


Article

Fault Tracing Method for Relay Protection System–Circuit Breaker Based on Improved Random Forest

Ning Shao, Qing Chen *, Chengao Yu, Dan Xie and Ye Sun

Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan 250061, China; 202020548@mail.sdu.edu.cn (N.S.); 202134719@mail.sdu.edu.cn (C.Y.); 202114600@mail.sdu.edu.cn (D.X.); 202134689@mail.sdu.edu.cn (Y.S.)

* Correspondence: qchen@sdu.edu.cn

Abstract: The incorrect operation of protective relays and circuit breakers will significantly compromise the safety and stability of power systems. To promptly detect the faults of the relay protection system and the circuit breakers in time and to ensure the operational reliability of these protective devices, this paper proposes a fault tracing method for a relay protection system–circuit breaker based on improved Random Forest. Firstly, an analysis is conducted to identify the causes of incorrect operation of the protective relay and the circuit breaker. The fault types and corresponding alarm messages for the relay protection system and the circuit breaker are categorized, and the alarm feature set is constructed. Then, the Random Forest is improved and trained to develop the fault tracking model. Finally, the operation evaluation process is developed to determine the incorrect operations of the protective relay and the circuit breaker, and the fault tracking model and fault tracking process are then employed to locate the faults of the relay protection system and the circuit breaker. The experimental results demonstrate the method’s capability to accurately track faults in the relay protection system and the circuit breaker, thereby assisting operation and maintenance personnel in troubleshooting and highlighting its promising practical potential.

Keywords: relay protection system; circuit breaker; improved Random Forest; fault tracing



Citation: Shao, N.; Chen, Q.; Yu, C.; Xie, D.; Sun, Y. Fault Tracing Method for Relay Protection System–Circuit Breaker Based on Improved Random Forest. *Electronics* **2024**, *13*, 582. <https://doi.org/10.3390/electronics13030582>

Academic Editor: Ahmed Abu-Siada

Received: 5 January 2024

Revised: 28 January 2024

Accepted: 29 January 2024

Published: 31 January 2024



Copyright: © 2024 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/>).

1. Introduction

The continuous development of power systems and the incremental improvement of infrastructure have significantly increased the requirements for reliability and maintenance in order to guarantee the uninterrupted and dependable operation of power systems [1]. Protective relays (PRs) and circuit breakers (CBs) serve as control and protection equipment in power systems. They detect faults or exceptions in the power grid and promptly isolate faulty equipment to minimize the impact of the fault, thereby reducing the power grid losses [2]. Thus, the reliable operation of the relay protection system and the CB is crucial for maintaining a safe and stable power system. When the PR or CB fails to operate or operates incorrectly, it is an important research topic to accurately diagnose the faults in the relay protection system and the CB.

The increasing prevalence and integration of the Internet of Things (IoT), blockchain, and other computer and communication technologies have accelerated the growth of smart grids and smart substations within the power system [3,4]. The secondary system has realized intelligence, informatization, networking, and communication standardization, which provides sufficient data support for fault diagnosis and location in the intelligent substation [5]. However, informatization and networking also lead to the diversity and complexity of alarm messages. Massive alarm messages reduce the efficiency of diagnosing the incorrect operation of PRs or CBs and locating faulty secondary devices [6]. Data mining techniques such as artificial neural networks and machine learning are employed in power system fault diagnosis and localization to effectively leverage alarm information

and enhance fault detection efficiency and accuracy [7]. A fault tracking architecture based on big data technology is proposed for smart substations [6]. The framework includes a big data platform that enables the data mining of various alarm messages, including device alarms, generic object-oriented substation event (GOOSE) alarms, sampled value (SV) alarms, and device self-checking alarms in the smart substation. This platform eliminates the message barriers between the safety isolation zone and multiple supervisory systems. In addition, fault diagnosis methods and handling methods have been proposed for merger units, intelligent terminals, and fiber optic links in the process layer of smart substations [8]. The FP-growth algorithm is applied using the Hadoop framework and MapReduce model to discover frequent items and strong correlations among anomaly signals, aiming to identify potential defects [9]. Previous studies [10–12] have conducted a correlation analysis between secondary device faults and alarm messages, enabling the matching of faulty devices and fault causes through the discovery of correlation rules between fault messages and faulty devices. However, the current matching approach faces challenges in handling uncertainty and incomplete messages, and it also lacks generalization. The long short-term memory (LSTM) network has been employed to automatically diagnosis faults during the relay protection test [13]. However, deep learning methods are less explanatory and have high requirements for training samples, which makes it challenging to obtain perfect samples.

Communication failures may occur in the secondary circuit due to transmission delays, network attacks, and other interferences, subsequently affecting the normal operation of the secondary equipment or potentially causing maloperation [3,14]. Moreover, communication among secondary devices occurs through the network, leading to a loss of one-to-one correspondence in signal transmission between equipment. Consequently, issues such as hidden logic circuits and challenging communication troubleshooting arise. In [15], a communication network model is proposed to establish a correlation between communication links and physical aspects among secondary devices. This model then identifies communication faults by analyzing message transmission paths. Moreover, a communication fault diagnosis model based on a deep confidence network is developed in [16], demonstrating high accuracy and fault tolerance when dealing with untrustworthy messages. In [17], a fault localization method for communication networks is proposed. This method utilizes deep neural networks and effectively performs localization, even when multiple faults and untrustworthy alarms are present. Matrix modeling is used in [18] to construct the connectivity state matrix and logical node model of the secondary system. The matrix algorithm is also integrated with a back-propagation neural network to propose a fault localization method for the secondary system. The secondary circuit topology is converted into fault map data, and the fault location portability of the secondary circuit is improved by training a graph neural network as a fault location model [19].

Contrary to the fault diagnosis and fault localization mentioned earlier, fault tracing refers to the process of finding the cause of incorrect operations in PRs and CBs using alarm messages in the substation [6,20]. A fault tracing method is proposed utilizing the information difference graph model [21]. By marking the components, this technique enables the tracing of interaction relationships within the graph model. Thus, it effectively overcomes the limitations of conventional fault tracing methods that depend heavily on expert experience and logical topological relations. The Bayesian suspected degree, as proposed in [22], addresses the issue of missed judgments caused by lost alarms by calculating the probability of faults in PRs and CBs. The method proposed in [23] offers comprehensive and intuitive state information, the timely identification of hidden system hazards, and accurate fault causality tracing. It enhances the operational state monitoring and risk control capabilities of communication networks. The Recurrent Neural Network (RNN) has been employed to localize secondary device faults by extracting the temporal features of the alarm messages [24]. This method does not consider the effect of timing disorder in alarm messages on their accuracy. The Decision Tree and Gradient Boosting

Decision Tree have been proposed in [25,26], respectively, as fault tracing models to trace the causes of device faults of the relay protection system.

The CB and relay protection system (composed of a merging unit, protective relay, and intelligent terminal) belong to the primary side and secondary side of the power system, respectively. However, their communication connection is established through optical fiber. During the fault isolation, the relay protection system detects faults and issues trip commands, while the CB acts as an actuator and trips upon receiving the tripping command. Therefore, the causes of PR and CB rejections or maloperations include device faults in the PR and CB, device faults in other secondary devices in the relay protection system, and communication faults between these devices. However, most existing methods study the fault diagnosis of CBs or relay protection systems separately, without considering the interactions between the relay protection system and the corresponding CB. Moreover, they fail to comprehensively diagnose the device and communication faults of the relay protection system and CB and explore the reasons for the incorrect operation of the PR and CB.

To ensure the operational reliability of the PR and CB, and considering the correlation between the relay protection system and the CB in terms of topology and action logic, this paper proposes a fault tracking method for the relay protection system and CB (denoted as the RPS-CB). This paper first analyzes the causes of incorrect operations in PRs and CBs. It then establishes a correspondence between fault types and alarm messages in the relay protection system and the CB to create an alarm feature set. Subsequently, an improved Random Forest model is developed by integrating the Re-Relief F algorithm and a weighted voting strategy. The improved Random Forest model is trained for fault tracking. Finally, the operation evaluation process is utilized to identify PRs and CBs that were operated incorrectly. A fault tracing process based on the improved Random Forests is proposed to determine fault types in relay protection systems and CBs.

This paper is structured as follows: Section 2 describes the fault types and alarm messages of relay protection systems and CBs; Section 3 introduces the fault tracing model and fault tracing process based on improved Random Forest; Section 4 presents the algorithm validation and case study; and Section 5 concludes.

2. Fault Types and Alarm Feature Sets for the Relay Protection System–Circuit Breaker

To facilitate fault tracking for the RPS-CB, a comprehensive analysis of the root causes behind the erroneous functioning of both the PR and CB components is imperative. Subsequently, based on the causes above, a systematic categorization of the corresponding relationships between fault types and alarm messages about the RPS-CB can be established. Lastly, a specialized representation method for effectively capturing and managing alarm messages is devised to facilitate fault tracking procedures.

2.1. Reasons for Incorrect Operation of Protective Relays and Circuit Breakers

The PR receives the SV message and compares it with the protection setting value to determine the fault and issue a trip signal. The transmission path for the SV message is as follows: electronic transformer → merging unit → SV message → protective relay. Therefore, besides the faults of the PR, the causes of PR rejection also include exceptions in SV messages, errors in protection setting value configuration, network errors, and faults in other related secondary devices. In contrast to the PR rejection, PR maloperation refers to the abnormal PR function, which is mainly caused by abnormal SV sampling messages, incorrect protection setting values, and algorithm logic errors. The main reasons for PR rejection and malfunction are illustrated in Figure 1.

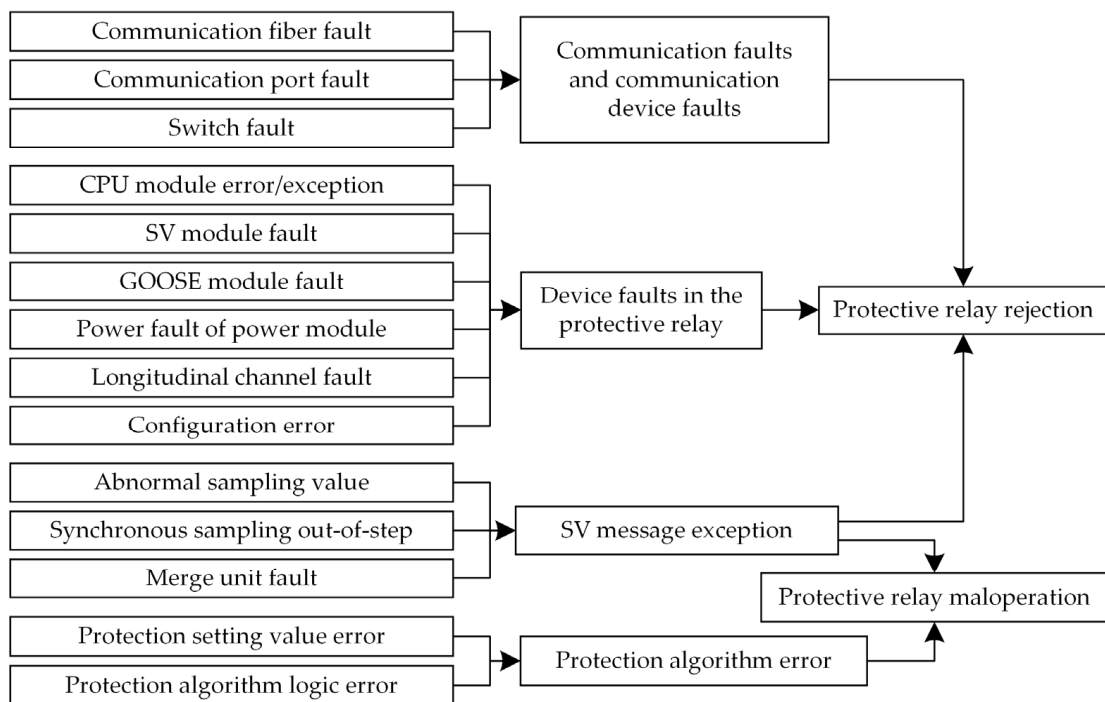


Figure 1. Reasons for PR rejection and malfunction.

The CB trips upon receiving the GOOSE message sent by the PR; the transmission path of the GOOSE message is as follows: protective relay → GOOSE message → intelligent terminal → circuit breaker. The operation state of the CB can be affected not only by the rejection and maloperation of the PR but also by CB faults, abnormal GOOSE messages, and communication faults. These factors contribute to the CB rejection and maloperation, as illustrated in Figure 2.

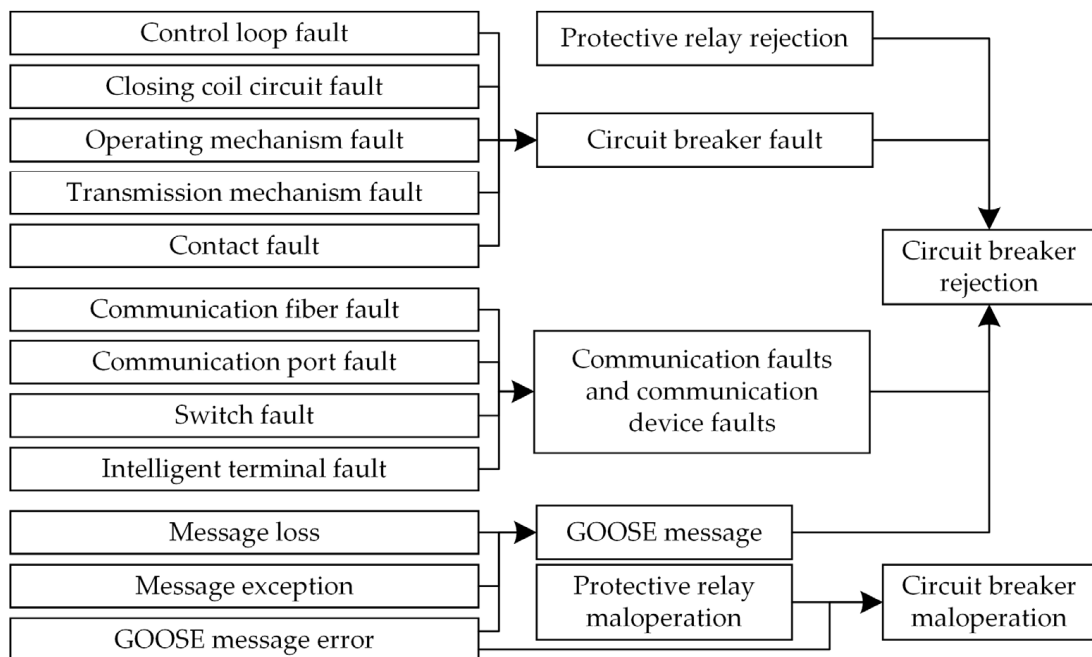


Figure 2. Reasons for CB rejection and malfunction.

The analysis above reveals that the incorrect operation of the PR and CB can be attributed to faults within the relay protection system and the CB. Specifically, CB rejection

and maloperation are directly linked to PR rejection and maloperation. This paper focuses on developing a fault tracking model and process for the RPS-CB (relay protection system and corresponding CB), aiming to investigate the relationship between system faults and the incorrect operation of the PR and CB.

2.2. Fault Types and Corresponding Alarm Messages of the RPS-CB

The fault tracing object mainly contains merging units, PRs, intelligent terminals, CBs, and the communication links between these devices. When the PR and CB fail to operate normally, corresponding alarm messages are issued by these devices to facilitate maintenance and repair. To achieve the fault tracking of the PR's and CB's incorrect operation, the faults of the protection relay and circuit breaker system (PR and CB) have been classified as device faults and communication faults, after consulting the literature [27,28] and relevant industry technical specifications. The fault types and their corresponding alarm messages are presented in Tables 1 and 2.

Table 1. Device fault types and corresponding alarm messages.

Devices	Alarm Messages	Fault Types	Number
Merging unit	Merging unit self-check alarm, sampling anomaly, merging unit synchronization anomaly, SV total alarm of the merging unit/PR, PR blocking, etc.	DSP module fault	000001
	Merging unit self-check alarm, protection SV total alarm, invalid protection SV data, protection blocking, etc.	AC input module fault	000010
	Merging unit self-check exception/alarm, input/output self-check circuit error, GOOSE/SV total alarm, GOOSE/SV communication interruption, chain break occurs when the merging unit receives GOOSE from measurement and control device, etc.	Input/output module fault	000011
	Power fault alarm of the merging unit	Power module fault	000100
	Abnormal synchronization of the merging unit, synchronization signal interruption of the merging unit, protection blocking, etc.	Sampling out-of-step	000101
	Merging unit GOOSE/SV configuration error, input/output configuration error, GOOSE/SV receiving plate error, etc.	Configuration error	000110
Protective relay	PR self-check alarm, CPU anomaly, fixed value verification error, device parameter sequence error, FLASH self-check anomaly, etc.	CPU module error/exception	000111
	PR self-check alarm, SV total alarm, protection SV sampling exception, protection blocking, etc.	SV module fault	001000
	PR self-check alarm, GOOSE total alarm, intelligent terminal GOOSE total alarm, reclosing lockout, etc.	GOOSE module fault	001001
	The power fault alarm of the protective relay	Power fault of the power module	001010
	Input/output communication interruption of the PR, the transmission state has not returned, input circuit exception, input circuit self-check error, device locking, etc.	Longitudinal channel fault	001011
	Inconsistent protection setting value, self-checking error of setting value, inconsistent panel configuration, inconsistent configuration file, SV/GOOSE configuration error, etc.	Configuration error	001100

Table 1. Cont.

Devices	Alarm Messages	Fault Types	Number
Intelligent terminal	Intelligent terminal self-check alarm, memory error, check error, GOOSE double receiving inconsistency, etc.	CPU module fault	001101
	Intelligent terminal self-check alarm, GOOSE total alarm, GOOSE/SV communication interruption, chain break occurs when the intelligent terminal receives GOOSE from merging unit, etc.	I/O module fault	001110
	The power fault alarm of the intelligent terminal	Power fault of the power module	001111
	Intelligent terminal self-check alarm, GOOSE configuration error, input configuration error, GOOSE receiving platen error, etc.	Configuration error	010000
Circuit breaker	Control circuit disconnection, control power fault	Control loop fault	010001
	Abnormal voltage of the closing coil, too little current of the closing coil, too short a current time of the closing coil, excessive current of the closing coils, excessive current time of the opening and closing coils	Closing coil fault	010010
	Loop locking of the switching circuit, abnormal total travel, the frequent start of the energy storage motor, long pressing time of the energy storage motor, wellhead pressure exception, energy storage anomaly of the spring	Operating mechanism fault	010011
	Long opening time of the contact, non-simultaneous movement of the contact, excessive shell temperature	Transmission mechanism fault	010100
	Excessive moisture content, SF6 gas leakage	Contact fault	010101
Switch	Switch communication interruption, data non-renovation, IED port communication fault, port indicator lights off, etc.	Communication link fault	010110
	Switch link interruption, port forwarding message inconsistency, broadcast storm, etc.	Port fault	010111
	StNum/SqNum jumping of GOOSE message, SV message counter jumping, etc.	Communication packet loss	011000
	The power fault alarm of the switch	Power fault	011001

Table 2. Communication fault types and corresponding alarm messages.

Communication Link	Alarm Messages	Fault Types	Number	
Merging unit– Protective relay	Merging unit device exception, self-test exception/alarms, etc.	Merging unit output port fault	011010	
	SV total alarm, SV sampling exception, input circuit exception, etc.	PR device exception, self-check alarm, input circuit exception, input circuit self-check error, etc.	Protective relay input port fault	011011
	-	Communication optical fiber fault	011100	

Table 2. Cont.

Communication Link	Alarm Messages	Fault Types	Number
Protective relay– Intelligent terminal	Intelligent terminal GOOSE communication interruption, no input message, input circuit exception, input circuit self-check error, etc.	PR output communication interruption, GOOSE communication interruption, self-check exception/alarm, etc.	Protective relay output port fault 011101
		Intelligent terminal device exception, self-check exception/alarm, etc.	Intelligent terminal input port fault 011110
	-	Communication optical fiber fault	011111
Bus merging unit– Line merging unit	Merging unit input circuit self-check error, input exception, self-check alarm, SV total alarm, Invalid/abnormal SV data, SV communication interruption, etc.	Bus merging unit device exception, self-check exception/alarm, etc.	Bus merging unit output port fault 100000
		Line merging unit device exception, self-check exception/alarm, etc.	Line merging unit input port fault 100001
	-	Communication optical fiber fault	100010
Intelligent terminal– Switch	Intelligent terminal exception, output test error, circuit breaker no action, intelligent terminal feedback	Intelligent terminal output port fault	100011

2.3. Alarm Feature Set

To make the expression of the alarm messages and the training of the fault tracking model more convenient, the alarm feature set is constructed according to the fault types and alarm messages in Tables 1 and 2, as shown in Equation (1).

$$X_i = \{X_{SVi}, X_{GOOSEi}, X_{SCi}, X_{COMi}\} \tag{1}$$

where X_i denotes the i th alarm feature set, which contains the SV alarm subset X_{SVi} , the GOOSE alarm subset X_{GOOSEi} , the device exception alarm subset X_{SCi} , and the communication alarm subset X_{COMi} .

The SV alarm subset X_{SVi} reflects the sampling state of the merging unit and PR, as shown in Equation (2).

$$\begin{cases} X_{SVi} = \{X_{SV_MU_i}, X_{SV_PR_i}\} \\ X_{SV_MU_i} = \{S_{MA}, S_{SA}, S_{SE}, S_{CE}, \dots\} \\ X_{SV_PR_i} = \{S_{MA}, S_{SA}, S_{EA}, S_{CE}, \dots\} \end{cases} \tag{2}$$

where $X_{SV_MU_i}$ and $X_{SV_PR_i}$ are the sampling alarm subsets of the merging unit and protective relay, including SV total alarm S_{MA} , abnormal sampling alarm S_{SA} , synchronization exception alarm S_{EA} , sampling configuration error S_{CE} , etc.

The GOOSE alarm subset X_{GOOSEi} reflects the switching quantity status of the merging unit, PR, and intelligent terminal, as shown in Equation (3).

$$\begin{cases} X_{GOOSEi} = \{X_{GOOSE_MU_i}, X_{GOOSE_PR_i}, X_{GOOSE_IT_i}\} \\ X_{GOOSE_MU_i} = \{G_{MA}, G_{DE}, G_{BC}, \dots\} \\ X_{GOOSE_PR_i} = \{G_{MA}, G_{DE}, G_{BC}, \dots\} \\ X_{GOOSE_IT_i} = \{G_{MA}, G_{DE}, G_{BC}, \dots\} \end{cases} \tag{3}$$

where $X_{GOOSE_MU_i}$, $X_{GOOSE_PR_i}$, and $X_{GOOSE_IT_i}$ are the subsets of the switching alarm features of the merging unit, PR, and intelligent terminal, respectively, which include GOOSE total alarm G_{MA} , GOOSE data exception G_{DE} , GOOSE interruption G_{BC} , etc.

The device exception alarm subset X_{SC_i} reflects the operation status of the merging unit, PR, intelligent terminal, and CB, as shown in Equation (4).

$$\begin{cases} X_{SC_i} = \{X_{SC_MU_i}, X_{SC_PR_i}, X_{SC_IT_i}, X_{SC_CB_i}, X_{SC_EX_i}\} \\ X_{SC_MU_i} = \{S_{CAN}, S_{PF}, S_{DL}, \dots\} \\ X_{SC_PR_i} = \{S_{CAN}, S_{PF}, S_{DL}, \dots\} \\ X_{SC_IT_i} = \{S_{CAN}, S_{PF}, S_{DL}, \dots\} \\ X_{SC_CB_i} = \{S_{CL}, S_{SC}, S_{SL}, \dots\} \\ X_{SC_EX_i} = \{S_{CAN}, S_{PF}, S_{BC}, \dots\} \end{cases} \quad (4)$$

where $X_{SC_MU_i}$, $X_{SC_PR_i}$, $X_{SC_IT_i}$, $X_{SC_CB_i}$, and $X_{SC_EX_i}$ are the subsets of the device self-test alarm features of the merging unit, PR, intelligent terminal, CB, and switch, respectively, which include device self-test S_{CAN} , power fault alarm S_{PF} , device lockout S_{DL} , etc. and CB control loop disconnection S_{CL} , closing coil voltage exception S_{SC} , closing coil loop blocking S_{SL} , and switch communication interruption S_{BC} , etc.

The communication exception alarm subset X_{COM_i} reflects the communication status of the merging unit, protective relay, intelligent terminal, and circuit breaker, as shown in Equation (5).

$$\begin{cases} X_{COM_i} = \{X_{COM_MU_i}, X_{COM_PR_i}, X_{COM_IT_i}, X_{COM_EX_i}\} \\ X_{COM_MU_i} = \{C_{SV}, C_{SOI}, C_{SIE}, \dots\} \\ X_{COM_PR_i} = \{C_{SV}, C_{GOOSE}, C_{SOI}, C_{SIE}, \dots\} \\ X_{COM_IT_i} = \{C_{GOOSE}, C_{SOI}, C_{SIE}, \dots\} \\ X_{COM_EX_i} = \{C_{BC}, C_{REF}, C_{IED}, \dots\} \end{cases} \quad (5)$$

where $X_{COM_MU_i}$, $X_{COM_PR_i}$, $X_{COM_IT_i}$, and $X_{COM_EX_i}$ are the subsets of the communication alarm features of the merging unit, PR, intelligent terminal, and switch, respectively, which include SV interruption C_{SV} , GOOSE interruption C_{GOOSE} , input/output communication interruption C_{SOI} , input circuit self-check error C_{SIE} , etc. and switch self-check fault C_{BC} , data non-renovation C_{REF} , port IED communication fault C_{IED} , etc.

When incorrect operation occurs in the PR or CB, the relevant devices in the corresponding spacing send out alarm messages. When the alarm message is received, the corresponding position element of the alarm feature set is set to 1; otherwise, the corresponding position element is 0.

3. Fault Tracking Based on Improved Random Forest

Fault tracing is determining the cause of incorrect operations in the PR and CB through alarm messages. Its essence is a multi-classification problem with alarm messages as classification features and fault types as classification results. However, fault types and alarm features in relay protection systems and CBs are diverse and complex. The alarm features of some fault types overlap, which may lead to inaccurate classification. In addition, certain faults may occur less frequently, causing data imbalance problems.

Random Forests are widely used as a multi-classification algorithm in the field of classification and prediction. The Random Forest is an integrated learning classification model consisting of multiple decision trees. High-dimensional datasets with complex features, such as alarm messages of the RPS-CB, can be effectively handled by random sampling. Training with partial samples and features can reduce the impact of noise on the overall performance of Random Forest and improve robustness. During the training process, the Random Forest will come to evaluate the contribution of each feature to the classification, in which the feature with the largest contribution is used as the node attribute

of the decision tree. Therefore, the classification process of the Random Forest has good interpretability, which is conducive to adjusting and improving the algorithm for fault tracing problems. In addition, the results of Random Forest are voted or averaged by multiple decision trees, which can reduce the variance of the model and improve the generalization ability. In summary, Random Forest is suitable for dealing with multi-classification problems in the RPS-CB.

This paper improves the Random Forest, establishes the mapping relationship between alarm messages and fault types through model training (as shown in Equation (6)), constructs the fault tracking model based on the improved Random Forest, and proposes the fault tracking process based on the improved Random Forest to achieve fault tracking of the causes of incorrect PR and CB operations.

$$Y = T(X) \quad (6)$$

where $T(\)$ is the fault tracking model with improved Random Forest, X is the alarm feature set, and Y is the set of classification labels, which consists of the fault numbers in Tables 1 and 2. To facilitate the input and output of the fault tracking model, Y is represented in the form of a vector. For example, the fault number 000001 can be represented as the vector $\{0, 0, 0, 0, 0, 0, 1\}$.

3.1. Improved Random Forest and Model Training

The Random Forest combines the Bootstrap method to form an ensemble learning model of multiple decision trees. Through the voting strategy, the robustness of the Random Forest is improved without feature scaling. The Random Forest is suitable for the fault tracing of the RPS-CB, which involves complex fault features with faults. This paper proposes an improved Random Forest that combines a feature selection algorithm and a weighted voting strategy. The feature selection algorithm eliminates the useless features and retains the features that impact model training and classification. The weighted voting strategy determines the weights of each decision tree according to the classification performance, which strengthens the impact of decision trees with good classification performance and further improves the accuracy of the fault tracking.

3.1.1. Feature Selection Algorithm

In constructing decision trees for Random Forests, each node randomly selects some features of the training samples to form a feature candidate set and then selects the optimal features from the feature candidate set as the node features. In the Random Forest, the samples are classified from the root node to the leaf nodes of each decision tree. After passing through each node, they will be classified according to the node feature and then go to the next child node to continue classification until they reach a leaf node. The samples are classified as the category of the leaf node. However, due to the wide variety of alarm messages of the RPS-CB and the high dimensionality of the alarm features, there is a likelihood that the feature candidate set will contain features that are not relevant to the samples assigned to the node. Supposing the node features are selected unreasonably, in such a case, the samples may be classified into the wrong leaf nodes along the wrong paths, leading to classification errors and reducing the classification performance of the Random Forest.

This paper addresses the problems by employing the Re-Relief F algorithm to evaluate alarm features within the training set. Weights are assigned based on their contribution to the classification, enhancing the selection probability of features with significant contributions and high weights. Furthermore, it reduces irrelevant and redundant features in the candidate set, improving the generalization ability and reducing overfitting. The Re-Relief F algorithm [29] is a feature selection algorithm applied to multi-classification problems. This algorithm first calculates the distances from the sample to the nearest neighbor samples of different classes and the distances to the nearest neighbor samples of the same class. Then, it takes the ratio of the two distances as the weights of the features. By emphasizing

features that exhibit strong proximity within classes and significant gaps between classes, the Re-Relief F algorithm improves the effectiveness and quality of the resulting feature subspace. The corresponding weight of the sample feature A of the Re-Relief F algorithm, $w[A]$, is calculated as follows:

$$w[A] = \frac{\sum_{i=1}^n \sum_{y \neq \text{class}(R_i)} \frac{P(y)}{1 - P(\text{class}(R_i))} \sum_{j=1}^k \text{diff}(A, R_i, M_j(y))}{\sum_{i=1}^n \sum_{j=1}^k \text{diff}(A, R_i, H_j)} \tag{7}$$

where n denotes the number of sample sampling, R_i is a random sample in the training set, k is the number of samples that are close neighbors to R , $M_j(y)$ denotes the j th nearest-neighbor sample of different category c , H_j denotes the j th nearest-neighbor sample in the same category as R , $P(y)$ denotes the ratio of the number of target samples of category y to the total number of samples, $\text{class}(R_i)$ denotes the category that R_i belongs to, and the function $\text{diff}(A, R_i, R_j)$ is to compute the distance of the sample instances R_i and R_j concerning feature A :

$$\text{diff}(A, R_i, R_j) = \begin{cases} 0, & R_i(A) = R_j(A) \\ 1, & R_i(A) \neq R_j(A) \end{cases} \tag{8}$$

where $R_i(A)$ and $R_j(A)$ are the feature values of the samples R_i and R_j corresponding to feature A .

The calculation process for the feature weights of the alarm features is as follows, based on the definition of the Re-Relief F algorithm:

1. Input training set D_{train} . The sample sampling number m is 10, and the number of feature dimensions is N . Initialize $n = 1$.
2. Initialize $i = 1$.
3. Select a random sample R_i from D_{train} .
4. Randomly select the k nearest neighbor sample H_j of R_i with different class samples and nearest neighbor samples M_j with different class samples.
5. If $i < m$, $i = i + 1$, return to step 3; otherwise, proceed to the next step.
6. According to Equation (7), the feature weight of the n th feature is calculated.
7. If $n < N$, $n = n + 1$, return to step 2; otherwise, output the feature weights of all alarm features.

The higher the weight of an alarm feature, the better the alarm feature is for distinguishing between similar and different types of faults in the immediate neighborhood. According to the feature weights, this paper improves the feature selection method in the training process of Random Forest. First, the Bootstrap method is used to obtain the training subsets required for decision tree training from the training set. Then, the features with a weight of 0 are excluded from the training subsets. The remaining features are sorted in descending order of weight and are evenly divided into high, medium, and low feature subsets (W_h , W_m , and W_l). Finally, the same number of alarm features from the feature subsets W_h , W_m , and W_l are randomly selected as the feature candidate set for the node of the decision trees.

The improved feature selection method increases the probability that features favorable for classification are selected, thus improving the classification performance of the decision tree. Simultaneously, this method maintains the randomness of feature selection and ensures robustness. In addition, the Re-Relief F algorithm compares the distances of the nearest neighbor samples of the same class and different classes, considering the correlation between alarm features, which is beneficial for distinguishing the types of faults with similar alarm features (e.g., communication-related device faults and communication faults), and thus reduces the probability of misclassification.

3.1.2. Weighted Voting Strategy

The average voting strategy ignores the differences in the classification performance of different decision trees in the Random Forest. It fails to give full play to the advantages of an ensemble learning model. Since fault tracking is a multi-classification problem, existing Random Forest-based methods primarily address binary classification problems and regression problems using weighted voting strategies. Therefore, this paper proposes a weighted voting strategy based on the Kappa coefficient to improve the overall performance of the Random Forest.

The Kappa coefficient is a metric for evaluating the performance of multi-classification models, which considers the classification accuracy and the consistency between classifications. It evaluates the multi-classification model performance more comprehensively by correcting for the expected accuracy of the classifications. The Kappa coefficient is computed as follows:

$$Kappa = \frac{N \sum_{i=1}^Y N_{ii} - \sum_{i=1}^Y N_{i+} N_{+i}}{N^2 - \sum_{i=1}^Y N_{i+} N_{+i}} \tag{9}$$

where N denotes the total number of test samples, Y is the total number of categories, N_{ii} is the number of correctly classified samples, and N_{+i} and N_{i+} are the number of true samples in category i and the number of samples classified as category i , respectively.

The process of calculating the voting weight of the decision tree in the improved random forest is as follows:

1. In addition to the training and validation sets needed for model training and validation, prepare a test set.
2. Train the Random Forest using the training set to obtain a fault tracking model consisting of T decision trees. Initialize $t = 1$.
3. Select the t th decision tree in the fault diagnosis model, and input the test set into the decision tree for classification. Calculate the weight $Kappa_t$ of the t th decision tree based on Equation (9) and the classification results of the test set.
4. If $t < T$, then $t = t + 1$; return to step 3; otherwise, proceed to the next step.

The classification result of the improved Random Forest depends on the weighted voting result of each decision tree classification result, as shown in Equation (10):

$$T(X_i) = \operatorname{argmax} \left\{ \sum_{i=1}^K Kappa_i \cdot I(t_i(X_i) = Y) \right\} \tag{10}$$

where $t_i(X_i)$ is the classification result of the i th decision tree; $Kappa_i$ is the Kappa coefficient corresponding to t_i ; $I(\)$ is a schematic function, and the function value is 1 when the classification result of the decision tree is related to a certain label in the classification label set Y , and 0 otherwise; $\operatorname{arg}(\)$ is the function of taking the autocorrelation, which serves the purpose of taking the classification with the highest number of votes.

3.1.3. Fault Tracing Model Training

Based on Random Forest training, combined with the Re-Relief F algorithm and weighted voting strategy, the training process of the fault tracking model based on the improved Random Forest is proposed as follows:

1. Sample the training set D_{train} T times using the Bootstrap method to obtain T sample subsets. Compute the feature weights for the training set D_{train} . Initialize $t = 1$.
2. Select the t th training subset. Refer to Section 3.1.1 to construct the high, medium, and low feature subsets (W_{ht} , W_{mt} , and W_{lt}) corresponding to the training subset.
3. Input the t th training subset into the root node of the t th decision tree, and start building the decision tree from the root node.

4. If the sample set in the current node is non-empty and all samples belong to multiple classes, this indicates that the current node is an internal node; proceed to step 4. Otherwise, the current node is a leaf node that is not further split. Other nodes are selected for further splitting. If all remaining nodes are leaf nodes, this indicates that all the samples have been trained and the construction of the decision tree is complete; proceed to step 8.
5. The current node contains samples with a total of M features. From W_{ht} , W_{mt} , and W_{lt} , m ($m = \sqrt{M/3}$, rounded down) features are randomly selected to form the candidate feature set D_t of the current node.
6. Calculate the *Gini* coefficient of each feature in the candidate feature set D according to Equation (11). The feature with the smallest *Gini* coefficient is removed from the candidate feature set D as the optimal feature of the current node.

$$Gini(p) = 1 - \sum_{k=1}^m p_k^2 \quad (11)$$

where $Gini(p)$ denotes the *Gini* coefficient of node p , k is the number of categories, and p_k denotes the classification probability.

7. Split the current node into two sub-nodes, and divide the samples contained in the current node into two parts according to the optimal features and input into the sub-nodes.
8. Repeat steps 4–8 for the sub-nodes.
9. If $t = T$, the Random Forest decision tree construction is completed; otherwise, $t = t + 1$, so return to step 3 and continue to construct the decision tree.
10. According to Section 3.1.2, the voting weights of the decision trees in the Random Forest are calculated, and the weighted voting result of this fault tracking model is shown in Equation (10).
11. The training of the fault tracking model based on the improved Random Forest is completed.

3.2. Fault Tracing Process Based on Improved Random Forest

3.2.1. Evaluation Process for PR and CB Operations

The classification object of the fault tracking model is the relay protection system and its corresponding CB. Therefore, before invoking the fault tracking model, it is necessary to identify incorrectly operated PRs and CBs. To this end, this paper develops the evaluation process for PR and CB operations. Firstly, the PRs and CBs are assigned confidence degrees according to alarm messages and fuzzy theory. Then, the confidence degrees of the PRs and CBs are fused and compared to determine the appropriate protective measure to isolate the faulty component. Finally, the incorrect operations of the main, near backup, and remote backup PRs and CBs are determined based on the configuration rules for relay protection. The specific steps of the evaluation process are as follows:

1. Suppose the fault diagnosis identifies R faulty components. S_{AM} represents the set of PRs and CBs that receive the alarm messages. Initialize $r = 1$.
2. Select the r th faulty component. Let C represent the number of CBs connected to the r th faulty component. Initialize $c = 1$.
3. Search for the main and near backup PR corresponding to the c th CB. Assume the r th faulty component has L neighboring lines, and search for the CB and the remote backup PR on the neighboring lines. Query the alarm message set S_{AM} and assign confidence degrees to the operation states of the PR and CB. The assignment rules for the confidence degrees of PRs and CBs [30] are shown in Table 3.
4. Calculate the fusion confidence degrees for the c th CB with the corresponding main and the near backup PR. Calculate the fusion confidence degrees for the remote CB and

remote backup PR of L adjacent lines. The calculation process of the fusion confidence degrees is shown in Equation (12).

$$\begin{cases} P_{MPCB} = \frac{1}{2}(P_{MPR} + P_{MCB}) \\ P_{NPCB} = \frac{1}{2}(P_{NPR} + P_{NCB}) \\ P_{RPCB} = \frac{1}{2l} \sum_1^L (P_{RPRI} + P_{RCBI}) \end{cases} \quad (12)$$

where P_{MPCB} , P_{NPCB} , and P_{RPCB} represent the fusion confidence degrees of the main, near backup, and remote backup PRs with the corresponding CBs, respectively. P_{MPR} , P_{NPR} , and P_{RPRI} represent the confidence degrees of the main, near backup, and remote backup PRs for the far end of the l th adjacent line, respectively. P_{MCB} , P_{NCB} , and P_{RCBI} represent the confidence degrees of CBs corresponding to the main, near backup, and remote backup PR in the far end of the l th adjacent line, respectively.

5. Compare the fusion confidence degrees of the main, near backup, and remote backup PRs. The protection that isolates the r th faulty component is the one with the highest fusion confidence degree. Based on the alarm messages and the action logic between different types of protection, evaluate the PR and CB operations. The operation evaluation rules for PRs and CBs are shown in Table 4.
6. According to the evaluation rules, the PRs and CBs configured for the r th faulty component are classified and placed into the set S_{NORM} of the normal operation PRs and CBs, the set S_{FP} (S_{FN}) of the PRs and CBs with no alarms (incorrect alarms), and the set S_{RO} (S_{FO}) of the rejected (incorrect operated) PRs and CBs.
7. If $c = C$, proceed to the next step after searching all connected CBs of the faulty components; otherwise, $c = c + 1$ and return to step 3.
8. If $r = R$, proceed to the next step after traversing all faulty components; otherwise, $r = r + 1$ and return to step 2.
9. If $S_{NORM} \cup S_{FP} \cup S_{FN} \cup S_{RO} \cup S_{FO} = S_{AM}$, the PRs and CBs with alarm messages have completed the operation evaluation; otherwise, there are PRs and CBs without corresponding faulty components. Considering that the PRs and corresponding CBs may be false alarms or maloperation, the PRs and CBs without corresponding faulty components are classified into S_{FN} and S_{FO} simultaneously. Since both the remote backup PR and the corresponding CB are backup protections for adjacent components, it is necessary to avoid the normally operated remote backup PRs and CBs as maloperation or rejection during the operation evaluation of the adjacent component. Therefore, the S_{FN} and S_{FO} require the following operations:

$$\begin{cases} S_{FN} = S_{FN} - (S_{FN} \cap S_{NORM}) \\ S_{FO} = S_{FO} - (S_{FO} \cap S_{NORM}) \end{cases} \quad (13)$$

10. Since maloperation and false alarms have alarm messages, while rejection and missing alarms do not, some of the PRs and CBs cannot be directly distinguished between the cases of maloperation or false alarms and the cases of rejections or missing alarms. Therefore, these PRs and CBs are divided into the pending set S_{UN} and evaluated after the fault tracking of the RPS-CB. The S_{UN} is as follows:

$$\begin{cases} S_{UN} = S_{FN} \cap S_{FO} + S_{FP} \cap S_{RO} \\ S_{FN} = S_{FN} - S_{FN} \cap S_{FO} \\ S_{FO} = S_{FO} - S_{FN} \cap S_{FO} \\ S_{FP} = S_{FP} - S_{FP} \cap S_{RO} \\ S_{RO} = S_{RO} - S_{FP} \cap S_{RO} \end{cases} \quad (14)$$

11. Conclude the operation evaluation and output the evaluation results S_{NORM} , S_{FP} , S_{FN} , S_{RO} , S_{FO} , and S_{UN} .

12. Either S_{RO} or S_{FO} is not an empty set; there are incorrectly operated PRs and CBs, and it is necessary to obtain the alarm messages of the corresponding spacings of the intelligent substation for fault tracing to determine the fault types of the relay protection system and circuit breakers. Either S_{FP} or S_{FN} is not an empty set; there are missing and false alarms of the PRs and CBs, and it is necessary to check the communication network between the substation and the control center or request the re-uploading of alarm messages. S_{UN} is not an empty set; there are PRs and CBs that are indistinguishable from those of the evaluation rules, and it is necessary to carry out secondary evaluation based on the fault tracking results.

Table 3. Assignment rules for confidence degrees of PRs and CBs.

PRs and CBs		Lines	Buses	Transformers
Main	PRs	0.9913 (0.2)	0.8564 (0.4)	0.7756 (0.4)
	CBs	0.9833 (0.2)	0.9833 (0.2)	0.9833 (0.2)
Near backup	PRs	0.80 (0.2)		0.75 (0.4)
	CBs	0.85 (0.2)		0.8 (0.2)
Remote backup	PRs	0.70 (0.2)	0.70 (0.4)	0.7 (0.4)
	CBs	0.75 (0.2)	0.75 (0.2)	0.75 (0.2)

Note: The values outside the parentheses represent the confidence degrees of PRs and CBs with alarm messages. The values inside parentheses indicate the confidence degrees of PRs and CBs without alarm messages.

Table 4. Operation evaluation rules for PRs and CBs.

Highest Fusion Confidence Degrees	Alarms of Main PR and CB	Operation Evaluation	Alarms of Near Backup PR and CB	Operation Evaluation	Alarms of Remote Backup PR and CB	Operation Evaluation
The main protection and corresponding CB	(1, 1)	Normal	(1, 1)	False alarm or maloperation of the near backup PR	(1, 1)	Remote backup PR and CB maloperation
					(1, 0)	False alarm of the remote backup PR
					(0, 1)	False alarm or maloperation of the CB
	(1, 0)	Missing alarm of the CB	(0, 0)	Missing alarm of the CB	(1, 0)	False alarm of the near backup PR
					(0, 1)	False alarm or maloperation of the CB
					(1, 0)	False alarm of the remote backup PR
(0, 1)	Missing alarm of the main protection	(0, 1)	Normal	(0, 1)	False alarm or maloperation of the CB	
				(1, 1)	Remote backup PR and CB maloperation	
				(1, 0)	False alarm of the remote backup PR	
The near backup protection and corresponding CB	(0, 1)	Main protection rejection	(1, 1)	Normal	(0, 1)	False alarm or maloperation of the CB
					(1, 0)	False alarm of the remote backup PR
					(0, 0)	False alarm of the remote backup PR
(0, 0)	Main PR rejection	(1, 0)	Missing alarm of the CB	(1, 0)	False alarm of the remote backup PR	
				(0, 1)	False alarm or maloperation of the CB	

Table 4. Cont.

Highest Fusion Confidence Degrees	Alarms of Main PR and CB	Operation Evaluation	Alarms of Near Backup PR and CB	Operation Evaluation	Alarms of Remote Backup PR and CB	Operation Evaluation
The remote backup protection and corresponding CB.	(0, 0)	Main PR rejection	(0, 0) (1, 0)	Near backup PR rejection CB rejection	(1, 1)	Normal
	(1, 0)	CB rejection	(1, 0) (0, 0)	CB rejection		
	(0, 1)	False alarm of the CB	(0, 1)	False alarm of the CB	(1, 0) (0, 1)	Missing alarm or rejection of the CB Missing alarm of the remote backup PR
	(0, 0)	Main PR rejection	(0, 0)	Near backup PR rejection		

Note: The binary group (p, q) indicates whether the alarm message exists or not; $p = 1$ ($q = 1$) indicates that the PR (CB) alarm message exists; $p = 0$ ($q = 0$) indicates that the PR (CB) alarm does not exist.

3.2.2. Fault Tracing Process

Since the improved Random Forest-based fault tracking model is a single-label and multiclassification model, this model is unable to track multiple faults simultaneously in the RPS-CB. In addition, faults in communication and neighboring devices may lead to alarms from multiple devices, so fault tracking must differentiate between communication faults and faults in adjacent neighboring devices. Furthermore, once faults are detected in the RPS-CB, it is crucial to ascertain whether the CB rejection (maloperation) results from PR rejection (maloperation) by examining the correlation between the operations of PR and CB.

To solve the above problems, this paper divides complex faults into multiple simple faults for fault tracking by dividing the alarm messages. Since communication faults simultaneously trigger alarms in adjacent devices, and the alarm messages for communication faults and device faults differ, it is prioritized to determine whether a communication fault is present. In the absence of a communication fault, the fault lies with adjacent devices and fault tracking is conducted based on individual device faults. If a communication fault exists, the alarm features related to the communication fault are removed from the alarm feature set. If the alarm feature set is not empty, then fault tracking is conducted according to multiple simple faults. Regarding multiple faults, the alarm messages are divided into individual device faults, which are further subdivided into multiple single faults for fault tracking.

According to the above problems and solutions, this paper proposes a fault tracking process based on an improved Random Forest, as shown in Figure 3.

The specific steps of the fault tracking process are as follows:

1. Once a fault occurs in the power grid, the fault diagnosis algorithm diagnoses the faulty component [30]. Subsequently, PRs and CBs are evaluated for their operation according to Section 2.1, and incorrectly operated PRs and CBs are identified.
2. Drawing from previous engineering research [6], the alarm messages of the PRs, CBs, and their associated secondary devices (merging units, protective relays, intelligent terminals, and switches) located within the device spacings are collected within 6 s following the fault. According to Section 2.3, these collected alarm messages form the corresponding alarm feature set, denoted as X.
3. If alarm messages are not issued from adjacent devices, this situation may indicate a single device fault or multiple faults of non-adjacent devices. The alarm feature set is divided according to the device to obtain the alarm feature subsets of single devices. If alarm messages are issued from adjacent devices, this situation may indicate communication faults or multiple faults. In such cases, the alarm feature set is divided based on the adjacent devices, resulting in subsets of alarm features for both the adjacent devices and individual devices.

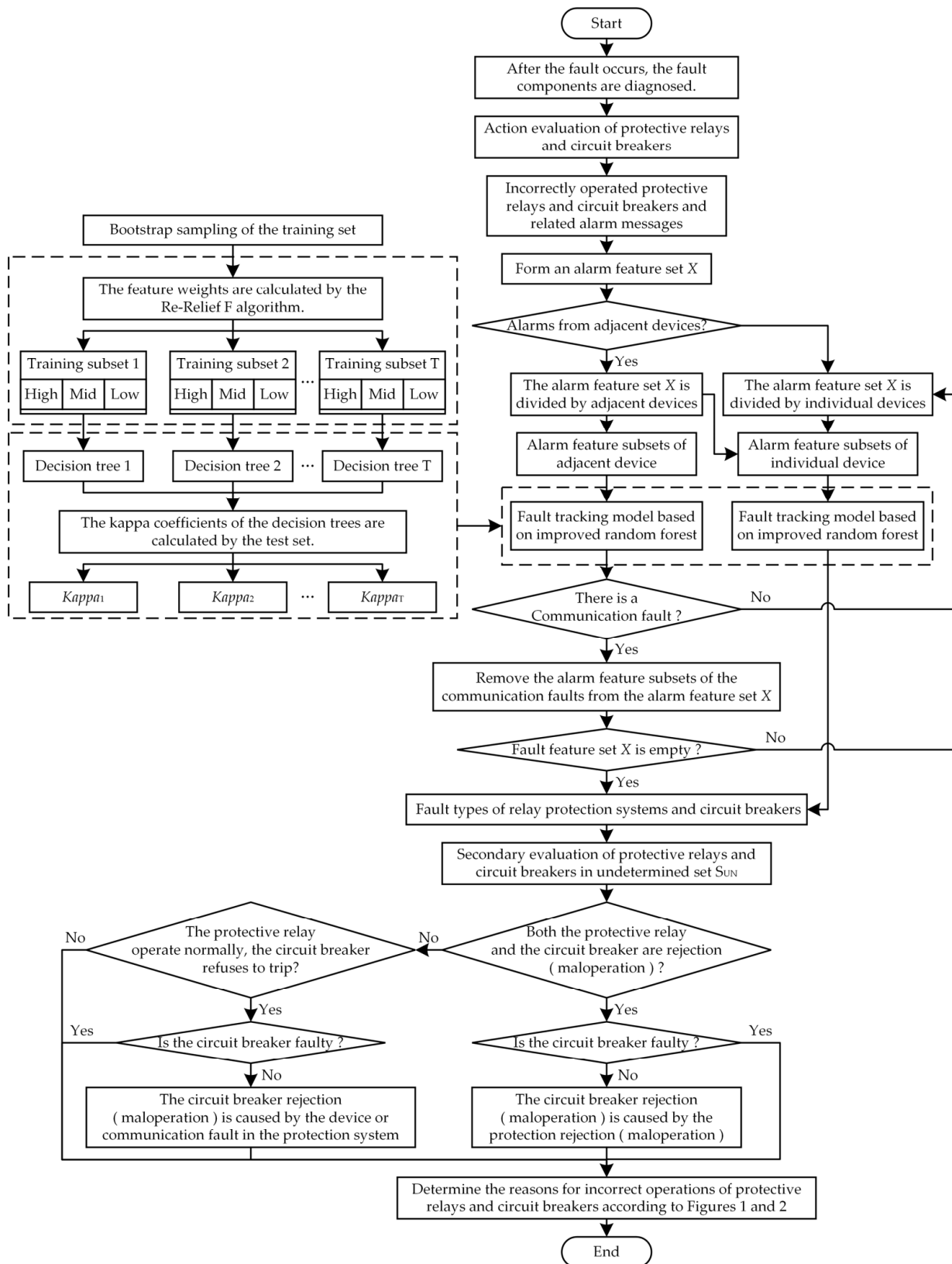


Figure 3. Fault Tracing Process Based on Improved Random Forest.

- The subsets of alarm features from adjacent devices are input to the fault tracking model. If a communication fault is found in the output result, in such cases, the subset

- of the alarm features corresponding to the communication fault is removed from the alarm feature set X to avoid repeated diagnosis.
5. If the alarm feature set X is empty, all alarm features have completed fault tracking and output the communication fault type of the RPS-CB. If not, there are other non-communication faults present, and it is then divided based on the individual device to obtain subsets for alarm features of individual devices.
 6. Alarm feature subsets of individual devices are input into a fault tracking model, outputting device fault types of the RPS-CB.
 7. For the PRs and CBs in the set S_{UN} , a secondary evaluation is necessary to assess whether they are rejected or missing (missing alarm or false alarm). The secondary evaluation rules are shown in Table 5.
 8. After combining the operation evaluation and the secondary evaluation results, whether the CB rejection (maloperation) is caused by the PR rejection (maloperation) or the faults in the relay protection system based on the fault tracking model results is determined. If both the PR and the CB are incorrectly operated, and the CB is not faulty, CB maloperation (CB rejection) is caused by the PR maloperation (PR rejection). If the PR operates and the corresponding CB fails to trip, it is considered that the CB rejection is caused by a device fault or a communication fault in the relay protection system.
 9. The relay protection system and CB faults that caused the PRs and CBs to operate incorrectly are determined based on Figures 1 and 2.

Table 5. Secondary evaluation rules.

Alarm Status of PRs and CBs in S_{UN}	Status of the Corresponding RPS-CBs	Operation Evaluation
Alarm messages are available	Fault	Maloperation
	No faults	False alarm
No alarm messages	Fault	Rejection
	No faults	Missing alarm

4. Case Verification

4.1. Verifications of Operation Evaluation Process

4.1.1. Fault Cases in the IEEE 39-Bus System

In this paper, the accuracy of the operation evaluation is verified by the fault cases of the IEEE 39-bus system (as shown in Figure 4). The fault cases and operation evaluation results are shown in Table 6. In Figure 4, B represents the bus, L represents the line, T represents the transformer, CB represents the circuit breaker, Lp represents the line protection, Bp represents the bus protection, Tp represents the transformer protection, and the suffixes m, p, and s represent the main, near backup, and remote backup protection, respectively. For example, the line connecting B_{03} and B_{18} is L_{0318} , and the transformer connecting B_{02} and B_{30} is T_{0230} ; the CB near the B_{03} of L_{0318} is CB_{0318} , and the circuit breaker near the B_{18} side is CB_{1803} . Lp_{0318m} , Lp_{0318p} , and Lp_{0318s} represent the corresponding main, near backup, and remote backup protection of CB_{0318} , respectively. Bp_{03} is the bus protection of B_{03} , Tp_{0230m} , and Tp_{0230p} are the main and near backup protection of the transformer.

Table 6 displays that Cases 1–10 simulate different single faults, including scenarios with incorrect PR and CB operations, as well as the absence of alarms and rejections. The proposed operation evaluation rules in this paper assess the operations of the PRs and CBs configured for the fault component based on the operational logic among the main, near backup, and remote backup protections, as well as between the PR and its corresponding CB. For example, in Cases 1 and 2, there are no alarm messages of CB_{1803} . Lp_{0318s} operates instead of CB_{1803} to isolate the fault in Case 2, while Case 1 lacks a backup protection

PR. Thus, Case 1 represents the absence of an alarm message for CB₀₃₁₈, whereas Case 2 indicates CB₀₃₁₈ rejection.

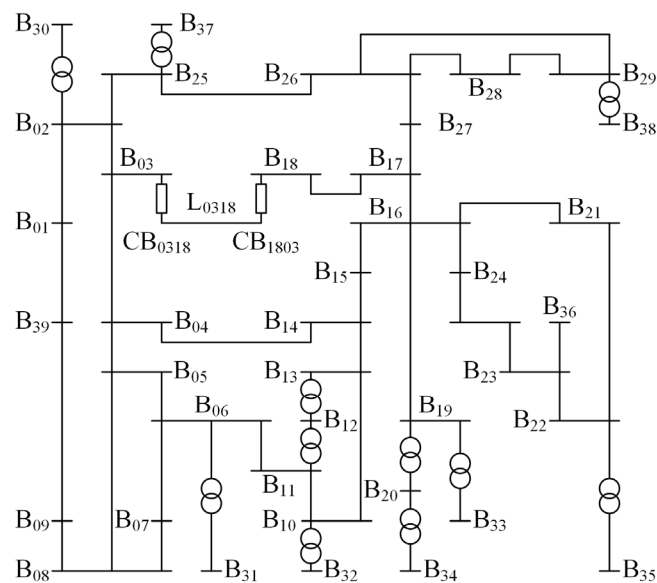


Figure 4. IEEE 39-bus power system.

Similarly, in Cases 3 and 4, it is possible to distinguish between the rejection or missing alarm of the PR by the operation state of the CB corresponding to the bus protection and the backup protections. In Cases 5–7, the PR and CB fail to operate, and the missing alarm occurs simultaneously. It can be seen from the evaluation results that the evaluation rules can correctly identify the simultaneous rejections and missing alarms in a single fault. In Cases 8–10, false alarms and the maloperations of PRs and CBs occur. It should be noted that since the main and near backup PRs correspond to the same CB, the maloperation and false alarm of the near backup PR cannot be directly judged by the corresponding CB operation state. Therefore, L_{p0203p} in Case 9 will be divided into the pending set S_{UN} and will wait for the fault tracking results for secondary evaluation, as per step 10 of the evaluation process in Section 3.2.1.

Unlike the single fault, the double faults in Cases 13–16 need to consider that the adjacent fault components are configured with the same backup PRs and CBs. When the PRs and CBs of a faulty component are evaluated, the remote backup PR and CB operation due to another faulty component will be judged as maloperation. For example, when a fault occurs in L_{0318} and L_{1718} in Case 13, and when evaluating the PRs and CBs operation of L_{0318} , CB_{1718} is a remote backup PR of L_{0318} and is operated due to the isolation of the L_{1718} fault. This situation is a false alarm or malfunction in the evaluation rules. When evaluating the PRs and CBs operation of L_{1718} , CB_{1718} functions as a normal CB. Therefore, during step 9 of the operation evaluation process, the evaluation results of PRs and CBs between different faulty components are compared, and the above double fault situation is corrected to avoid incorrect evaluation results.

The operation evaluation process proposed in this study accurately evaluates the operating states of PRs and CBs, reducing the scope of fault tracking for subsequent relay protection systems and CBs. Furthermore, this process can be applied not only to single faults but also to false alarms or missing alarms, maloperation or rejection, adjacent components configured with the same PRs and CBs, and other complex situations. The proposed method makes the operation evaluation more comprehensive and practical, and it is applicable to complex real-world scenarios in power systems.

Table 6. Evaluation results for fault cases in the IEEE 39-bus power system.

No.	Fault Details	Brief Description of the Alarm Messages	Faulty Components	Operation Evaluation
1	A fault occurs at B ₁₈ , and the alarm message of CB ₁₈₀₃ is lost.	Bp ₁₈ , CB ₁₈₁₇	B ₁₈	Missing alarm message of CB ₁₈₀₃
2	A fault occurs at B ₁₈ , CB ₁₈₀₃ fails to trip, Lp _{0318s} operates, and CB ₀₃₁₈ trips.	Bp ₁₈ , CB ₁₈₁₇ , Lp _{0318s} , CB ₀₃₁₈	B ₁₈	CB ₁₈₀₃ rejection
3	A fault occurs at B ₀₃ , and the alarm message of Bp ₀₃ is lost.	CB ₀₃₀₂ , CB ₀₃₀₄ , CB ₀₃₁₈	B ₀₃	Missing alarm message of Bp ₀₃
4	A fault occurs at B ₀₃ , Bp ₀₃ fails to operate, and remote backup PRs operate.	Lp _{1803s} , Lp _{0403s} , Lp _{0203s} , CB ₁₈₀₃ , CB ₀₄₀₃ , CB ₀₂₀₃	B ₀₃	Bp ₀₃ rejection
5	A fault occurs at L ₀₃₁₈ , Lp _{1803m} fails to operate, and Lp _{1803p} operates.	Lp _{0318m} , CB ₀₃₁₈ , Lp _{1803p} , CB ₁₈₀₃	L ₀₃₁₈	Lp _{1803m} rejection
6	A fault occurs at L ₀₃₁₈ , CB ₁₈₀₃ fails to trip, and the alarm message of Lp _{0318m} is lost.	Lp _{1803m} , CB ₀₃₁₈ , Lp _{1718s} , CB ₁₇₁₈	L ₀₃₁₈	CB ₁₈₀₃ rejection and missing alarm message of Lp _{0318m}
7	A fault occurs at L ₀₃₁₈ , and the alarm message of CB ₀₃₁₈ is lost.	Lp _{0318m} , Lp _{1803m} , Lp _{1718s} , CB ₁₇₁₈	L ₀₃₁₈	CB ₁₈₀₃ rejection and missing alarm message of CB ₀₃₁₈
8	A fault occurs at L ₀₃₁₈ , CB ₁₈₀₃ fails to trip, and the alarm message of Lp _{0203s} is wrong.	Lp _{1803m} , Lp _{0318m} , CB _{1817s} , CB ₁₇₁₈ , Lp _{0203s}	L ₀₃₁₈	CB ₁₈₀₃ rejection and false alarm message of Lp _{0203s}
9	A fault occurs at B ₀₃ , CB ₀₃₁₈ fails to trip, and Lp _{0203p} operates incorrectly.	Bp ₀₃ , CB ₀₃₀₂ , CB ₀₃₀₄ , Lp _{0203p} , CB ₀₂₀₃ , Lp _{1803s} , CB ₁₈₀₃	B ₀₃	malfunction or false alarm of Lp _{0203p} and CB ₀₃₁₈ rejection
10	A fault occurs at L ₀₃₁₈ , Lp _{1718s} operates incorrectly, and CB ₁₇₁₈ trips.	Lp _{0318m} , Lp _{1803m} , CB ₀₃₁₈ , CB ₁₈₀₃ , Lp _{1817s} , CB ₁₈₁₇	L ₀₃₁₈	Lp _{1718s} and CB ₁₇₁₈ malfunction
11	A fault occurs at T ₁₂₁₃ , and the alarm message of CB ₁₃₁₂ is lost.	Tp _{1213m} , CB ₁₂₁₃	T ₁₂₁₃	Missing alarm message of CB ₁₃₁₂
12	A fault occurs at T ₁₂₁₃ , CB ₁₃₁₂ fails to trip, Lp _{1413s} and Lp _{1013s} operate, and CB ₁₄₁₃ and CB ₁₀₁₃ trip.	Tp _{1213m} , CB ₁₂₁₃ , Lp _{1413s} , Lp _{1013s} , CB ₁₄₁₃ , CB ₁₀₁₃	T ₁₂₁₃	CB ₁₃₁₂ rejection
13	A fault occurs at L ₀₃₁₈ , and the alarm message of CB ₀₃₁₈ is lost; a fault occurs at L ₁₇₁₈ .	Lp _{0318m} , Lp _{1803m} , CB ₁₈₀₃ , Lp _{1817m} , Lp _{1718m} , CB ₁₈₁₇ , CB ₁₇₁₈	L ₀₃₁₈ , L ₁₇₁₈	Missing alarm message of CB ₀₃₁₈
14	Faults occur at L ₀₃₁₈ and B ₁₇ , CB ₁₈₀₃ fails to trip, and the alarm message of Lp _{1817s} is lost.	Lp _{1803m} , Lp _{0318m} , CB ₀₃₁₈ , CB ₁₇₁₈ , Bp ₁₇ , CB ₁₇₂₇ , CB ₁₇₁₆	L ₀₃₁₈ , B ₁₇	CB ₁₈₀₃ rejection and missing alarm message of Lp _{1817s}
15	A fault occurs at B ₀₃ , and the alarm message of CB ₀₃₁₈ is lost; a fault occurs at B ₁₄ .	Bp ₀₃ , CB ₀₃₀₂ , CB ₀₃₀₄ , Bp ₁₄ , CB ₁₄₀₄ , CB ₁₄₁₅ , CB ₁₄₁₃	B ₀₃ , B ₁₄	Missing alarm message of CB ₀₃₁₈
16	A fault occurs at B ₁₈ , Bp ₁₈ fails to operate, and remote backup PRs operate. A fault occurs at T ₁₂₁₃ , and the alarm message of Tp _{1213m} is lost.	Lp _{0318s} , Lp _{1718s} , CB ₀₃₁₈ , CB ₁₇₁₈ , CB ₁₃₁₂ , CB ₁₂₁₃	B ₁₈ , T ₁₂₁₃	Bp ₁₈ rejection and missing alarm message of Tp _{1213m}

4.1.2. Fault Cases in the IEEE 118-Bus System

For the fault tracing method to be applied to larger and more complex power systems, the evaluation rules must be capable of evaluating the operations of the PRs and CBs in this power system to identify failed relay protection systems and CBs. This paper verifies the applicability of the evaluation rules by testing them on fault cases in the IEEE 118-bus power system (as shown in Figure 5). The circles in Figure 5 represent busbars, and the lines connecting them are transmission lines. The numbering rules for lines, buses, CBs, and PRs match those in Figure 4.

Table 7 presents the results of the operation evaluation for fault cases within the IEEE 118-bus power system. Cases 17–19 represent line faults characterized by CB rejections, the missing alarm message, and the false alarm from the PR. Cases 20–22 involve bus faults with CB rejections, missing alarm messages from CBs, and the false alarm of the PR. Cases 23 and 24 encompass double faults, including the rejection and missing alarm messages from the bus PR. The evaluation results demonstrate the accurate identification of improperly functioning PRs and CBs for both single faults (Cases 17–22) and complex faults (Cases 22 and 24).

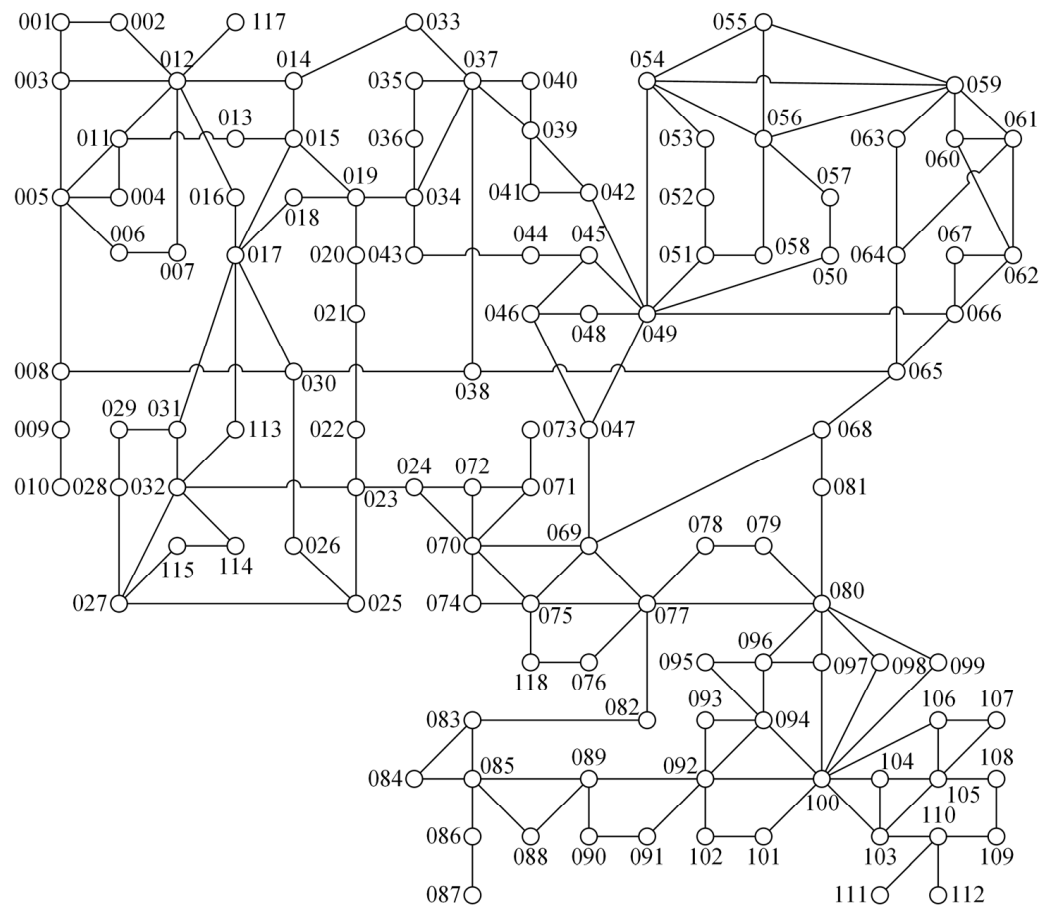


Figure 5. IEEE 118-bus power system.

Table 7. Evaluation results for fault cases in the IEEE 118-bus power system.

No.	Fault Details	Brief Description of the Alarm Messages	Faulty Components	Operation Evaluation
17	A fault occurs at L ₀₂₀₀₂₁ , CB ₀₂₁₀₂₀ fails to trip, and the alarm message of Lp _{020021m} is lost.	Lp _{021020m} , CB ₀₂₀₀₂₁ , Lp _{022021s} , CB ₀₂₂₀₂₁	L ₀₂₀₀₂₁	CB ₀₂₁₀₂₀ rejection and missing alarm message of Lp _{020021m}
18	A fault occurs at L ₀₂₀₀₂₁ , and Lp _{021020m} fails to operate.	Lp _{020021m} , CB ₀₂₀₀₂₁ , Lp _{022021s} , CB ₀₂₂₀₂₁	L ₀₂₀₀₂₁	Lp _{021020m} rejection
19	A fault occurs at L ₀₂₀₀₂₁ , and the alarm message of Lp _{021020p} is wrong	Lp _{021020m} , Lp _{020021m} , Lp _{022021p} , CB ₀₂₀₀₂₁ , CB ₀₂₂₀₂₁	L ₀₂₀₀₂₁	False alarm of Lp _{021020p}
20	A fault occurs at B ₀₃₈ , CB ₀₃₈₀₃₀ fails to trip, and the alarm message of CB ₀₃₈₀₃₇ is lost.	Bp ₀₃₈ , Lp _{030038s} , CB ₀₃₈₀₆₅ , CB ₀₃₀₀₃₈	B ₀₃₈	CB ₀₃₈₀₃₀ rejection and missing alarm message of CB ₀₃₈₀₃₇
21	A fault occurs at B ₀₄₅ , CB ₀₄₅₀₄₄ fails to trip, and the alarm message of CB ₀₄₆₀₄₅ is wrong.	Bp ₀₄₅ , Lp _{044045s} , CB ₀₄₄₀₄₅ , CB ₀₄₅₀₄₆ , CB ₀₄₅₀₄₉ , CB ₀₄₆₀₄₅	B ₀₄₅	CB ₀₄₅₀₄₄ rejection and false alarm or maloperation of CB ₀₄₆₀₄₅
22	A fault occurs at B ₀₄₅ , CB ₀₄₅₀₄₄ fails to trip, and the alarm message of Lp _{046045s} is wrong.	Bp ₀₄₅ , Lp _{044045s} , CB ₀₄₄₀₄₅ , CB ₀₄₅₀₄₆ , CB ₀₄₅₀₄₉ , Lp _{046045s}	B ₀₄₅	CB ₀₄₅₀₄₄ rejection and false alarm of Lp _{046045s}
23	A fault occurs at L ₀₂₀₀₂₁ , and the alarm message of Lp _{020021s} is wrong. A fault occurs at B ₀₇₂ , and Bp ₀₇₂ fails to trip.	Lp _{020021m} , Lp _{021020m} , CB ₀₂₀₀₂₁ , CB ₀₂₁₀₂₀ , Lp _{020021s} , Lp _{024072s} , Lp _{070072s} , Lp _{071072s} , CB ₀₂₂₀₂₁ , CB ₀₂₂₀₂₁ , CB ₀₂₂₀₂₁	L ₀₂₀₀₂₁ B ₀₇₂	False alarm of Lp _{020021s} and Bp ₀₇₂ rejection
24	A fault occurs at B ₀₃₉ , and the alarm message of Bp ₀₃₉ is lost. A fault occurs at L ₀₄₃₀₄₄ , and the alarm message of Lp _{044043m} is lost.	CB ₀₃₉₀₃₇ , C _{B039040} , CB ₀₃₉₀₄₁ , CB ₀₃₉₀₄₂ , Lp _{043044m} , CB ₀₄₃₀₄₄ , CB ₀₄₄₀₄₃	B ₀₃₉ L ₀₄₃₀₄₄	Missing alarm message of Bp ₀₃₉ and Lp _{044043m}

False alarms of CBs and PRs occur in Cases 19, 21, and 22, respectively. The operation evaluation rule determines the protection type that isolates the faulty components by comparing the status of the main, near backup, and far backup PRs. Whether it is classified as a maloperation or a false alarm depends on whether the corresponding CB (PR) of the falsely alarmed PR (CB) operates or not. Since the corresponding PRs and CBs do not operate in these cases, they are judged as false alarms.

In power systems of varying sizes and topologies, components are configured with main, near backup, and remote backup PRs and corresponding CBs. Based on the action logic of different types of protection, the complex PR and CB configurations are divided into several mutually independent combinations of main, near backup, and remote backup PRs and corresponding CBs. Operation evaluation rules can be applied to evaluate each combination. Thus, the operation evaluation rules can evaluate the operational status of the PRs and CBs, even in larger and more complex power systems.

The above cases prove that the operation evaluation rules can be applied in larger and more complex power systems, which lays the foundation for the application of the fault tracking model and fault tracking process in large-scale and complex power systems.

4.2. Improved Random Forest Verification

4.2.1. Data Preprocessing and Sample Set Construction

In this paper, a representative line interval from a 220 kV intelligent substation (illustrated in Figure 6) is selected as a case study. Alarm messages are gathered from historical records and simulation experiments to create sample datasets of relay protection systems and CBs. The effectiveness and reliability of the improved Random Forest proposed in this paper are assessed.

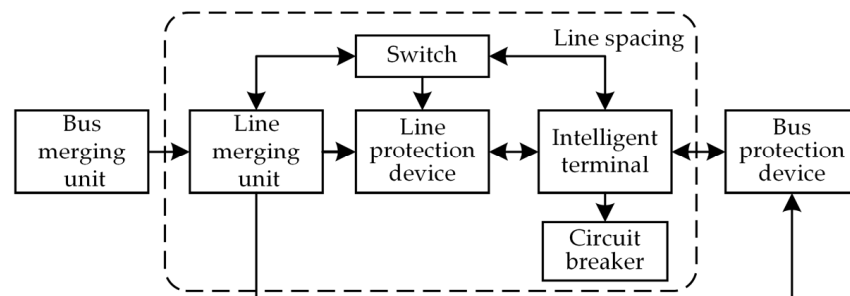


Figure 6. Topological structure of line spacing.

The steps for preprocessing and sample preparation for alarm messages are as follows:

1. Collect the alarm messages of the relay protection systems or CBs from the secondary system of the substation, and extract the names of stations, intervals, and devices. Then, group the alarm messages according to the plant name and the interval name.
2. According to the engineering experience [6], take 6 s as a time window and calculate the frequency of the alarm message within the time window. If an alarm appears only 1–2 times in a time window, regard it as a false alarm; regard an alarm that appears three times or more as a real alarm.
3. According to the definition of the alarm feature set, map alarm messages to the corresponding alarm feature set X_i . If the alarm message exists, set the alarm feature at the corresponding position of the alarm feature set X_i set to 1; otherwise, set it to 0.
4. Label the alarm feature set X_i with fault type Y_i ; the numbering of the fault types is detailed in Tables 1 and 2. Combine the alarm features X_i and the corresponding fault type Y_i as (X_i, Y_i) and input into the sample set of fault tracking.

Based on the above steps, this paper constructs a fault sample set of relay protection systems and circuit breakers with a total number of 16,107 samples, of which 11,183 are training samples, 1604 are test samples, and 3320 are validation samples. The details of the samples are shown in Tables 8 and 9.

Table 8. Samples of device faults.

Devices	Fault Types	Total Number of Samples	Number of Training Samples	Number of Test Samples	Number of Validation Samples
Merging unit	DSP module fault	477	335	44	98
	AC input module fault	371	257	37	77
	Input/output module fault	530	366	54	110
	Power module fault	310	214	31	65
	Sampling out-of-step	197	137	19	41
	Configuration error	366	259	31	76
Protective relay	CPU module error/exception	247	175	21	51
	SV module fault	406	286	35	85
	GOOSE module fault	572	406	49	117
	Power fault of the power module	591	405	65	121
	Longitudinal channel fault	662	450	78	134
	Configuration error	320	181	85	54
Intelligent terminal	CPU module fault	256	178	24	54
	I/O module fault	651	457	59	135
	Power fault of the power module	641	448	57	136
	Configuration error	546	387	48	111
Circuit breaker	Control loop fault	303	213	26	64
	Closing coil fault	676	464	72	140
	Operating mechanism fault	702	486	77	139
	Transmission mechanism fault	635	443	62	130
	Contact fault	306	213	29	64
Switch	Communication link fault	533	377	48	108
	Port fault	666	442	90	134
	Communication packet loss	500	351	44	105
	Power fault	299	213	25	61

Table 9. Samples of communication faults.

Communication Links	Fault Types	Total Number of Samples	Number of Training Samples	Number of Test Samples	Number of Validation Samples
Merging unit– Protective relay	Merging unit output port fault	454	318	40	96
	Protective relay input port fault	362	253	33	76
	Communication optical fiber fault	524	367	48	109
Protective relay– Intelligent terminal	Protective relay output port fault	310	217	27	66
	Intelligent terminal input port fault	412	288	40	84
	Communication optical fiber fault	262	183	25	54
Bus merging unit– Line merging unit	Bus merging unit output port fault	500	350	46	104
	Line merging unit input port fault	466	326	44	96
	Communication optical fiber fault	511	358	44	109
Intelligent terminal– Circuit breaker	Intelligent terminal output port fault	543	380	47	116

4.2.2. Evaluation Indicators

Considering that the fault tracking of the RPS-CB is a multiclassification problem, this paper chooses producer accuracy, user accuracy, overall accuracy, and the Kappa coefficient to evaluate the classification performance. User and producer accuracy are used to evaluate the classification performance of a certain category, and the overall accuracy and Kappa coefficient are used to evaluate the overall classification performance of the algorithm. The

Kappa coefficient is defined as shown in Equation (9), and the producer accuracy (PA), user accuracy (UA), and overall accuracy (OA) are defined as follows:

$$PA = N_{ii} / N_{+i} \quad (15)$$

$$UA = N_{ii} / N_{i+} \quad (16)$$

$$OA = \sum_{i=1}^Y N_{ii} / N \quad (17)$$

4.2.3. Parameter Optimization of the Improved Random Forest

The number and maximum depth of decision trees are the main parameters affecting the classification performance of the improved Random Forest. Reasonable parameter settings can ensure good accuracy and generalization ability and prevent the fault tracking model from being too complex, and therefore being large, which affects the training and operation efficiency. For this reason, this paper adopts the hierarchical cross-validation method. The overall accuracy OA is the evaluation index to tune the number of decision trees ($n_estimators$) and the maximum depth (max_depth) of the improved Random Forest.

The process of 10-fold hierarchical cross-validation is as follows:

1. The training set is divided into ten groups, ensuring that the proportion of each fault type in each group is consistent with the training set.
2. Nine groups are selected as the cross-validated training set to train the improved Random Forest, and the remaining group as the cross-validated validation set to verify the training effect. This is repeated ten times to ensure that each group is selected as the validation set for cross-validation.
3. The average of the overall accuracies of the ten cross-validations is counted as the result of the cross-validation.

The maximum number of decision trees is determined through the optimization process using 10-fold hierarchical cross-validation. The number of decision trees for the improved Random Forest starts from one, and a hierarchical cross-validation is conducted for each additional decision tree. The relationship between the number of decision trees and the cross-validation results is visualized in Figure 7. As the number of decision trees increases, the average overall accuracy shows an upward trend until it stabilizes at 50 trees.

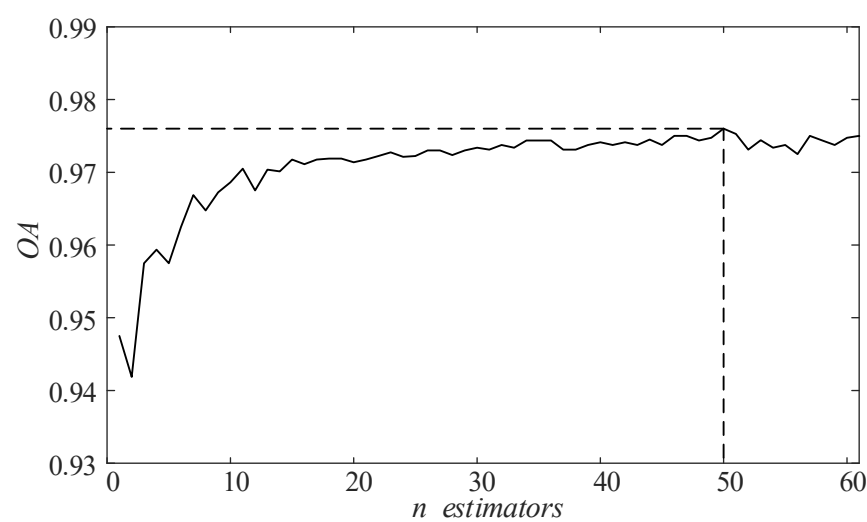


Figure 7. Verification curve of the number of decision trees.

Following the selection of the optimal number of decision trees, the maximum depth of the decision trees is optimized using 10-fold hierarchical cross-validation. The improved Random Forest initializes its decision tree with a root node and performs hierarchical

cross-validation for each subsequent node added incrementally. The relationship between the maximum depth of the decision tree and the cross-validation results is visualized in Figure 8. As the maximum depth of the decision tree increases, there is a tendency for the average overall accuracy to rise until it stabilizes at a maximum depth of approximately 20.

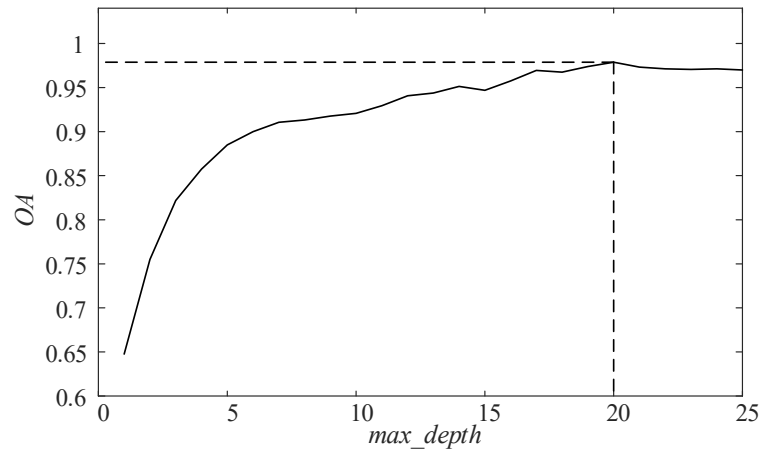


Figure 8. Verification curve of the max depth of decision trees.

Decreasing the size of the Random Forest can reduce the computational burden and improve the efficiency of operation while maintaining a high overall accuracy. Moreover, an improved Random Forest-based model with large scale and deep depth can lead to overfitting. Therefore, this paper sets the number of decision trees for the improved Random Forest-based fault tracking model to 50 and sets the maximum depth to 20.

4.2.4. Comparisons of Fault Tracing Methods

In order to evaluate the classification effectiveness of the improved Random Forest, this paper compares it with the existing fault tracking methods, which include the Decision Tree [6,25], Gradient Boosting Decision Tree (GBDT) [26], Recurrent Neural Network (RNN) [24], and Random Forest.

The comparison experiment of fault tracking methods is conducted as follows:

Firstly, the improved Random Forest and other existing methods are trained ten times using the training set. Subsequently, the validation set is employed to validate each method after every training iteration. The running time T_{test} , the overall accuracy (OA), and the Kappa coefficient ($kappa$) of the classification results of the validation set are recorded for each validation result, as shown in Table 10. Finally, the optimal validation results of each method are selected and compared with the user accuracy (UA) and producer accuracy (PA) for each type of fault, as shown in Tables 11 and 12. All comparison experiments are performed on a computer with a 2.2 GHz main frequency, a six-core processor (Intel Core i7-8750H), and 32 G of RAM, utilizing the PyCharm programming tool.

Table 10. Statistical results.

No	[6,25]			Random Forest			[26]			[24]			This Paper		
	T_{test} (ms)	OA (%)	Kappa (%)	T_{test} (ms)	OA (%)	Kappa (%)	T_{test} (ms)	OA (%)	Kappa (%)	T_{test} (ms)	OA (%)	Kappa (%)	T_{test} (ms)	OA (%)	Kappa (%)
1	7	91.60	91.32	23	93.77	93.57	27	95.84	95.70	73	96.42	96.30	20	96.81	96.71
2	6	91.20	90.93	30	93.31	93.12	23	95.78	95.65	71	96.45	96.33	22	97.20	97.11
3	6	91.14	90.87	22	93.37	93.18	23	95.66	95.52	74	96.20	96.08	21	97.65	97.57
4	7	91.54	91.26	23	94.01	93.81	24	94.61	94.43	67	96.36	96.23	23	97.83	97.79
5	9	91.51	91.26	22	94.58	94.40	20	95.12	94.96	75	96.54	96.42	25	97.08	96.98
6	10	89.91	89.64	23	93.49	93.01	24	95.39	95.24	77	96.11	95.98	23	97.20	97.11
7	10	91.08	90.81	24	93.37	93.18	23	95.30	95.15	69	96.30	96.17	20	97.44	97.36
8	7	90.75	90.48	22	93.67	93.48	26	95.51	95.37	73	96.23	96.11	26	97.08	96.98
9	6	90.57	90.30	22	94.49	94.22	25	94.41	94.21	72	96.66	96.54	24	97.35	97.23
10	7	91.27	90.99	23	92.95	92.76	26	95.00	94.84	75	95.99	95.86	23	97.11	97.01

Table 11. Verification results of device faults.

Devices	Fault Types	[6,25]		Random Forest		[26]		[24]		This Paper	
		UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Merging unit	DSP module fault	84.62	89.80	91.09	93.88	92.16	95.92	92.41	95.92	95.00	96.94
	AC input module fault	85.19	89.61	87.80	93.51	90.00	93.51	94.50	94.81	97.37	96.10
	Input/output module fault	88.60	91.82	90.27	92.73	92.04	94.55	100.00	93.64	96.36	96.36
	Power module fault	100.00	95.38	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Sampling out-of-step	94.59	85.37	100.00	87.80	97.37	90.24	100.00	95.12	97.56	97.56
	Configuration error	100.00	94.74	100.00	96.05	100.00	94.74	91.84	98.68	100.00	98.68
Protective relay	CPU module error/exception	96.08	96.08	98.00	96.08	98.00	96.08	100.00	98.04	100.00	98.04
	SV module fault	83.70	90.59	91.86	92.94	94.05	92.94	98.23	95.29	100.00	96.47
	GOOSE module fault	92.59	85.47	94.83	94.02	94.74	92.31	100.00	94.87	97.44	97.44
	Power fault of the power module	100.00	98.37	100.00	100.00	100.00	100.00	97.76	100.00	100.00	100.00
	Longitudinal channel fault	98.40	91.79	98.47	96.27	98.48	97.01	98.18	97.76	97.76	97.76
	Configuration error	96.23	94.44	100.00	94.44	98.11	96.30	88.57	100.00	100.00	100.00
Intelligent terminal	CPU module fault	92.59	92.59	96.36	98.15	98.18	100.00	96.97	100.00	98.15	98.15
	I/O module fault	85.92	90.37	89.51	94.81	93.57	97.04	100.00	94.81	96.32	97.04
	Power fault of the power module	95.04	98.53	99.27	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Configuration error	99.06	94.59	100.00	97.30	100.00	98.20	100.00	97.30	100.00	98.20
Circuit breaker	Control loop fault	100.00	100.00	100.00	100.00	100.00	100.00	97.79	100.00	100.00	100.00
	Closing coil fault	93.38	90.71	97.08	95.00	97.12	96.43	92.96	95.00	97.86	97.86
	Operating mechanism fault	81.17	89.29	87.84	93.53	91.72	95.68	96.95	94.96	97.12	97.12
	Transmission mechanism fault	89.76	87.69	92.97	91.54	96.83	94.57	100.00	97.69	99.23	99.23
	Contact fault	100.00	89.06	100.00	93.75	98.41	96.88	91.96	96.88	100.00	100.00
Switch	Communication link fault	82.35	90.74	85.22	90.74	89.29	92.59	96.24	95.37	95.50	98.15
	Port fault	92.06	86.57	92.31	89.55	93.98	93.28	100.00	95.52	98.48	97.01
	Communication packet loss	100.00	97.14	100.00	97.17	100.00	97.14	100.00	99.05	100.00	99.05
	Power fault	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 12. Verification results of communication faults.

Communication Links	Fault Types	[6,25]		Random Forest		[26]		[24]		This Paper	
		UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
Merging unit–Protective relay	Merging unit output port fault	87.13	91.67	87.88	90.63	91.92	94.79	91.14	93.75	92.08	96.88
	Protective relay input port fault	87.84	85.53	90.79	90.79	91.03	93.42	97.17	94.74	96.05	96.05
	Communication optical fiber fault	95.05	88.07	96.23	92.73	99.04	94.50	94.29	94.50	99.06	96.33
Protective relay–Intelligent terminal	Protective relay output port fault	77.63	89.39	88.41	92.42	89.86	93.94	91.86	93.94	95.52	96.97
	Intelligent terminal input port fault	86.05	88.10	91.76	92.86	91.86	94.05	91.23	94.05	94.19	96.43
	Communication optical fiber fault	80.36	83.33	94.12	88.89	94.34	92.59	98.18	96.30	98.11	96.30
Bus merging unit–Line merging unit	Bus merging unit output port fault	88.57	89.42	93.33	94.23	91.92	94.23	94.85	95.19	95.24	95.24
	Line merging unit input port fault	89.69	90.63	93.81	94.79	91.03	95.83	95.37	95.83	95.88	95.88
	Communication optical fiber fault	92.59	91.74	95.37	94.50	99.04	96.33	100.00	94.50	97.22	97.22
Intelligent terminal–Circuit breaker	Intelligent terminal output port fault	100	94.83	100	96.55	100	97.41	100.00	100.00	100	100

The average computing times of the validation process for the Decision Tree, Random Forest, GDBT, RNN, and the proposed method are 7.5 ms, 23.4 ms, 24.1 ms, 72.6 ms, and 22.7 ms, respectively. Their highest overall accuracies are 91.60%, 94.58%, 95.84%, 96.54%, and 97.83%; the corresponding Kappa coefficients are 91.32%, 94.40%, 95.70%, 96.43%, and 97.79%, respectively. Based on the overall performance, the improved Random Forest has high accuracy and computational efficiency, making it suitable for real-time systems. Moreover, the improved Random Forest demonstrates superior user accuracy and producer accuracy in classifying different fault types. Producer accuracy is the ratio of correctly classified samples to all samples belonging to a category. In contrast, user accuracy is the ratio of correctly classified samples to all samples classified as that category. Therefore, the enhanced scheme presented in this paper effectively reduces the probability of the misclassification of fault types.

Some faults occur infrequently, resulting in an imbalance in sample sets, with certain fault types classified as minority or majority classes. For example, the sampling out of step of the merging unit, the CPU module error/exception and the configuration error of the PR, the control loop fault and the contact fault of the CB, and the communication optical fiber fault between the protection device–intelligent terminal are categorized as minority class fault types. The comparison demonstrates that the proposed method achieves high *UA* and *PA* in classifying minority classes. It supports the claim that the proposed method exhibits excellent classification accuracy when dealing with fault tracking in unbalanced data. While the other methods achieve 100% *UA* in the minority class, this merely indicates that the other faults are not misclassified as the minority class. Furthermore, the lower *PA* of the other methods in the minority class suggests that more minority class faults are misclassified as other majority classes, resulting in a decrease in the *UA* of those majority classes and ultimately lowering the overall classification performance.

When dealing with imbalanced classification problems, relying solely on accuracy and recall metrics is insufficient to evaluate the performance of the classifier due to the influence of the majority class. Therefore, this paper incorporates Kappa coefficients as voting weights to assess the consistency between predicted and actual classes in the classification results and to evaluate the classification bias of the decision trees. By employing a weighted average strategy based on the Kappa coefficient, the results of decision trees with more balanced classification performance carry greater significance, ensuring that the final voting outcomes of the improved Random Forest exhibit high classification consistency and low classification bias. Additionally, the Re-Relief F-based feature selection algorithm measures the distance ratio between each sample and its nearest neighboring samples to determine the importance of each feature for classification. Even in the presence of class imbalance within the training set, the Re-Relief F algorithm selects features that effectively differentiate between minority and majority classes, assigning them higher feature weights. Since features with higher weights are more likely to be selected as node features, the ability of the Random Forest to distinguish between minority and majority classes is enhanced.

Based on the above analysis, the approach proposed in this paper demonstrates strong performance in terms of classification results both overall and for each fault type. Additionally, it ensures a high accuracy rate for the classification of minority classes within unbalanced data.

4.2.5. Comparison of Feature Selection Methods

In this paper, the Re-Relief F algorithm is utilized to evaluate feature importance in classification, with the objective of eliminating redundant features and enhancing classification accuracy. To assess the impact of different feature selection methods on the classification performance of fault tracking models, the Spearman correlation coefficient [31], Kendall correlation coefficient [32], information gain (InfoGain) [33], and maximum correlation minimum redundancy (mRMR) [34] are selected for conducting comparative experiments. Among them, Spearman and Kendall correlation coefficients measure the importance of features for classification by calculating the correlation between categories and features; InfoGain uses information entropy to describe the degree of influence of features on the classification results; and mRMR searches for the combination of features that have the maximum correlation with the categories and the minimum redundancy among them.

The experimental procedure for comparing feature selection methods is as follows:

First, the feature weights of the alarm features are calculated according to the definitions of Spearman, Kendall, InfoGain, and mRMR, respectively. Then, these weights are incorporated into the training process of the Random Forest. The features in the training subset are divided into three corresponding high, medium, and low feature subsets (W_h , W_m , and W_l) based on their corresponding feature weights. The same number of features from the three subsets are randomly selected as the candidate features of nodes. Finally, each feature selection method is incorporated into the Random Forest algorithm and repeated five times for training. The validation set is used to verify and compare the

effects of different feature selection algorithms when combined with Random Forest. The experimental results are presented in Table 13.

Table 13. Results of different feature selection algorithms.

No	Random Forest		Spearman		Kendall		InfoGain		mRMR		This Paper	
	OA (%)	Kappa (%)	OA (%)	Kappa (%)	OA (%)	Kappa (%)	OA (%)	Kappa (%)	OA (%)	Kappa (%)	OA (%)	Kappa (%)
1	93.37	93.18	94.34	94.15	95.12	94.96	95.42	95.26	96.08	95.95	96.81	96.70
2	94.01	93.81	94.67	94.49	94.79	94.61	95.66	95.53	96.02	95.91	97.83	97.66
3	94.58	94.40	94.52	94.33	95.30	95.47	95.60	95.45	96.23	96.10	97.08	96.98
4	93.49	93.01	94.40	94.21	94.91	94.74	95.81	95.77	96.11	95.97	97.20	97.10
5	93.37	93.18	94.46	94.27	95.00	94.84	95.87	95.73	95.87	95.75	97.35	97.23

The Spearman and Kendall correlation coefficients measure the extent to which alarm features contribute to classification by assessing the correlation between alarm features and fault types. A comparison reveals that feature selection algorithms based on Spearman and Kendall coefficients enhance both classification stability and accuracy to a certain degree. Notably, the Kendall-based feature selection method demonstrates the best performance, achieving an OA of 95.30% and a Kappa of 95.47% when combined with the Random Forest.

There is a correlation between fault types and alarm features in relay protection systems and CBs. Moreover, duplication exists among features of different fault types. InfoGain calculates the information gain for each alarm feature that corresponds to a specific fault type. The higher the gain, the greater the impact on classification. Additionally, mRMR comprehensively considers the correlation between alarm features and fault types, as well as the redundancy within alarm features. In comparison with the correlation coefficient, InfoGain and mRMR have a more prominent effect on the classification results.

Both the feature selection algorithm mentioned earlier and the proposed method enhance the stability and accuracy of classification compared with the original Random Forest. Additionally, the proposed method outperforms the other methods in terms of OA and Kappa, demonstrating higher classification accuracy and consistency. The Re-Relief F-based feature selection algorithm accentuates features with strong intra-class proximity and significant inter-class differentials. This refined feature subset enables a more accurate representation of the variations in fault types and their distances. In contrast to other feature selection methods, the Re-Relief F-based approach proves to be more effective in improving fault tracking outcomes for relay protection systems and CBs.

In addition, the feature selection method is a sample preprocessing method employed before model training and does not participate in the process of training and running the fault diagnostic model, so it does not affect the complexity and computational efficiency of the proposed method.

4.2.6. Fault Tolerance in Case of Unreliable Alarms

The false and missing alarm message can result in errors and the loss of features in the alarm feature set, thereby reducing the accuracy of the fault tracking model. To assess the fault tolerance of the proposed method in handling unreliable alarms, a specific number of alarm features in the validation set are chosen to undergo an inverse operation. This operation involves setting the original “0” feature to “1” or the original “1” feature to “0”, simulating false and missed alarms in the alarm messages.

The verification process is as follows:

1. An alarm subset with non-zero feature values is selected from the alarm feature set of each validation set sample. Then, a certain number (u) of alarm features with a value of “0” are randomly chosen, and their values are changed to “1”. This process simulates the occurrence of false alarms.

2. A certain number (v) of alarm features with a value of “1” are randomly chosen from the alarm feature set of each validation set sample, and their values are changed to “0”. This process simulates the occurrence of missed alarms.
3. The above methods are also utilized in the training and test sets to simulate false alarms and missed alarms.
4. In this paper, the values of u and v are set in the range of [1, 2, 3] to simulate the occurrence of false alarms and missed alarms for 1 to 3 alarm messages. Each value of u and v is simulated five times to simulate the occurrence of false alarms and missed alarms for different alarm features.
5. The processed training and test sets are used to train the fault tracking model, and the processed validation set is fed into the fault tracking model to compare the classification results in the case of false alarms and missed alarms. The *OA* and *Kappa* of the classification results are counted, as shown in Table 14.

Table 14. Results in case of unreliable alarms.

No	$u = 1 (v = 0)$		$u = 2 (v = 0)$		$u = 3 (v = 0)$		$v = 1 (u = 0)$		$v = 2 (u = 0)$		$v = 3 (u = 0)$	
	<i>OA</i> (%)	<i>Kappa</i> (%)	<i>OA</i> (%)	<i>Kappa</i> (%)	<i>OA</i> (%)	<i>Kappa</i> (%)	<i>OA</i> (%)	<i>Kappa</i> (%)	<i>OA</i> (%)	<i>Kappa</i> (%)	<i>OA</i> (%)	<i>Kappa</i> (%)
1	97.20	97.11	96.14	96.02	95.87	95.74	96.81	96.71	95.36	95.21	94.10	93.91
2	97.35	97.26	96.63	96.52	95.78	95.65	96.90	96.80	95.24	95.10	94.52	94.34
3	96.90	96.80	96.42	96.30	96.08	95.96	96.66	96.55	95.39	95.24	94.01	93.81
4	97.05	96.95	96.20	96.08	95.84	95.71	96.63	96.52	95.45	95.31	93.89	93.69
5	96.99	96.89	96.48	96.36	95.51	95.37	96.72	96.61	95.63	95.49	94.22	94.03

According to Table 14, despite the false and missing alarm messages, the proposed method achieves a high overall accuracy and Kappa coefficient. This suggests that the method has strong anti-interference and fault-tolerant capabilities. By comparing the classification results with the same number of false and missing alarm messages, it is found that missing alarms have a greater impact on the classification results. This phenomenon is caused by the following two reasons: firstly, missing alarm messages may result in crucial fault features being missed, preventing the fault tracking model from accurately distinguishing between fault types; secondly, some fault types have fewer alarm features, which further increases the missing probability of important alarm features.

Due to the unreliability of alarm messages, there may be some noises or missing alarm features. In this paper, the Re-Relief F-based feature selection method is employed to filter out the features with stronger relevance to the classification results and reduce the interference of noise and redundant features. Randomly selecting features for training can reduce the impact of missing or wrong features on a single decision tree. Even if a feature is missing or wrong, it will only affect part of the decision tree, not all of the decision tree.

This paper employs weighted voting based on Kappa coefficients to assign higher weights to decision trees that demonstrate greater reliability, thereby reducing the impact of missing and erroneous alarm features on the final classification results. Moreover, the result of the improved Random Forest is determined through weighted voting based on the results of multiple decision trees, which can make up for the incorrect classification of certain decision trees, thus improving the robustness of the model.

4.3. Case Analysis of Complex Faults

This section provides a detailed analysis of the fault tracing process based on the improved Random Forest in various fault cases. The fault cases contain communication faults in protection devices and intelligent terminals and double faults in neighboring and non-neighboring devices. Through this case study, it can be proved that the proposed method can accurately trace the complex faults of relay protection systems and CBs and identify the causes of incorrect operation of PRs and CBs.

The fault case takes a 220 kV substation in Shandong as an example, and its main wiring is shown in Figure 9. Among them, 220 kV double buses (B_{11} and B_{12}) connect two outgoing lines (L_{11} and L_{12}), two 110 kV double buses (B_{21} and B_{22} , B_{31} and B_{32}) connect four outgoing lines (L_{21} and L_{22} , L_{31} and L_{32}), and two transformers (T_1 and T_2) are connected between the buses.

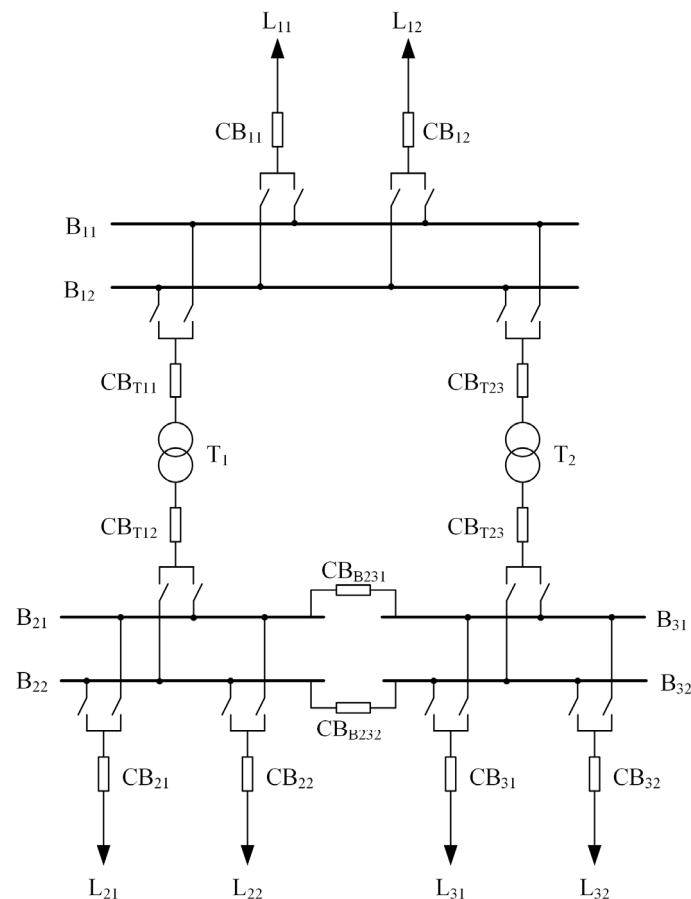


Figure 9. Main wiring diagram of a 220 kV substation in Shandong, China.

4.3.1. Case 1

Case 1 is a real fault case of a 220 kV substation with a line fault and CB rejection. The fault scenario is as follows: a two-phase ground fault occurs on the 110 kV line L_{22} in Figure 9, and the main protection L_{p22m} detects the fault and sends out the tripping signal. However, because of the output port fault of the protection device, the CB_{22} does not receive the tripping signal and fails to trip. Therefore, CB failure protection L_{p22f} is used as a backup protection to trip CB_{21} , CB_{T12} , CB_{B231} , and CB_{B232} to isolate the faulty line L_{22} .

The alarm messages of the PRs and CBs are shown in Table 15. The alarm messages of the secondary system of the substation are shown in Table 16.

According to Table 15, L_{22} is diagnosed as a faulty component. The corresponding main and near backup protections for CB_{22} are L_{p22m} and L_{p22p} , and the CB failure protection and corresponding CBs are L_{p22f} , CB_{T12} , CB_{B231} , and CB_{B232} . The operation states of the main, near backup, and remote backup PRs and their corresponding CBs are (1, 0), (0, 0), and (1, 1). Their corresponding fusion confidence degrees are 0.5957, 0.2, and 0.72. According to the operation evaluation rules, it is concluded that L_{p22m} , L_{p22f} , CB_{21} , CB_{T12} , CB_{B231} , and CB_{B232} are normal, while the CB_{22} fails to trip.

The alarm features of the relay protection system and CB are extracted from the secondary system alarm information in Table 16, as shown in Table 17. Alarm messages

are issued by adjacent devices (PRs and intelligent terminals) at the same time. Based on the fault tracking process in Figure 3, the alarm feature subset $X_1 = \{X_{GOOSE1}, X_{SC1}, X_{COM1}\}$ of the PRs and intelligent terminals is divided from alarm feature set X according to the adjacent devices. The corresponding element in the alarm feature subset X_1 of the alarm messages in Table 9 is one, and the rest are zero, as shown in Equations (18)–(20) (due to the large dimension of the alarm feature set, only the alarm feature set is shown with non-zero elements).

$$\begin{cases} X_{GOOSE1} = \{X_{GOOSE_IT1}\} \\ X_{GOOSE_IT1} = \{1, 0, 1, 0, \dots, 0\} \end{cases} \quad (18)$$

$$\begin{cases} X_{SC1} = \{X_{SC_PR1}\} \\ X_{SC_PR1} = \{1, 0, 0, \dots, 1, 0, \dots, 0\} \end{cases} \quad (19)$$

$$\begin{cases} X_{COM1} = \{X_{COM_PR1}, X_{COM_IT1}\} \\ X_{COM_PR1} = \{0, 0, 1, 0, \dots, 0\} \\ X_{COM_IT1} = \{1, 0, 1, 0, \dots, 1, 1, 0, \dots, 0\} \end{cases} \quad (20)$$

Table 15. Alarm messages of PRs and CBs in Case 1.

Substation	Devices	Statuses
HZ Station	Line L ₂₂ protection Lp _{22m}	Operated
HZ Station	Line L ₂₁ failure protection Lp _{22f}	Operated
HZ Station	Line L ₂₁ Switch CB ₂₁	Tripped
HZ Station	Transformer T ₁ switch CB _{T12}	Tripped
HZ Station	Busbar B ₂₁ Switch CB _{B231}	Tripped
HZ Station	Busbar B ₂₂ Switch CB _{B232}	Tripped

Table 16. Alarm messages of the secondary system in Case 1.

Substation	Interval	Devices	Alarms
HZ Station	Interval 22	Line L ₂₂ Protection Lp _{22m}	Self-check alarm
HZ Station	Interval 22	Line L ₂₂ Protection Lp _{22m}	Exception operation
HZ Station	Interval 22	Line L ₂₂ Protection Lp _{22m}	Output communication interruption
HZ Station	Interval 22	Intelligent terminal	GOOSE total alarm
HZ Station	Interval 22	Intelligent terminal	GOOSE interruption
HZ Station	Interval 22	Intelligent terminal	No GOOSE input messages
HZ Station	Interval 22	Intelligent terminal	GOOSE communication interruption
HZ Station	Interval 22	Intelligent terminal	Input circuit self-check error
HZ Station	Interval 22	Intelligent terminal	Input exception

Table 17. Alarm features of Case 1.

No.	Feature Categories	Alarm Features
1	GOOSE Alarm	GOOSE total alarm of the intelligent terminal
2		GOOSE interruption of the intelligent terminal
3	Device self-check alarm	The self-check alarm of the protective relay
4		Exception operation of the protective relay
5	Communication alarm	Output communication interruption of the protective relay
6		No GOOSE input messages of the intelligent terminal
7		GOOSE communication interruption of the intelligent terminal
8		Input circuit self-check error of the intelligent terminal
9		Input exception of the intelligent terminal

The subset of alarm features X_1 is input into the fault tracking model based on the improved Random Forest, and the output is:

$$Y_1 = \{0, 1, 1, 1, 0, 1\} \tag{21}$$

In looking at Table 2, the fault type is the output port fault of the PR, and there is a communication fault in the fault tracing result. The alarm feature set X is empty after removing the alarm feature subset X_1 , and the fault tracing result is the output port fault of the PR. According to the fault tracing process in Figure 3, the CB refuses to trip without the CB fault, and a communication fault occurs in the relay protection system. Therefore, the CB rejection is caused by the output port fault of the PR.

In Case 1, the output port fault of the PR will cause the PR and the intelligent terminal to issue alarm messages. The fault tracking process collects alarm messages from adjacent devices that may have communication faults for fault tracking, which is conducive to distinguishing device faults and communication faults.

4.3.2. Case 2

The adjacent device faults or communication faults in relay protection systems and CBs may cause adjacent devices to issue alarm messages. Based on Case 1, this paper simulates the simultaneous occurrence of an adjacent device fault and a communication fault. It proves that the proposed method can distinguish between adjacent device faults and communication faults. The details of Case 2 are as follows: a fault occurs at L_{22} , and the DSP module of the merging unit and output port fail at the same time, resulting in L_{p22m} rejection. By the time the near backup protection (L_{p22p}) operates, CB_{22} starts tripping and isolates the fault line L_{22} .

The alarm messages of PRs and CBs in Case 2, based on the typical monitoring information table of the 220 kV substation, are presented in Table 18. The alarm messages of the DSP module fault are fused with Case 1 to obtain the alarm messages of the secondary system in Case 2, as shown in Table 19.

Table 18. Alarm messages of PRs and CBs in Case 2.

Substation	Devices	Statuses
HZ Station	Line L_{22} Protection L_{p22p}	Operated
HZ Station	Line L_{21} Switch CB_{22}	Tripped

Table 19. Alarm messages of the secondary system in Case 2.

Substation	Interval	Devices	Alarms
HZ Station	Interval 22	Merging unit	Total SV alarm
HZ Station	Interval 22	Merging unit	Sampling exception
HZ Station	Interval 22	Merging unit	Sampling exception
HZ Station	Interval 22	Merging unit	Device exception
HZ Station	Interval 22	Line L_{22} Protection L_{p22m}	Self-check alarm
HZ Station	Interval 22	Line L_{22} Protection L_{p22m}	Exception operation
HZ Station	Interval 22	Line L_{22} Protection L_{p22m}	Output communication interruption
HZ Station	Interval 22	Intelligent terminal	GOOSE total alarm
HZ Station	Interval 22	Intelligent terminal	GOOSE interruption
HZ Station	Interval 22	Intelligent terminal	No GOOSE input messages
HZ Station	Interval 22	Intelligent terminal	GOOSE communication interruption
HZ Station	Interval 22	Intelligent terminal	Input circuit self-check error
HZ Station	Interval 22	Intelligent terminal	Input exception

According to Table 18, the operation states of the main, near backup, and remote backup PRs and their corresponding CBs are (0, 1), (1, 1), and (0, 0). Their corresponding fusion confidence degrees are 0.5917, 0.825, and 0.2. According to the operation evaluation rules, it is concluded that Lp_{22p} and CB_{22} are normal, while Lp_{22m} fails to operate.

The alarm features of the relay protection system and the CB are extracted from the secondary system alarm messages in Table 19, as shown in Table 20. Alarm messages are issued by adjacent devices (merging unit with PR and PR with intelligent terminal) simultaneously. Based on the fault tracking process of Figure 3, the alarm feature subset $X_1 = \{X_{GOOSE1}, X_{SC1}, X_{COM1}\}$ (Equations (18)–(20)) of the PR and the intelligent terminal, and the alarm feature subset $X_2 = \{X_{SV2}, X_{SC2}, X_{COM2}\}$ of the merging unit and the PR are divided according to the adjacent device, as shown in Equations (22)–(24).

$$\begin{cases} X_{SV2} = \{X_{SV_MU2}\} \\ X_{SV_MU2} = \{1, 1, 1, 0, \dots, 0\} \end{cases} \quad (22)$$

$$\begin{cases} X_{SC2} = \{X_{SC_MU2}, X_{SC_PR2}\} \\ X_{SC_MU2} = \{1, 0, 0, \dots, 1, \dots, 0\} \\ X_{SC_PR2} = \{1, 0, 1, \dots, 1, \dots, 0\} \end{cases} \quad (23)$$

$$\begin{cases} X_{COM2} = \{X_{COM_MU2}, X_{COM_PR2}\} \\ X_{COM_MU2} = \{1, 1, 0, \dots, 0\} \\ X_{COM_PR2} = \{0, 0, 1, 0, \dots, 0\} \end{cases} \quad (24)$$

Table 20. Alarm features of Case 2.

No.	Feature Categories	Alarm Features
1	SV alarm	Total SV alarm of the merging unit
2		Sampling exception of the merging unit
3		Synchronization exception of the merging unit
4	GOOSE alarm	Intelligent terminal GOOSE total alarm
5		Intelligent terminal GOOSE interruption
6	Device self-check alarm	Device exception of the merging unit
7		The self-check alarm of the merging unit
8		The self-check alarm of the protective relay
9		Exception operation of the protective device
10		Output communication interruption of the merging unit
11	Communication alarm	SV communication interruption of the merging unit
12		Output communication interruption of the protective relay
13		No input GOOSE messages of the intelligent terminal
14		Communication interruption of the intelligent terminal GOOSE
15		Input circuit self-check error of the intelligent terminal
16		Input exception of the intelligent terminal

The alarm feature subsets X_1 and X_2 of adjacent devices are input into the fault tracking model based on the improved Random Forest, and the output is as follows:

$$\begin{cases} Y_1 = \{0, 1, 1, 1, 0, 1\} \\ Y_2 = \{0, 0, 0, 0, 0, 1\} \end{cases} \quad (25)$$

In looking at Tables 1 and 2, the fault types are the DSP module fault of the merging unit and the output port fault of the protective relay.

There is a communication fault in the fault tracking results. The alarm feature set X is not empty after excluding the alarm feature subset X_1 . Firstly, the communication fault is

the output port fault of the PR, and then the alarm feature subset $X_3 = \{X_{SV3}, X_{SC3}, X_{COM3}\}$ of the merging unit is divided according to a single device.

$$\begin{cases} X_{SV3} = \{X_{SV_MU3}\} \\ X_{SV_MU3} = \{1, 1, 1, 0, \dots, 0\} \end{cases} \quad (26)$$

$$\begin{cases} X_{SC3} = \{X_{SC_MU3}, X_{SC_PR3}\} \\ X_{SC_MU3} = \{1, 0, 0, \dots, 1, \dots, 0\} \end{cases} \quad (27)$$

$$\begin{cases} X_{COM3} = \{X_{COM_MU3}\} \\ X_{COM_MU3} = \{1, 1, 0, \dots, 0\} \end{cases} \quad (28)$$

The alarm feature subset X_3 is input into the fault tracking model based on the improved Random Forest, and the output is as follows:

$$Y_3 = \{0, 0, 0, 0, 0, 1\} \quad (29)$$

In looking at Table 1, the fault type is the DSP module fault of the merging unit, consistent with the results of the subset X_2 . The result of the fault tracking is the output port fault of the PR and the DSP module fault of the merging unit. According to the fault tracking process in Figure 3, the PR rejection is caused by the DSP module fault of the merging unit.

During the fault tracking process, the alarm feature subset of adjacent devices is input into the fault tracking model to determine if there are any communication faults. The alarm feature set is re-divided according to a single device for fault tracking after eliminating the alarm feature subset of the communication fault. After two divisions of the alarm feature set and fault tracking, the device faults and communication faults are distinguished. Through the verification of Case 2, the fault tracking process does not incorrectly divide the alarm features of the device and communication faults. It can accurately track the complex scenario of device and communication faults.

4.3.3. Case 3

Case 3 simulates a complex fault scenario in which both the PR and the CB refuse to operate, and faults in non-adjacent devices occur in the relay protection system. The details of Case 3 are as follows: a fault occurs in L_{22} , and the DSP module of the merging unit fails, resulting in the L_{p22m} rejection. The L_{p22p} operates, but due to the configuration error of the intelligent terminal, the CB_{22} does not receive the trip signal to refuse to trip. Therefore, CB failure protection L_{p22f} is operated, and CB_{21} , CB_{T12} , CB_{B231} , and CB_{B232} trip and isolate the faulty line L_{22} .

The alarm messages of PRs and CBs in Case 3, based on the typical monitoring information table of the 220 kV substation, are presented in Table 21. By fusing the alarm messages of the DSP module fault of the merging unit and the configuration error of the intelligent terminal, the alarm messages of the secondary system of the Case 3 substation are shown in Table 22.

Table 21. Alarm messages of PRs and CBs in Case 3.

Substation	Devices	Statuses
HZ Station	Line L_{22} Protection L_{p22p}	Operated
HZ Station	Line L_{21} failure protection L_{p22f}	Operated
HZ Station	Line L_{21} Switch CB_{21}	Tripped
HZ Station	Transformer T_1 switch CB_{T12}	Tripped
HZ Station	Busbar B_{21} Switch CB_{B231}	Tripped
HZ Station	Busbar B_{22} Switch CB_{B232}	Tripped

Table 22. Alarm messages of the secondary system in Case 3.

Substation	Interval	Devices	Alarms
HZ Station	Interval 22	Merging unit	Total SV alarm
HZ Station	Interval 22	Merging unit	Sampling exception
HZ Station	Interval 22	Merging unit	Sampling exception
HZ Station	Interval 22	Merging unit	Device exception
HZ Station	Interval 22	Intelligent terminal	GOOSE total alarm
HZ Station	Interval 22	Intelligent terminal	GOOSE interruption
HZ Station	Interval 22	Intelligent terminal	No GOOSE input messages
HZ Station	Interval 22	Intelligent terminal	GOOSE communication interruption
HZ Station	Interval 22	Intelligent terminal	Input circuit self-check error
HZ Station	Interval 22	Intelligent terminal	Input exception

According to Table 21, the operation states of the main, near backup, and remote backup PRs and their corresponding CBs are (0, 0), (1, 0), and (1, 1). Their corresponding fusion confidence degrees are 0.2, 0.5, and 0.725. According to the operation evaluation rules, it is concluded that L_{p22p} , L_{p22f} , CB_{21} , CB_{T12} , CB_{B231} , and CB_{B232} are all normal, while L_{p22m} and CB_{22} fail to operate.

The alarm features of the relay protection system and CB are extracted from the secondary system alarm messages in Table 22, as shown in Table 23. There are no alarm messages issued by neighboring devices at the same time in the alarm messages. According to the fault tracking process of Figure 3, the alarm feature subset $X_3 = \{X_{SV3}, X_{SC3}, X_{COM3}\}$ of the merging unit (such as Equations (26)–(28)) and the alarm feature subset $X_4 = \{X_{GOOSE4}, X_{COM4}\}$ of the intelligent terminal are divided according to a single device, as shown in Equations (30) and (31).

$$\begin{cases} X_{GOOSE4} = \{X_{GOOSE_IT4}\} \\ X_{GOOSE_IT4} = \{0, 0, 0, \dots, 1, \dots, 0\} \end{cases} \quad (30)$$

$$\begin{cases} X_{COM4} = \{X_{COM_IT4}\} \\ X_{COM_IT4} = \{0, 0, 0, \dots, 1, \dots, 1, \dots, 0\} \end{cases} \quad (31)$$

Table 23. Alarm features of Case 3.

No.	Alarm Messages	
1	SV alarm	Total alarm of the merging unit
2		Sampling exception of the merging unit
3		Synchronization exception of the merging unit
4	GOOSE alarm	GOOSE configuration error of the intelligent terminal
5	Device self-check alarm	Device exception of the merging unit
6		The self-check alarm of the merging unit
7	Communication alarm	Output communication interruption of the merging unit
8		SV communication interruption of the merging unit
9		GOOSE panel configuration error of the intelligent terminal
10		Input configuration error of the intelligent terminal

The alarm feature subsets X_3 and X_4 of a single device are input into the fault tracking model based on improved Random Forest, and the output is as follows:

$$\begin{cases} Y_3 = \{0, 0, 0, 0, 0, 1\} \\ Y_4 = \{0, 1, 0, 0, 0, 0\} \end{cases} \quad (32)$$

In Table 1, the fault types are the faulty DSP module of the merging unit and the misconfiguration of the intelligent terminal, which are identified as fault tracing results. According to the fault tracing process in Figure 3, the protection rejection is caused by the fault of the DSP module of the merging unit, and the protection rejection causes the CB rejection.

Since the fault tracing model is designed as a single-input and single-output multi-classification model, directly using alarm messages of complex faults as input does not enable simultaneous identification of all fault types. Developing separate fault tracking models for individual devices and communication faults results in multiple outputs and inaccuracy results. In Case 3, the fault tracking process divides the alarm feature subset based on individual devices, breaks down complex faults into multiple simpler faults for tracking, and simultaneously determines the fault types of multiple faults occurring at the same time. This study utilizes single-input and single-output fault tracking models to achieve multi-label classification results, effectively addressing the challenge of achieving complex fault tracking with a single model and mitigating the inaccuracies associated with multiple models.

4.3.4. Scalability Analysis

Based on the fault tracing process in the above case study, it is evident that the fault tracing in this paper consists of two parts: firstly, the identification of improperly operated PRs and CBs, and secondly, the determination of fault types within the relay protection systems and CBs associated with them.

The operation valuation rules are established based on the fundamental configurations of main, near backup, and remote backup PRs. In Section 4.1, it is demonstrated that these rules are applicable to both the IEEE 39-bus and 118-bus power systems. They effectively identify incorrectly operated PRs and CBs in fault cases, thereby narrowing down the fault tracing process to the corresponding substation intervals. This paper proposes a fault tracing model and process for complex faults occurring in relay protection systems and CBs. The alarm feature set is divided based on individual devices and neighboring devices, allowing for the separation of complex faults into simple device faults and communication faults. The above case study validates the feasibility of this division approach. Other types of relay protection systems, such as busbar protection and transformer protection, also incorporate essential components like merging units, protection devices, and intelligent terminals within an intelligent substation. Consequently, these relay protection systems can be further broken down into individual or adjacent devices for fault tracing.

In summary, this paper utilizes operation evaluation rules and a fault tracing process to gradually convert the fault tracing problem into a fault categorization problem for basic devices. It demonstrates that the proposed method is not only applicable to larger and more complex power systems but also has good scalability for the fault classification of different types of relay protection systems and CBs.

5. Conclusions

This paper presents a fault tracking method for the RPS-CB based on an improved Random Forest algorithm. Firstly, the reasons for the incorrect operation of PRs and CBs are analyzed, the fault types of a relay protection system–circuit breaker and corresponding alarm messages are demonstrated, and the method of characterizing the alarm messages is provided; secondly, the feature selection and voting strategy of the Random Forest algorithm is improved, and the fault tracking model based on the improved Random Forest is trained; then, the incorrectly operated PRs and CBs are identified through the operation evaluation; finally, the fault tracking model and the process are utilized to derive the reasons for the faults of the relay protection system and the CB and determine the case of incorrect operation. The main contributions of this paper contain the following points:

1. A fault tracking model based on an improved Random Forest is constructed. This model combines the Re-Relief F algorithm and the weighted voting strategy. The

feature selection is optimized to ensure the stable classification performance of the Random Forest. Higher weights are assigned to the decision trees with strong classification performance, improving overall classification performance and accuracy.

2. An operation evaluation process is proposed for identifying PR and CB that are incorrectly operated during complex faults, providing a target for fault tracking. The complex faults verify the evaluation results, and it is confirmed that the evaluation results are still reliable in the case of missing or false alarms.
3. A fault tracking process based on improved Random Forest is developed to realize accurate tracking of complex faults such as device faults, communication faults and multiple faults of the RPS-CB by dividing the alarm feature set and complex faults into multiple simple faults.

After case validation and analysis, it is proved that the proposed method achieves the fault tracing of the incorrect operations of PRs and CBs. This work not only provides an important supplement to the fault diagnosis but also provides accurate and effective references and suggestions, which helps to realize accurate operation and maintenance.

While fault tracing can accurately pinpoint faults in relay protection systems and circuit breakers, it cannot prevent or anticipate them. Thus, our future research aims to employ device operational data to evaluate the performance of the relay protection system and CBs, predicting probable device failures and promptly alerting maintenance personnel to perform necessary repairs or replacements. This approach aims to enhance the reliability of both PRs and CBs.

Author Contributions: Conceptualization, N.S. and Q.C.; data curation, N.S. and C.Y.; formal analysis, N.S., D.X. and Y.S.; funding acquisition, Q.C.; methodology, N.S., C.Y. and D.X.; project administration, Q.C.; software, N.S.; validation, N.S. and Q.C.; visualization, N.S. and Y.S.; writing original draft, N.S.; writing review and editing, N.S., C.Y., D.X. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 51877123.

Data Availability Statement: The data presented in this study are available in the article.

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

References

1. Alvarez-Alvarado, M.S.; Donaldson, D.L.; Recalde, A.A.; Noriega, H.H.; Khan, Z.A.; Velasquez, W.; Rodriguez-Gallegos, C.D. Power System Reliability and Maintenance Evolution: A Critical Review and Future Perspectives. *IEEE Access* **2022**, *10*, 51922–51950. [[CrossRef](#)]
2. Wang, S.-P.; Zhao, D.-M. A Hierarchical Power Grid Fault Diagnosis Method Using Multi-Source Information. *IEEE Trans. Smart Grid* **2020**, *11*, 2067–2079. [[CrossRef](#)]
3. Goudarzi, A.; Ghayoor, F.; Waseem, M.; Fahad, S.; Traore, I. A Survey on IoT-Enabled Smart Grids: Emerging, Applications, Challenges, and Outlook. *Energies* **2022**, *15*, 6984. [[CrossRef](#)]
4. Waseem, M.; Adnan Khan, M.; Goudarzi, A.; Fahad, S.; Sajjad, I.A.; Siano, P. Incorporation of Blockchain Technology for Different Smart Grid Applications: Architecture, Prospects, and Challenges. *Energies* **2023**, *16*, 820. [[CrossRef](#)]
5. Dong, X.Z.; Wang, D.L.; Zhao, M.Y.; Wang, B.; Shi, S.X.; Apostolov, A. Smart Power Substation Development in China. *Csee J. Power Energy* **2016**, *2*, 1–5. [[CrossRef](#)]
6. Wang, L.; Chen, Q.; Gao, H.; Ma, Z.; Zhang, Y.; He, D. Framework of Fault Trace for Smart Substation Based on Big Data Mining Technology. *Autom. Electr. Power Syst.* **2018**, *42*, 84–91.
7. Moradzadeh, A.; Mohammadi-Ivatloo, B.; Pourhossein, K.; Anvari-Moghaddam, A. Data Mining Applications to Fault Diagnosis in Power Electronic Systems: A Systematic Review. *IEEE Trans. Power Electron.* **2022**, *37*, 6026–6050. [[CrossRef](#)]
8. Zhang, Y.; Li, D.; Wen, F.; Zhang, Y.; Dong, Y.; Zheng, J. Process Level Fault Diagnosis in IEC 61850 Based Smart Substations. *Electr. Power Constr.* **2018**, *39*, 42–48.
9. Fang, X.; Huang, W.; Ye, D.; Huang, Y. Application of a distributed parallel FP-growth algorithm in secondary device defects monitoring. *Power Syst. Prot. Control* **2021**, *49*, 160–167.
10. Wang, M.; Li, T.; Ren, J.; Xu, Y. Realization of Fault Diagnosis of Substation Secondary System Using Apriori Algorithm. *Proc. CSU-EPSA* **2021**, *33*, 145–150.

11. Wang, M.; Xu, Y.; Fan, W.; Li, T.; Ren, J. Fault Correlation Analysis of Substation Secondary System Based on H-mine Algorithm. *J. N. China Electr. Power Univ.* **2021**, *48*, 36–44.
12. Xu, Y.; Wang, M.; Fan, W. Defect Data Association Analysis of the Secondary System Based on AFWA-H-Mine. *Energies* **2021**, *14*, 4228. [[CrossRef](#)]
13. Chen, G.; Dong, X.; Zheng, Y.; Xu, H. Fault diagnosis of a relay protection test based on a long short-term memory network. *Power Syst. Prot. Control* **2022**, *50*, 65–73.
14. Higgins, M. *Operational Moving Target Defences for Improved Power System Cyber-Physical Security*; University of Oxford: Oxford, UK, 2022.
15. Zhang, Y.; Cai, Z.; Long, P.; Li, X.; Su, Z. Real-Time Fault Diagnosing Models and Method for Communication Network in Smart Substation. *Power Syst. Technol.* **2016**, *40*, 1851–1857.
16. Sun, Y.; Cai, Z.; Guo, C.; Ma, G.; Dai, G. Fault Diagnosis and Positioning for Communication Network in Intelligent Substation Based on Deep Learning. *Power Syst. Technol.* **2019**, *43*, 4306–4314.
17. Ren, B.; Li, J.; Zheng, Y.; Chen, X.; Zhao, Y.; Zhang, H.; Zheng, C. Research on Fault Location of Process-Level Communication Networks in Smart Substation Based on Deep Neural Networks. *IEEE Access* **2020**, *8*, 109707–109718. [[CrossRef](#)]
18. Dai, Z.; Geng, H.; Han, J.; Li, J.; Fang, W. Fault Location Method of Secondary System in Smart Substation Based on Matrix Algorithm and BP Neural Network. *J. N. China Electr. Power Univ.* **2022**, *49*, 1–10.
19. Zhang, C.; Zheng, Y.; Lu, J.; Zhang, H.; Ren, H.; Yang, Z. Fault location of secondary circuits in a smart substation based on a graph neural network. *Power Syst. Prot. Control* **2022**, *50*, 81–90.
20. Zhang, X.; Chen, Q.; Sun, M.; Huang, W.; Wang, L.; Liu, B. Fault tracking of high-voltage circuit breakers in case of secondary circuit faults in intelligent substations. *Electr. Power Autom. Equip.* **2020**, *40*, 212–217+224.
21. Gao, X.; Ren, B.; Zhang, H.; Liu, M.; Li, J.; Xu, J. Component Fault Tracing of Power Dispatching Automation System Based on Information Difference Graph Model. *Power Syst. Technol.* **2021**, *45*, 4808–4817.
22. Ji, J.J.; Chen, Q.; Jin, L.; Zhou, X.T.; Ding, W. Fault Diagnosis System of Power Grid Based on Multi-Data Sources. *Appl. Sci.* **2021**, *11*, 7649. [[CrossRef](#)]
23. Wang, H.; Ma, Y.; Wu, Y.; Zhao, L.; Liu, T.; Li, X.; Wang, Y.; Zhang, B. Call chain monitoring and distributed tracking method for the communication bus of a power grid control system. *Power Syst. Prot. Control* **2021**, *49*, 29–37. [[CrossRef](#)]
24. Ren, B.; Zheng, Y.; Wang, Y.; Sheng, S.; Li, J.; Zhang, H.; Zheng, C. Fault Location of Secondary Equipment in Smart Substation Based on Deep Learning. *Power Syst. Technol.* **2021**, *45*, 713–721.
25. Cui, C.; Ren, W.; Feng, S.; Qu, L.; Liu, D.; Wang, Y. Remote Power Failure Tracing Based on Classification Decision Tree. In Proceedings of the 2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 26–28 May 2023; pp. 873–877.
26. Ding, W.; Chen, Q.; Dong, Y.; Shao, N. Fault Diagnosis Method of Intelligent Substation Protection System Based on Gradient Boosting Decision Tree. *Appl. Sci.* **2022**, *12*, 8989. [[CrossRef](#)]
27. Jiao, K. Research on Intelligent Substation Fault Diagnosis Technology Based on Relay Protection Test Technology. Master's Thesis, North China Electric Power University, Baoding, China, 2020.
28. Ren, B. Research on Fault Location of Secondary System in Smart Substation Based on Deep Learning. Master's Thesis, North China Electric Power University, Baoding, China, 2021.
29. Yang, Z.; Nurbol; Jia, X.; Hu, L. Intrusion Feature Selection Methods Based on ReliefF. *J. Jilin Univ. Sci. Ed.* **2015**, *53*, 505–510. [[CrossRef](#)]
30. Wang, T.; Wei, X.; Wang, J.; Huang, T.; Peng, H.; Song, X.; Valencia Cabrera, L.; Perez-Jimenez, M.J. A weighted corrective fuzzy reasoning spiking neural P system for fault diagnosis in power systems with variable topologies. *Eng. Appl. Artif. Intel.* **2020**, *92*, 103680. [[CrossRef](#)]
31. Liu, S.X.; Li, Y.K.; Gao, S.Y.; Xing, C.J.; Li, J.; Cao, Y.D. Prediction of Residual Electrical Life in Railway Relays Based on Convolutional Neural Network Bidirectional Long Short-Term Memory. *Energies* **2023**, *16*, 6357. [[CrossRef](#)]
32. Iqbal, T.; Elahi, A.; Wijns, W.; Amin, B.; Shahzad, A. Improved Stress Classification Using Automatic Feature Selection from Heart Rate and Respiratory Rate Time Signals. *Appl. Sci.* **2023**, *13*, 2950. [[CrossRef](#)]
33. Chen, Y.; Ma, L.X.; Yu, D.S.; Zhang, H.D.; Feng, K.Y.; Wang, X.; Song, J. Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests. *Ecol. Indic.* **2022**, *135*, 108545. [[CrossRef](#)]
34. Shaffie, A.; Soliman, A.; Eledkawy, A.; Fu, X.A.; Nantz, M.H.; Giridharan, G.; van Berkel, V.; El-Baz, A. Lung Cancer Diagnosis System Based on Volatile Organic Compounds (VOCs) Profile Measured in Exhaled Breath. *Appl. Sci.* **2022**, *12*, 7165. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.