



# A Random Forest and Current Fault Texture Feature–Based Method for Current Sensor Fault Diagnosis in Three-Phase PWM VSR

Lei Kou<sup>1,2</sup>, Xiao-dong Gong<sup>1,2,3</sup>, Yi Zheng<sup>1,2\*</sup>, Xiu-hui Ni<sup>1,2,3</sup>, Yang Li<sup>4</sup>, Quan-de Yuan<sup>5</sup> and Ya-nan Dong<sup>6</sup>

<sup>1</sup>Institute of Oceanographic Instrumentation, Qilu University of Technology (Shandong Academy of Sciences), Qingdao, China, <sup>2</sup>Joint China-Ukrainian Scientific and Innovation Laboratory for Hydroacoustics, Qingdao, China, <sup>3</sup>National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine, <sup>4</sup>School of Electrical Engineering, Northeast Electric Power University, Jilin, China, <sup>5</sup>School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun, China, <sup>6</sup>Qingdao Branch, China Merchants Bank, Qingdao, China

## OPEN ACCESS

### Edited by:

Bo Yang,  
Kunming University of Science and  
Technology, China

### Reviewed by:

Jun Yin,  
North China University of Water  
Resources and Electric Power, China  
Jia Cui,  
Shenyang University of Technology,  
China

### \*Correspondence:

Yi Zheng  
zhengy@sdas.org

### Specialty section:

This article was submitted to  
Smart Grids,  
a section of the journal  
Frontiers in Energy Research

**Received:** 12 May 2021

**Accepted:** 25 May 2021

**Published:** 23 June 2021

### Citation:

Kou L, Gong X, Zheng Y, Ni X, Li Y,  
Yuan Q and Dong Y (2021) A Random  
Forest and Current Fault Texture  
Feature–Based Method for Current  
Sensor Fault Diagnosis in Three-Phase  
PWM VSR.  
*Front. Energy Res.* 9:708456.  
doi: 10.3389/fenrg.2021.708456

Three-phase PWM voltage-source rectifier (VSR) systems have been widely used in various energy conversion systems, where current sensors are the key component for state monitoring and system control. The current sensor faults may bring hidden danger or damage to the whole system; therefore, this paper proposed a random forest (RF) and current fault texture feature–based method for current sensor fault diagnosis in three-phase PWM VSR systems. First, the three-phase alternating currents (ACs) of the three-phase PWM VSR are collected to extract the current fault texture features, and no additional hardware sensors are needed to avoid causing additional unstable factors. Then, the current fault texture features are adopted to train the random forest current sensor fault detection and diagnosis (CSFDD) classifier, which is a data-driven CSFDD classifier. Finally, the effectiveness of the proposed method is verified by simulation experiments. The result shows that the current sensor faults can be detected and located successfully and that it can effectively provide fault locations for maintenance personnel to keep the stable operation of the whole system.

**Keywords:** data-driven, random forests, current fault texture features, current sensors, fault diagnosis, three-phase PWM VSR

## 1 INTRODUCTION

Three-phase PWM VSR systems have been incrementally applied in the field of renewable energy systems, uninterruptible power supplies (UPSs), motor drivers, and so forth and have become an indispensable energy conversion device (Yang et al., 2016; Li and Yang, 2017; Bueno and Pomilio, 2018; Zhang et al., 2021). However, most of the renewable energy equipment is built in the mountains or remote areas away from the coast as the probability of sensor faults is very high under harsh climates or conditions (Bidadfar et al., 2021). Once the fault occurs in the key sensors related to the control system, it may cause the whole system to crash or run inefficiently (Peng et al., 2018). Therefore, the sensor fault diagnosis for three-phase PWM VSR systems is of great significance to ensure the reliability of the whole system.

According to the fault degree, the sensor fault types can be divided into hard faults and soft faults (Huang and Tan, 2008; Li et al., 2011; Darvishi et al., 2021). Hard faults are usually caused by sensor components' damage, or electrical system short circuit or open circuit, and the measured value will change greatly. Soft faults generally refer to the aging of sensor components. The harm of soft faults is not as great as that of hard faults, but long-term operation will reduce the efficiency and accelerate the aging of systems (Fravolini et al., 2019). For example, if the fault output signals of the current sensors are used as the input signals of the control system, it will affect the closed-loop feedback control, reduce the control performance, accelerate the aging of other equipment, and even lead to safety accidents. If the early soft fault features are monitored, the potential hazards can be found in time, the maintenance can be carried out in time, other health equipment can be protected, and the stability of the whole system can be ensured. Therefore, an accurate and effective fault diagnosis method for sensor faults is especially essential to enhance the reliability of the system, and it would be better if the fault locations can be detected.

Sensor fault detection and diagnosis (SFDD) methods can be broadly divided into data-driven and model-based methods (Reppa et al., 2015; Lee et al., 2021). The model-based methods are usually easy to integrate into control systems, but they also need to set complex thresholds, and the model-based methods are more difficult to apply in different fields, especially for a nonlinear system since the fault models are more difficult to establish (Kou et al., 2020; Wang et al., 2020). The data-driven methods can only use the historical data to establish a black-box model and use the mature black-box classifier to realize fault diagnosis and location, without the requirement of understanding the fault mathematical model about the sensor systems (Li F. et al., 2021; Shi et al., 2008). Therefore, the data-driven methods do not rely on mathematical models and have attracted attention of many scholars (Ojo et al., 2021). Lee et al. (2021) proposed a convolutional neural network (CNN)-based FDD method for battery energy storage systems to detect and classify false battery sensor data. Ojo et al. (2021) proposed a long short-term memory recurrent neural network (LSTM-RNN)-based thermal fault diagnosis method for lithium-ion batteries, which is an easy-to-implement way and does not need to pay attention to the complex mathematical modeling and parameters of battery physics. Chen et al. (2021a) proposed a hypergrid and statistical analysis-based FDD method, which can identify the sensor faults in wireless sensor networks. Jana et al. (2021) developed a distributed SFDD framework, in which a fuzzy deep neural network (FDNN) was adopted to detect and identify the sensor faults. Hajer and Okba (2020) proposed an interval-valued data-driven method, which was adopted to detect and locate the sensor faults in chemical industrial fields. Gao et al. (2020) proposed a CNN-based FDD method for micro-electromechanical system (MEMS) inertial sensors, in which the time-domain features of temperature-related sensor faults were adopted to train the data-driven FDD classifier. Haldimann et al. (2020) proposed a disentangled RNN and residual analysis-based SFDD method and

developed a novel procedure to identify the fault sensors. Chen et al. (2021b) proposed an NN-based fault estimation method, which can obtain the accurate estimation of sensor faults. Li L. et al. (2021) proposed an LSTM-based CSFDD method, which can learn data features automatically and predict the acceleration responses from the measured data. According to the previous discussion, data-driven methods have been widely used for sensor fault diagnosis in various fields. Data-driven methods can effectively set up the nonlinear model between input features and fault modes, and the historical fault data under both normal and fault modes can be obtained from the simulation tools. However, the research on fault data extraction is very important, which is helpful to improve the fault diagnosis accuracy of the whole diagnosis system.

Many scholars have made good achievements in sensor fault diagnosis, which have given us many experiences. Although the artificial neural network (ANN) is a popular supervised learning algorithm, it is vulnerable to causing over-fitting and affecting the generalization ability and the diagnosis results. Therefore, the random forest (RF) algorithm is adopted to train the CSFDD classifier, which is not easy to fall into over-fitting due to the introduction of two randomness (random samples, random features) (Roy et al., 2020; Fezai et al., 2021). Meanwhile, the current fault texture features are proposed to train the RF CSFDD classifier, which can improve the feature diversity and the diagnosis accuracy. Hard faults can cause severe damage, and it usually takes a long time to change from soft faults to hard faults. Therefore, if soft faults can be detected and diagnosed in time, the equipment can be maintained in time and hard faults can be avoided.

As shown in **Figure 1**, the proposed method can be applied to the sensor fault diagnosis of power-electronics energy conversion systems in the photovoltaic system, the wind generation system, and so forth. Therefore, this paper mainly takes the soft faults and hard faults of current sensors in the three-phase PWM VSR as examples to verify the proposed CSFDD method. The detailed contributions of this study are summarized as follows:

- (1) Only three-phase alternating currents (ACs) are selected as the input signals, additional sensors will not be introduced, and external interference can be avoided.
- (2) It presents a current fault texture feature-based method, which can retain important fault features.
- (3) A mature data-driven CSFDD classifier is trained by the RF algorithm and extracted current fault texture features, which can realize fault diagnosis and location for current sensors.

The remainder of this paper is organized as follows. **Section 2** describes the current sensor fault data and fault features. In **Section 3**, the current fault texture feature-based method is proposed and the mature RF CSFDD classifier is trained and compared with the classifier trained by original samples. In **Section 4**, the proposed method is verified by simulation experiments. Conclusions are drawn in **Section 5**.

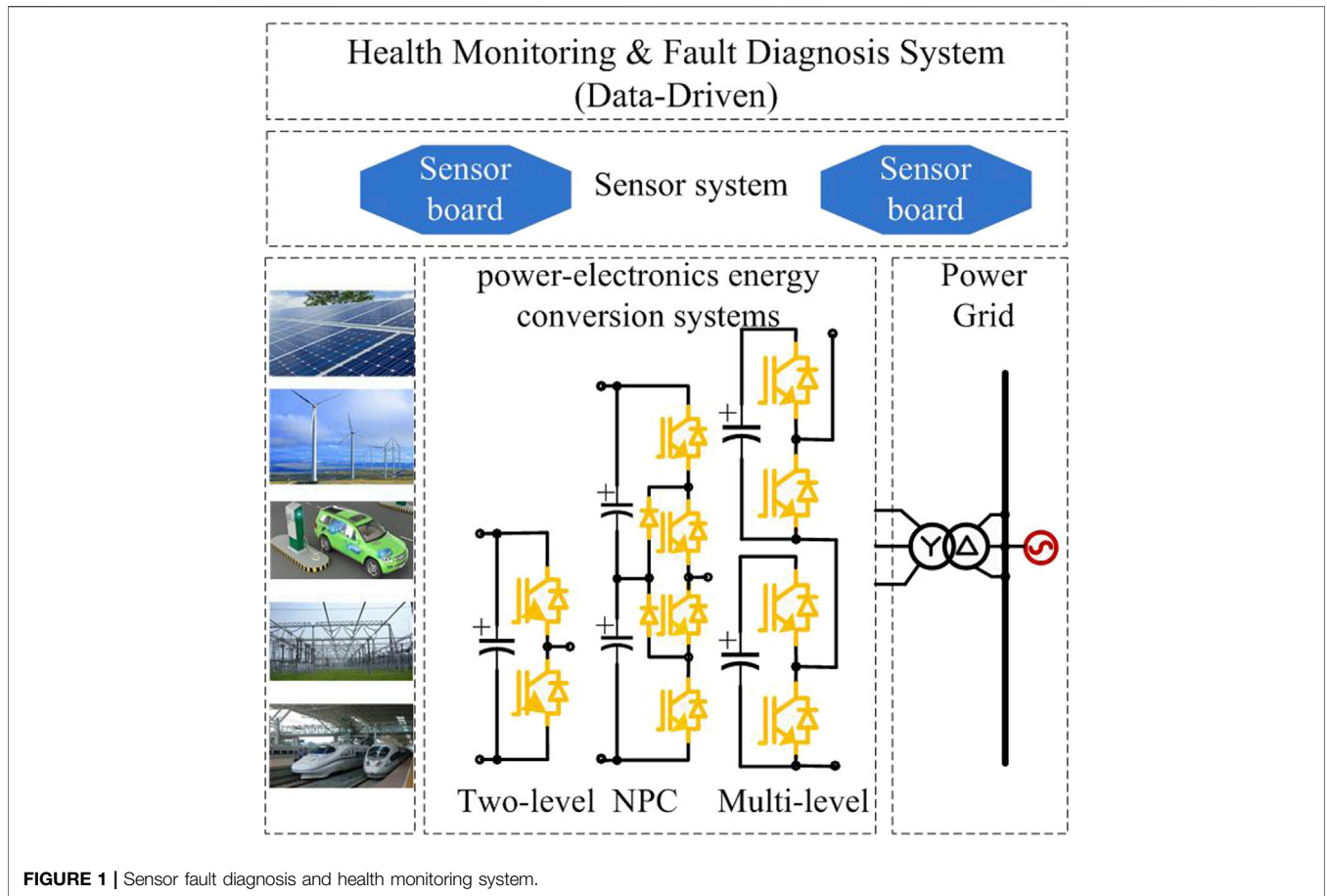


FIGURE 1 | Sensor fault diagnosis and health monitoring system.

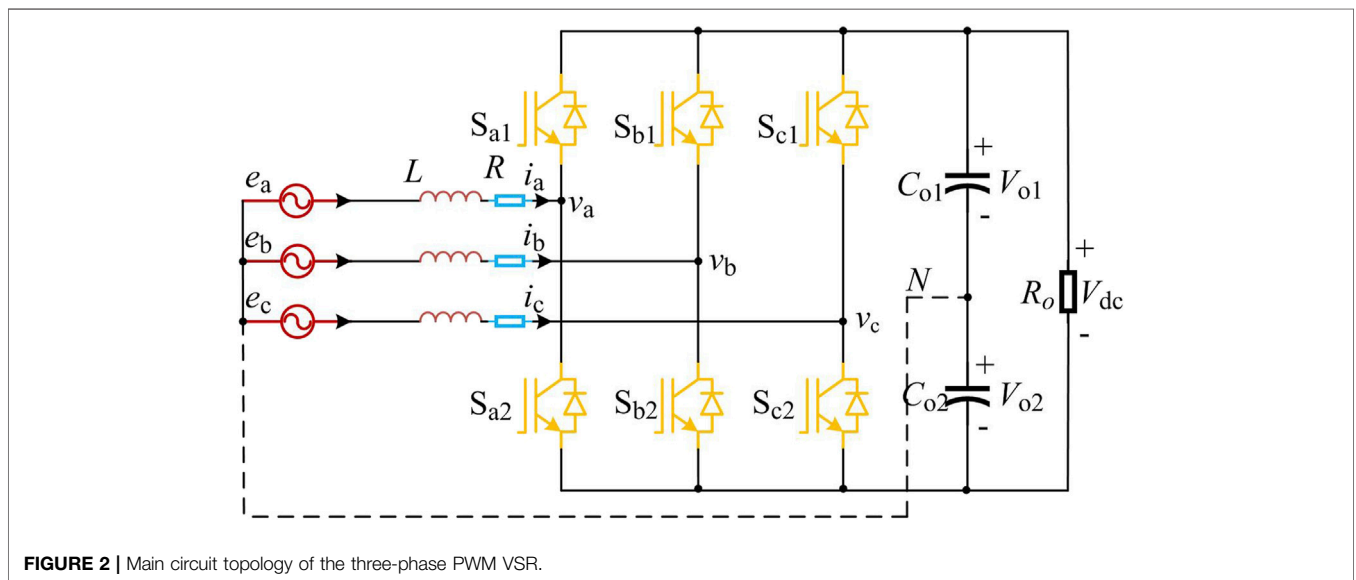


FIGURE 2 | Main circuit topology of the three-phase PWM VSR.

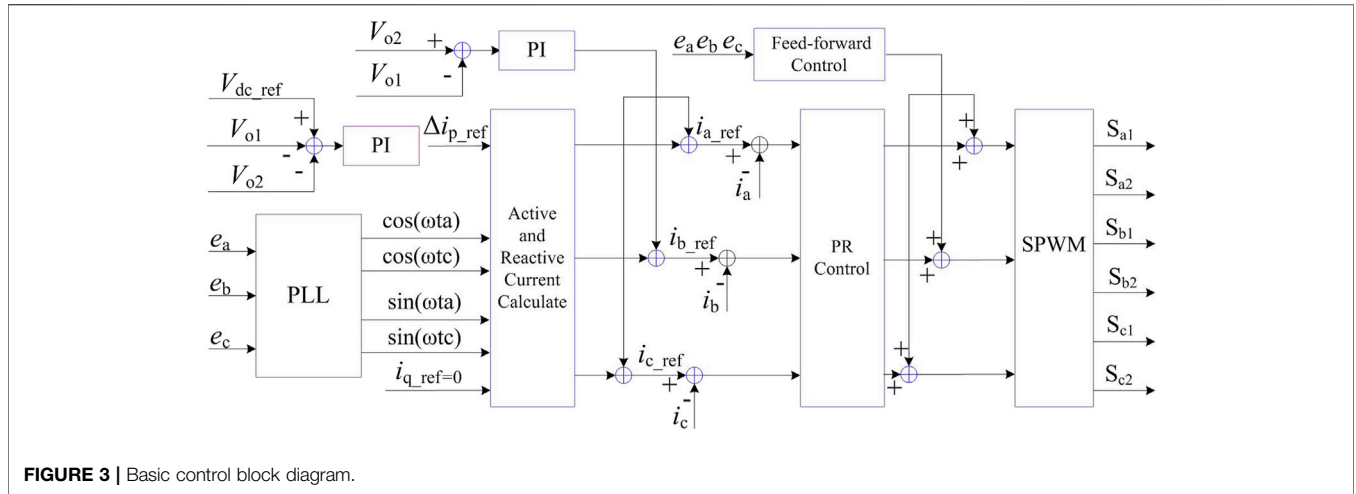


FIGURE 3 | Basic control block diagram.

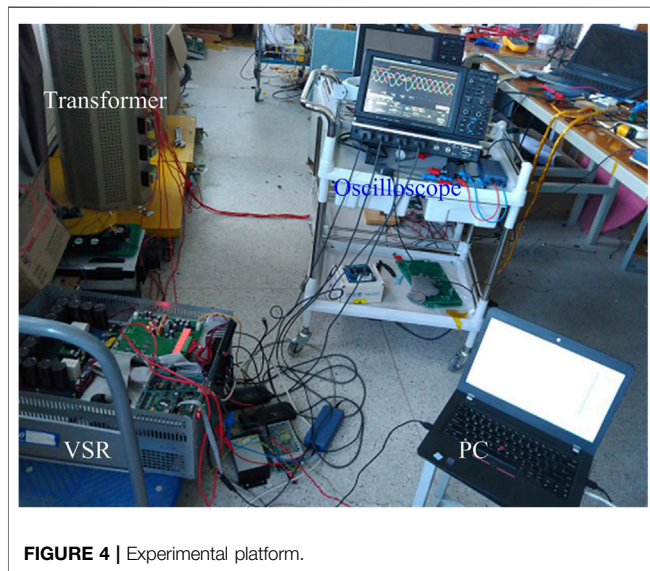


FIGURE 4 | Experimental platform.

## 2 CURRENT SENSOR DATA OF THREE-PHASE PWM VSR

### 2.1 Structure of Three-Phase PWM VSR

The main circuit topology of the three-phase PWM VSR is depicted in Figure 2. As shown in Figure 3, the control strategy of the three-phase PWM VSR in this article is proportional resonance (PR), and it can be referred from Rocha et al. (2018).

Under ideal conditions,  $S_k$  stand for the switching functions, whose equation can be expressed as

$$S_k = \begin{cases} 1 (S_{k1} = 1, S_{k2} = 0) \\ 0 (S_{k1} = 0, S_{k2} = 1) \end{cases} (k = a, b, c) \quad (1)$$

where  $S_{k1}$  and  $S_{k2}$  stand for the upper and lower IGBTs' switching state of each phase, respectively.

The expression of  $v_k$  can be described as

$$\begin{cases} v_a = S_a V_{o1} - (1 - S_a) V_{o2}, \\ v_b = S_b V_{o1} - (1 - S_b) V_{o2}, \\ v_c = S_c V_{o1} - (1 - S_c) V_{o2}. \end{cases} \quad (2)$$

where the relational expressions of  $V_{o2}$  and the output voltage  $V_{dc}$  can be expressed as

$$V_{o1} = V_{o2} = \frac{1}{2} V_{dc}. \quad (3)$$

The dynamic model of the three-phase PWM rectifier can be expressed as follows:

$$\begin{cases} e_a = L \frac{di_a}{dt} + Ri_a + v_a, \\ e_b = L \frac{di_b}{dt} + Ri_b + v_b, \\ e_c = L \frac{di_c}{dt} + Ri_c + v_c. \end{cases} \quad (4)$$

According to Eqs 1–4, the mathematical models in the a–b–c frame for the three-phase PWM rectifier can be expressed as

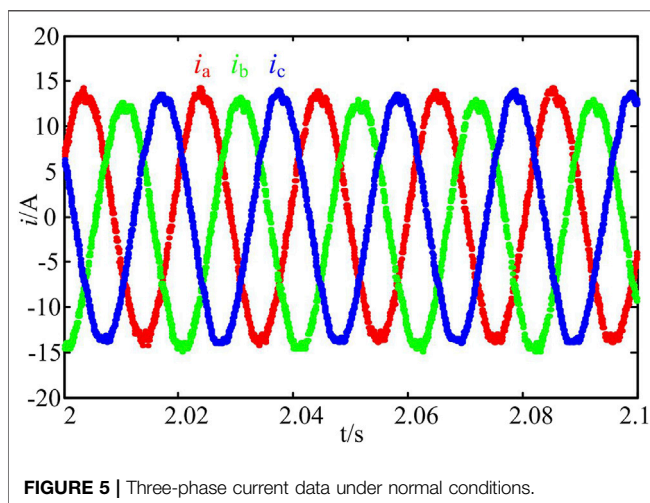


FIGURE 5 | Three-phase current data under normal conditions.



$$\begin{cases} L \frac{di_a}{dt} = e_a - Ri_a - \frac{1}{2}V_{dc}(1 - 2S_a), \\ L \frac{di_b}{dt} = e_b - Ri_b - \frac{1}{2}V_{dc}(1 - 2S_b), \\ L \frac{di_c}{dt} = e_c - Ri_c - \frac{1}{2}V_{dc}(1 - 2S_c). \end{cases} \quad (5)$$

### 2.2 Current Sensor Fault Data

The three-phase currents of the three-phase PWM VSR are collected. **Figure 4** shows the experimental platform of the three-phase PWM VSR, where the input phase voltage is 40 V, the grid voltage frequency is 50 Hz, the load is 16Ω, the sampling frequency is 25.6 kHz, and the output voltage is 100 V. **Figure 5** shows the three-phase AC data under normal conditions.

Suppose the expression of normal three-phase currents is expressed as

$$\begin{cases} i_a = A \sin(\omega t), \\ i_b = A \sin\left(\omega t - \frac{2}{3}\pi\right), \\ i_c = A \sin\left(\omega t + \frac{2}{3}\pi\right), \end{cases} \quad (6)$$

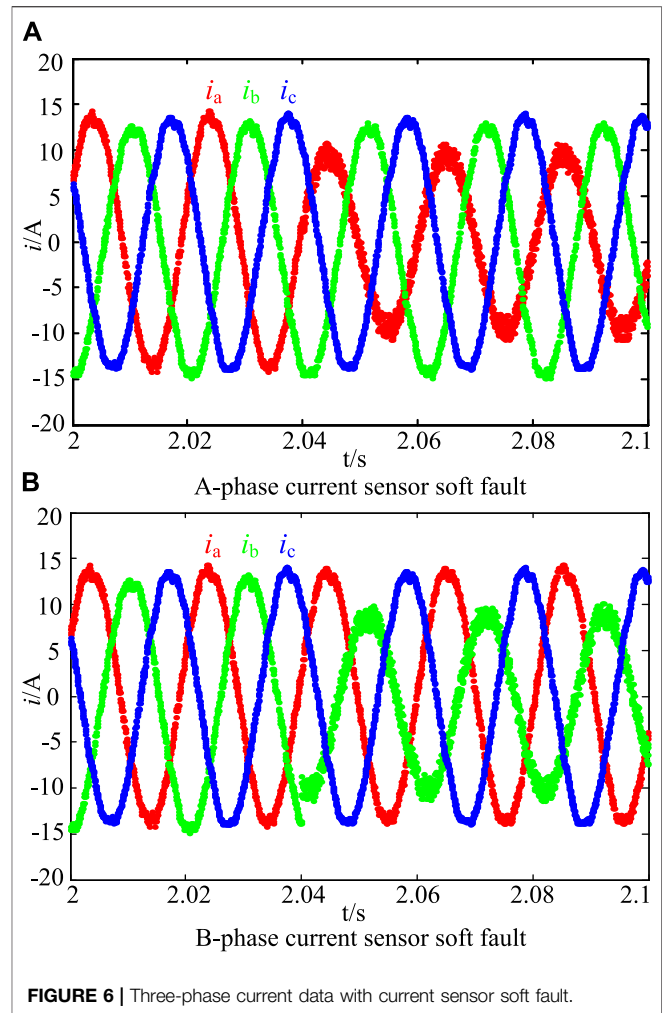
where  $i_a$ ,  $i_b$ , and  $i_c$  are the three-phase currents,  $A$  is the amplitude of three-phase currents, and  $\omega = 100\pi$ .

When the faults occur in the current sensors, the measured three-phase currents will be randomly increased by a number of different deviations. The measured three-phase currents can be expressed as

$$\begin{cases} i'_a = i_a + \sigma, \\ i'_b = i_b + \sigma, \\ i'_c = i_c + \sigma, \end{cases} \quad (7)$$

where  $i'_a$ ,  $i'_b$ , and  $i'_c$  are the measured three-phase currents with different deviations and  $\sigma$  are the different deviations.

According to the research by Li et al. (2011) and Huang and Tan (2008), the soft faults and hard faults are simulated. Taking the soft fault and the hard fault of the A-phase current sensor as examples, the soft fault is simulated by reducing the measured A-phase current to 70% of the normal value (as shown in **Figure 6**) and the hard fault is simulated by reducing the measured value to 40% of the normal value (as shown in **Figure 7**). Meanwhile, the random noises are also superimposed. **Figure 8** shows the three-phase current data with multiple current sensor faults, **Figure 8A** shows the three-phase current data when soft faults occur in both A-phase and B-phase current sensors, and **Figure 8B** shows the three-phase current data when soft fault occurs in the A-phase current sensor and hard fault occurs in the B-phase current sensor. According to **Figures 6–8**, in some cases, when faults occur in multiple current sensors at the same time, the fault phase currents will have superposition effect of fault features. According to **Figures 6, 7**, soft faults have little harm in a short time, but they will bring great harm when these faults develop into hard faults. Therefore, the soft faults of current sensors should be monitored in time to avoid hard faults.



**FIGURE 6 |** Three-phase current data with current sensor soft fault.

## 3 CSFDD CLASSIFIER BASED ON RFS AND CURRENT FAULT TEXTURE FEATURES

In this section, the current fault texture features are proposed, and then the fault diagnosis method based on RFs and current fault texture features is described. And meanwhile compared with the CSFDD classifier trained with original samples, the CSFDD classifier trained with current fault texture feature samples obtained a higher accuracy.

### 3.1 Current Fault Texture Features

Current fault texture features are important attributes of three-phase ACs, which can reduce the data dimension, retain most of the main features, and improve the ability of anti-interference. The main current fault texture features include the mean value, standard deviation, smoothness, skewness, kurtosis, and root mean square value. The current fault texture features retain most of the important features of three-phase ACs in a cycle.

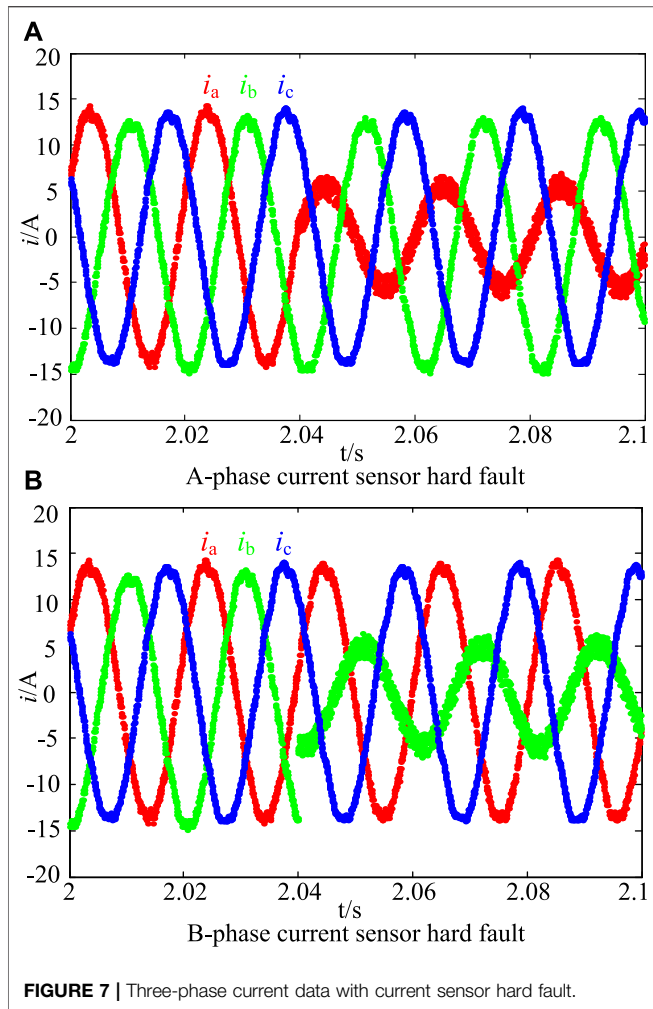


FIGURE 7 | Three-phase current data with current sensor hard fault.

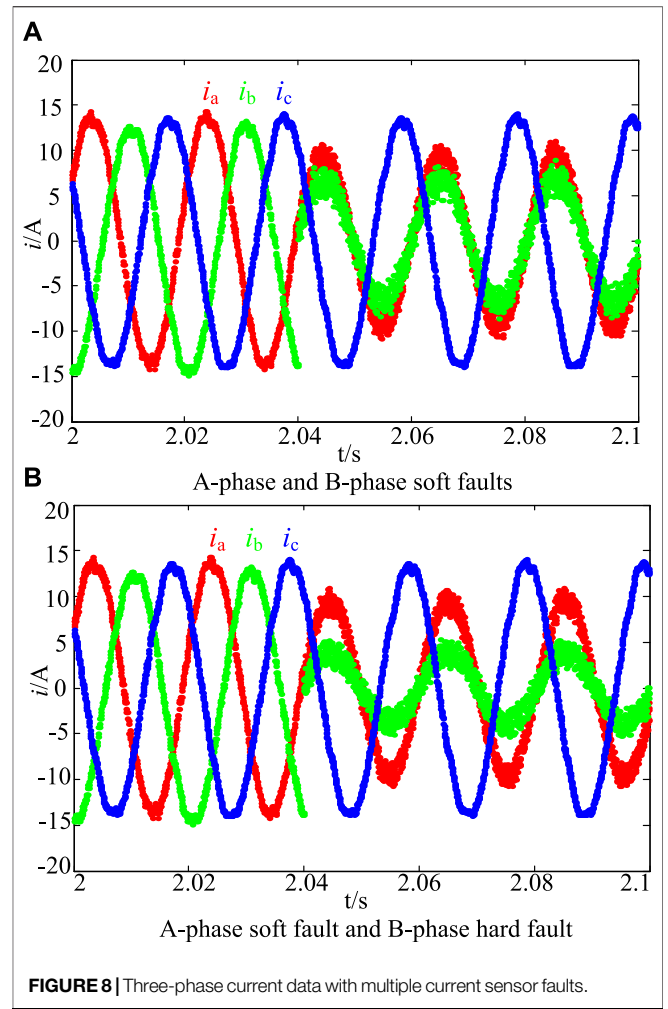


FIGURE 8 | Three-phase current data with multiple current sensor faults.

The expression of the mean value is

$$\bar{i} = \frac{1}{N} \sum_{j=0}^{N-1} i(j), \quad (8)$$

where  $N$  is the number of measured three-phase current values in a cycle and  $i(j)$  is the  $j^{th}$  measured current value.

The expression of standard deviation is

$$\sigma_D = \sqrt{\frac{1}{N} \sum_{j=0}^{N-1} [i(j) - \bar{i}]^2}. \quad (9)$$

The expression of smoothness is

$$\sigma_P = 1 - \frac{1}{1 + \sigma_D^2}. \quad (10)$$

The expression of skewness is

$$\sigma_S = \frac{1}{(N-1)\sigma_D^3} \sum_{j=0}^{N-1} [i(j) - \bar{i}]^3. \quad (11)$$

The expression of kurtosis is

$$\sigma_K = \frac{1}{(N-1)\sigma_D^4} \sum_{j=0}^{N-1} [i(j) - \bar{i}]^4. \quad (12)$$

The expression of the root mean square value is

$$\sigma_R = \sqrt{\frac{1}{N} \sum_{j=0}^{N-1} i(j)^2}. \quad (13)$$

Equations 8–13 are adopted to calculate the fault texture features of ABC three-phase currents, and the current fault texture features are used as the input signals of RFs. The input signals include  $(\bar{i}_a, \sigma_{aD}, \sigma_{aP}, \sigma_{aS}, \sigma_{aK}, \sigma_{aR})$ ,  $(\bar{i}_b, \sigma_{bD}, \sigma_{bP}, \sigma_{bS}, \sigma_{bK}, \sigma_{bR})$ , and  $(\bar{i}_c, \sigma_{cD}, \sigma_{cP}, \sigma_{cS}, \sigma_{cK}, \sigma_{cR})$ . If 200 points are sampled in each cycle, the dimension of the original samples is 600 and the dimension of current fault texture feature samples will be reduced to 18. Through this method, the important features of the signal are extracted and the dimension of the training samples is reduced, which will be more conducive to improving the accuracy of the diagnosis classifier.

**TABLE 1** | Process of the bagging classification algorithm.

- (1) Input: Sample training set  $S=(X^1, Y^1), (X^2, Y^2), \dots, (X^N, Y^N)$ ,  $X^N$  represents the feature attribute,  $Y^N$  represents the sample label; The total number of training is  $M$ ,  $\xi$  is the function of the weak classifier,  $S_{sub-j}$  is the selected sample training set;
- (2) for  $j = 1, 2, 3, \dots, M$
- (3)  $h_j(X) = \xi(S, S_{sub-j})$ ;
- (4) end
- (5) Output:  $H(X) = \operatorname{argmax} \left( \sum_{j=1}^M \prod h_j(X) = Y \right)$

### 3.2 Training and Evaluation of the CSFDD Classifier

As a popular data-driven classifier design approach, RFs are applied to train the CSFDD classifier with the features  $(\bar{i}_a, \sigma_{aD}, \sigma_{aP}, \sigma_{aS}, \sigma_{aK}, \sigma_{aR})$ ,  $(\bar{i}_b, \sigma_{bD}, \sigma_{bP}, \sigma_{bS}, \sigma_{bK}, \sigma_{bR})$ , and  $(\bar{i}_c, \sigma_{cD}, \sigma_{cP}, \sigma_{cS}, \sigma_{cK}, \sigma_{cR})$ . In general, over-fitting always is a troubled problem for most machine learning algorithms, but this problem does not easily occur in the RF classifier (Phan et al., 2020). That is because the RF algorithms can overcome the problem of over-fitting with the help of enough decision trees. The excellent performance of RFs in supervised machine learning is the reason why they are adopted to train the CSFDD classifier.

The RF algorithm is a widely used ensemble learning algorithm, which can make joint decision through multiple weak classification models to improve the accuracy of decision (Gu et al., 2020). In the training process of RFs, samples and features are randomly selected in order to increase the difference between each weak model; the tree model can be more diversified by this way. The bagging method is usually used in the RF algorithm (Song et al., 2016). In the bagging method, the independence of different models is improved by using different independent training datasets. Through random sampling with the replacement method, the training sample set is extracted from the original sample set, and  $n$  training samples sets are extracted from the original sample set each time.

**TABLE 2** | Fault sensors and class labels.

Fault sensors	Class labels					
	S1	S2	S3	H1	H2	H3
Normal state	0	0	0	0	0	0
A-phase soft fault	1	0	0	0	0	0
A-phase hard fault	0	0	0	1	0	0
B-phase soft fault	0	1	0	0	0	0
B-phase hard fault	0	0	0	0	1	0
C-phase soft fault	0	0	1	0	0	0
C-phase hard fault	0	0	0	0	0	1
A-B-phase soft faults	1	1	0	0	0	0
A-B-phase hard faults	0	0	0	1	1	0
A-phase soft fault and B-phase hard fault	1	0	0	0	1	0

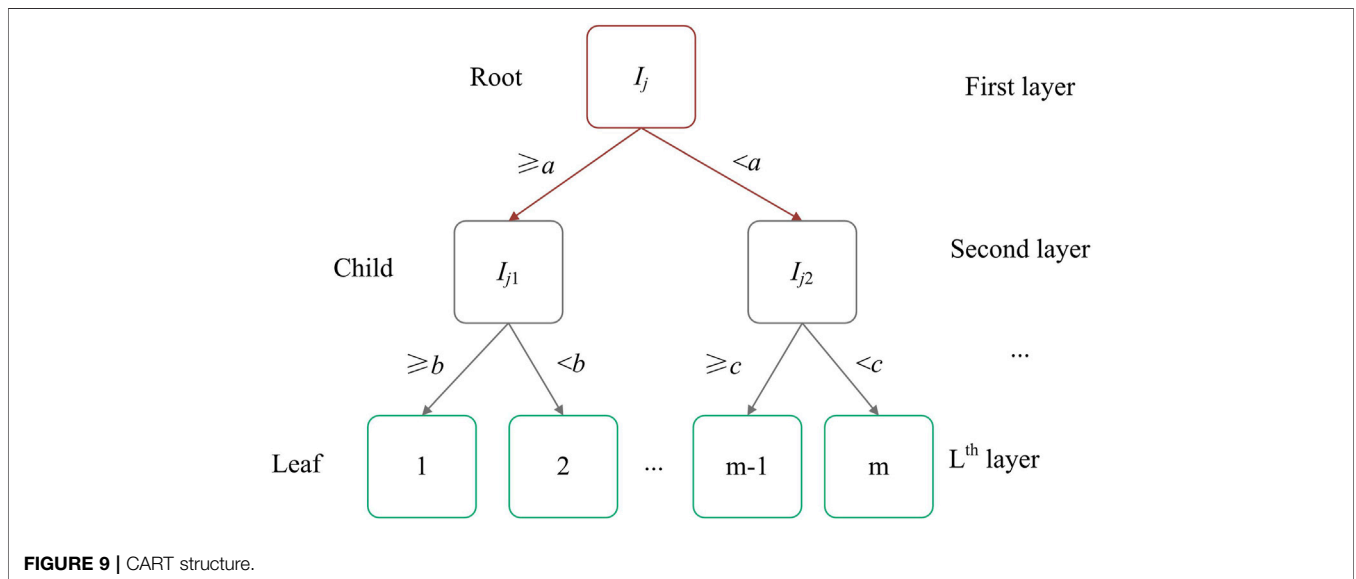
In the training sets, some samples may have been repeatedly extracted and some samples may have never been selected. Using  $n$  training sets to train  $n$  models,  $n$  models are independent of each other, and then the integration model is obtained by voting.

Suppose that the relationship between input samples  $X$  and output class labels  $Y$  is  $Y = f(X)$ . As shown in Figures 10, 11, for the models  $h_1(X), h_2(X), \dots, h_N(X)$  of  $N$  different weak classifiers, the expected error of each model is

$$\begin{aligned} \operatorname{err}(h_k) &= E_X \left[ (h_k(X) - f(X))^2 \right] \\ &= E_X \left[ \varepsilon_k^2(X) \right], \end{aligned} \tag{14}$$

where  $\varepsilon_k(X) = h_k(X) - f(X)$  is the  $k^{\text{th}}$  error of the input samples  $X$ . And the average error of all weak models is

$$\begin{aligned} \overline{\operatorname{err}}(h_k) &= \frac{1}{N} \sum_{k=1}^N E_X \left[ (h_k(X) - f(X))^2 \right] \\ &= \frac{1}{N} \sum_{k=1}^N E_X \left[ \varepsilon_k^2(X) \right]. \end{aligned} \tag{15}$$



**FIGURE 9** | CART structure.

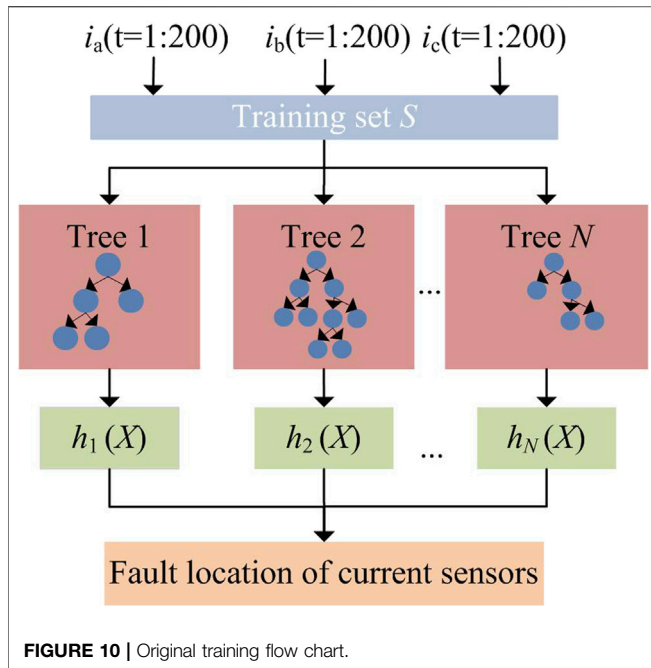


FIGURE 10 | Original training flow chart.

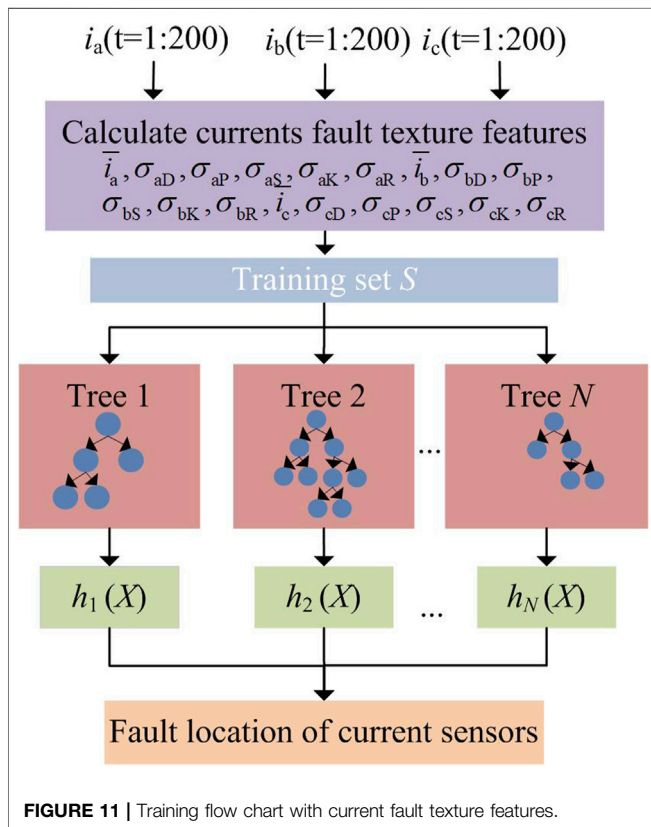


FIGURE 11 | Training flow chart with current fault texture features.

The direct average method is also adopted to integrate all weak classification models, and the expression of the integrated model is

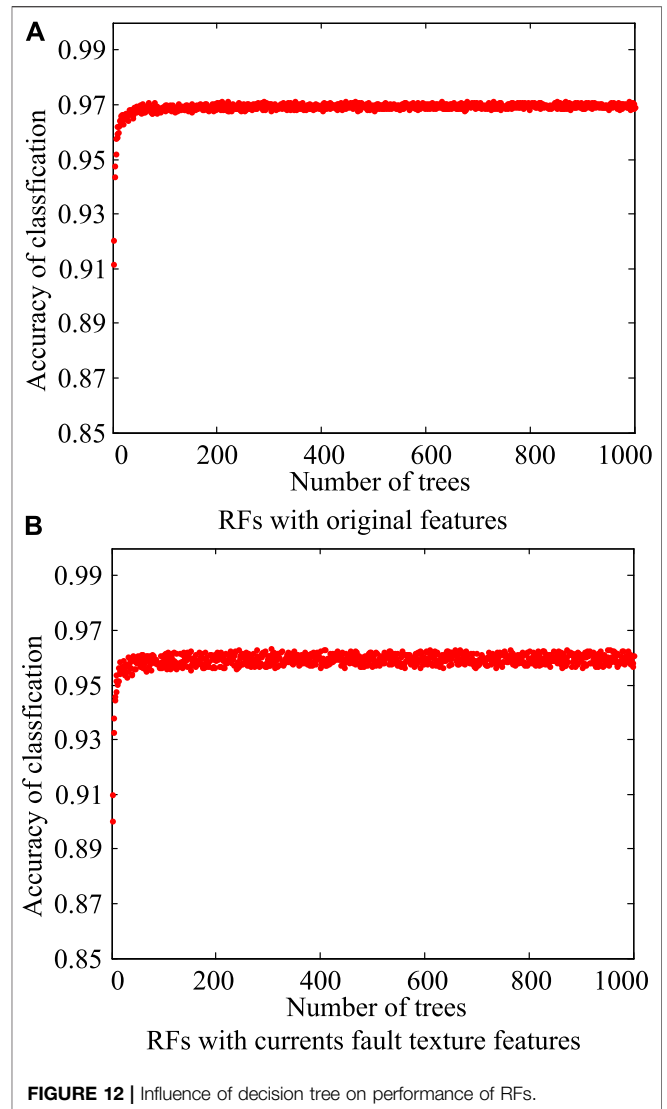


FIGURE 12 | Influence of decision tree on performance of RFs.

$$H(X) = \frac{1}{N} \sum_{k=1}^N h_k(X). \tag{16}$$

The expected error of the integrated model is

$$\begin{aligned} \text{err}(H) &= E_X \left[ \left( \frac{1}{N} \sum_{k=1}^N h_k(X) - f(X) \right)^2 \right] \\ &= \frac{1}{N^2} E_X \left[ \left( \sum_{k=1}^N \varepsilon_k(X) \right)^2 \right] \\ &= \frac{1}{N^2} \sum_{k=1}^N \sum_{i=1}^N E_X [\varepsilon_k(X) \varepsilon_i(X)], \end{aligned} \tag{17}$$

where  $E_X [\varepsilon_k(X) \varepsilon_i(X)]$  is the correlation of two different models' errors. When the error of each weak classification model is irrelevant, for any  $n \neq i$ ,  $E_X [\varepsilon_k(X) \varepsilon_i(X)] = 0$ . When the error of each weak classification model is the same, for any  $n \neq i$ ,



**TABLE 3** | Accuracy of RFs with original features.

Fault sensors	Diagnosis accuracy	Misdiagnosis rate
Normal state	0.9671	0.0329
A-phase soft fault	0.9606	0.0394
A-phase hard fault	0.9652	0.0348
B-phase soft fault	0.9486	0.0514
B-phase hard fault	0.9673	0.0327
C-phase soft fault	0.9526	0.0474
C-phase hard fault	0.9611	0.0389
A–B–phase soft faults	0.9643	0.0357
A–B–phase hard faults	0.9606	0.0394
A-phase soft fault and B-phase hard fault	0.9666	0.0334

$\epsilon_k(X) = \epsilon_i(X)$ . Because  $\epsilon_k(X) \geq 0$  for all weak classification models, the following relationships should be satisfied:

$$\overline{\text{err}}(h) \geq \text{err}(H) \geq \frac{1}{N} \overline{\text{err}}(h). \tag{18}$$

According to the above derivation, the greater the difference between all weak models is, the better the performance of the integrated model will be. At the same time, the error will gradually decrease and tend to zero with the increase in the number of models.

The decision tree is a widely used classification model in the classification algorithm, where the classification and regression tree (CART) has the advantages of fast training speed, good stability, and supporting multiple segmentation of continuous data and feature data. Bagging algorithm plays a key role in the construction of random forest classifier, and its process is shown in **Table 1**. As shown in **Figure 9**, the CART is often used as the basic learning algorithm of RFs, whose structure is a binary classification tree.

As shown in **Table 2**, the class labels  $S_k(k = 1-3)$  mean the case when the soft faults are in A-phase, B-phase, and C-phase current sensors,  $H_k(k = 1-3)$  mean the case when the hard faults are in A-phase, B-phase, and C-phase current sensors, and  $S_k = 1$  or  $H_k = 1$  means the fault occurs in the current sensor. Each sample in the training datasets consists of several current fault texture features and one class label, for example,  $X=(\bar{i}_a, \sigma_{aD}, \sigma_{aP}, \sigma_{aS}, \sigma_{aK}, \sigma_{aR}, \bar{i}_b, \sigma_{bD}, \sigma_{bP}, \sigma_{bS}, \sigma_{bK}, \sigma_{bR}, \bar{i}_c, \sigma_{cD}, \sigma_{cP}, \sigma_{cS}, \sigma_{cK}, \sigma_{cR})$ ,  $Y=(S1, S2, S3, H1, H2, H3)$ . The purpose of the RF classifier is adopted to make a useful model with better classification performance, and it will formulate some set of rules when

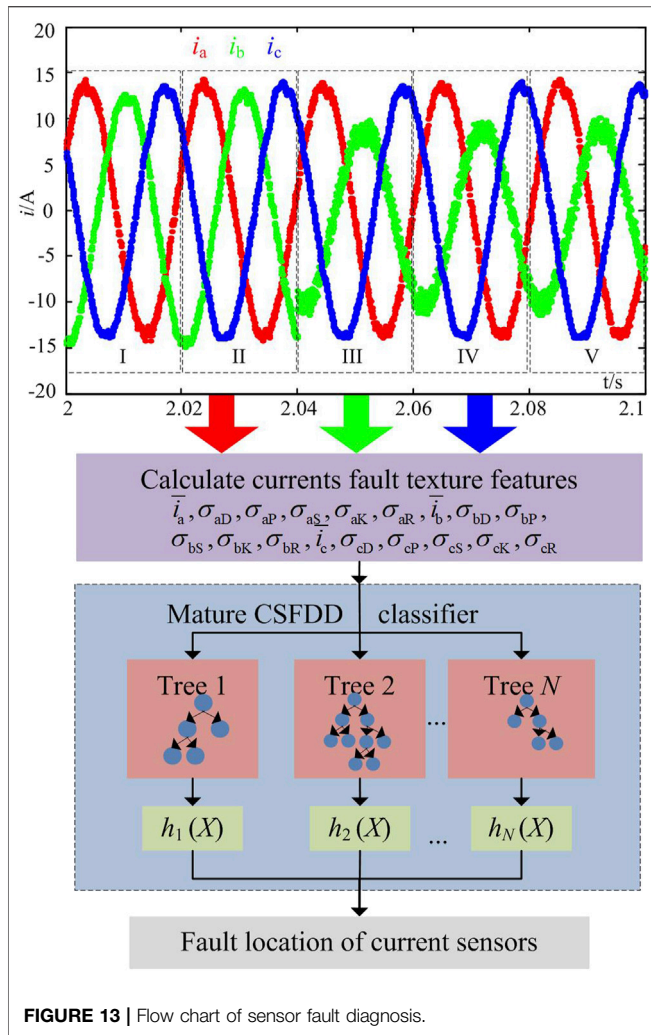
inputting a training dataset with features  $X$  and labels  $Y$  into the RF algorithm (as shown in **Figure 10**). What is most mysterious is that it can correctly predict the target class label of the sample in the testing datasets, which are given only the features without class labels. **Figure 10** shows the original training flow chart of the RF algorithm with original features, and **Figure 11** shows the training flow chart of the RF algorithm with current fault texture features. The biggest difference between them is that the latter introduces the current fault texture feature-based method to transform the original data features, where the current fault texture features are obtained after the transform of three-phase currents. The current fault texture feature-based method is adopted to perform feature extraction, and the data-driven method is used to train the CSFDD classifier with the extracted features.

In the training process of the CSFDD classifier, the training set is formed by randomly selecting 16,800 samples from 24,000 samples under different operating conditions, and the rest samples are used as the test samples. Here, the repeated cross test method is adopted to train the RF classifier. **Figure 12A** shows the trend of diagnosis results varying with trees when the CSFDD classifier was trained with original features, and when the number of RF decision trees is set as 305, the highest accuracy is 0.9636. **Figure 12B** shows the trend of diagnosis results varying with trees when the CSFDD classifier was trained with current fault texture features, and when the RF trees' number is set as 209, the highest accuracy is 0.9715.

**Table 3** shows the diagnosis accuracy of RFs with original features, in which the CSFDD classifier was trained with the original fault current samples (as shown in **Figure 10**).

**TABLE 4** | Accuracy of RFs with current fault texture features.

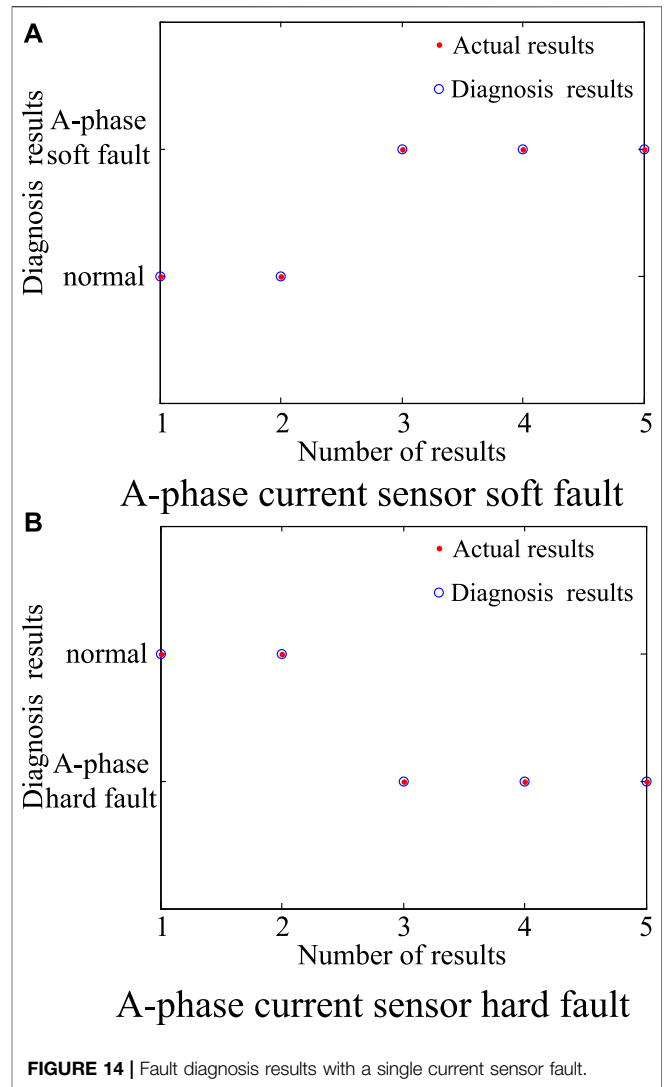
Fault sensors	Diagnosis accuracy	Misdiagnosis rate
Normal state	0.9745	0.0255
A-phase soft fault	0.9743	0.0257
A-phase hard fault	0.9650	0.0350
B-phase soft fault	0.9663	0.0337
B-phase hard fault	0.9752	0.0248
C-phase soft fault	0.9715	0.0285
C-phase hard fault	0.9725	0.0275
A–B–phase soft faults	0.9735	0.0265
A–B–phase hard faults	0.9675	0.0325
A-phase soft fault and B-phase hard fault	0.9730	0.0270



**Table 4** shows the diagnosis accuracy of the proposed method, in which the CSFDD classifier was trained with current fault texture features (as shown in **Figure 11**). According to **Tables 3, 4**, since the original fault current samples contain more noise, and the current fault texture features retain a large number of important features and remove noise, the accuracies of the proposed method are higher than those of RFs with original features. The proposed method offers high accuracy even with a training dataset of 70% of total samples compared with the test/validation dataset, which gives more than 97.15% average accuracy for different types of faults.

### 4 SIMULATION EXPERIMENTS

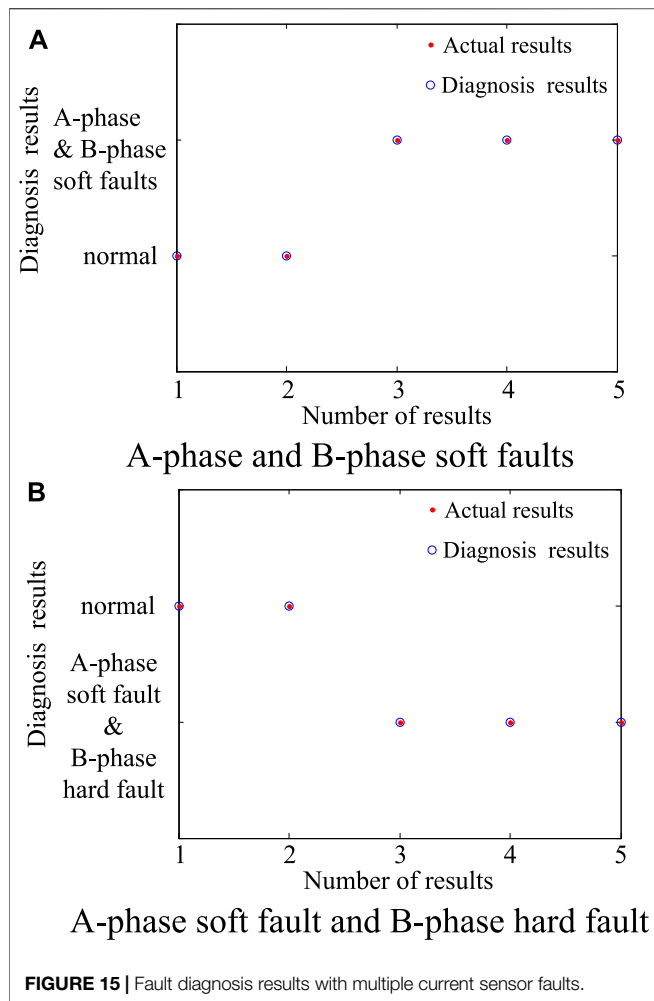
To verify the validity of the CSFDD classifier based on RFs and current fault texture features, the simulation experiments of soft and hard faults are carried out, and the soft fault and hard fault of the A-phase current sensor and multiple current sensor faults are studied as examples. **Figure 13** shows the flow chart



of sensor fault diagnosis, where the current samples are five cycles of fault samples, and it contains five sets of samples. When a set of fault samples are input each time, they will be processed by the current fault texture feature method and then input into the mature CSFDD classifier to get the final diagnosis results. The data of **Figures 6A, 7A, 8** are input into the mature CSFDD classifier (samples of one cycle are input at a time, as shown in **Figure 13**), and then the location and type of the fault sensor can be diagnosed.

**Figure 14** shows the diagnosis results of five cycles. **Figure 14A** shows the diagnosis results of **Figure 6A**; at the beginning of diagnosis, the results are normal and the sensors work normally. When the soft fault occurs in the A-phase current sensor, the diagnosis results turn to A-phase soft faults, and the diagnosis results are consistent with the actual results. **Figure 14B** shows the diagnosis results of **Figure 7A**; the early diagnosis results are normal, the diagnosis results turn to A-phase hard faults when the hard fault occurs in the A-phase current sensor, and the diagnosis results are consistent with the actual results.

**Figure 15** shows the fault diagnosis results with multiple current sensor faults. **Figure 15A** shows the diagnosis results



of **Figure 8A**, and **15B** shows the diagnosis results of **Figure 8B**. According to the diagnosis results, the proposed method can locate multiple faults and different types of faults.

According to the diagnosis results of **Figures 14, 15**, the current sensor faults can be effectively diagnosed and located by the proposed method.

## REFERENCES

- Bidadfar, A., Saborio-Romano, O., Cutululis, N. A., and Sorensen, P. E. (2021). Control of Offshore Wind Turbines Connected to Diode-Rectifier-Based HvdC Systems. *IEEE Trans. Sustain. Energ.* 12, 514–523. doi:10.1109/TSTE.2020.3008606
- Bueno, A. G., and Pomilio, J. A. (2018). “Balancing Voltage in the Dc Bus with Split Capacitors in Three-phase Four-Wire Pwm Boost Rectifier,” in 2018 13th IEEE International Conference on Industry Applications (INDUSCON), 523–529. doi:10.1109/INDUSCON.2018.8627210
- Chen, L., Li, G., and Huang, G. (2021a). A Hypergrid Based Adaptive Learning Method for Detecting Data Faults in Wireless Sensor Networks. *Inf. Sci.* 553, 49–65. doi:10.1016/j.ins.2020.12.011
- Chen, L., Liu, M., Shi, Y., Zhang, H., and Zhao, E. (2021b). Adaptive Fault Estimation for Unmanned Surface Vessels with a Neural Network Observer Approach. *IEEE Trans. Circuits Syst.* 68, 416–425. doi:10.1109/TCSI.2020.3033803
- Darvishi, H., Ciunozzo, D., Eide, E. R., and Rossi, P. S. (2021). Sensor-fault Detection, Isolation and Accommodation for Digital Twins via Modular Data-Driven Architecture. *IEEE Sensors J.* 21, 4827–4838. doi:10.1109/JSEN.2020.3029459
- Fezai, R., Dhibi, K., Mansouri, M., Trabelsi, M., Hajji, M., Bouzrara, K., et al. (2021). Effective Random forest-based Fault Detection and Diagnosis for Wind Energy Conversion Systems. *IEEE Sensors J.* 21, 6914–6921. doi:10.1109/JSEN.2020.3037237
- Fravolini, M. L., Del Core, G., Papa, U., Valigi, P., and Napolitano, M. R. (2019). Data-driven Schemes for Robust Fault Detection of Air Data System Sensors. *IEEE Trans. Contr. Syst. Technol.* 27, 234–248. doi:10.1109/TCST.2017.2758345
- Gao, T., Sheng, W., Zhou, M., Fang, B., and Zheng, L. (2020). Mems Inertial Sensor Fault Diagnosis Using a Cnn-Based Data-Driven Method. *Int. J. Patt. Recogn. Artif. Intell.* 34, 2059048. doi:10.1142/S021800142059048X

## 5 CONCLUSION

This paper proposed an RF and current fault texture feature-based method for current sensor fault diagnosis in three-phase PWM VSR systems. The proposed method employed only three-phase ACs without other additional hardware. Firstly, the current fault texture feature-based method is adopted for feature extraction, which can retain a large number of important features and remove noise. After that, the current fault texture features are used to train the CSFDD fault classifier by the RF algorithm, and the accuracy of the proposed method is higher than those of RFs with original features. Finally, the results of simulation experiments show that the current sensor faults can be effectively diagnosed and located by the proposed method, and the experimental results prove the effective and robust performance of the proposed method.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

LK and X-DG involved in data curation and writing the original draft. LK, X-DG, and YL performed formal analysis. LK, X-HN, and Y-ND ran the software. LK, YZ, X-HN, YL, and Q-DY validated the data. LK, X-DG, YZ, X-HN, YL, Q-DY, and Y-ND reviewed and edited the paper. YZ administered the project and obtained the resources.

## FUNDING

This research is funded by Special Program on Self-Innovation Ability Improvement (Smart Ocean Project) Central Infrastructure Investment Program 2020 under grant number Z135060000070.

- Gu, K., Zhang, Y., and Qiao, J. (2020). Random forest Ensemble for River Turbidity Measurement from Space Remote Sensing Data. *IEEE Trans. Instrum. Meas.* 69, 9028–9036. doi:10.1109/TIM.2020.2998615
- Lahdhiri, H., and Taouali, O. (2021). Interval Valued Data Driven Approach for Sensor Fault Detection of Nonlinear Uncertain Process. *Measurement* 171, 108776. doi:10.1016/j.measurement.2020.108776
- Haldimann, D., Guerriero, M., Maret, Y., Bonavita, N., Ciarlo, G., and Sabbadin, M. (2020). A Scalable Algorithm for Identifying Multiple-Sensor Faults Using Disentangled Rnns. *IEEE Trans. Neural Netw. Learn. Syst.* 1, 1–14. doi:10.1109/TNNLS.2020.3040224
- Huang, S. N., and Kok Kiang Tan, K. K. (2008). Fault Detection, Isolation, and Accommodation Control in Robotic Systems. *IEEE Trans. Automat. Sci. Eng.* 5, 480–489. doi:10.1109/TASE.2008.917009
- Jan, S. U., Lee, Y. D., and Koo, I. S. (2021). A Distributed Sensor-Fault Detection and Diagnosis Framework Using Machine Learning. *Inf. Sci.* 547, 777–796. doi:10.1016/j.ins.2020.08.068
- Kou, L., Liu, C., Cai, G.-w., Zhang, Z., Zhou, J.-n., and Wang, X.-m. (2020). Fault Diagnosis for Three-phase Pwm Rectifier Based on Deep Feedforward Network with Transient Synthetic Features. *ISA Trans.* 101, 399–407. doi:10.1016/j.isatra.2020.01.023
- Lee, H., Kim, K., Park, J.-H., Bere, G., Ochoa, J. J., and Kim, T. (2021). Convolutional Neural Network-Based False Battery Data Detection and Classification for Battery Energy Storage Systems. *IEEE Trans. Energ. Convers.*, 1. doi:10.1109/TEC.2021.3061493
- Li, Y., and Yang, Z. (2017). Application of Eos-Elm with Binary Jaya-Based Feature Selection to Real-Time Transient Stability Assessment Using Pmu Data. *IEEE Access* 5, 23092–23101. doi:10.1109/ACCESS.2017.2765626
- Li, H., Zhao, M., Zhao, B., Tang, X., and Liu, Z. (2011). Fault Diagnosis Methods for Key Sensors of Doubly Fed Wind Turbine. *Proc. CSEE* 31, 73–78. doi:10.13334/j.0258-8013.pcsee.2011.06.011
- Li, F., Li, Q., Zhang, J., Kou, J., Ye, J., Song, W., et al. (2021a). Detection and Diagnosis of Data Integrity Attacks in Solar Farms Based on Multilayer Long Short-Term Memory Network. *IEEE Trans. Power Electron.* 36, 2495–2498. doi:10.1109/TPEL.2020.3017935
- Li, L., Liu, G., Zhang, L., and Li, Q. (2021b). Fs-lstm-based Sensor Fault and Structural Damage Isolation in Shm. *IEEE Sensors J.* 21, 3250–3259. doi:10.1109/JSEN.2020.3022099
- Ojo, O., Lang, H., Kim, Y., Hu, X., Mu, B., and Lin, X. (2021). A Neural Network Based Method for thermal Fault Detection in Lithium-Ion Batteries. *IEEE Trans. Ind. Electron.* 68, 4068–4078. doi:10.1109/TIE.2020.2984980
- Peng, Y., Qiao, W., Qu, L., and Wang, J. (2018). Sensor Fault Detection and Isolation for a Wireless Sensor Network-Based Remote Wind Turbine Condition Monitoring System. *IEEE Trans. Ind. Appl.* 54, 1072–1079. doi:10.1109/TIA.2017.2777925
- Phan, D., Nguyen, N., Pathirana, P. N., Horne, M., Power, L., and Szmulewicz, D. (2020). A Random forest Approach for Quantifying Gait Ataxia with Truncal and Peripheral Measurements Using Multiple Wearable Sensors. *IEEE Sensors J.* 20, 723–734. doi:10.1109/JSEN.2019.2943879
- Reppa, V., Papadopoulos, P., Polycarpou, M. M., and Panayiotou, C. G. (2015). A Distributed Architecture for Hvac Sensor Fault Detection and Isolation. *IEEE Trans. Contr. Syst. Technol.* 23, 1323–1337. doi:10.1109/TCST.2014.2363629
- Rocha, M. A., de Souza, W. G., Serni, P. J. A., Andreoli, A. L., Clerice, G. A. M., and da Silva, P. S. (2018). “Control of Three-phase Pwm Boost Rectifiers Using Proportional-Resonant Controllers,” in 2018 Simposio Brasileiro de Sistemas Eletricos (SBSE), 1–6. doi:10.1109/SBSE.2018.8395803
- Roy, S. S., Dey, S., and Chatterjee, S. (2020). Autocorrelation Aided Random forest Classifier-Based Bearing Fault Detection Framework. *IEEE Sensors J.* 20, 10792–10800. doi:10.1109/JSEN.2020.2995109
- Shi, Z.-b., Yu, T., Zhao, Q., Li, Y., and Lan, Y.-b. (2008). Comparison of Algorithms for an Electronic Nose in Identifying Liquors. *J. Bionic Eng.* 5, 253–257. doi:10.1016/S1672-6529(08)60032-3
- Song, H., Liu, Z., Xie, Y., Wu, L., and Huang, M. (2016). Rgb Co-saliency Detection via Bagging-Based Clustering. *IEEE Signal. Process. Lett.* 23, 1722–1726. doi:10.1109/LSP.2016.2615293
- Wang, B., Li, Z., Dai, Z., Lawrence, N., and Yan, X. (2020). Data-driven Mode Identification and Unsupervised Fault Detection for Nonlinear Multimode Processes. *IEEE Trans. Ind. Inf.* 16, 3651–3661. doi:10.1109/TII.2019.2942650
- Yang, H., Zhang, Y., Zhang, N., Walker, P. D., and Gao, J. (2016). A Voltage Sensorless Finite Control Set-Model Predictive Control for Three-phase Voltage Source Pwm Rectifiers. *Chin. J. Electr. Eng.* 2, 52–59. doi:10.23919/CJEE.2016.7933126
- Zhang, M., Li, J., Li, Y., and Xu, R. (2021). Deep Learning for Short-Term Voltage Stability Assessment of Power Systems. *IEEE Access* 9, 29711–29718. doi:10.1109/ACCESS.2021.3057659

**Conflict of Interest:** Y-ND is employed by Qingdao Branch, China Merchants Bank.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kou, Gong, Zheng, Ni, Li, Yuan and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.