237–12244, 2023.

[7] J. Purohit and R. Dave, "Leveraging deep learning techniques to obtain efficacious segmentation results," *Archives of Advanced Engineering Science*, vol. 1, no. 1, pp. 11–26, 2023.

[8] W. Du, G. Yang, M. Tian, W. Hu, and C. Ma, "Active distribution network fault location method based on improved multiverse algorithm," *International Journal of Power and Energy Systems*, vol. 45, no. 2, pp. 78–88, 2025.

[9] G. V. Raju and N. V. Srikanth, "A novel protection scheme for transmission lines connected to solar photovoltaic and wind turbine farms using fuzzy logic systems and bagged ensemble learning," *Electrical Engineering*, vol. 106, no. 6, pp. 7509–7529, 2024.

[10] Solimun and A. A. R. Fernandes, "Ensemble bagging discriminant and logistic regression in classification analysis," *New Mathematics and Natural Computation*, vol. 21, no. 1, pp. 91–111, 2025.

[11] N. Fassina, F. Ranzato, and M. Zanella, "Robustness verification of k-nearest neighbors by abstract interpretation," *Knowledge and Information Systems*, vol. 66, no. 8, pp. 4825–4859, 2024.

[12] Z. K. Abdul, A. K. A. Talabani, C. M. Rahman, and S. M. Asaad, "Electrocardiogram heartbeat classification using convolutional neural network-k nearest neighbor," *ARO-The Scientific Journal of Koya University*, vol. 12, no. 1, pp. 61–67, 2024.

[13] R. Sukshitha, "Empirical likelihood ratio based k-nearest neighbours regression," *International Journal of Agricultural & Statistical Sciences*, vol. 20, no. 2, pp. 421–428, 2024.

[14] S. Mulewa, A. Parmar, and A. De, "A novel Bagged-CNN architecture for short-term wind power forecasting," *International Journal of Green Energy*, vol. 21, no. 12, pp. 2712–2723, 2024.

[15] Y. Xue, Y. Chang, Y. Zhang, J. Sun, Z. Ji, H. Li, Y. Peng, and J. Zuo, "UAV signal recognition of heterogeneous integrated KNN based on genetic algorithm," *Telecommunication Systems*, vol. 85, no. 4, pp. 591–599, 2024.

[16] A. Chanane and H. Houassine, "Toward unique electrical ladder network model synthesis of a transformer winding high-frequency modeling using K-means and metaheuristic-based method," *COMPEL*, vol. 43, no. 1, pp. 247–266, 2024.

[17] S. Mantach, M. Partyka, V. Pevtsov, A. Ashraf, and B. Kordi, "Unsupervised deep learning for detecting number of partial discharge sources in stator bars," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 30, no. 6, pp. 2887–2895, 2023.

[18] C. Zhu, H. Yang, X. Jin, K. Xu, and W. Shen, "Locality preserving projections-based spatiotemporal modeling of the temperature distribution of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 1, pp. 179–189, 2023.

[19] B. Despodov, D. Stojanov, and C. M. Bande, "Evaluating handwritten character recognition with Hu moments and k-nearest neighbors algorithm," *TEM Journal*, vol. 13, no. 3, p. 1813, 2024.

[20] M. Bindi, A. Luchetta, G. M. Lozito, C. F. M. Carobbi, F. Grasso, and M. C. Piccirilli, "Frequency characterization of medium voltage cables for fault prevention through multi-valued neural networks and power line communication technologies," *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp. 3227–3237, 2023.

[21] W. Jiang, D. Wang, B. Liu, Y. Hu, and L. Zhou, "Fault diagnosis for shielded cable in EMUs based on TLS-ESPRIT and 3D-BIS images," *IEEE Transactions on Transportation Electrification*, vol. 11, no. 1, pp. 2230–2242, 2024.

[22] C. Zhang, M. Chen, Y. Zhang, W. Deng, Y. Gong, and D. Zhang, "Partial discharge pattern recognition algorithm of overhead covered conductors based on feature optimization and bidirectional LSTM-GRU," *IET Generation, Transmission & Distribution*, vol. 18, no. 4, pp. 680–693, 2024.

[23] Q. Huang, Z. Li, Z. Fu, Y. Hu, Q. Fang, and Y. Wei, "Complex wired network fault diagnosis based on distributed reflectometry and multi-channel 1D-CNN," *IEEE Sensors Journal*, vol. 25, no. 11, pp. 19415–19427, 2025.

Biographies

Yinchun Xu received the bachelor's degree in Electrical Engineering from South China University of Technology in 2011. She is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. She has published articles in internationally reputed peer-reviewed journals and conference proceedings. and her research interests are power grid planning, intelligent monitoring of distribution networks and substation intelligent inspection technology.

Xiangmao Cheng received the bachelor's degree in Electrical Engineering from Guangdong University of Technology. He is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. He has in-depth research on distribution networks, relay protection, and power equipment disaster prevention and mitigation technologies. He has published 12 technical papers, applied for 32 invention patents, and his research interests are electrical engineering automation, electric power intelligent operation, and power grid production management.

Peifeng Huang received the bachelor's degree in Electrical Engineering from Guangdong University of Technology in 2015. He is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. He has published articles in internationally reputed peer-reviewed journals and conference proceedings. His research interests are power grid planning, intelligent distribution systems and distribution network reliability.

Runjie Lin received the Master's degree in Electrical Engineering from Xi'an Jiaotong University in Shaanxi in 2011. He is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. He has published articles in internationally reputed peer-reviewed journals and conference proceedings, and his research interests are power grid safety production, dispatching and operation, and power grid planning.

Yangrui Lin received the Master's degree in Electrical Engineering from Guangdong University of Technology in 2017. He is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. He has published articles in internationally reputed peer-reviewed journals and conference proceedings. His research interests are power grid planning, power grid safety production, intelligent distribution systems and distribution network reliability.

Jia Weng received the bachelor's degree in Electrical Engineering from North China University of Water Resources and Electric Power in Henan. He is working in the Chaozhou Power Supply Bureau of Guangdong Power Grid Co., Ltd. He has published articles in internationally reputed peer-reviewed journals and conference proceedings, and his research interests are power grid adaptive development strategies, power grid equipment construction quality improvement, and power grid equipment defect identification.