Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 8423571, 12 pages
https://doi.org/10.1155/2020/8423571

Research Article

A Complex Fault Diagnostic Approach of Active Distribution Network Based on SBS-SFS Optimized Multi-SVM

Keyan Liu,[1] **Weijie Dong**,[1] **Huanyu Dong**,[2] **Jia Wei**,[2] **and Shiwu Xiao**[2]

[1]*China Electric Power Research Institute, Beijing 100192, China*
[2]*North China Electric Power University, Beijing 102206, China*

Correspondence should be addressed to Huanyu Dong; 120192201280@ncepu.edu.cn

Received 4 January 2020; Revised 14 March 2020; Accepted 24 April 2020; Published 22 May 2020

Academic Editor: Yang Li

Copyright © 2020 Keyan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

After renewable energy distributed generator (DG) is connected to the power grid, traditional diverse-electric-information-based fault diagnosis approaches are not suitable for an active distributed network (ADN) due to the weak characteristics of fault current. Thus, this paper proposes a comprehensive nonformula fault diagnostic approach of ADN using only voltage as input. In the preprocess, sequential forward selection (SFS) and sequential backward selection (SBS) are utilized to optimize the input feature matrix of the sample in order to reduce the information redundancy of multiple measuring points in ADN. Then, a single "1-a-1" support vector machine (SVM) classifier is used for fault identification, and a multi-SVM, with radial basis function (RBF) as the kernel function, is applied to identify the location and fault type. To prove the proposed method is adaptable for ADN, two direct drive fans are used as a DG to test the IEEE 33 node model at every 10% of the line under three operating conditions that include all cases of distributed power generation in ADN. Results comparing real-time and historical data show that the proposed multi-SVM model reaches an average fault type diagnosis accuracy of 97.27%, with a fault identification accuracy of 96%. A backpropagation neural network is then compared to the proposed model. The results show the superior performance of the SBS-SFS optimized multi-SVM. This model can be usefully applied to the fault diagnosis of new energy sources with distributed power access to distribution networks.

1. Introduction

As a significant part of the Ubiquitous Power Internet of Things (UPIOT), a distribution network is a complex framework that enables electric power to be distributed to numerous powered clients according to their voltage level. Any complex faults, such as short circuits among lines, that occur in the distribution network would inhibit practical work, increasing the cost of maintenance [1]. Thus, there is an urgent need to develop a fault detection system to prevent the deterioration of the stability and security of these distribution networks. Several experts [2–5] have been working on the prediction of distribution networks. Clean energy sources, including photovoltaic, wind power, and hydropower, have been developed as aggressive sources to solve the lasting reliance on fossil fuels [6]. These developments have stimulated the implementation of wind turbines (WTs) into the active distribution network (ADN) in recent years [7]. However, WTs and other equipment that are connected to a distribution network act as distributed generators (DGs) [8], which have complex structures, flexible operation, and uncertain outputs. The main reason is that the control strategy adopted by the power electronic device of which limits the magnitude of the fault current, which is relatively small and difficult to express with formulas. These features undermine the utility of traditional fault diagnosis methods based on one-way power flow characteristics, creating an urgent demand for other diagnostic methods that are suitable for multi-DG accessed ADNs [9].

An ideal fault diagnostic system should include state identification, fault type identification, and the determination of fault locations [10]. State identification in a distributed network relies on threshold setting or logical judgment utilized by three-stage current [11] or low voltage

protection, and none of these methods take the effect of DGs in ADNs into consideration. Fault type identification in distribution grids generally contains single-phase, two-phase, and three-phase short circuits to ground, as well as two-phase short circuits. Several researchers use current, direct, and other information to determine fault type, while their judgment process requires cutting off the entire power network [12], which is not suitable for ADNs with renewable DGs. There are two primary fault location methods for an ADN: one is to improve the traditional three-segment current protection using information uploaded by a distribution automation system, and the other is to use wide-area measurement information for fault location [13]. Existing approaches cover a broader range of fault areas in distribution grids and increase the accuracy of fault location. Though progress has been made in previous reports, state identification, fault classification, and fault identification all use different factors of the distribution grid, and fault diagnosis in ADNs can be time-consuming. There is a lack of a comprehensive evaluation method to realize ADN fault status discrimination, fault type classification, and fault location with less electrical quantity.

Machine learning-based theories are emerging techniques combining mass data-driven from sensors and have previously been applied to grid fault diagnosis [14]. Commonly used artificial intelligence methods include artificial neural networks (ANNs) [15], Petri nets [16], and extreme learning machine (ELM) [17]. Support vector machine (SVM) [18] is a major algorithm of classification learning that separates the entire sample zone into several parts and then uses the boundaries to classify the data correctly. The samples nearest to the boundaries are named support vectors [19]. The rule for determining the boundaries using SVM is "1-a-1" voting, which is quite suitable for two-state classification problems like state identification in distributed networks. Compared with other artificial intelligence algorithms, the final decision function of SVM is determined by only a few support vectors, which outperform those algorithms whose convergence process is closely related to the dimension of the sample space. This approach is probably not feasible for fault type and fault location diagnosis owing to the high-dimensional features of the data. To solve this problem, measures should be taken to reduce the data feature dimensions. Selection of forward sequential (SFS) and selection of backward sequential (SBS) are processes that select out factors that best reflect the characteristics of the data. Using both SFS and SBS is optimal because the data undergo multiple screening processes. Inspired by machine learning theory, this paper proposes a complex fault diagnostic approach for ADNs using only voltage information. Specifically, single SVM classifiers are applied to state identification, while multi-SVM classifiers optimized by SFS and SBS are used to categorize fault type and locate faults. Subsequently, the proposed approach is employed in an IEEE 33-node model combined with two DGs to test its reliability.

This paper is organized as follows: Section 2 presents the complex methodologies of using SBS and SFS to select the fittest feature matrix of SVM and a related flowchart of the proposed model. In Section 3, the implantation of the proposed method for ADN fault diagnosis is described, including a structural flowchart. Section 4 employs SBS/SFS-multi-SVM to an ADN with 2 DGs and then tests this model with simulation data of voltage information from an IEEE 33-node model to demonstrate its effectiveness. The fault location of diverse fault types is also presented in this section. Section 5 provides the results of the experiments.

2. Proposed Method

2.1. SVM. Proposed by Vapnik [20], SVM is a machine learning method that is based on statistical theory. SVM can map trained input data into a high-dimensional feature space using the inner product function [21]. For a single SVM classifier, a linear decision surface with a maximum margin of separation between the two input classes among the mapping process is generated. The next group of input data can be matched to their category according to the feature matrix. The matrix that best reflects the characteristics of the input data is selected.

A sample dataset $X = \{X_1, X_2, \ldots, X_N\}$ contains two values, described by a label $Y_i \in \{-1, 1\}$, representing the corresponding binary class label. Thus, the training data can be written as $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)\}$. For linearly separable data, the regression function can be given by Vapnik [22]:

$$f(X) = w \cdot X + b \geq 1. \tag{1}$$

Here, w is the normal vector of hyperplane and b is a constant. Those two coefficients determine the location of the optimal hyperplane that meets the maximum interval of two classes. For point X_1, X_2 from $w \cdot X + b = 1$ and $w \cdot X + b = -1$ individually, the interval among them is $w/\|w\| \cdot (X_1 - X_2) = 2/\|w\|$.

Finding the optimal hyperplane (maximum interval) can be considered a quadratic programming problem, which can be described as follows:

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|w^2\| + C\sum_{i=1}^{N}\xi_i, \quad i = 1, \ldots, N \\ &\text{Subject to} \quad \begin{cases} f(X_i) = w \cdot X + b \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}. \end{aligned} \tag{2}$$

where ξ is a positive slack variable and C is the loss factor. The above equations can be solved with the Lagrange method, whose function is given by

$$L(w,b,a) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left[Y_i\left(w \cdot X_i + b\right) - 1 + \xi_i\right] - \mu_i\xi_i, \tag{3}$$

where a and μ are Lagrange multipliers. The function should meet the following constraints:

$$\text{Maximize} \quad \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j Y_i Y_j \left(X_i \cdot X_j\right)$$
$$\text{Subject to} \quad \sum_{i=1}^{N} a_i Y_i = 0, a_i \geq 0. \tag{4}$$

The kernel function is used to measure the distance between two input vectors in the nonlinearly transformed feature space, which requires the mapping process, as shown in Figure 1.

During the nonlinear classifying process $\varphi(X)$, the problem can be restated as follows:

$$\text{Maximize} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j Y_i Y_j \varphi\left(X_i\right)^T \varphi\left(X_j\right)$$
$$\text{Subject to} \quad \sum_{i=1}^{N} a_i Y_i = 0, 0 \leq a_i \leq C. \tag{5}$$

Let $K(X_i, X_j) = \varphi(X_i)^T \varphi(X_j)$ and $Q_{ij} = Y_i Y_j K(X_i, X_j)$. The above equations can then be written in a brief form:

$$\text{Minimize} \quad \frac{1}{2} a^T Q a - e^T a$$
$$\text{Subject to} \quad Y^T a = 0, 0 \leq a_i \leq C. \tag{6}$$

Current research has focused on finding better kernel functions as the main improvement direction for SVM. The most commonly used kernel function is the radial basis function (RBF).

2.2. SBS and SFS

2.2.1. Feature Selection Rules of SVM. Before data classification for SVM, the original data should be expressed in a brief but accurate way due to its feature of high dimensions. The universal method to solve this problem is feature selection and extraction, which searches all properties to reduce the number of dimensions. The determining criterion in SVM is based on the classification accuracy; however, it is not applicable for cases with limited samples and high feature dimensions. For a hyperplane, $f(X) = w \cdot X + b$ is one of the key factors to determine classification effects. The value of this function is determined by a within-class scatter matrix S_w and a between-class scatter matrix S_b [23]:

$$S_w = \frac{1}{N}\sum_{i=1}^{k}\sum_{l=1}^{N_i}\left(X_l^i - \mu_i\right)\left(X_l^i - \mu_i\right)^T,$$
$$S_b = \frac{1}{N}\sum_{i=1}^{k} N_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T, \tag{7}$$

where μ_i represents the sample mean of a class and μ is the mean of the total sample.

$$J_1 = \operatorname{tr}\left(S_w^{-1} S_b\right). \tag{8}$$

By calculating the value of J_1, higher criterion means better classification performance of w. Thus, the best feature matrix is ensured.

2.2.2. SFS and SBS. SFS and SBS are data traversal processes based on the leave-one-out cross-validation theory, which can select features to improve model classification performance. Namely, SFS starts its searching process from an empty set, adds one property to the feature set each time, tests the classification performance of this set, and repeats until the best classification accuracy is reached. The pseudocode of SFS is shown in Algorithm 1 [24].

In contrast to SFS, SBS starts with the entire sample set and deletes one property from the set after each test until optimal classification performance is reached.

2.3. SBS/SFS Modified SVM. From the above description of SFS, SBS, and SVM, our proposed method for ADN fault diagnosis is built. After the original data are loaded into the model, the classification accuracy of SVM is calculated, if it does not meet the demand. SBS and SFS processes are employed to ensure an optimal M-dimensional matrix. For each test, one type of property is added or deleted from the original dataset; then, the new dataset is used to retest its SVM accuracy. The whole process continues until the accuracy is more than 95%. The flowchart of SBS and SFS optimized SVM is drawn in Figure 2.

3. Implementation and Results of SBS/SFS Optimized SVM for ADN Fault Diagnosis

3.1. Model Establishment. The proposed method achieves the three aims of ADN fault diagnosis separately. Specifically, the operation state of an ADN contains fault and normal states. The most common fault types are single-phase short to ground, two-phase short, two-phase short to ground, and three-phase short. The fault location depends on the number of nodes in the ADN.

In this work, the input information originates from voltage instead of the switching state of the circuit breakers or other information for achieving the aforementioned goals. The fault states along with fault types and their location are outcomes of this model.

3.2. Data Reduction. Sensors in ADNs can obtain different varieties of voltage information, including phase voltage, fundamental frequency, and harmonic components. The proposed model uses this diverse information to aid fault diagnosis.

3.2.1. State Identification. To figure out the relationship between input and output information, we establish an example ADN as shown in Figure 3. Here, V_s denotes the power on the system side, the numbers represent the serial

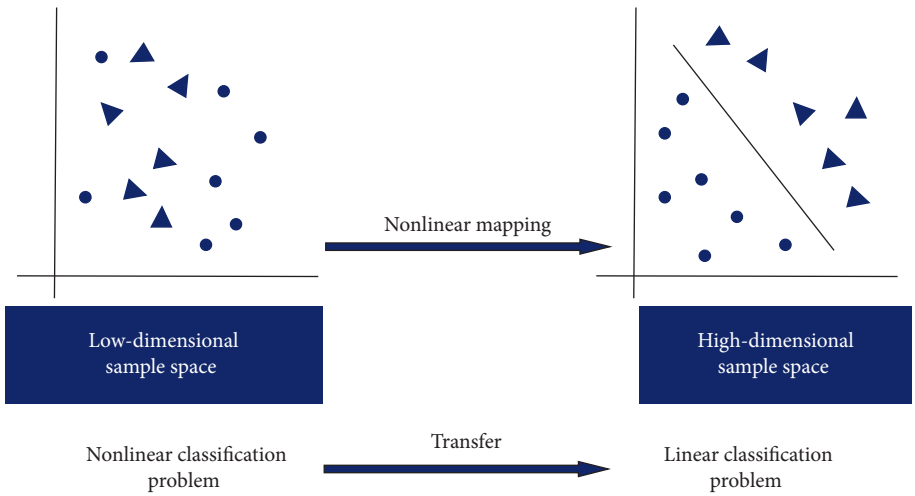

Figure 1: Nonlinear data mapping process.

(1) Initialize $P_k = \{\varnothing\}$, $k = 0$
(2) $J_0 = 0$;
(3) $X_N = \{X_1, X_2, \ldots, X_N\}$
(4) For each X left in the X_{N-k}, select the X with the greatest J according to accuracy evaluated by SVM
(5) $J_{k-1} = J$
(6) **if** $J_{k-1} > J_k$ **then**
(7) $P_{k+1} = P_k + X$
(8) $X_{N-k-1} = X_{N-k} - X$
(9) $k = k + 1$
(10) **if** $k <= N$ **then**
(11) back to the selection process
(12) $P_c = \{P_1, P_2, \ldots, P_C\}$

Algorithm 1: Pseudocode of SFS.

numbers of the measurement points, and f1, f2, and f3 mark fault locations on the line.

When a fault occurs at location f1, the amplitude of low-frequency voltage components is far from normal operation. ADN's fundamental frequency voltage component has a difference of 2 kV under normal operation and low-frequency operation, which is clearly seen in Figure 4.

There is no obvious difference in amplitudes at other frequencies. Due to differences in the accuracy of sensors and the selection of measurement points in an ADN, in practice, it is necessary to select measurement points that can accurately describe the fault and nonfault conditions, and use the obtained voltages as input information to determine operational status. Because the amplitude of the fundamental frequency voltage in the nonfault state of the ADN is much larger than that of the fault state, it is selected as the characteristic quantity to judge the fault status.

3.2.2. Fault Type Diagnosis. Single-phase short to ground, two-phase short, two-phase short to ground, and three-phase short are closely related to each phase voltage component. Taking phase A as an example, when a short circuit between AB phases occurs, the voltages of A and B phases are equal, i.e., the voltage between two phases is 0. Similarly, the difference between A, B, and C phases when an ABC three-phase short circuit occurs is 0. The voltage difference for each fault type is shown in Figure 5. The zero-sequence voltage is higher than 0 when a short to ground occurs, allowing for easy identification. Therefore, the phase difference value and the zero-sequence voltage can be used to determine the fault type.

3.2.3. Fault Location. When the same fault occurs at f1, f2, and f3, respectively, the low-frequency voltage component of the fault phase shows a significant difference from the same measuring point, as shown in Figure 5. It can be concluded from the figure that the low-frequency voltage amplitude components generated by the fault show the same trend at the three measurement points, that is, Uf1 > Uf2 > Uf3. Therefore, based on the judgment of the operating state and the type of fault, the low-frequency component of the fault phase voltage can be used to locate the fault (Figure 6).

3.3. SBS/SFS-SVM for ADN Fault Diagnosis Platform. The proposed complex fault diagnosis platform for ADNs is

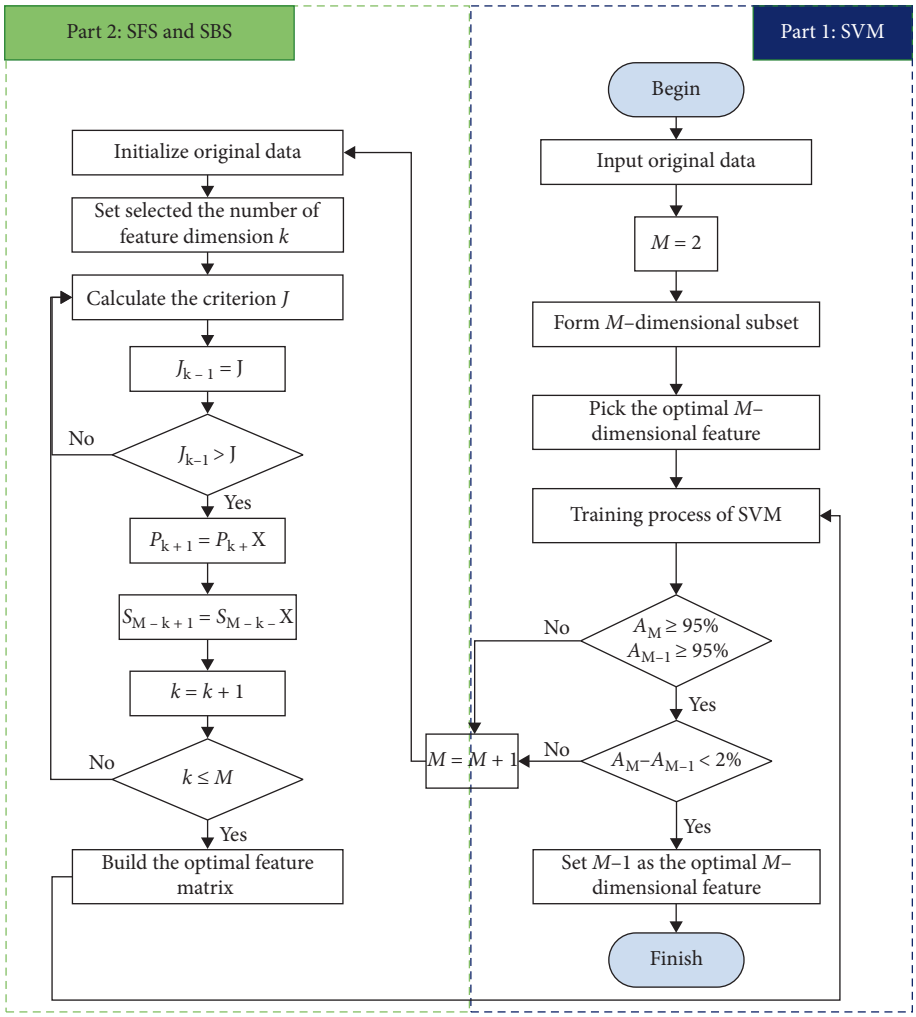

Figure 2: Flowchart of SFS/SBS-SVM.

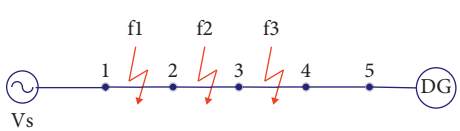

Figure 3: Schematic diagram of example ADN.

shown in Figure 7. After data are collected from the sensors, the data selection process is performed on the voltage information. Then, SBS and SFS are used to add or delete unrelated features, forming the optimal M-dimensional feature matrix. The fundamental frequency voltage is then used to monitor the operating status and differences in phase and zero-sequence voltages are analyzed to distinguish fault types. Finally, the low-frequency voltage components are used to determine the location of the fault.

4. Validation of SBS/SFS-SVM Model for ADN Fault Diagnosis

4.1. Model Startup

4.1.1. Establishment of Simulated Model. Based on PSCAD/EMTDC software, the IEEE 33-node ADN was selected as

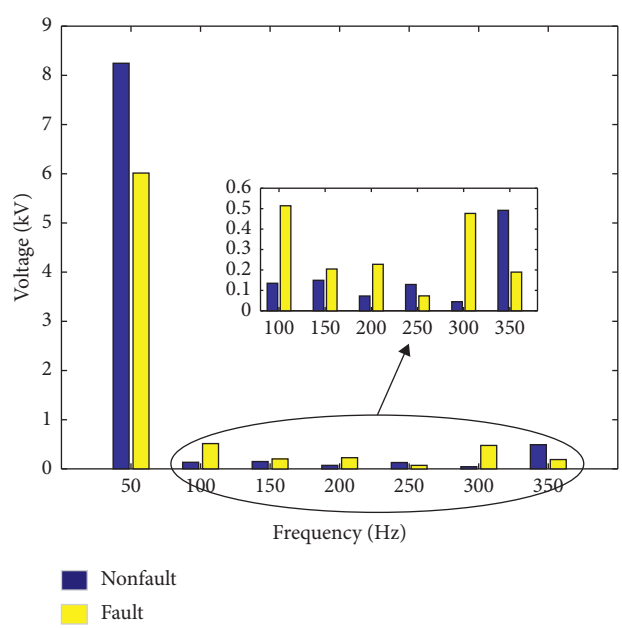

Figure 4: Voltage amplitude of each frequency component in fault and nonfault state of ADN.

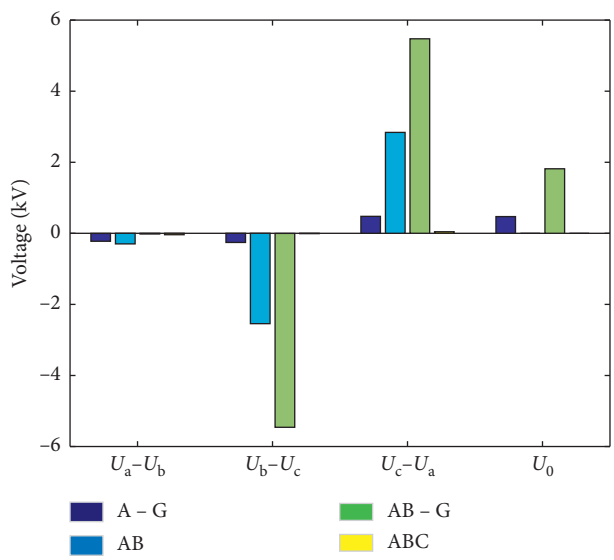

Figure 5: Phase voltage difference and zero-sequence voltage value under different fault types of ADN.

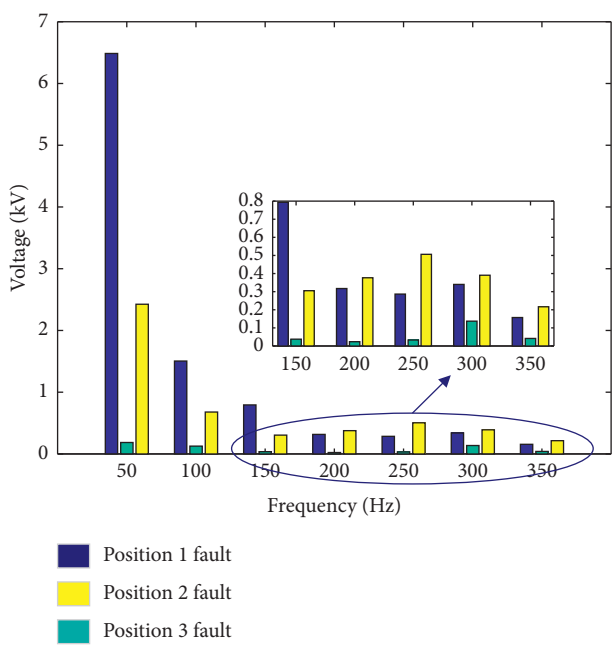

Figure 6: Voltage amplitude of each frequency component at the same measuring point when ADN fails at different positions.

the simulation model. The entire model was connected to two inverter-type WTs, with 1 kW output total. The model sets 34 measurement points, which are divided into two parts: (1) the head of each segment of the line and the DG access point and (2) the DG1 and DG2 access points. The sequence number setting rules are that the measurement point sequence number is the same as the node number at the end of the line, and the measurement point numbers at the DG1 and DG2 access points are 33 and 34, respectively (Figure 8).

4.1.2. Data Acquisition and Processing

(1) State Identification. Simulation of five types of faults at 50% of the 32 segments of the ADN model was performed, namely, AB, AC, AB-G, AC-G, and ABC, and the three-phase voltages at each measurement point were extracted within one week after the occurrence of the fault. Under the analysis of 160 kinds of fault states, the A-phase fundamental frequency and 2–7th harmonic voltage amplitudes of 34 measurement points were obtained. Similarly, 160 kinds of normal operating states were simulated, and 320 sets of voltage data were obtained.

(2) Fault Type Identification. By applying Fourier and symmetrical component method analysis to the obtained

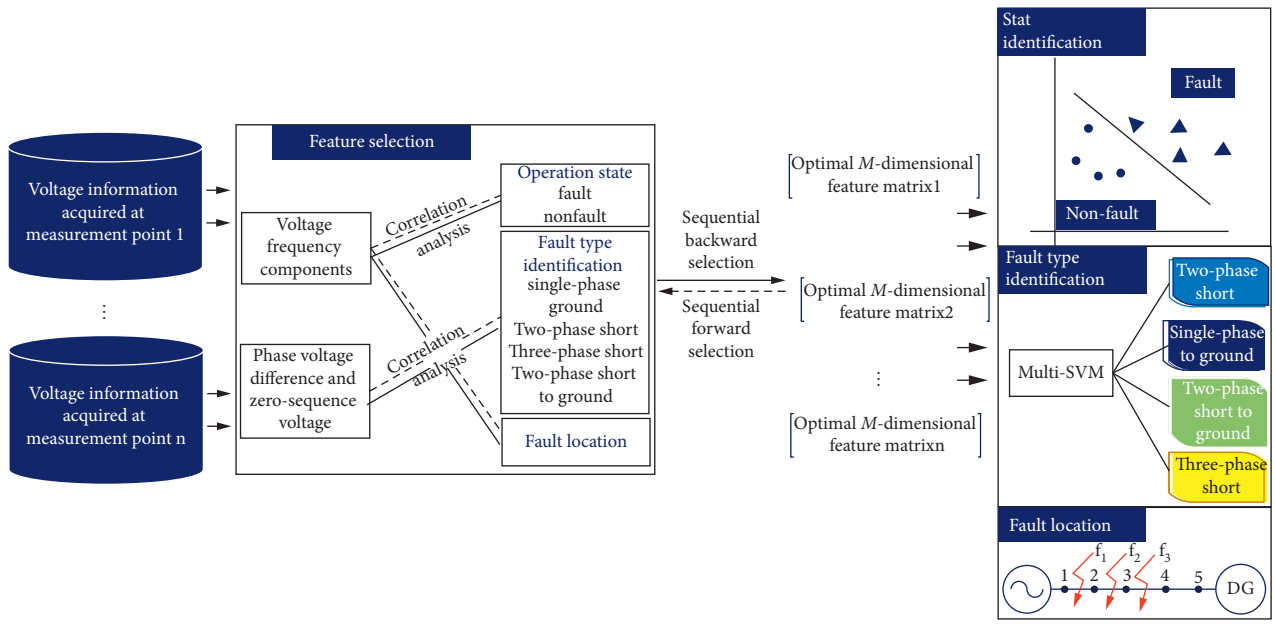

Figure 7: The proposed complex fault diagnostic approach for an ADN based on SBS-SFS optimized multi-SVM.

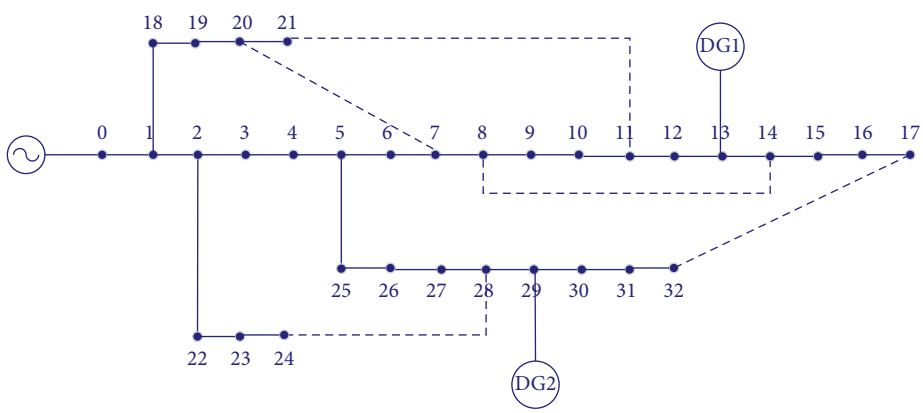

Figure 8: IEEE 33-node with 2 DGs ADN PSCAD model.

160 sets of fault data, the three-phase fundamental and zero-sequence fundamental voltage amplitudes at 34 measurement points under each operating state were obtained. The difference between the fundamental phase voltages U_a–U_b, U_b–U_c, and U_c–U_a was then calculated. Those values, together with the zero-sequence fundamental voltage, form a four-dimensional characteristic vector at each measurement point.

(3) Fault Location. The AB two-phase short-circuit fault was simulated at every 1% of the line length in the 40%–60% section of each line of the ADN model. Subsequently, phase A voltages at various measurement points within one week after the occurrence of fault are extracted. Fourier analysis is then used to obtain the phase A fundamental frequency and 2–7th harmonic voltage amplitudes at 34 measurement points in each operating state. This is the original dataset. Finally, the AB two-phase short-circuit fault simulation is performed on the unit line at points every 10% of the length, and the same data processing is performed as with the original multi-SVM sample set.

4.2. Experimental Results

4.2.1. State Identification. The state identification based on the fundamental frequency component of the voltage is shown in Figure 9. For each set of voltage fundamental frequencies, the SFS/SBS-SVM forms optimized one-dimensional and two-dimensional feature vectors. Then, the combination of those two vectors is used as input for the SVM state identification. The two operation states are well separated on both sides of the hyperplane, highlighting the excellent state identification performance of the SFS/SBS-SVM method.

4.2.2. Fault Type Identification. The result of the fault type identification test is listed in Table 1. The average diagnosis

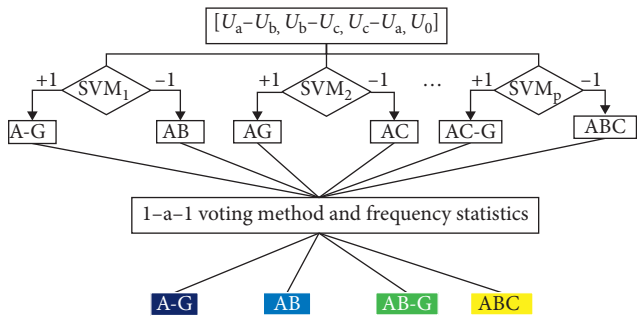

Figure 9: Fault type identification of an integrated SVM ADN based on the "1-a-1" method.

Table 1: Result of fault type identification accuracy.

Fault type	A-G (%)	AB (%)	AB-G (%)	AC (%)	AC-G (%)	ABC (%)
Fault type diagnosis accuracy	97.12	96.28	98.46	91.65	96.84	97.31

accuracy is 97.27%, which is higher than the presetting value of 95%. These results suggest that SFS and SBS play a significant role in adding and deleting unrelated factors from the original feature matrix. In addition, the coding program results show the optimal feature matrix is four-dimensional, most likely taking all the phase voltage differences and zero-sequence voltage into account for the calculation (Figure 10).

To prove the efficiency of the proposed SFS/SBS modified multi-SVM method, we compare the results of the fault type classification with a backpropagation neural network, as shown in Figure 11.

The backpropagation neural network does not perform well in fault type recognition. Several groups of data do not match their original fault type, and the average fault type classification of backpropagation is 85.24%, which is much less than our model.

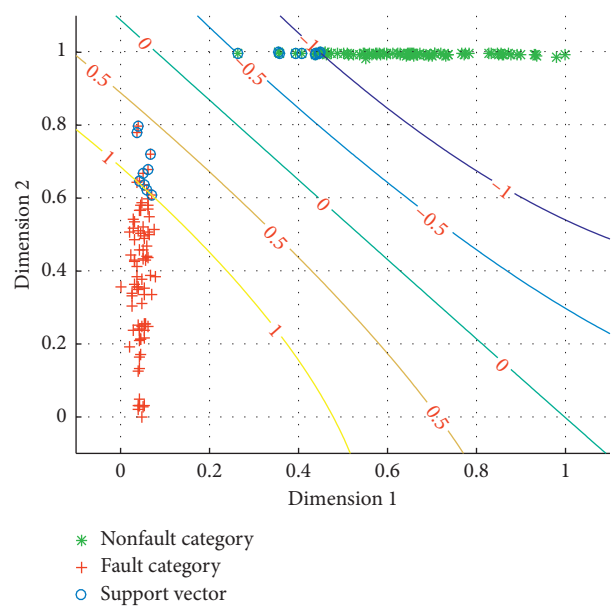

Figure 10: IEEE 33-node with 2 DG model running state discrimination results.

4.2.3. Fault Location

(1) Operation Condition 1: DG1 and DG2 Are Both Connected to the Grid. Based on the simulation data with both DGs connected to the grid, the feature quantities are selected, and a multi-SVM classification model is established to locate the fault.

Taking the AB two-phase short-circuit fault as an example, because the model contains 32 segments, the samples need to be divided into 32 categories. Each type of sample set contains 21 training samples.

The optimal feature matrix is two-dimensional: it can be seen from Figure 12 that when different line faults occur, the sample points of the measurement data are gathered in different regions, enhancing the classification effect. Therefore, the fault location can be distinguished by differences found in this two-dimensional data. The training sample set composed of the two-dimensional feature quantity is used as input for the SVM classifier to obtain a fault discrimination SVM model. The sample data of different line faults are separated by the hyperplane in the model.

The test sample is loaded as input into the SVM, the label of the sample is obtained, and then, the fault section corresponding to the label is determined. The fault section is extended by a section of the line upstream and downstream to obtain a suspected fault area. The location accuracy of the suspected fault area was found to be 95% that meets the presetting value; therefore, a three-dimensional optimal feature matrix was developed to relocate the fault.

The optimal feature matrix is three-dimensional: it can be seen from Figure 13 that when different line faults occur, the measurement data sample points are found in different regions, and a better fault location effect can be obtained. The training sample set composed of the three-dimensional feature quantity is loaded into the SVM classifier to obtain a fault localization multi-SVM model, again producing a location accuracy of 95%, identical to the 2D feature matrix

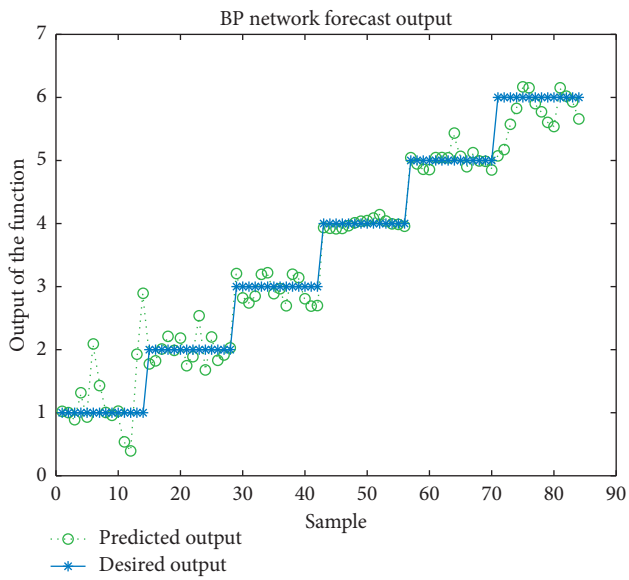

Figure 11: Backpropagation neural network fault type diagnosis results.

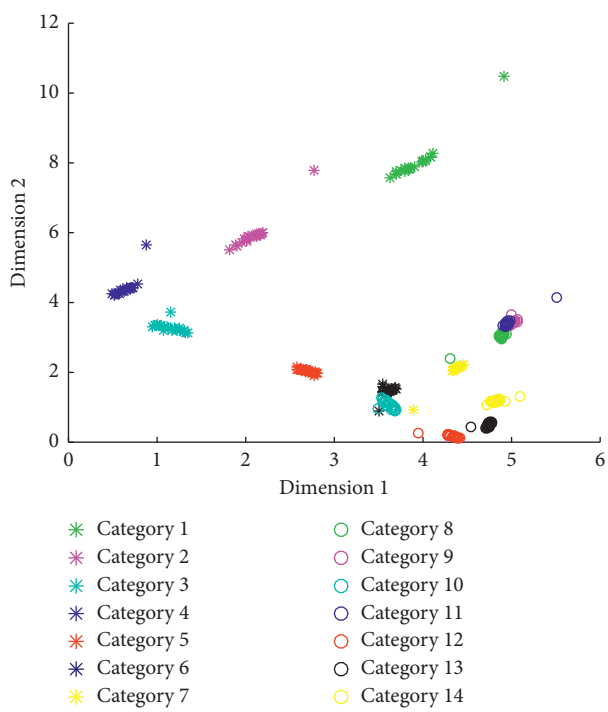

Figure 12: Two-dimensional sample data distribution diagram of fault location when DG1 and DG2 are both connected to the grid.

results. Therefore, we suggest that the fault location limit of the SFS/SBS-SVM model for this kind of working condition has been reached without increasing the dimension of the optimal feature matrix.

Comparison with back-propagation neural network: the fault location accuracy of the backpropagation neural network is shown in Figure 14. The vertical axis represents the fault zone. Although the fault type identification is over 95%, the fault location accuracy is only 66.33%. Notably, the

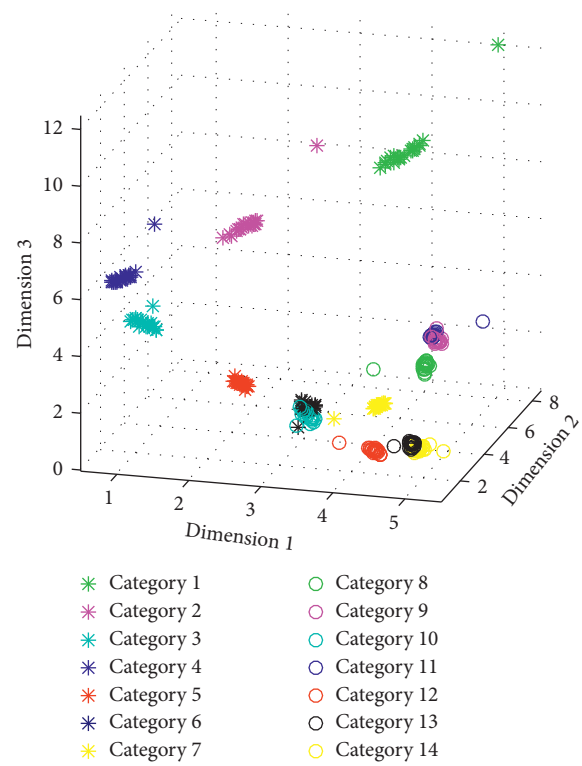

Figure 13: Three-dimensional sample data distribution diagram of fault location when DG1 and DG2 are both connected to the grid.

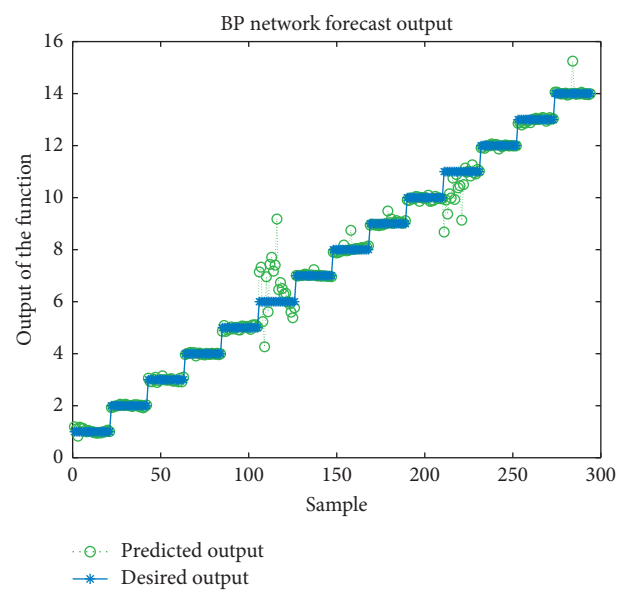

Figure 14: Backpropagation neural network fault location diagnosis results.

backpropagation neural network has a distinct difficulty identifying faults that occur at length 7 and also misjudges a fault occurring from section 8 to section 11 as a fault occurring in section 9, which does not meet the demand.

(2) Operation Condition 2: DG1 On-Grid and DG2 Off-Grid. Under the condition that DG1 is connected to the grid and

DG2 is off-grid, the AB two-phase short-circuit fault is simulated at every 1% of the line length in each line segment. Subsequently, the amplitude of the A-phase fundamental voltage at measurement points 4 and 15 is extracted. Combining the two into a two-dimensional feature quantity, this sample is input to the DG1 grid connection, and the multi-SVM under the DG2 off-grid condition is used for training and testing. The test result of the location accuracy of the suspected fault area is 100% (Figure 15).

To avoid erroneous location results, we enlarge the dimension of the optimal matrix from 1 to 2. We then simulate faults at 0%, 10%, 20%, 30%, 70%, 80%, 90%, and 100% of each section of the line and extract phase A fundamental wave voltage amplitude at measurement points 4 and 15 which are combined into a two-dimensional feature to form a test sample set. This set is used as input for the multi-SVM classifier to obtain a suspected fault area, and the location accuracy of the suspected fault area is again found to be 100%.

(3) Operation Condition 3: DG1 Off-Grid and DG2 On-Grid. Under the conditions that DG1 is off-grid and DG2 is on-grid, the AB two-phase short-circuit fault is simulated at every 1% of the line length in the 40%–60% section of each line. Phase A fundamental wave voltage amplitude at measurement points 4 and 15 are extracted to form a two-dimensional feature, and both are combined into a sample set. Subsequently, the sample was input to a multi-SVM under the conditions that DG1 was off-grid and DG2 was on-grid, and the suspected fault area localization accuracy was 100% (Figure 16).

As above, we simulated the occurrence of a fault at 0%, 10%, 20%, 30%, 70%, 80%, 90%, and 100% of each section of the line and extracted the A-phase fundamental voltage amplitudes at test points 4 and 15. The location accuracy rate of this test after SVM optimization is 96%.

In summary, the SVM classification model trained on the training sample composed of the two-dimensional feature set provides accurate classification results, even when DGs are removed.

4.2.4. Comparison of Results. Table 2 lists the fault diagnosis performance of the proposed model and the backpropagation neural network. Both algorithms have accurate operation identification performance. Notably, SBS/SFS-SVM has better performance in fault type classification and fault location, which are both over 95%, respectively.

Optimizing the structural parameters of the classification algorithm can certainly increase the fault diagnosis rate. However, in an active power distribution network connected to multiple wind turbines, only a few measurement points have significant characteristics in characterizing ADN fault anomalies. The input matrix dimension reduction process of SFS/SBS can select key measurement points from hundreds of measurement points and form SVM support vectors from them, which greatly improves the accuracy of fault diagnosis. Although BPNN is an excellent multiclass classification algorithm, it is applied to the fault diagnosis of ADN. The input matrix is huge and there is a risk of falling into a dimensional disaster.

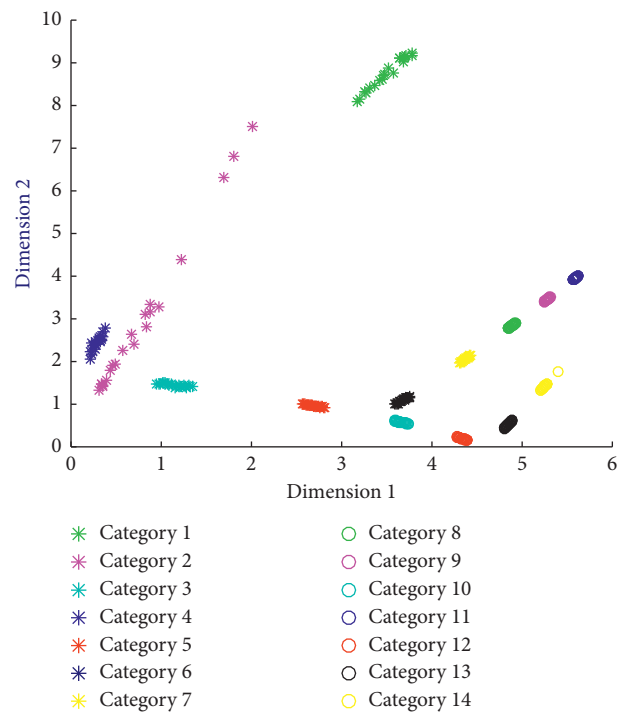

Figure 15: Two-dimensional sample data distribution diagram of fault location with DG1 on-grid and DG2 off-grid.

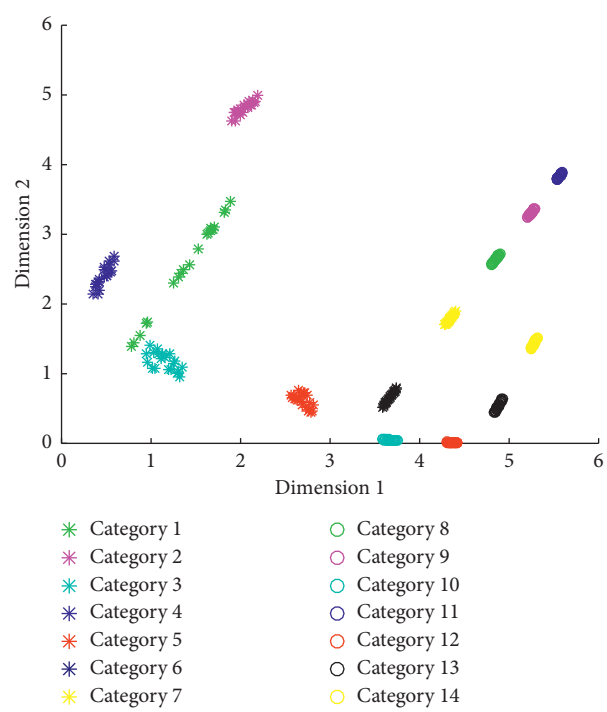

Figure 16: Two-dimensional sample data distribution diagram of fault location with DG1 off-grid and DG2 on-grid.

Table 2: Comparison of the proposed SBS/SFS-SVM and backpropagation neural network.

Algorithm	Probability of diagnostic accuracy		
	Running state discrimination (%)	Fault type discrimination (%)	Fault location (%)
Backpropagation neural network	100	85.24	66.33
SBS/SFS-SVM	100	97.27	95

5. Conclusions

Fault diagnosis of distributed networks is vital to ensure the reliability of the power grid. Numerous related diagnostic models have been applied in this area. However, they lack the combination of functions required for optimal performance, namely, state identification, fault type identification, and fault location. Existing approaches focus solely on distribution network fault diagnosis, ignoring the addition of DGs to the grid.

In addition, the existing methods need to combine different types of electrical quantity information in order to achieve fault diagnosis and location, among which increases the difficulty in information processing. Our complex ADN fault diagnostic approach performs exceptionally in-state identification, fault type identification, and fault location, while relying upon only voltage information.

Specifically, single SVM is applied to distinguish the operation state of faults in the ADN, and multi-SVM, obeying "1-a-1" rule, was used to identify fault type and locate faults. To enable better extraction of feature matrices for SVM, SFS and SBS were successfully employed to improve accuracy. We validated the optimized model by establishing an IEEE 33-node model with two WTs as DGs and rigorously testing with 160 sets of simulated data. The results show that the SFS/SBS-SVM method can reliably and accurately locate the fault and determine its type.

In practice, a large number of electronic devices are connected to an ADN, which limits the fault current and obscures the failure characteristics. Future work will focus on extending fault identification and location in ADNs under more realistic, weak current conditions.

This method is proposed for the problem that it is of significance to capture the short-circuit current caused by the modulation strategy of the power electronic device in the ADN of the AC system. Therefore, this method is also applicable to a DC system where a large number of power electronic devices cause weak currents to be difficult to capture. Considering the difference between the IEEE 33-node model and the actual ADN model, the following two situations should be discussed in applying this method to practical work: (a) The newly created ADNs. Due to the lack of historical operation information, it is impossible to use a large amount of data to establish an SFS/SBS-SVM model. Therefore, an ADN model consistent with the actual system can be constructed through the simulation platform (PSCAD), and a large number of fault samples can be obtained through simulation, so as to train the required fault diagnosis model. (b) For the ADN that has been put into operation. The training process of the SFS/SBS-SVM model can be directly based on the historical data obtained from the measurement points and the phase voltage components selected in section 3.

In conclusion, the SBS/SFS-SVM model is well suited for ADN fault diagnosis problems. It is also adaptable for classification and prediction problems in other areas such as electrical devices monitoring.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the State Grid Corporation of China Research Program "Research on Intelligent Fault Diagnosis and Prediction Technology of Active Distribution Network Based on Data Relevance (PDB17201800193)."

References

[1] C. Y. Ma, C. X. Li, X. Z. Zhang, G. X. Li, and Y. G. Han, "Reconfiguration of Distribution Networks with Distributed Generation Using a Dual Hybrid Particle Swarm Optimization Algorithm," *Mathematical Problems in Engineering*, vol. 2017, Article ID 1517435, 10 pages, 2017.

[2] M. Sedghi, A. Ahmadian, and M. Aliakbar-Golkar, "Optimal storage planning in active distribution network considering uncertainty of wind power distributed generation," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 304–316, 2016.

[3] S. Jamali and H. Borhani-Bahabadi, "Protection method for radial distribution systems with DG using local voltage measurements," *IEEE Transactions on Power Delivery*, vol. 34, no. 2, pp. 651–660, 2019.

[4] Y. C. Sun, N. L. Tai, and X. D. Zheng, "Research on single-phase grounding fault location of distribution network with DC," *Journal of Electric Power Science and Technology*, vol. 31, pp. 73–80, 2016.

[5] M. Y. Gao, B. Y. Xu, and X. H. Zhang, "Fault location based on fault current amplitude comparison for active distribution network," *Electric Power Automation Equipment*, vol. 35, pp. 21–25, 2015.

[6] C. A. Frate and C. Brannstrom, "How do stakeholders perceive barriers to large-scale wind power diffusion? A Q-method case study from ceará state, Brazil," *Energies*, vol. 12, no. 11, p. 2063, 2019.

[7] I. Aboudrar, S. El Hani, M. S. Heyine, and N. Naseri, "Dynamic modeling and robust control by ADRC of grid-connected hybrid PV-wind energy conversion system,"

Mathematical Problems in Engineering, vol. 2019, Article ID 8362921, 19 pages, 2019.

[8] A. Bahmanyar and S. Jamali, "Fault location in active distribution networks using non-synchronized measurements," *International Journal of Electrical Power & Energy Systems*, vol. 93, pp. 451–458, 2017.

[9] Y. Li, B. Feng, G. Li, J. Qi, D. Zhao, and Y. Mu, "Optimal distributed generation planning in active distribution networks considering integration of energy storage," *Applied Energy*, vol. 210, pp. 1073–1081, 2018.

[10] B. Xu, Z. Zhu, and Yu Chen, "Christoph Brunner Application of IEC 61850 for distribution network automation with distributed control," in *Proceedings of the DPSP 2018 The 14th International Conference on Developments in Power System Protection*, Belfast, UK, March 2018.

[11] M. Monadi, C. Koch-Ciobotaru, A. Luna et al., "Implementation of the differential protection for MVDC distribution systems using real-time simulation and hardware-in-the-loop," in *Proceedings of the IEEE Energy Conversion Congress and Exposition*, pp. 3380–3385, Montreal, Canada, September 2015.

[12] J. Ren, H. Liang, and Y. Gao, "Research on evaluation of power supply capability of active distribution network with distributed power supply with high permeability," *Energies*, vol. 12, no. 11, p. 2223, 2019.

[13] Y. Chen, D. Liu, B. Xu, and Y. Huang, "Wide area travelling wave fault location in the transmission network," in *Proceedings of the CICED 2010 Proceedings*, pp. 1–6, Nanjing, China, September 2010.

[14] A. F. Novelo, E. Q. Cucarella, E. G. Moreno, and F. M. Anglada, "Fault diagnosis of electric transmission lines using modular neural networks," *IEEE Latin America Transactions*, vol. 14, no. 8, pp. 3663–3668, 2016.

[15] F. Soares Vitor Petite and R. Caneloi dos Santos, *Teixeira Leite Asano P A Scheme Based on ANNs for Single-phase Fault Location in Distribution Systems with DG IEEE*, IEEE, Piscataway, NJ, USA, 2017.

[16] Santoyo-Sanchez A, Ruiz-Beltran E, L. I. Aguirre-Salas, and Ortiz-Muro, *VH Fault Diagnosis of Electrical Systems Using Interpreted Petri Nets*, IEEE, Piscataway, NJ, USA, 2018.

[17] Y. Zhang, T. Li, Na Guangyu, G. Li, and Li Yang, "Optimized extreme learning machine for power system transient stability prediction using synchrophasors," *Mathematical Problems in Engineering*, vol. 2015, Article ID 529724, 8 pages, 2015.

[18] M. L. Ding, D. M. Yang, and X. B. Li, "Fault diagnosis for wireless sensor by twin support vector machine," *Mathematical Problems in Engineering*, vol. 2013, Article ID 718783, 5 pages, 2013.

[19] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electric Power Systems Research*, vol. 83, no. 1, pp. 73–79, 2012.

[20] V. N. Vapnik, "Adaptive and learning systems for signal processing communications, and control," in *Neural-Based Orthogonal Data Fitting: The EXIN Neural Networks*, Wiley-IEEE Press, Piscataway, NJ, USA, 1998.

[21] X. Shen, L. Niu, Z. Qi, and Y. Tian, "Support vector machine classifier with truncated pinball loss," *Pattern Recognition*, vol. 68, pp. 199–210, 2017.

[22] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, Berlin, Germany, 2013.

[23] G. Jun, F.-l. Chung, and S. Wang, "Matrix pattern based minimum within-class scatter support vector machines," *Applied Soft Computing*, vol. 11, no. 8, pp. 5602–5610, 2011.

[24] B. Liu, S. Li, Y. Wang, L. Lu, Y. Li, and Y. Cai, "Predicting the protein SUMO modification sites based on properties sequential forward selection (PSFS)," *Biochemical and Biophysical Research Communications*, vol. 358, no. 1, pp. 136–139, 2007.