

Article

A Robust Accuracy Weighted Random Forests Algorithm for IGBTs Fault Diagnosis in PWM Converters without Additional Sensors

Gen Qiu , Fan Wu *, Kai Chen  and Li Wang

School of Automation, University of Electronic Science and Technology of China, Chengdu 611731, China; qgen615@uestc.edu.cn (G.Q.); kaichen@uestc.edu.cn (K.C.); colorsky@uestc.edu.cn (L.W.)

* Correspondence: wufan_xihua@163.com

Abstract: When an insulated-gate bipolar transistor (IGBT) open-circuit fault occurs, a three-phase pulse-width modulated (PWM) converter can usually keep working, which will lead to system instability and more serious secondary faults. The fault detection and diagnosis of the converter is extremely necessary to improve the reliability of the power supply system. In order to solve the problem of fault misdiagnosis caused by parameters disturbance, this paper proposes a robust accuracy weighted random forests online fault diagnosis model to accurately locate various IGBTs open-circuit faults. Firstly, the fault signal features are preprocessed by using the three-phase current signal and normalization method. Based on the test accuracy of the perturbed out-of-bag data and the multiple converters test data, a robust accuracy weighted random forests algorithm is proposed for extracting a mapping relationship between fault modes and current signal. In order to further improve the fault diagnosis performance, a parameter optimization model is built to optimize hyper-parameters of the proposed method. Finally, comparative simulation and online fault diagnosis experiments are carried out, and the results demonstrate the effectiveness and superiority of the method.



Citation: Qiu, G.; Wu, F.; Chen, K.; Wang, L. A Robust Accuracy Weighted Random Forests Algorithm for IGBTs Fault Diagnosis in PWM Converters without Additional Sensors. *Appl. Sci.* **2022**, *12*, 2121. <https://doi.org/10.3390/app12042121>

Academic Editor: Jordi Cusido

Received: 20 January 2022

Accepted: 15 February 2022

Published: 17 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: three-phase PWM converter; IGBT open-circuit fault; fault diagnose; robust accuracy weighted; random forests

1. Introduction

In recent years, due to the increase in nonlinear connected to the utility grid, the problems of low quality and harmonic distortion in power systems have increasingly attracted widespread attention [1]. A large amount of harmonic current will cause power grid voltage distortion and affect the normal operation of electrical equipment. It may also cause parallel or series resonance of the power grid [2]. Three-phase PWM converter, as the interface for energy conversion, has become an important device to improve this problem due to its excellent performance and potential advantages, such as sinusoidal input current and adjustable high input power factor. It is widely used in photovoltaic power generation, wind power generation, traction drive system and other occasions that require strict equipment reliability.

However, the reliability of converters used in high-reliability applications is not high, and semiconductor power devices are the weakest components in converters. According to the semiconductor device survey report of more than 200 products from 80 companies, the converter failure caused by semiconductor devices accounts for about 60% of all failure data [3]. In order to obtain high-performance power quality, the system needs to work at a very high switching frequency, generally above 10 kHz. Working at high frequency and high temperature for a long time further increases the damage probability of the IGBT. Since the fault of IGBT has always been the main reason for the fault of the three-phase PWM converter, the fault diagnosis and location of IGBT becomes particularly important [4].

The fault types of IGBT are divided into short-circuit fault and open-circuit fault. Extremely high inrush currents from short-circuit conditions can cause permanent damage to the system, and this type of fault is catastrophic and easy to measure. Therefore, the protection circuit is usually integrated in the inverter system [5]. On the other hand, the open-circuit fault is recessive, and when the fault occurs, it is not enough to trigger the hardware protection circuit. Excessive electrical stress and thermal stress will lead to secondary failures of other equipment. If failure detection methods are not taken, serious safety and property accidents will occur. Therefore, the open-circuit fault diagnosis of IGBT has attracted a lot of research attention [6].

Generally, fault diagnosis methods can be divided into model-based, signal-based and data-driven [7–9]. Model-based and signal-based inverter fault diagnosis for drive systems have been extensively studied in recent years. Reference [10] proposes a fast diagnosis method for a senseless inverter open-circuit fault by analyzing the switching function model of the inverter in healthy state and fault state. Subsequently, they proposed a fault diagnosis method based on the existing residual vector to remove the influence of the load [11]. Reference [12] proposed an open-circuit fault diagnosis method for voltage source inverters in PMSM drive systems using model reference adaptive system technology. References [13,14] proposed a simple single-switch and double-switch OC fault diagnosis method based on the three-phase current distortion of a vector-controlled induction motor driven by voltage source inverter. Reference [15] proposed a fault detection technology for an IGBT open-circuit fault in an induction motor driven by voltage source inverter. This technology detects the IGBT open-circuit fault by analyzing the PWM switching signal and the line-to-line voltage level during the switching time.

In recent years, benefiting from the rapid development of machine learning theory, the machine learning data-driven fault diagnosis method, which can directly realize fault diagnosis without logical or mathematical description of the inspection object, has received special attention due to its significant advantages [16]. The method is independent of the system model and signal image, and provides a potential fault diagnosis scheme to improve the accuracy of fault diagnosis. Reference [17] proposed a fault diagnosis and reconstruction method for multi-level inverters using neural networks. The output phase voltage of the inverter is used as a diagnostic signal to detect the fault and its location. Reference [18] proposed a fuzzy logic-based PWM voltage source inverter induction motor drive fault detection and diagnosis method. This technique requires measuring the output current of the inverter to detect the intermittent loss of ignition pulses in the inverter power switch. Reference [19] proposed a machine learning technique for fault diagnosis of induction motors by using structured neural networks. This method can detect and isolate common fault types such as single switch open-circuit fault, back short-circuit fault, short-circuit fault and unknown fault. Reference [20] proposed a fault diagnosis strategy for cascaded H-bridge multilevel inverter systems based on principal component analysis and multi-class correlation vector machines. Reference [21] proposes an OC fault diagnosis and online monitoring scheme for grid-connected single-phase inverters using an adaptive neuro-fuzzy inference system algorithm. Reference [22] proposes OC fault diagnosis and sliding-window classification based on hybrid ensemble learning. Higher fault diagnosis accuracy is achieved and the fault diagnosis speed is improved. Reference [23] proposes a data-driven fault diagnosis method for three-phase PWM inverter in induction motor drives to realize simultaneous diagnosis of IGBT and current sensor fault of three-phase PWM inverter in induction motor drives. Reference [24] proposes a IGBT open-circuit fault diagnosis method based on transferable data driven to improve the generalization performance of fault diagnosis model.

Random forests (RFs) algorithms are suitable for regression and classification, due to their excellent performance in pattern recognition, state monitoring, system control and other fields [25–29]. Reference [27] proposed a RFs-based fault identification model for power transformer by radio frequency identification and the feature extraction of differential measurement current data, which can effectively distinguish internal faults

and disturbances. Integrating XGBoost in RFs, a data-driven based wind turbine fault detection method is proposed in [28], which prevents over-fitting when dealing with multidimensional data. Reference [30] proposed a RFs regression based implementation of SVPWM for a two-level inverter, which can improve the performance of the three-phase induction motor (TIM) drive. Reference [31] proposed a DAWRF algorithm for IoT fault detection based on edge computing and blockchain, which improved the traditional RFs algorithm effective application in IoT fault detection.

In practical applications, most data-driven methods pay great attention to the robustness and generalization ability of fault diagnosis. Reliable diagnosis needs to adapt to as many operating conditions as possible, which is not fully addressed in most literatures. In addition, data-driven algorithms often suffer from algorithmic accuracy bottlenecks and huge training burdens. Aiming at the above problems, this paper proposes a data-driven IGBT open-circuit fault diagnosis method. Its salient features and advantages are as follows:

- (1) In order to extract the mapping relationship between fault features and failure modes, a robust accuracy weighted random forests algorithm is proposed, which uses out-of-bag datasets and object datasets with random disturbances to evaluate the performance of the fault diagnosis model. In accordance with the evaluation results, the weights of the decision trees are changed accordingly to improve the anti-noise ability of RFs.
- (2) Several hyper-parameters of the trained diagnostic model are optimized by an optimization programming framework (OPF). The OPF's aims are to improve the accuracy of the proposed fault diagnosis model. By this optimization, a number of hyper-parameter sets can be provided to satisfy the test object limitations and difficulties of dataset collection.
- (3) This method is a non-invasive fault diagnosis method, which only needs three-phase current signal as input without additional sensors. Compared with most fault diagnosis methods, the proposed method has higher fault diagnosis accuracy and less computational burden.

The method has been validated in simulation and in-circuit testing. Compared with the existing fault identification methods, this method has the advantages of high fault diagnosis accuracy and better robustness.

The rest of this paper is organized as follows: Section 2 introduces the inverter system and fault types in this study. Section 3 briefly introduces the general framework of the method. The theoretical basis and detailed description of the improved robust accuracy weighted random forests algorithm are presented in Section 4. Section 5 details the simulation and experimental verification and discussion. Finally, Section 6 presents the overall conclusion of this paper.

2. System Description and Fault Identification

The object of this paper is a three-phase PWM converter widely used in electric drive systems. Its circuit topology is a typical three-phase full-bridge structure as shown in Figure 1, consisting of six IGBTs (T_1 – T_6) and corresponding anti-parallel connected diodes (D_1 – D_6). I_a , I_b , I_c are the output currents of the converter, which are controlled by the control system with 100 kHz sampling rate, 20 kHz control sampling rate. The IGBT is controlled by the corresponding drive signals.

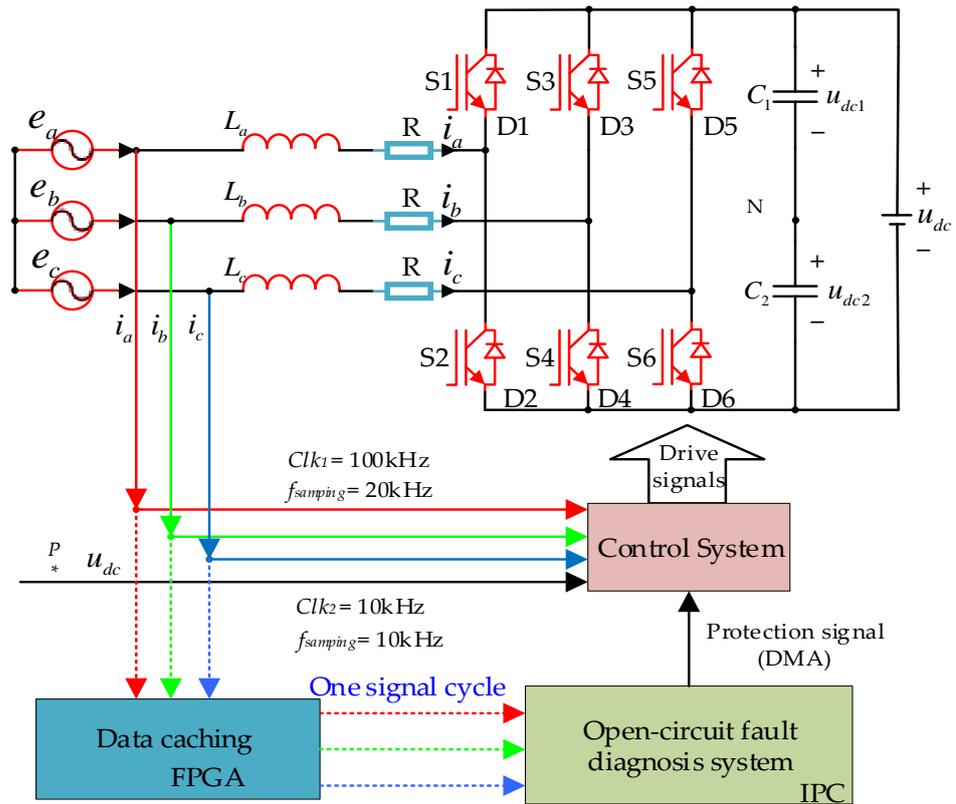


Figure 1. Fault diagnosis system of three-phase PWM converter.

In this paper, the proposed open-fault diagnosis system is as shown in Figure 1, which is composed of data caching field programmable gate array (FPGA) and industrial personal computer IPC. FPGA resamples the high speed three-phase current signals at a sampling rate of 10 kHz to caches one cycle of current signals. IPC executes the fault diagnosis algorithm according to the current signal of one cycle to obtain the fault diagnosis result. Once a fault signal is detected, IPC gives instructions via direct memory access (DMA) to shut down the control system. By this way, the fast IGBT open fault protection is realized. The main advantage of this system is that it reduces the computational burden by down sampling, so that one fault diagnosis system can protect multiple power sources simultaneously.

However, in the actual operation of the fault diagnosis system, the three-phase current signal is affected by many disturbance factors, which may lead to misdiagnosis and power shutdown. Therefore, the reliability and anti-interference ability of fault diagnosis are very important. Usually, the IGBT open-circuit fault mainly includes the following two situations:

- (1) When the chip S_n fails, the body diode D_n also fails at the same time.
- (2) When the chip S_n fails, the body diode D_n works normally.

Case 1 usually causes over-voltage to trigger the hardware protection, so this type is not the research scope of open-circuit fault diagnosis in this study.

Different IGBT open-circuit fault scenarios will generate different output current signals. In total, there are six single IGBT fault types and 15 dual IGBT fault types. Considering normal operating conditions, there are a total of 22 fault tags. All labels are listed in Table 1. Based on these fault labels, this paper proposed a data-driven fault diagnosis strategy. The input of the method is a periodic three-phase output current signal, and the output is a fault label representing the type of fault and the location of the fault.

Table 1. IGBT open-circuit fault label.

Fault IGBT/IGBTs	Label	Fault IGBT/IGBTs	Label
/	1	S1 and S6	12
S1	2	S2 and S3	13
S2	3	S2 and S4	14
S3	4	S2 and S5	15
S4	5	S2 and S6	16
S5	6	S3 and S4	17
S6	7	S3 and S5	18
S1 and S2	8	S3 and S6	19
S1 and S3	9	S4 and S5	20
S1 and S4	10	S4 and S6	21
S1 and S5	11	S5 and S6	22

3. Generic Method Framework

This paper proposed a data-driven IGBT open-circuit fault diagnosis method. The three-phase current sampling of the inverter is used as the diagnostic input, and finally the fault label is the diagnostic output to realize the fault diagnosis and location. The framework of the whole method is shown in Figure 2, which is divided into two stages: offline model training and online fault diagnosis.

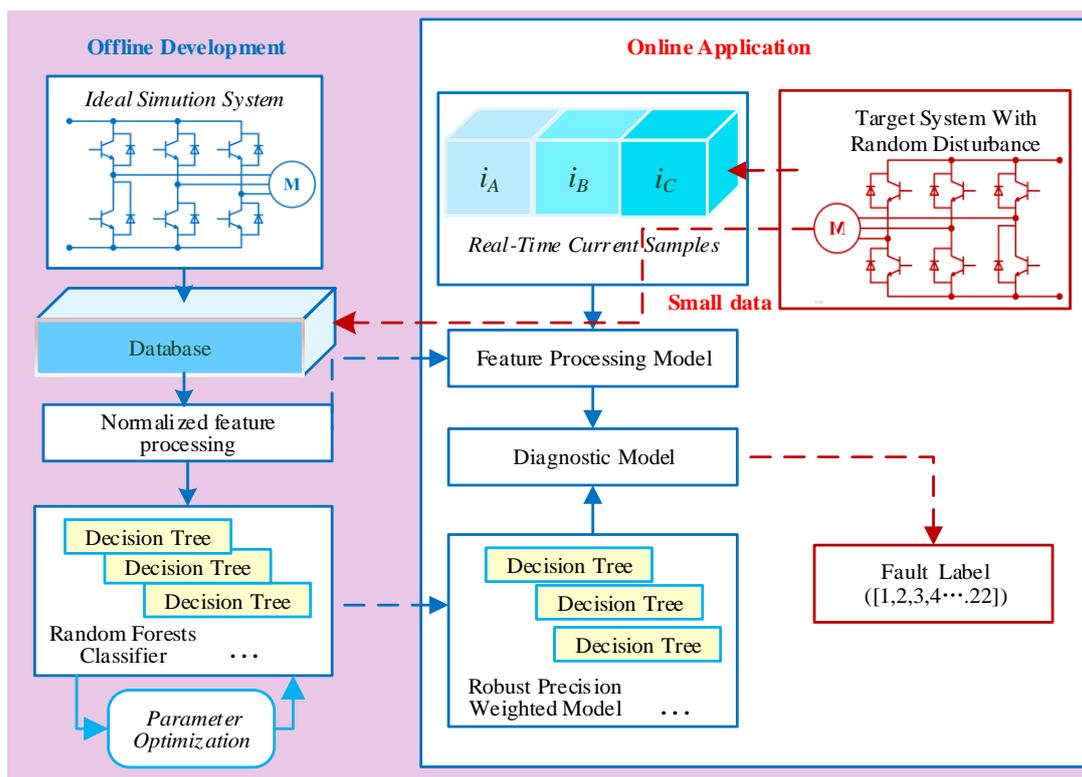


Figure 2. Framework of fault diagnosis method.

In the offline development process, the random forests fault diagnosis model is trained using a historical fault database composed of ideal model samples. In order to improve the performance of fault diagnosis, standardization is used to perform feature processing on the sampled three-phase current data. Random forest is an ensemble learning algorithm composed of multiple decision trees. It has strong anti-disturbance ability and robust characteristics, and is one of the most widely used fault classifiers. In this paper, an improved random forests algorithm is used to train the fault diagnosis model to obtain the mapping relationship between input features and fault labels. Different from the

traditional random forests method, each decision tree in the improved random forests will vote weighted according to its fault diagnosis accuracy. Overall, they provide more accurate and robust fault diagnosis results.

In online application, the measured object is a non-ideal system affected by various factors. The fault diagnosis method proposed in this paper, the trained fault diagnosis model can be directly applied to the non-ideal power system with disturbance factors, and the measured three-phase current information input into the fault diagnosis model to acquire the fault label, no additional model training is required, and excessive dependence on the fault data of the tested object is avoided.

4. Random Forests Theory

The essence of fault diagnosis and localization of power source is a classifier, which classifies the operating state of the power source inputting characteristic signals. The random forests algorithm is a supervised ensemble learning algorithm that integrates many decision trees for prediction and classification. The core idea is that weak classifiers formed by multiple decision trees are integrated into a strong classifier, and multiple weak classifiers give classification results through certain rules. The decision tree algorithm uses the labeled samples to construct a fault diagnosis classifier, and obtains the mapping relationship between input features and fault labels. It enables the classifier to classify and discriminate unlabeled samples. The random forests classifier trains each decision tree classifier with random samples and uses different training sets to increase the difference between the decision tree models, thereby improving the generalization ability and robustness of the classifier. This characteristic is beneficial to solve various disturbance misdiagnosis of IGBT open-circuit fault diagnosis.

4.1. Decision Tree

Decision tree is a binary recursive partitioning technique that splits the current sample set into two subsets at each node (except leaf nodes). The attribute selection measure adopted by the algorithm is the Gini index. Assuming that the dataset D contains m categories, the formula for calculating the Gini index G_D is:

$$G_D = 1 - \sum_{j=1}^m p_j^2 \quad (1)$$

In the formula: P_j is the frequency of the j -th type of faults in the training database. The Gini index needs to consider the binary division of each feature. Assuming that the binary division of feature A divides dataset D into D_1 and D_2 , the Gini index of sample set D divided by feature A at the child node this time is:

$$G_{D,A} = \frac{|D_1|}{D} G_{D_1}(D_1) + \frac{|D_2|}{D} G_{D_2}(D_2) \quad (2)$$

For each attribute, every possible binary partition is considered, and finally the subset that yields the smallest Gini index for that attribute is selected as its split subset. Therefore, the smaller the Gini index G_D on attribute A , the better the division effect on attribute A is. Under this rule, the division continues from top to bottom until the entire decision tree grows:

4.2. REF Algorithm

Definition 1. The random forests model f is a set of decision trees $\{h(X, \theta_k), k = 1, 2, \dots, N_{tree}\}$, and the classifier $h(X, \theta_k)$ is an unpruned decision tree constructed with a decision tree algorithm.

θ_k is a random vector independent and identically distributed with the k th decision tree, representing the growth process of the tree; the final classification value of the random forests is obtained by the majority voting method.

Definition 2. For the input vector X , it contains at most J different categories, let Y be the correct classification category, and for the input vector X and output Y , define the edge function as:

$$K(X, Y) = \begin{aligned} & a_k I(h(X, \theta_k) = Y) \\ & - \max_{j \neq Y} a_k I(h(X, \theta_k) = j) \end{aligned} \quad (3)$$

In the formula: j is one of the J categories; $I(\cdot)$ is the indicator function; a_k is the average function; $k = 1, 2, \dots, n$. The larger the edge function, the higher the confidence of the correct classification. The generalization error of random forests is thus defined as:

$$E^* = P_{X,Y}(K(X, Y) < 0) \quad (4)$$

In the formula: $P_{X,Y}$ is the classification error probability function for a given input vector X . When the number of decision trees in the forest is large, the following theorem is obtained by using the law of large numbers:

Theorem 1. When the number of decision trees is large enough, for all sequences θ_k , E^* converges almost everywhere:

$$\lim_{k \rightarrow \infty} E^* = P_{X,Y}(P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X, \theta) = j) < 0) \quad (5)$$

In the formula: $P_\theta(c)$ is the probability of satisfying the condition c for a given sequence θ . This theorem shows that the generalization error of random forests will not cause overfitting as the number of trees increases, but will tend to an upper bound.

Theorem 2. The upper bound of the random forests generalization error is:

$$E^* \leq \frac{\rho(1 - s^2)}{s^2} \quad (6)$$

In the formula: ρ and s are the average correlation coefficient and average strength of the tree. It can be seen from Theorem 2 that with the decrease in the correlation of decision trees and the increase in the strength of a single decision tree, the upper bound of the generalization error of random forests will decrease, and its generalization error will be effectively controlled.

Guaranteed by the law of large numbers, random forests has a high classification accuracy without overfitting, which is extremely suitable for power source fault diagnosis. From the above analysis, there are two main ways to improve the diagnostic accuracy performance of random forests fault diagnosis model, namely, reducing the correlation ρ of decision trees and improving the fault diagnosis performance s of a single decision tree. In addition, an important feature of random forests is out-of-bag estimation. When a training subset is generated by bagging, for each decision tree, nearly 37% of the samples in the original sample set S will not appear in the other trees. In the training subset, these samples are called out of bag (OOB) samples. OOB samples can be used to estimate the generalization error of RFs and also to calculate the importance of each feature. The above theorems are the key factors to reduce the fault misdiagnosis in this paper.

4.3. Improved Random Forest Algorithm

The RFs algorithm uses bootstrap sampling, which is a simple random sampling method with permutation. When extracting each training subset, about a third of the samples are not selected, and these data are called out-of-bag data. These data are of high research value and can be used as an alternative to cross-validation methods for datasets. Based on the traditional RFs algorithm, this study designs a weighted random forests

algorithm based on precision prediction with perturbed OOB datasets and a few object datasets. In the training phase of the algorithm, the test dataset and the out-of-bag dataset with additional disturbance factors are used to predict the fault diagnosis accuracy of the decision tree on the tested object. The out-of-bag weight and test weight are calculated according to the above precision. In the decision-making stage, the voting weights of the decision tree are adjusted according to the out-of-bag weight and test weight. The detailed implementation process is described below.

First, the test data, P , and the training dataset, T , are collected at the test sample rate X_{TP} (the ratio of test data to training data). The training dataset, T , is from an ideal system simulation, and the relatively small test dataset, P , is from the test object with uncertain perturbation factors. In the training phase, bootstrap sampling is used to separate the training data subset S and the out-of-bag data O from the training dataset, T , at the out-of-bag data rate X_{Os} (the ratio of the OOB data to the training subset data), and add a certain random Gaussian white noise to the OOB data O . The fault diagnosis accuracy of the out-of-bag dataset and test dataset with disturbance factors reflects the fault classification ability of the decision tree. The decision trees with higher classification accuracy and better classification effect will have heavier weight.

After training the k th decision tree with the training subset S_k , use the out-of-bag dataset O_k with variance M gaussian white noise to predict the fault diagnosis accuracy of the k th decision tree. For the k th decision tree, the weights for OOB data are as follows:

$$w_{O_k} = \frac{X_{O_k}^{corr}}{X_O}, k = 1, 2, \dots, K \quad (7)$$

In this formula: X_O is the total number of samples, and $X_{O_k}^{corr}$ is the number of samples correctly classified by the k -th decision tree. Use the test dataset P to predict the fault diagnosis accuracy of the k th decision tree. The ratio of the number of correct test samples to the number of predicted test samples is the weight of the predicted test.

$$w_{P_k} = \frac{X_P^{corr}}{X_P}, k = 1, 2, \dots, K \quad (8)$$

In the decision-making stage, the OOB data weights are combined with the predicted test weights to determine the final weights. When voting for fault diagnosis, the vote result of each decision tree is multiplied by its corresponding weight W_k .

The development of the proposed Algorithm 1 is as follows:

Algorithm 1: Improved Random Forest Algorithm

Begin:

- (1) Determine the out-of-bag data rate, X_{Os} , the test data rate, X_{TP} , the number of decision trees, K , and the Gaussian white noise variance, M ;
- (2) According to X_{TP} , determine the training data, T , and predict the test data, P ;
- for** $k = 1$ **to** K **do**
- (3) Using bootstrap sampling and according to X_{OS} , the training set, T , is divided into training subset, S_k , and out-of-bag data, O_k ;
- (4) According to the C4.5 algorithm, N features are randomly selected as node classification features, and S_k is used to generate a decision tree;
- (5) Add random Gaussian white noise of M decibels to O_k ;
- (6) Take O_k, P as the test set;
- (7) According to Equations (7) and (8), calculate the weights of the k th decision tree as w_{O_k} and w_{P_k} ;
- (8) Calculate the final weight, w_k , of the k th decision tree by Equation (9);
- end for**
- (9) The test data are classified by the decision tree set, and the final classification result is determined by Equation (11);

END

Finally, the type with the most votes is the fault diagnosis result of random forests. The final weights of the RFs model, the fault diagnosis result of each decision tree, and the fault diagnosis result of random forests are as follows:

$$w_k = \frac{w_{P_k} + w_{O_k}}{2}, k = 1, 2, \dots, K \tag{9}$$

$$h_k^{DT}(x) = y_k \tag{10}$$

$$H'(x) = \operatorname{argmax}_{y \in Y} \sum_{k=1}^K w_k I(h_k^{DT}(x) = y_k) \tag{11}$$

In order to obtain a better fault diagnosis effect, an optimization model is constructed to optimize the hyperparameters X_{TP} , X_{OS} , K and M . The optimization objective function is defined as:

$$\max F_{obj} = X_{ORF} + X_{PRF} \tag{12}$$

$$X_{ORF} = \frac{\sum_{k=1}^K X_{O_k}^{corr}}{K} \tag{13}$$

X_{ORF} is the average number of samples correctly classified by the random forests model in the out-of-bag data test set. X_{PRF} predicts the correct number of samples for the random forests in the test dataset P .

Boundary Condition:

$$s.t \left\{ \begin{array}{l} X_{TP}T \leq P \\ M \leq a_h \max(THD_I), 1 \leq a_h \leq 2 \\ X_{OS} \leq 0.5 \\ K \in N \end{array} \right. \tag{14}$$

In the formula: $\max(THD_I)$ is the maximum total harmonic ratio of the measured object current.

In recent years, particle swarm optimization algorithm has been widely used in various optimization problems because of its simplicity, easy implementation and fast convergence speed. The optimization objective is solved by the particle swarm optimization algorithm. For details, please refer to [29].

5. Simulation and Experimental

In order to verify the effectiveness of the proposed data-driven method in practical applications, simulation tests and experiments are carried out. The performance of the ANN, SVM, and RFs, weighted RFs, PSO weighted RFs fault diagnosis model was evaluated in the simulation phase. The commonly used evaluation indicators of data-driven method fault diagnosis, accuracy, precision, recall, and model training time are used to measure the performance of the proposed method. The definition of the fault diagnosis accuracy, precision, and recall are as follows:

$$Accu = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$Prec = \frac{TP}{TP + FP} \tag{16}$$

$$Reca = \frac{TP}{TP + FN} \tag{17}$$

TP , TN , FP , and FN denote true positive, true negative, false positive and false negative, respectively.

5.1. Database Generation

In order to generate fault diagnosis models, a comprehensive and informative database is required. Such databases can be obtained from historical measurements or simulations. Considering that the measured object has a variety of random disturbance factors, which are caused by sensors disturbance, converter parameters and so on, therefore, the fault data are not easy to precisely obtain from object system. In order not to lose generality, the data are simulated in this paper, and DC voltage, output voltage and output current assignment are considered. The fault characteristics will be different under different operating conditions. Therefore, in order to contain more fault information in the fault database, we simulate various converter operation conditions. The data acquisition process is shown in Table 2, and the three-phase PWM converter simulation parameters are shown in Table 3. The simulation obtains a total of 88,000 sets of data as the model training database, DT . Due to the risk of damage from real system open failure, only a few test databases, DP (2200 sets), are obtained through two real three-phase converter systems with different model parameters, as shown in Table 4. Of which, the parameters of the real system 1 are consistent with those of simulation as shown in Table 3.

Table 2. Database acquisition process.

Data Acquisition Condition	
DC Voltage	400:700/20 V
Output voltage (RMS)	100:300/10 V
Output voltage frequency	30:130/10 Hz
Output current (RMS)	10:100/10 A
Open-circuit fault type	22

Table 3. Simulation parameters and real system 1 parameters of three-phase PWM converter.

Contents	Values
DC-link voltage U_{dc} .	700 V
Rated output voltage	220 V (RMS)
Rated output current	40 A (RMS)
Rated voltage frequency	50 HZ
Switching frequency	12.8 kHz
Filter inductance	1.2 mH
Capacitance	480 μ F

Table 4. Parameters of real system 2.

Contents	Values
DC-link voltage U_{dc}	400 V
Rated output voltage	110 V (RMS)
Rated output current	100 A (RMS)
Rated voltage frequency	400 HZ
Switching frequency	5 kHz
Filter inductance	0.82 mH
Capacitance	480 μ F

The training dataset, T , is extracted from DT according to coefficient X_{OS} . Extract the test dataset P from the DP according to coefficient X_{OS} . The out-of-bag data in the DP is used to calculate the evaluation index Equations (15)–(17).

5.2. Simulation and Comparison Results

In order to verify and analyze the performance of the fault diagnosis model, the dataset T is used to train the proposed fault diagnosis model. In addition, Bayesian network (BN) fault diagnosis model, support vector machine (SVM) fault diagnosis model, RFs fault

diagnosis model, ensemble extreme learning machine (ELM) fault diagnosis model and the proposed fault diagnosis model, are compared through the same training dataset, T , and validation dataset, Vd , which is the part of the database DP that excludes dataset P . All algorithms adopt the same general fault diagnosis framework as shown in Figures 1 and 2. In order to search for the optimal fault diagnosis performance, the hyperparameters of all algorithms are optimized including architecture, learning rate, kernel function, the number of decision tree, etc. The optimization results and some decision tree weights of the proposed method are shown in Table 5.

Table 5. The optimization model results and some decision tree weights of the proposed method.

Hyperparameters	Results	Decision Tree Label	Weight	Decision Tree Label	Weight
X_{OS}	0.2693	100	0.6237	500	0.2678
X_{TP}	0.0151	200	0.9764	600	0.7884
K	1246	300	0.1366	700	0.8256
M	0.0217	400	0.8247	800	0.9314

In order to verify the training performance of the fault diagnosis algorithms, we conducted 10-fold cross-validation on the above model in 88,000 training datasets. The results are as shown in Table 6. After the hyperparameter optimization, all the fault diagnosis algorithms have a good fault diagnosis performance index (above 99.5%) in the training datasets. However, integrated fault diagnosis algorithms including RFS, ensemble ELM, and the proposed method have an excellent fault diagnosis performance index of nearly 100% in the OOB of training datasets.

Table 6. The 10-fold cross-validation results.

Diagnosis Method	Accuracy	Precision	Recall
BN	95.79%	96.10%	95.45%
SVM	96.33%	96.09%	96.59%
RFS	99.64%	99.74%	99.55%
Ensemble ELM	99.76%	99.74%	99.75%
The proposed method	99.55%	99.33%	99.77%

In order to verify the generalization performance of the fault diagnosis model, we performed performance verification on real system 1 and real system 2. Finally, 880 sets of validation dataset, Vd , were selected from the DP to calculate the performance index. The comparison results are as shown in Table 7. The simulation results show that the proposed algorithm has the highest accuracy (96.25%) and recall (97.73%). This shows that the proposed method can check all fault types well and avoid fault miss detection. However, the ensemble ELM has the highest (96.15%), which reflects that the algorithm has the best performance in avoiding false positives. The offline test time of the proposed method is about 0.2837 s reflecting the proposed method has less computation time while achieving the same diagnosis performance.

Table 7. Validation and comparison with other methods.

Diagnosis Method	Accuracy	Precision	Recall	Offline Test Time
BN	81.36%	85.45%	78.99%	1.1527 s
SVM	84.93%	91.4%	77.27%	3.0729 s
RFS	88.98%	90.15%	89.58%	0.1374 s
Ensemble ELM	94.55%	96.15%	93.75%	1.9273 s
The proposed method	96.25%	94.92%	97.73%	0.2837 s

Due to the training data being mainly from ideal simulation system, the performance of BN, SVM, and RFs fault diagnosis is poor when applied to the real system with disturbance factors. However, the proposed methods and the proposed algorithm improve the adaptability of the diagnostic model by modifying the weights of the model through disturbed data. Therefore, the above results show that the proposed method has good generalization performance

Besides, in order to simulated random disturbance, we add Gaussian white noise with different variances, M , to the validation dataset, Vd , to verify the robustness of the proposed method as shown in Figure 3. With the increase in M , the fault diagnosis accuracy of other methods decreases. Especially when $M = 6\%$, the fault diagnosis accuracy of these fault diagnosis methods is lower than 75.12% due to signal disturbance. As a comparison, the fault diagnosis accuracy of the proposed method is above 92.16% when $M < 6\%$. It shows that the proposed method has excellent robustness.

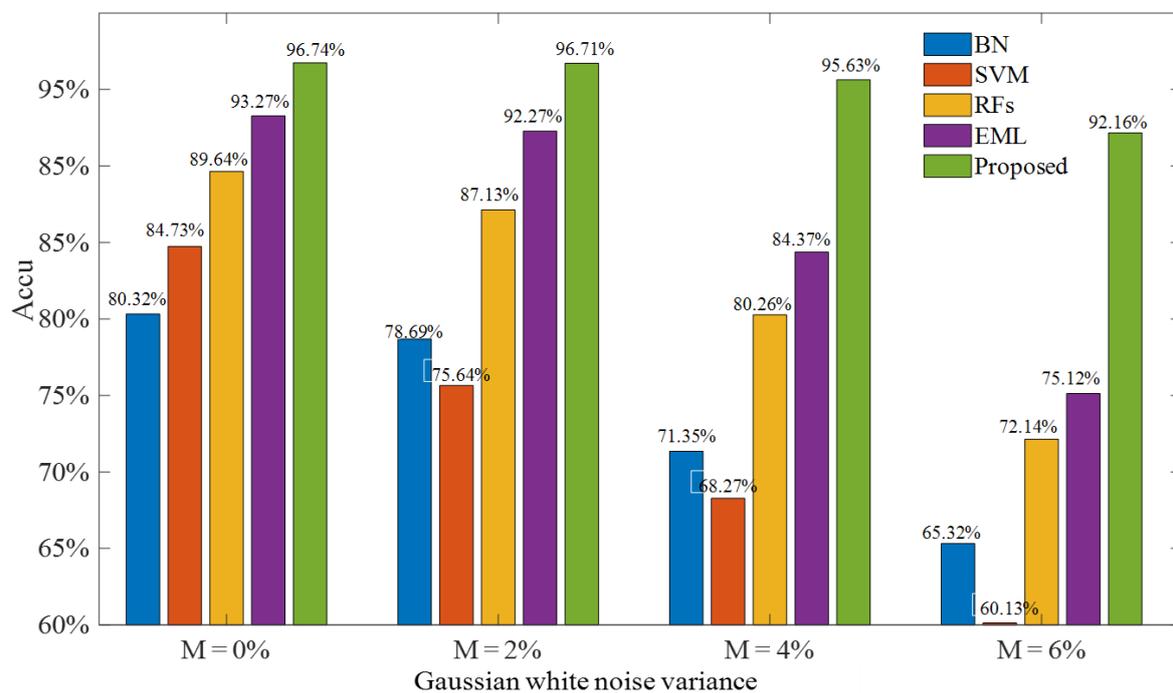


Figure 3. Robustness test result.

5.3. Experimental Verification

To verify the feasibility of the method, an online fault diagnosis experiment is carried out on the three-phase PWM inverter. The three-phase PWM inverter online fault diagnosis system is shown in Figure 4. The fault diagnosis system consists of an IPC and data caching system realized by FPGA, in which the underlying controller is used for open-circuit fault monitoring, the IPC can be replaced by an ARM chip, and the proposed RFs can run on the ARM chip. The sampling clock of the closed-loop control system is 100 kHz, and the sampling frequency is 20 kHz. At the same time, the current signal is resampled with a clock of 10 kHz and a resampling frequency of 10 kHz. Therefore, it only sends 200 points to the IPC per cycle (20 ms) to reduce the pressure of computational burden. Once an open-circuit fault is detected, the online fault diagnosis system will send a protection signal to the controller to turn off all IGBT control signals.

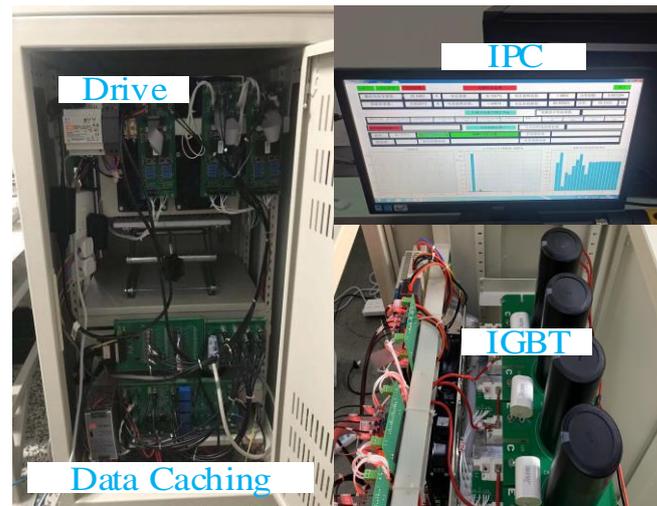
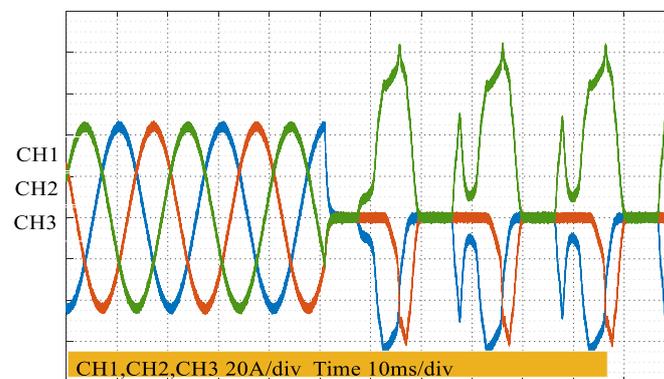
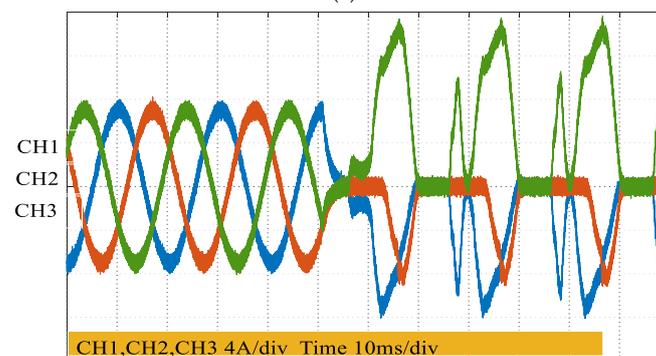


Figure 4. Experimental platform.

The output current fault characteristics are different when the converter runs under different working conditions, mainly including DC Voltage, output voltage, output voltage frequency, output current. For example, the current fault characteristics are different, as shown in Figure 5 (S1 open-circuit fault) and Figure 6 (S1 and S2 open-circuit fault), when the output current is 44 A (Figure 6a) and 8.8 A (Figure 6b). Hence, it is difficult to achieve high precision fault diagnosis by setting thresholds or traditional data-driven fault diagnosis methods. It can be seen from the experimental waveform that the occurrence of faults is often accompanied by overstress. If there is no high-performance fault diagnosis algorithm to quickly detect the fault and shut down the system when the IGBT open-circuit fault occurs, the overstress will lead to a more destructive secondary failure.

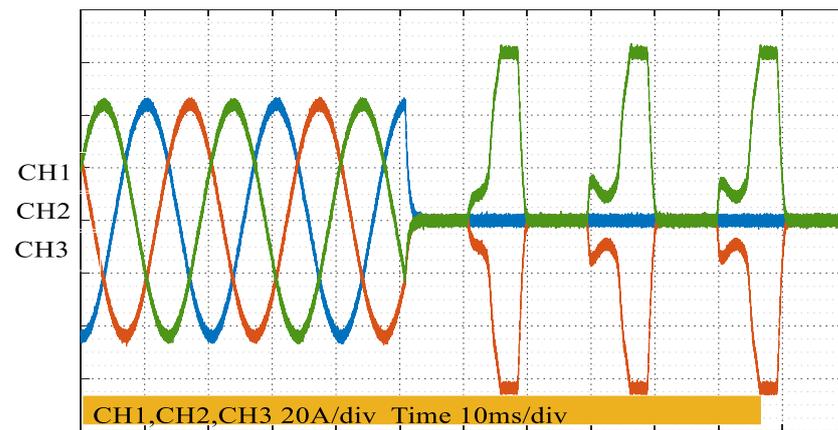


(a)

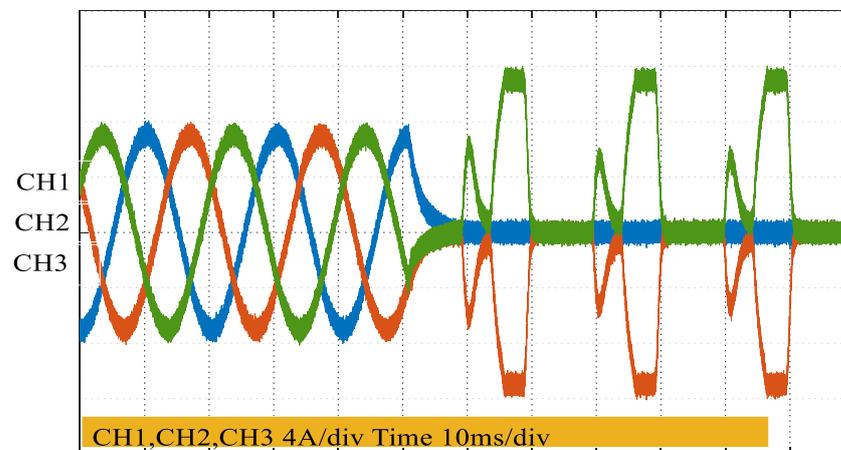


(b)

Figure 5. S1 open-circuit faults: (a) output current is 44 A, (b) output current is 8.8 A.



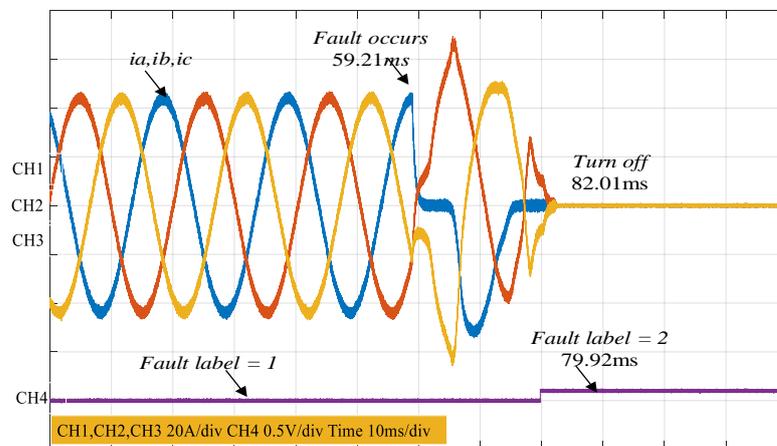
(a)



(b)

Figure 6. S1 and S2 open-circuit faults: (a) output current is 44 A, (b) output current is 8.8 A.

This paper illustrates the effectiveness of the method by taking S1 open-circuit faults, both S1 and S3 open-circuit faults as examples. Once an open-circuit fault is detected, the controller will immediately turn off all IGBTs to protect the system. This method can locate the fault while ensuring the safety of the system, and the diagnosis result is shown in Figure 7.



(a)

Figure 7. Cont.

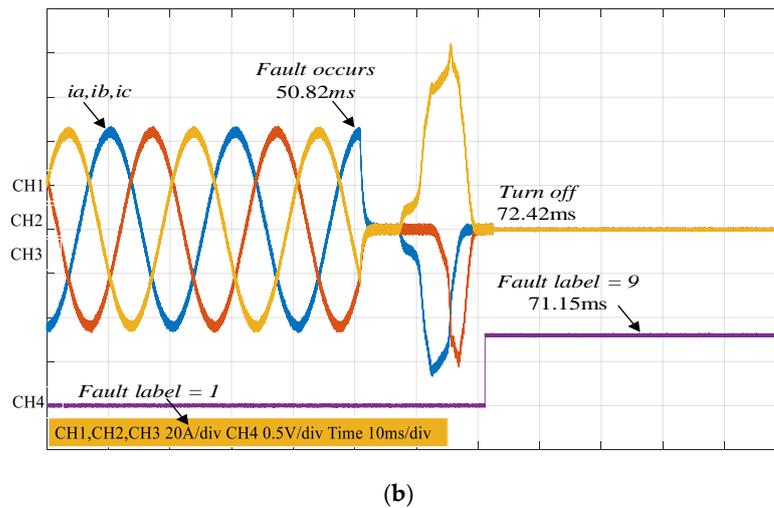


Figure 7. Fault diagnosis results: (a) S1 open-circuit faults, (b) S1 and S3 open-circuit faults.

Figure 7a illustrates the experimental results of S1 open-circuit fault. S1 open-circuit fault occurs at 59.21 ms; the fault diagnosis system detected and located the fault at 79.92 ms, of which the fault label is as shown in Table 1. After detecting the IGBT open-circuit fault, the converter shut down at 82.01 ms to protect the device from secondary failures.

Figure 7b illustrates the experimental results of both the S1 and S2 open-circuit fault. Both the S1 and S2 open-circuit fault occur at 50.82 ms. The fault diagnosis system detected and located the fault at 71.15 ms, of which the fault label is as shown in Table 1. After detecting the IGBT open-circuit fault, the converter shut down at 73.09 ms to protect the device from secondary failures.

The above experiment data show that using the proposed method, IGBT open-circuit fault can be identified in around one current cycle (20 ms). The online calculation time is minor, around 0.46 ms. Once IGBT open-circuit failure occurs, the proposed method can quickly turn off the power supply within about 23 ms, so as to avoid more serious secondary accidents.

In addition, reliable open-circuit fault diagnosis should ensure that there is no error trigger when the converter is adjusted in the transient process. Therefore, the transient sensitivity of the proposed algorithm is analyzed in this paper. The analysis result is shown in Figure 8. The output current of the converter is regulated from 44 A to 36 A at 50.92 ms, while the fault label maintain 1 (Fault label = 1 indicates that the converter is operating normally as shown in Table 1) all the time. The above results show that the fault diagnosis of the proposed algorithm is independent of the converter regulation transient.

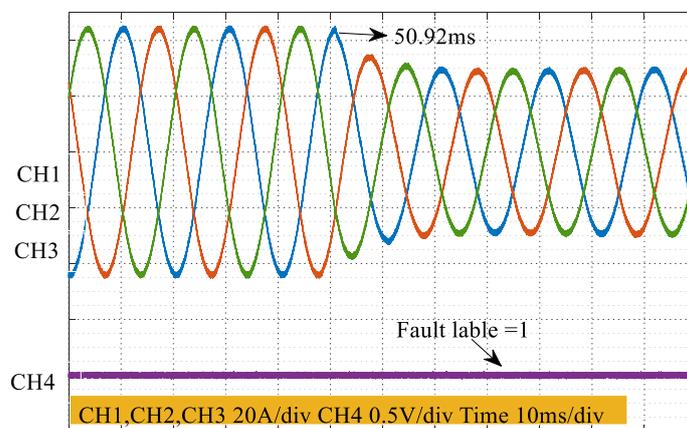


Figure 8. Transient sensitivity analysis of fault diagnosis, when the output current of the converter is regulated from 44 A to 36 A.

The above results show that the model trained by the proposed method under ideal data can be applied to real systems with disturbance factors very well. Therefore, this method has good robustness and generalization ability.

6. Conclusions

In this paper, a robust accuracy weighted random forests fault diagnosis method for three-phase PWM converters is proposed. The proposed method takes the three-phase output current as the input signal and uses the normalization method to preprocess the data without additional sensors. Based on the test accuracy of the perturbed out-of-bag data and the multi-source model test data on the model, an accuracy weighted random forests algorithm is proposed for extracting mapping relationship between fault modes and current signal. In order to further improve the fault diagnosis performance, the hyper-parameters of the parameter optimization model are constructed.

Compared with the BN, SVM, RFs, ensemble ELM, the RFs algorithm has better performance in terms of training time, computational burden, diagnostic accuracy and robustness. Finally, comparison simulation and online fault diagnosis experiment are carried out. The comparison simulation and experimental results show that the method can accurately and rapidly locate the open-circuit fault of the IGBT under the premise of ensuring the safety of the system. In addition, this method is not limited to typical three-phase PWM converters, other converters are still applicable through the establishment of database and model retraining in the same way.

The limitations of this method are that in order to perfectly adapt to different application scenarios, it is necessary to study the disturbance distribution type of the tested object. On the other hand, the training efficiency of the model needs to be improved.

Author Contributions: Conceptualization, G.Q.; methodology, F.W. and G.Q.; software, K.C.; formal analysis, L.W.; resources, K.C.; writing—original draft preparation, G.Q.; writing—review and editing, F.W.; validation L.W.; supervision, K.C.; project administration, F.W.; funding acquisition, K.C. and G.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the Fundamental Research Funds for the Central Universities under Grants No. ZYGX2019J063, No. ZYGX2020ZB004 and No. ZYGX2020ZB001, the Sichuan Science and Technology Project under Grants No. 2019ZDZX0045 and the second batch of industry-university cooperation collaborative education projects of the Ministry of Education in 2021(202102371056).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this research are available on request from the corresponding author. The data are not publicly available due to the follow-up project has not been finalized.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Subjek, J.S.; Mcquilkin, J.S. Harmonics-causes, effects, measurements and analysis: An Update. *IEEE Trans. Ind. Appl.* **1990**, *26*, 1034–1042. [[CrossRef](#)]
2. Duarte, L.H.S.; Alves, M.F. The Degradation of Power Capacitors under the Influence of Harmonics. In Proceedings of the 10th International Conference on Harmonics and Quality Power, Rio de Janeiro, Brazil, 6–9 October 2002; Volume 1, pp. 334–339.
3. Yang, S.; Xiang, D.; Bryant, A.; Mawby, P.; Ran, L.; Tavner, P. Condition Monitoring for Device Reliability in Power Electronic Converters: A Review. *IEEE Trans. Power Electron.* **2010**, *25*, 2734–2752. [[CrossRef](#)]
4. Ge, J.; Zhao, Z.; Yuan, L.; Lu, T.; He, F. Direct power control based on natural switching surface for three-phase PWM rectifiers. *IEEE Trans. Power Electron.* **2015**, *30*, 2918–2922. [[CrossRef](#)]
5. Krishnamoorthy, H.S.; Rana, D.; Garg, P.; Enjeti, P.N.; Pitel, I.J. Wind turbine generator-battery energy storage utility interface converter topology with Medium frequency transformer link. *IEEE Trans. Power Electron.* **2014**, *29*, 4146–4155. [[CrossRef](#)]
6. Jung, S.M.; Park, J.S.; Kim, H.W.; Cho, K.Y.; Youn, M.J. An MRAS-based diagnosis of open-circuit fault in PWM voltage-source inverters for PM synchronous motor drive systems. *IEEE Trans. Power Electron.* **2013**, *28*, 2514–2526. [[CrossRef](#)]

7. Estima, J.O.; Cardoso, A.J.M. A new algorithm for real-time multiple open-circuit fault diagnosis in voltage-fed PWM motor drives by the reference current errors. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3496–3505. [[CrossRef](#)]
8. Freire, N.M.A.; Estima, J.O.; Cardoso, A.J.M. Open-circuit fault diagnosis in PMSG drives for wind turbine applications. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3957–3967. [[CrossRef](#)]
9. Freire, N.M.A.; Estima, J.O.; Cardoso, A.J.M. A voltage-based approach without extra hardware for open-circuit fault diagnosis in closed-loop PWM AC regenerative drives. *IEEE Trans. Ind. Electron.* **2014**, *61*, 4960–4970. [[CrossRef](#)]
10. Reddy, K.M.; Singh, B. Multi-objective control algorithm for small hydro and SPV generation-based dual mode reconfigurable system. *IEEE Trans. Smart Grid* **2018**, *9*, 4942–4952. [[CrossRef](#)]
11. An, Q.T.; Sun, L.Z.; Zhao, K.; Sun, L. Switching function model-based fast-diagnostic method of open-switch faults in inverters without sensors. *IEEE Trans. Power Electron.* **2011**, *26*, 119–126. [[CrossRef](#)]
12. An, Q.T.; Sun, L.; Sun, L.Z. Current residual vector-based open-switch fault diagnosis of inverters in PMSM drive systems. *IEEE Trans. Power Electron.* **2015**, *30*, 2814–2827. [[CrossRef](#)]
13. Alavi, M.; Wang, D.W.; Luo, M. Short-circuit fault diagnosis for three-phase inverters based on voltage-space patterns. *IEEE Trans. Ind. Electron.* **2014**, *61*, 5558–5569. [[CrossRef](#)]
14. Zhang, J.H.; Zhao, J.; Zhou, D.H.; Huang, C.G. High-performance fault diagnosis in PWM voltage-source inverters for vector-controlled induction motor drives. *IEEE Trans. Power Electron.* **2014**, *29*, 6087–6099. [[CrossRef](#)]
15. Wu, F.; Zhao, J. A real-time multiple open-circuit fault diagnosis method in voltage-source-inverter fed vector controlled drives. *IEEE Trans. Power Electron.* **2016**, *31*, 1425–1437. [[CrossRef](#)]
16. Khomfoi, S.; Tolbert, L.M. Fault diagnostic system for a multilevel inverter using a neural network. *IEEE Trans. Power Electron.* **2007**, *22*, 1062–1069. [[CrossRef](#)]
17. Trabelsi, M.; Boussak, M.; Gossa, M. PWM-switching pattern-based diagnosis scheme for single and multiple open-switch damages in VSI-fed induction motor drives. *ISA Trans.* **2012**, *51*, 333–344. [[CrossRef](#)]
18. Khomfoi, S.; Tolbert, L.M. Fault diagnosis and reconfiguration for multilevel inverter drive using AI-based techniques. *IEEE Trans. Power Electron.* **2007**, *54*, 2954–2968. [[CrossRef](#)]
19. Zidani, F.; Diallo, D.; Benbouzid, M.E.H.; Rachid, N.S. A fuzzy-based approach for the diagnosis of fault modes in a voltage-fed PWM inverter induction motor drive. *IEEE Trans. Ind. Electron.* **2008**, *55*, 586–593. [[CrossRef](#)]
20. Masrur, M.A.; Chen, Z.; Murphey, Y. Intelligent diagnosis of open and short circuit faults in electric drive inverters for real-time applications. *IET Power Electron.* **2010**, *3*, 279–291. [[CrossRef](#)]
21. Wang, T.Z.; Xu, H.; Han, J.G.; Elbouchikhi, E.; Benbouzid, M.E.H. Cascaded H-bridge multilevel inverter system fault diagnosis using a PCA and multiclass relevance vector machine approach. *IEEE Trans. Power Electron.* **2015**, *30*, 7006–7018. [[CrossRef](#)]
22. Kamel, T.; Biletskiy, Y.; Chang, L.C. Fault diagnosis and on-line monitoring for grid-connected single-phase inverters. *Electr. Power Syst. Res.* **2015**, *126*, 68–77. [[CrossRef](#)]
23. Xia, Y.; Xu, Y.; Gou, B. A Data-Driven Method for IGBT Open-Circuit Fault Diagnosis Based on Hybrid Ensemble Learning and Sliding-Window Classification. *IEEE Trans. Ind. Inform.* **2019**, *16*, 5223–5233. [[CrossRef](#)]
24. Gou, B.; Xu, Y.; Xia, Y.; Deng, Q.; Ge, X. An Online Data-driven Method for Simultaneous Diagnosis of IGBT and Current Sensor Fault of 3-Phase PWM Inverter in Induction Motor Drives. *IEEE Trans. Power Electron.* **2020**, *35*, 13281–13294. [[CrossRef](#)]
25. Xia, Y.; Xu, Y. A Transferrable Data-Driven Method for IGBT Open-Circuit Fault Diagnosis in Three-Phase Inverters. *IEEE Trans. Power Electron.* **2021**, *36*, 13478–13488. [[CrossRef](#)]
26. Ma, C.; Luo, G.; Wang, K. Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images. *IEEE Trans. Med. Imaging* **2018**, *37*, 1943–1954. [[CrossRef](#)]
27. Faqhruldin, O.N.; El-Saadany, E.F.; Zeineldin, H.H. A universal islanding detection technique for distributed generation using pattern recognition. *IEEE Trans. Smart Grid* **2014**, *5*, 1985–1992. [[CrossRef](#)]
28. Shah, A.M.; Bhalja, B.R. Fault discrimination scheme for power transformer using random forest technique. *IET Gener. Transm. Distrib.* **2016**, *10*, 1431–1439. [[CrossRef](#)]
29. Zhang, D.; Qian, L.; Mao, B.; Huang, C.; Huang, B.; Si, Y. A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access* **2018**, *6*, 21020–21031. [[CrossRef](#)]
30. Hannan, M.A.; Abd Ali, J.; Mohamed, A.; Uddin, M.N. A random forest regression based space vector PWM inverter controller for the induction motor drive. *IEEE Trans. Ind. Electron.* **2017**, *64*, 2689–2699. [[CrossRef](#)]
31. Zhang, W.; Wang, J.; Han, G.; Huang, S.; Feng, Y.; Shu, L. A Data Set Accuracy Weighted Random Forest Algorithm for IoT Fault Detection Based on Edge Computing and Blockchain. *IEEE Internet Things J.* **2021**, *8*, 2354–2363. [[CrossRef](#)]