

Article

Multi-Branch Line Fault Arc Detection Method Based on the Improved Northern Goshawk Optimization Adaptive Base Class LogitBoost Algorithm

Xue Wang and Yu Zhao * 

Department of Electrical Engineering, North China Electric Power University, Baoding 071003, China; wangxuedl@126.com

* Correspondence: 220212213060@ncepu.edu.cn

Abstract: In low-voltage AC distribution systems, when a series arc fault occurs in a branch with multiple loads operating in parallel, it will be significantly more difficult to identify. Existing arc fault detection methods make it difficult to effectively detect faults occurring in the lower-level branch. This study introduces a novel series arc fault detection approach based on the improved northern goshawk optimization adaptive base class LogitBoost (INGO-ABCLogitBoost) algorithm. Considering the zero-rest, intermittent, and random fluctuation and high-frequency features of the arc current, the zero-rest coefficient, discrete coefficient, harmonic amplitude, and wavelet entropy are proposed to establish the high-dimensional feature matrix of the arc current. The ReliefF feature selection algorithm is used to optimize feature quality and decrease feature dimensionality. Subsequently, the ABCLogitBoost fault detection model is proposed, with the INGO algorithm applied to optimize the model parameters, thus enhancing the model's diagnostic capabilities. The efficacy of the proposed diagnostic model is validated through the construction of a multi-load arc simulation system. The simulation results show that the overall fault diagnosis accuracy of the proposed method reaches 99.01% and can effectively identify the fault load types, which helps to locate the fault location.

Keywords: series fault arc; branch circuit fault; feature extraction; feature selection; ABCLogitBoost



Citation: Wang, X.; Zhao, Y.

Multi-Branch Line Fault Arc Detection Method Based on the Improved Northern Goshawk Optimization Adaptive Base Class LogitBoost Algorithm. *Energies* **2024**, *17*, 954. <https://doi.org/10.3390/en17040954>

Academic Editor: Santi A. Rizzo

Received: 17 January 2024

Revised: 10 February 2024

Accepted: 15 February 2024

Published: 19 February 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

With the rapid development of the electric power system, the household power distribution system presents a wide variety of electrical equipment and line connection points. For this reason, the rate of occurrence of electrical fires has greatly increased, causing significant economic losses and human casualties and seriously threatening the safety of people's lives and properties. During the use of domestic appliances, arc faults may be induced due to such problems as line aging, poor contact, external damage, and so on. In the discharge process of the arc, a large amount of heat will be emitted, further deteriorating the insulation performance of electrical equipment [1]. When series arc faults occur, the current is generally low, and the current waveforms of certain electrical appliances are similar in normal and faulty conditions. That makes it difficult to detect arc faults [2]. Numerous studies have shown that arcing faults in low-voltage distribution lines are the main cause of electrical fires. Therefore, the effective detection of series arc faults is of great practical significance.

In recent years, scholars from various countries have extensively discussed the field of arc faults from different perspectives. In terms of feature extraction of arc fault signals, the use of analytical methods based on time and frequency domains to process arc currents are the main means currently used. To name a few, reference [3] used improved complete ensemble empirical mode decomposition adaptive noise (ICEEMDAN) to obtain the IMF component of the signal and constructed detection variables to distinguish between arc fault and normal states. Reference [4] applied the entropy feature of the variational modal

decomposition (VMD) as the indicator for arc fault detection. Reference [5] utilized the correlation coefficient of the signal as the indicator for the detection of an arc with different loads. Reference [6] identified arc faults by comparing the peak or average values of the arc and normal currents. Reference [7] detected arc faults based on the Fourier transform FFT of the fifth harmonic value of the current.

Feature extraction is followed by classifier selection. Specifically, the time and frequency domain feature signals serve as model inputs, and then machine learning, neural networks, or data-driven related algorithms are utilized to distinguish the sample types and output classification results [8]. Reference [9] used the light gradient boosting machine (LightGBM) to process the redundant features in the feature set and input the optimized feature set into a decision tree (DT) and k-nearest neighbor (KNN) to achieve fast detection of arc faults. Reference [10] used the maximum mutual information coefficient for mining highly recognizable features and implemented series arc fault detection based on a support vector machine (SVM). Reference [11] applied raw data as the input to the lightweight residual network (LRN) to detect arc faults. Reference [12] utilized a discrete wavelet transform and color map indexing to obtain the image features of arc signals, and then a deep residual network (DRN) was used to identify arc faults. Reference [13] proposed the ECMC feature selection algorithm to construct the optimal set of features, and the stochastic configuration network (SCN) was used to realize the detection of arc faults. Reference [14] proposed a fault detection method based on sparse representation and fully connected neural networks (SRFCNN). This method extracts the features of current signals through dictionary learning and sparse coding, then combines them with neural networks to identify arc faults.

Currently, numerous detection methodologies have been developed to effectively identify series arc faults within single-load circuits. However, these methods exhibit diminished diagnostic efficacies when applied to arc faults in circuits with multiple loads. This study presents a new method for detecting series arc faults in multi-load circuits based on the INGO-ABCLogitBoost algorithm. This method meticulously considers the fault characteristics of arc currents across diverse load operations, achieving the acquisition of the current signal's fault characteristics through the application of time–frequency domain feature extraction techniques. The ReliefF feature selection algorithm is utilized to filter the best subset of features and reduce the feature dimension. The INGO algorithm is introduced to optimize the ABCLogitBoost model parameters, which improves the diagnostic performance. The proposed method is not only suitable for single-load fault detection but also has a high diagnostic accuracy for arc faults occurring in multi-load circuits. Finally, the simulation results verify the feasibility of the algorithm.

2. Arc Signal Feature Extraction

An arc is a form of gas discharge. The discharge process is related to the type of gas, the material and geometry of the electrode, and the parameters of the load. The current in the branch circuit where the arc is located is affected by many factors, making the time and frequency domain characteristics of the arc current complex and diverse. Figure 1 shows the current waveforms at different loads when arc faults occur. Analyzing the waveform characteristics of the arc current, it is found that the current waveforms of different loads after series arc faults contain obvious features such as zero rest, the absence of a semi-period waveform, random fluctuations of the arc current [15–17], and high-frequency currents with high amplitudes appearing during the zero rest of inductive load faults [18,19].

For linear loads, there is a standard sinusoidal waveform during normal operation, and the zero-rest phenomenon at the over-zero point during arc faults is observed, along with the absence of semi-period waveforms at the intermittent extinguishing of the arc. This leads to a low current amplitude in this period, which is approximatively close to zero. The zero rest coefficient z is proposed to denote the arc characteristics; specifically, the ratio

of the number of sampled points to the total number of sampling points in the threshold interval $[-\varepsilon, \varepsilon]$.

$$z = \sum_{j=1}^n k_j / n, \quad k_j = \begin{cases} 1, & |I_j| \leq \varepsilon \\ 0, & |I_j| > \varepsilon \end{cases} \quad (1)$$

In this formula, k_j is the comparison value of the current $|I_j|$ with the threshold value ε , which is used to determine whether the current of this sample point is in the zero-rest period. n is the number of sampling points of the current, I_j is the instantaneous value of the current after normalization, and the threshold ε needs to be set by considering different load current fluctuations.

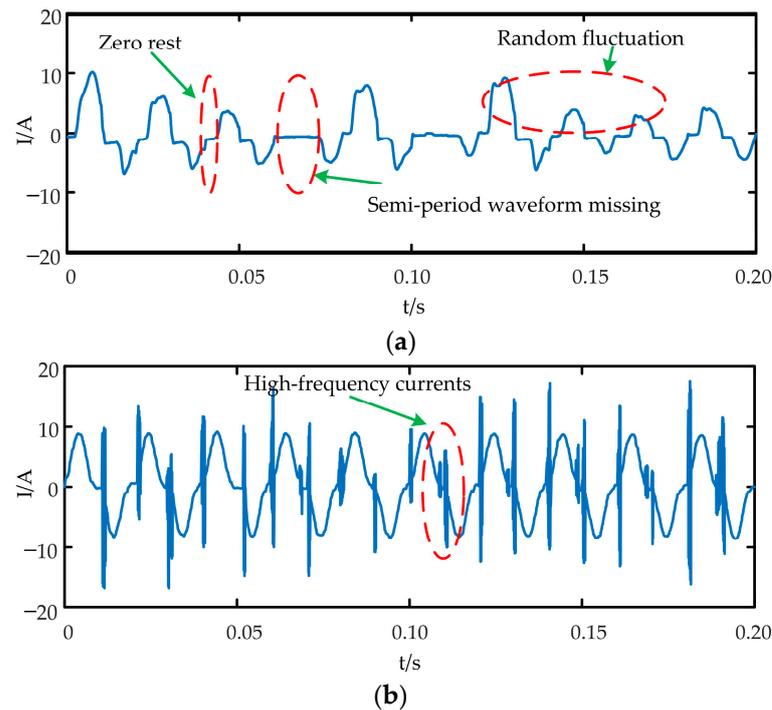


Figure 1. Arc current waveforms after arc faults under different loads. (a) Arc current waveform of an air compressor; (b) arc current waveform of a vacuum cleaner.

Usually, the waveform is stable and unchanged during the normal operation of electrical equipment. The current amplitude fluctuates randomly in different periods with the dynamic change of the arc during the arc fault. The discrete coefficient d is proposed to represent the arc characteristics; specifically, counting the current amplitude in each half-period and using the variance of the amplitude to express the degree of discretization of the arc current.

$$d = \sum_{h=1}^m (X_h - \overline{X_h})^2 / m \quad (2)$$

In this formula, X_h is the half-period amplitude of the current, $\overline{X_h}$ is the average value of the current amplitude, and m is the number of half-periods.

Due to the different settings of various load parameters, the arc current data are normalized to avoid being affected by variations in the parameters. A time–frequency analysis can be used to obtain the time domain and frequency domain information of the signal at the same time, which is an effective method to analyze the characteristics of nonlinear and nonsmooth signals. In this paper, the maximum overlap discrete wavelet transform (MODWT) is used to extract the time–frequency domain features.

The discrete wavelet transform (DWT) is commonly used to analyze the time–frequency characteristics of non-smooth signals [20]. The DWT can be expressed as follows:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|\gamma_0^j|}} \psi\left(\frac{t - ku_0\gamma_0^j}{\gamma_0^j}\right) \quad (3)$$

In this formula, γ_0 , u_0 , and ψ are the scaling parameter, the translation parameter, and the mother wavelet, respectively, and the variables j and k are integer values of the control scaling and translation parameters, respectively.

MODWT is the deformed extension of the DWT, which is a highly redundant non-orthogonal transform [21,22]. Compared with DWT, MODWT has no down-sampling process, which maintains the integrity of the coefficients and avoids the omission of important information. It can quickly detect faults and transients in electrical quantities. MODWT is translation invariant, which allows it to more accurately locate the moment of the fault.

To better describe the differences between different loads during normal operation and arc faults, this paper utilizes the advantages of entropy feature theory in dealing with uncertain signals and carries out multi-scale entropy calculations on the five-layer high-frequency wavelet coefficients obtained from decomposition. The calculated entropy features are fuzzy entropy, envelope entropy, and permutation entropy.

Fuzzy entropy, envelope entropy, and permutation entropy have physically similar meanings. They are used to measure the probability of generating new patterns in a time series. If the probability of generating new patterns is higher, then the complexity of the sequence is higher [23–25]. Fuzzy entropy introduces a fuzzy affiliation function based on sample entropy to deal with the similarity measure of a time series. Envelope entropy reflects the sparse characteristics of the signal, and its size is inversely correlated with the periodicity of the signal. The stronger the periodicity of the signal, then the smaller the value of the envelope entropy. Permutation entropy is based on the probability of permutation of the sub-sequence of the entropy calculation, and it has a high degree of sensitivity to the changes in the time series. Five layers of wavelet coefficients are used to seek the three kinds of entropy for a total of fifteen dimensional features.

To comprehensively reflect the complex characteristics of the arc signal, different computational methods are combined to fully explore the potential characteristics of arc faults from different perspectives. The time–domain features are selected with zero-rest coefficients and discrete coefficients. The frequency–domain features are selected based on the amplitude of the first twenty harmonics. The time–frequency features are selected with fifteen dimensional wavelet entropies, and a total of 37 dimensional features are extracted.

3. ReliefF Feature Selection

Not all features in the feature space of the arc current are favorable. If all of them are used as inputs during classification without selection, the subsequent fault diagnosis will be time consuming and will affect the accuracy of the diagnostic results. Optimizing the original feature set can reduce the amount of computation needed and improve the diagnostic accuracy of the model.

The ReliefF algorithm is a typical filtered feature selection method that is computationally simple and widely used. When analyzing the classification problem, each time a sample R is randomly taken out from the training sample set, k -nearest neighbor samples K are identified from the set of similar samples, k -nearest neighbor samples M are identified from the set of dissimilar samples, and the score of each feature is updated according to Equation (4) [26]. Feature selection follows the principle of “aggregation within classes and dispersion between classes.” If the features have small differences among samples of the same class and large differences among samples of different classes, the features have a strong discriminatory ability, and their feature scores are also larger.

$$W_{m+1}(A) = W_m(A) - \sum_{j=1}^k d(A, R, H_j) / (m \cdot k) + \sum \left[\frac{p(c)}{1-p(c(R))} \sum_{j=1}^k d(A, R, M_j) \right] / (m \cdot k) \quad (4)$$

In this formula, c denotes the category to which the sample R belongs, $d(A, R_1, R_2)$ denotes the distance between the samples R_1 and R_2 on the feature A , $p(c)$ denotes the prior probability of the category c , and m is the number of samples.

4. Arc Fault Detection

4.1. ABCLogitBoost Algorithm

The ensemble learning (EL) algorithm combines weak learners, resulting in a better performance than single models in most cases. ABCLogitBoost is one of the Boosting ensemble learning algorithms, proposed by Ping Li, which has certain advantages in solving problems with noise or with the presence of misclassified labels. It is mainly used for solving the problem of multi-class classification [27–29].

Compared with other algorithms, ABCLogitBoost uses the adaptive binning algorithm (ABA) to preprocess the feature values and discretize the continuous floating-point feature values into n integers from small to large, which not only reduces the memory consumption but also decreases the time complexity. At the same time, based on the commonly used loss function, ABCLogitBoost improves the loss function so that the sum of the loss function is 0 as a constraint. The ABCLogitBoost algorithm adaptively selects the class with the smallest loss value as the base class according to the training loss to improve the training efficiency and model performance.

The performance of the ABCLogitBoost algorithm is mainly affected by three key parameters, including the number of decision trees N , the maximum number of splits J , and the shrinkage rate η . Typically, if the number of decision trees N is too small, it may lead to difficulty in the convergence of the model, which weakens its generalization ability. Conversely, if N is too large, the model may be overfitted. The maximum number of splits J and the contraction rate η have a greater impact on the generalization ability of the model.

These three parameters have a large impact on the model performance, and the optimization algorithm can optimize the objective function to train a better model [30]. To reasonably configure these parameters, this paper uses the improved northern goshawk optimization algorithm to perform parameter optimization on the ABCLogitBoost model to improve the diagnostic accuracy of the model.

4.2. Optimization of ABCLogitBoost Parameters Based on the INGO Algorithm

The northern goshawk optimization (NGO) algorithm is an intelligent optimization algorithm established by Mohammad Dehghani et al. in 2021 based on the hunting behavior of northern goshawks [31]. In the original NGO algorithm, the initial population position is randomly distributed, and the position distribution is not uniform. Cubic chaotic mapping can generate sequences with a more uniform distribution. In this paper, the NGO algorithm is improved, and cubic chaotic mapping is chosen to initialize the population. The iterative optimization process of the algorithm is divided into a prey identification phase and a pursuit phase, and its mathematical model is as follows:

In the first phase of northern goshawk hunting, it will randomly select a prey and then quickly attack it. During this phase, its position is updated with the following formula:

$$P_i = X_k, i = 1, 2, \dots, k = 1, 2, \dots, i - 1, i + 1, \dots, N \quad (5)$$

$$x_{i,j}^{new,P1} = \begin{cases} x_{i,j} + r(p_{i,j} - Ix_{i,j}), & F_{P_i} < F_i \\ x_{i,j} + r(x_{i,j} - p_{i,j}), & F_{P_i} \geq F_i \end{cases} \quad (6)$$

$$X_i = \begin{cases} X_i^{new,P1}, & F_i^{new,P1} < F_i \\ X_i, & F_i^{new,P1} \geq F_i \end{cases} \quad (7)$$

In this formula, P_i is the position of the prey chosen by the i th northern goshawk, F_{P_i} is the fitness value, k is a random number belonging to $[1, N]$, $X_i^{new,P1}$ is the new state of the i th northern goshawk, $x_{i,j}^{new,P1}$ is its new state in the j th dimension, and $F_i^{new,P1}$ is the fitness value corresponding to it. r is a random number belonging to $[0, 1]$, and the value of I is 1 or 2. r and I are used to generate random numbers in the search and update phase.

After a northern goshawk attacks its prey, the prey will attempt to escape. Assuming this prey is in the attack position at radius R , the position update formula for the second phase is:

$$x_{i,j}^{new,P2} = x_{i,j} + R(2r - 1)x_{i,j} \quad (8)$$

$$R = 0.02(1 - \frac{t}{T}) \quad (9)$$

$$X_i = \begin{cases} X_i^{new,P2}, & F_i^{new,P2} < F_i \\ X_i, & F_i^{new,P2} \geq F_i \end{cases} \quad (10)$$

In this formula, t is the current number of iterations, T is the maximum number of iterations, $X_i^{new,P2}$ is the new state of the i th northern goshawk in the second hunting phase, $x_{i,j}^{new,P2}$ is its new state in the j th dimension, and $F_i^{new,P2}$ is the value of fitness corresponding to it.

The process of series arc fault detection is shown in Figure 2:

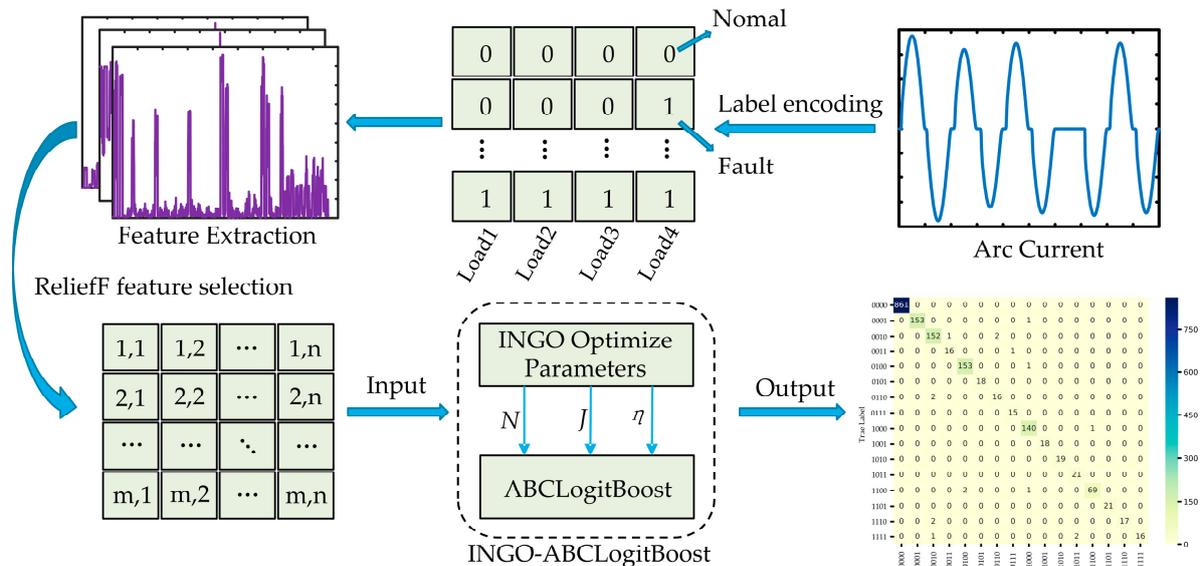


Figure 2. Process of series arc fault detection based on INGO-ABCLoGitBoost.

5. Simulation Verification

5.1. Dynamic Arc Model

The Cassie arc model takes into account the convective heat dissipation effect of the arc. It is assumed that the energy dissipated from the electrode and the diffusion of energy caused by the process of arc column change is negligible. With the change in energy, the larger the cross-sectional area of the arc column, the greater the rate of energy diffusion [32]. The model equation is as follows:

$$\frac{1}{g} \frac{dg}{dt} = \frac{1}{\tau} \left(\frac{U_{arc}^2}{U_c^2} - 1 \right) \quad (11)$$

In this formula, g is the arc conductance, τ is the arc time constant, U_c is the arc voltage constant, and U_{arc} denotes the arc instantaneous voltage.

The normal resistive load and fault arc current waveforms are shown in Figures 3 and 4. The normal current is a stable sinusoidal waveform, while the current waveform after the fault exhibits distortion, and the phenomenon of current crossing the zero point becomes apparent during the zero-rest period.

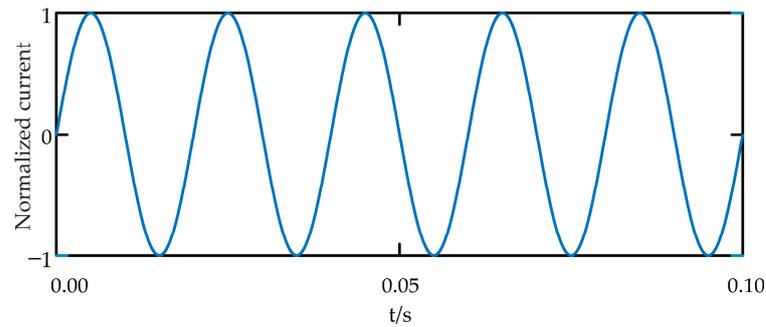


Figure 3. Analysis of normal working current of resistive load.

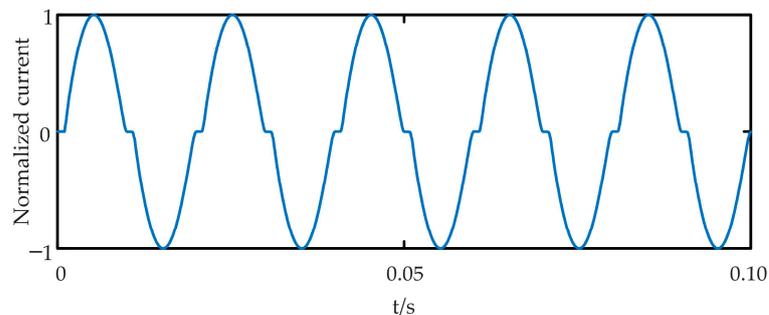


Figure 4. Analysis of arc fault current of resistive load.

The classical arc model exhibits a continuous and unchanging arc waveform, which cannot reflect the characteristics of the actual arc such as randomness, intermittency, and the appearance of high-frequency components when inductive load faults occur [15,16,18,19]. This makes it more in line with the actual situation. The following three improvements are made to the arc model based on the Cassie model:

1. The arc parameters of the Cassie model are fixed constants, while the magnetic field, temperature, external electric field, air resistance, and other factors affect the arc length through random nonlinear changes. To fully describe the arc length variation process, the change in the arc length introduced to improve the Cassie model, which can more truly reflect the combustion process of the arc and fault characteristics [33]. The expression is as follows:

$$U'_c = U_c \cdot r(t) \quad (12)$$

$$U_c = E \cdot L \quad (13)$$

In this formula, U_c is the arc voltage constant, $r(t)$ is a dynamically varying random number, E is the static arc voltage drop, and L is the static arc length.

2. The actual arc is not always stable. The arc burning period is accompanied by an intermittent re-ignition phenomenon of the arc, resulting in the current waveform missing a semi-period waveform. This will affect the accurate identification of arc faults. In this paper, the phenomenon of intermittent re-ignition of the arc is simulated by modifying the duration of the arc combustion time.

3. The high-frequency oscillatory component appears during the zero-rest period when an arc fault occurs in the inductive load under the influence of an inductive component. The high-frequency oscillation characteristics are simulated by modeling the

high-frequency oscillation of the arc and controlling the moment of appearance of the high-frequency component.

5.2. Load Modeling

Considering the use of electrical equipment in ordinary households and according to the working characteristics of different electrical appliances, this paper builds simulation models for four types of loads: resistive loads, inductive loads, phase-angle controllable loads, and switching power supply loads [34]. Resistive loads mainly include cooking, heating, and lighting electrical devices, which are represented using resistive elements. Inductive loads include some electrical devices that work based on the principle of electromagnetic induction. To simplify the model, inductive loads are composed of resistors and inductors in series. Phase-angle controllable loads regulate the load current by controlling the trigger angle of controllable devices such as thyristors, power-adjustable lighting appliances, heaters, and fans. The specific model of phase-angle controllable loads is shown in Figure 5.

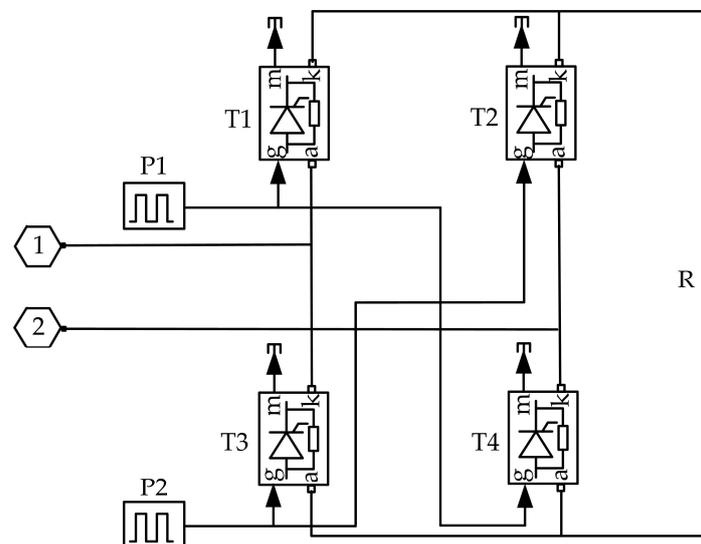


Figure 5. Phase-angle controllable load simulation model.

Switching power supply loads usually require a DC power supply for downstream devices, so a rectifier link is needed to convert AC power into DC power. Such loads mainly include computer monitors, cell phone chargers, etc. The specific model for switching power supply loads is shown in Figure 6.

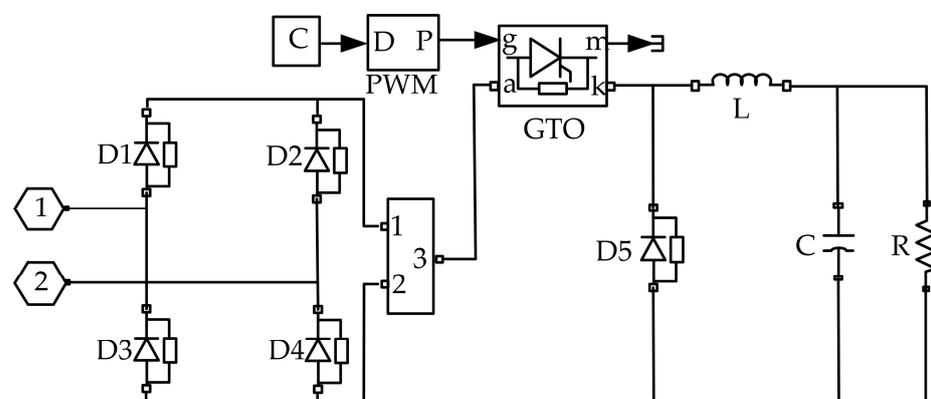


Figure 6. Switching power supply load simulation model.

To simulate the occurrence of arc faults in residential households, four types of loads are arranged in different branches, and each load is connected in series with an ideal switch to control the switching state of the load through the on–off of the ideal switch. When the load is running, the arcing fault occurs at different branches. The trunk current will appear with different degrees of distortion, with each load and its upstream branch modeled in series, denoted as Arc 1~6. When the arc fault occurs, only one arc model is put into operation, while the remaining arc models are short-circuited. The total circuit simulation model is shown in Figure 7. The simulation step size is 1×10^{-5} s. The supply voltage $U_s = 220$ V. The internal resistance of the power supply $R_s = 0.2 \Omega$. In the arc model, $U_C = 50$ V, $g(0) = 1.7 \times 10^{-4}$ [35]. The solver is set to auto (automatic solver selection) because auto allows for the selection of different numerical methods in different situations, improving the performance and accuracy of the simulation compared to setting a fixed solver.

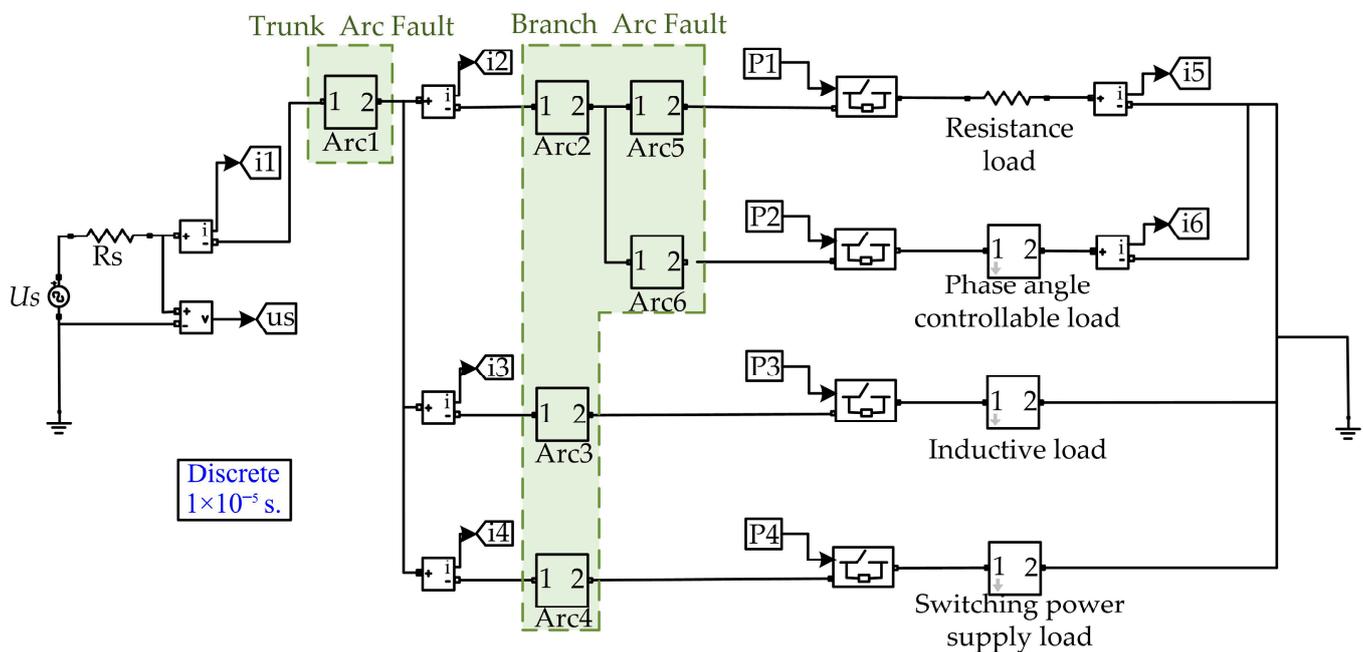


Figure 7. Circuit simulation model.

5.3. Simulation Program

Fault arc simulation programs are carried out in single-load, two-load, three-load, and four-load operations, respectively. With arc faults occurring in the trunk and branch locations, the load operating states are obtained using the Latin hypercube sampling (LHS) algorithm, which collects the trunk currents of different loads in the normal and fault states, with a total of 5768 samples collected.

The trunk currents are analyzed, and the output statuses are coded according to whether the current sample contains fault information for a particular type of load. The indication “1” signifies that an arc fault has occurred in that type of load, and “0” indicates that that type of load is normal or not in operation. The output status is a four-digit binary number composed of four types of loads in sequence, namely a resistive load, phase-angle controllable load, inductive load, and switching power supply load. There are 16 output states in total. Among them, the code 0000 indicates the normal state, and the remaining 15 output states indicate the sample code of the arc fault.

5.4. Analysis of the Simulation Results

For the single-load current waveforms shown in Figure 8, the resistive load and inductive load are to linear loads, and the normal current waveform is an ideal sinusoidal waveform. When the arc fault occurs, in addition to random fluctuations in current and

intermittent disconnection, the resistive load exhibits a significant zero-rest phenomenon at each zero crossing. Due to the influence of the inductive components, the zero-rest phenomenon of the inductive load is not obvious, and a high-frequency pulse occurs near the zero crossing point. The phase-angle controllable load and switching power supply load belong to nonlinear loads. Affected by the load structures, their normal current waveform appears similar to the zero-rest phenomenon of the linear loads arc fault, and the time-domain waveforms undergo different degrees of aberration during the arc fault.

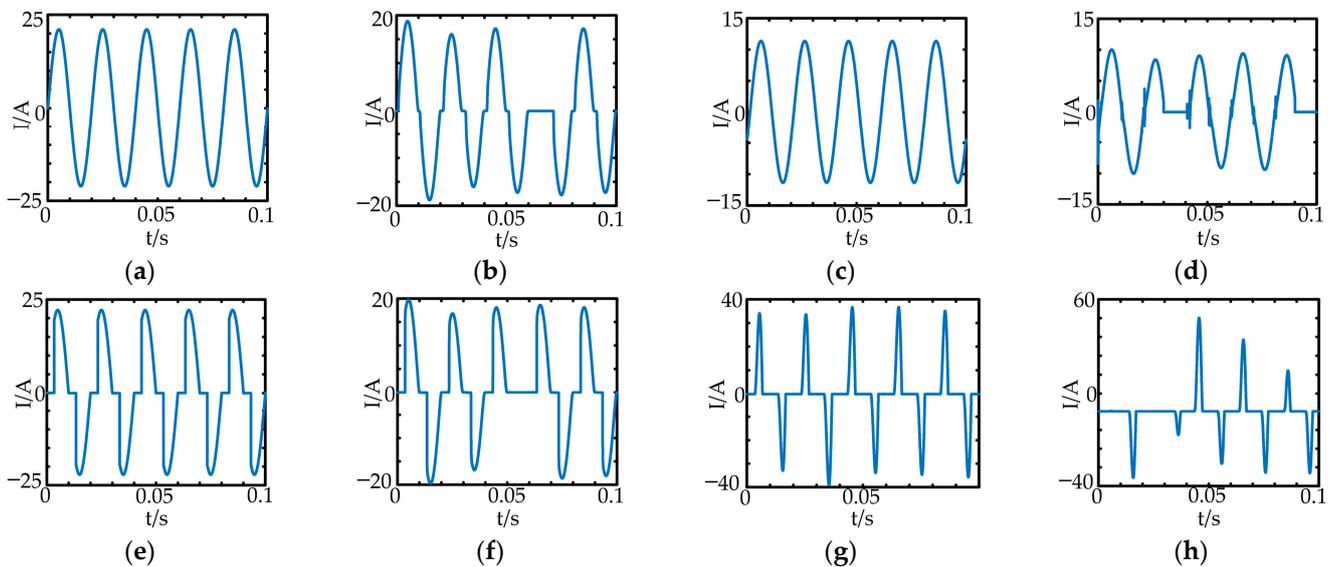


Figure 8. Single-load operation current waveforms. (a) Resistive load normal operation; (b) resistive load arc fault; (c) inductive load normal operation; (d) inductive load arc fault; (e) phase-angle controllable load normal operation; (f) phase-angle controllable load arc fault; (g) switching power supply load normal operation; (h) switching power supply load arc fault.

The waveforms of the trunk-circuit current during the simultaneous operation of the four types of loads are shown in Figure 9. During normal operation, the current waveform of each cyclicity is stable and unchanged. The zero-rest phenomenon occurs at the point of current over zero after resistive branch failure, but due to the influence of other branch currents, the zero-rest feature is weakened, which is manifested as a reduction in the slope of the current during the zero-rest period. The current amplitude of the section is lower due to the intermittent arcing at 0.04–0.05 s. The fault current of the branch of the phase-angle controllable load appears to not be significantly changed. The high-frequency component increases near the zero-crossing point when the inductive load branch circuit faults. The trunk arc fault is equivalent to the simultaneous fault of all the branch loads, and the waveform distortion is the most obvious.

Taking the normal and phase-angle controllable load branch fault signals during four-load operation as examples, the sym5 wavelet basis function is selected for its five-layer MODWT decomposition, and the results are shown in Figure 10. From Figure 10, it can be seen that both the normal and fault current signals contain obvious high-frequency components. Compared with the current signals during normal operation, the degree of irregularity of the signals in each frequency band changes after the arc fault.

5.5. Validation of the Effectiveness of the ReliefF Algorithm

The ReliefF algorithm is used to sort and filter the features of the training set samples and to determine the optimal number of features k . The feature sets under different k values are input into the ABCLogitBoost model, and the optimal k value is selected by comparing the effects of the k value on the accuracy of the model. Meanwhile, in order

to compare the performance of the ReliefF algorithm, this paper selects different feature selection algorithms for comparison, and the results are shown in Figure 11 and Table 1.

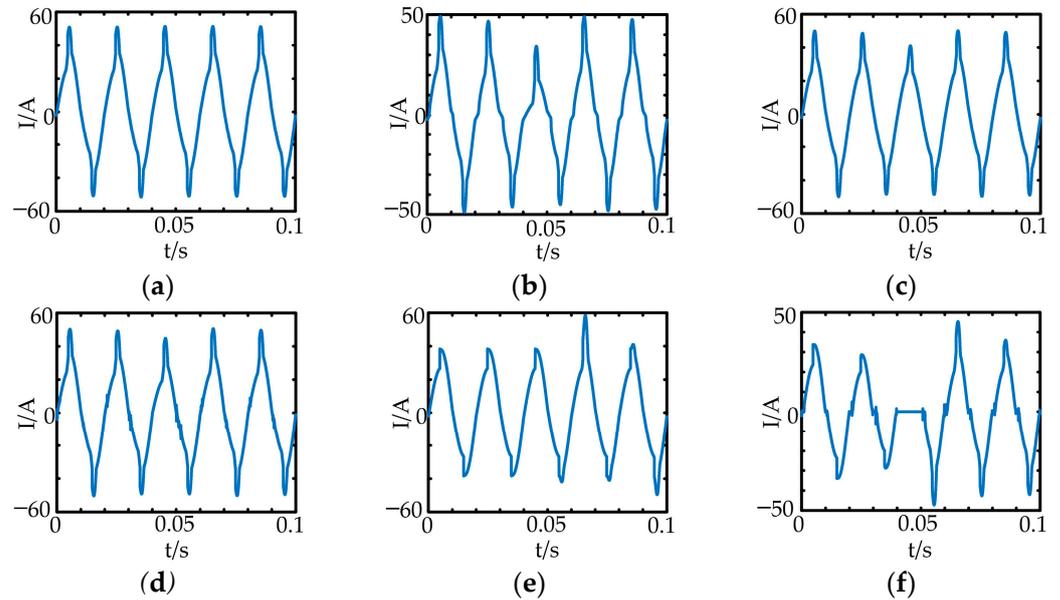


Figure 9. Four-load simultaneous operation current waveforms. (a) Normal operation; (b) resistance branch arc fault; (c) phase-angle controllable load branch arc fault; (d) inductive load branch arc fault; (e) switching power supply load branch arc fault; (f) trunk arc fault.

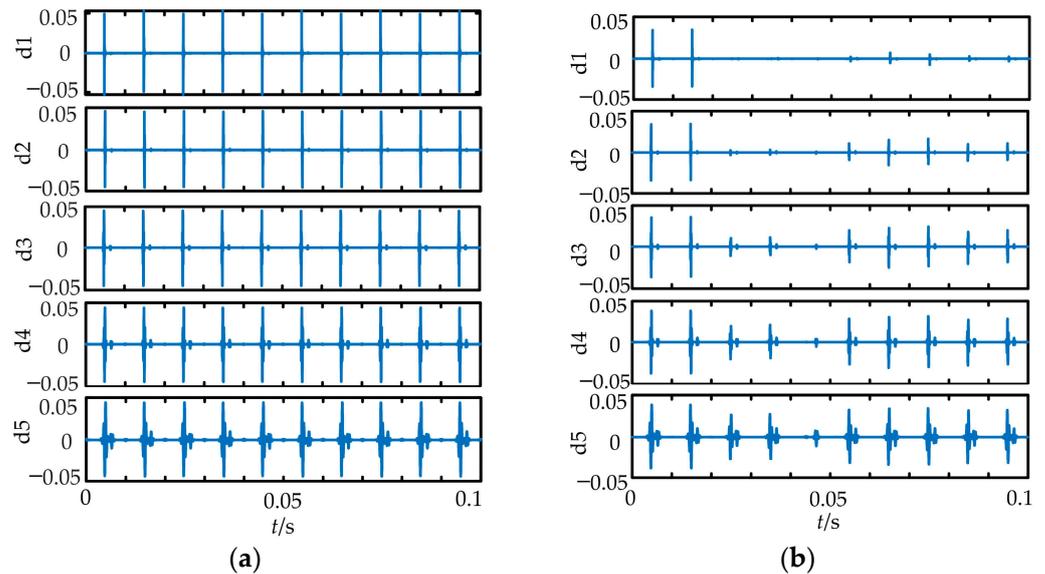


Figure 10. MODWT decomposition of the arc current signal and wavelet coefficients of each layer. (a) Normal operation; (b) phase-angle controllable load branch arc fault.

As can be seen in Figure 11, the trend of model accuracy with the k value is similar under the three algorithms: ReliefF, mRMR, and LDA. Under the ReliefF algorithm, when the number of features k is 2~15, the model accuracy rate gradually increases with increases in the k value. And when $k > 15$, the model accuracy rate tends to stabilize and fluctuates near 93%. When k is 17, the accuracy rate is the highest. The 17 features consist of 2 time-domain features, 4 frequency-domain features, and 11 time-frequency domain entropy features.

Table 1 presents the computed results of the maximum accuracy using various feature selection algorithms and the corresponding number of features. The superior performance

of the feature selection algorithms results in a higher accuracy coupled with lower count of features. The ReliefF algorithm is significantly better than the mRMR and the LDA feature selection algorithms, which also verifies the validity of the algorithms selected in this paper.

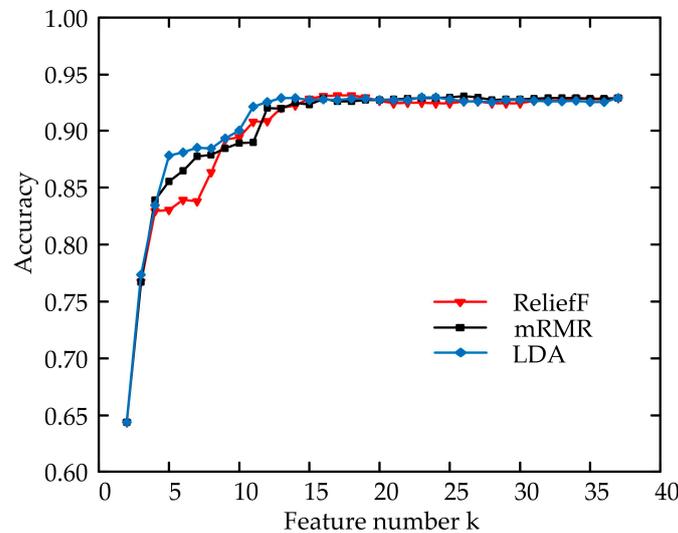


Figure 11. Influence of feature number k on the accuracy of the model.

Table 1. Comparison of calculation results of different feature selection algorithms.

Feature Selection Algorithm	k	Accuracy
ReliefF	17	93.03%
mRMR	27	93.03%
LDA	23	92.92%

5.6. Performance Evaluation of INGO-ABCLogitBoost Fault Diagnosis Model

The number of decision trees N , the maximum number of splits J , and the shrinkage rate η in the ABCLogitBoost model are optimized by using the INGO algorithm. The initial parameter settings of the INGO algorithm are shown in Table 2.

Table 2. INGO parameters.

INGO Parameters	Value
Population	10
Number of iterations	50
Optimization dimensions	3

Considering the degree of influence of the above hyperparameters on the model, their optimization ranges are set, and the optimization searches are carried out within the allowable range. The optimization results are shown in Table 3.

Table 3. Hyperparameter optimization results.

Hyperparameters	Optimization Range	Value
Decision trees (N)	[1500]	470
Splits (J)	[1, 20]	13
Shrinkage rate (η)	[0.001, 0.1]	0.0364

Figure 12 provides a visualization of the confusion matrix of the INGO-ABCLogitBoost arc fault diagnostic results. The diagonal elements of the matrix indicate the number of samples correctly predicted for each class. The sum of each row indicates the total number

of samples of that class, and the sum of each column indicates the total number of samples predicted to be of that class.

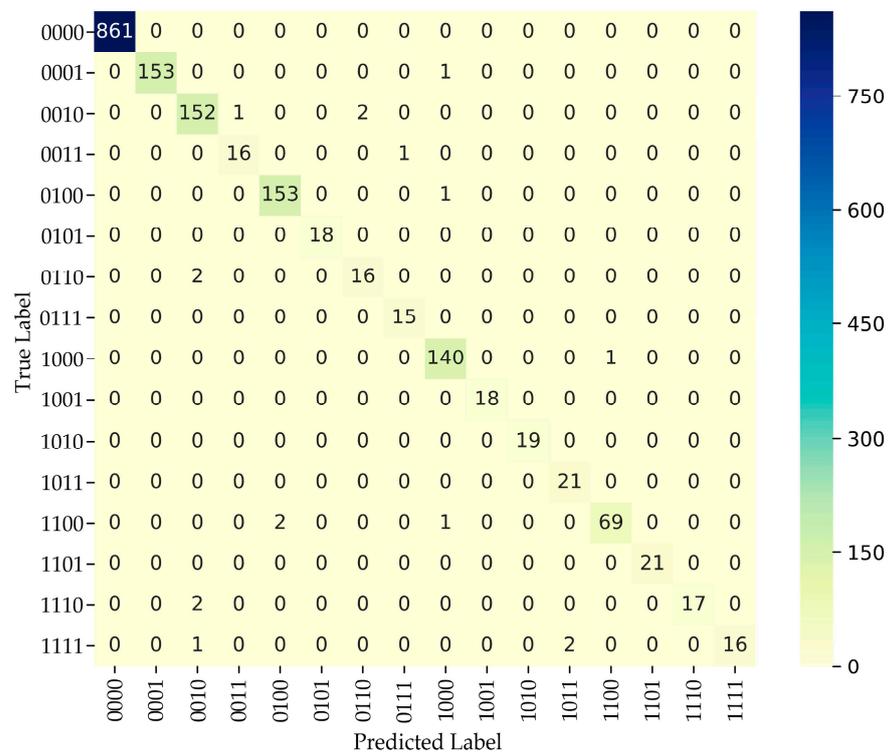


Figure 12. Fault diagnosis classification confusion matrix.

The diagnostic results have shown that the accuracy rate of all types of labels is above 84.21%, and the overall accuracy rate is as high as 99.01%. Label 0000 indicates that all the loads are working in the normal state. The diagnostic accuracy of the samples in this category is 100%, with no misjudgment or omission, indicating that the model can accurately distinguish between the normal and faulty states of the system. When specifically determining which category of load is located in the branch circuit failure, the diagnostic accuracy decreases due to the influence of different load failure degrees and fault characteristics. Label 1111 has the lowest category accuracy of 84.21%. This is because label 1111 indicates the case of arcing faults on the trunk. In this condition, four categories of loads are running simultaneously, which contains more loads and is prone to misjudgment.

5.7. Comparison of Diagnostic Effects of Different Models

To verify the excellent performance of the proposed INGO-ABCLogitBoost arc fault diagnosis model, the class accuracy and overall accuracy of the model under each class are calculated as evaluation indexes, and the model is analyzed and compared with the ELM and XGBoost models that have performed well in the arc fault diagnosis field. The results are shown in Table 4. The overall accuracy of both the ELM and XGBoost models is relatively poor, and they have a wide range of variation in accuracy across different classes of samples (0~97.33%). This indicates that these algorithms are only suitable for the detection of some classes of samples, and that they perform badly on others. The original ABCLogitBoost model has a high diagnostic accuracy for the overall samples, which is up to 93.09%, but the effects of class accuracy for the corresponding samples are weak. After using the INGO optimization algorithm, the overall accuracy and class accuracy are significantly improved, which indicates that the optimization algorithm can effectively improve the diagnostic performance of the model. The classification ability of the INGO-ABCLogitBoost arc fault diagnosis model proposed in this paper has been verified.

Table 4. Comparison of diagnostic results of different models.

Model	Accuracy	
	Class	Overall
ELM	0~100%	59.18%
NGO-ELM	0~100%	67.94%
XGBoost	0~97.33%	62.02%
NGO-XGBoost	0~99.88%	71.72%
ABCLogitBoost	11.11~100%	93.09%
NGO-ABCLogitBoost	84.21~100%	98.84%
INGO-ABCLogitBoost	84.21~100%	99.01%

6. Conclusions

The escalation in power equipment complexity and load branches exert a substantial influence on the diagnostics of series arc faults. This study introduces a novel approach for series arc fault diagnosis based on the INGO-ABCLogitBoost algorithm, grounded in a data-driven perspective.

When an arc fault occurs in a power system, the arc current signal undergoes varying degrees of distortion. In this paper, the time–frequency domain features of the arc current signal are extracted to fully exploit its signal characteristics. The ReliefF feature dimensionality reduction algorithm is employed to efficiently eliminate non-essential features, thereby enhancing recognition efficacy. Arc fault detection is conducted by utilizing the ABCLogitBoost model, with the model parameters being meticulously optimized through integration with the INGO algorithm. This optimization enhances both the detection accuracy and the generalization capacity of the model. The efficacy of the proposed methodology is corroborated through simulation analyses.

Arc fault diagnosis is an important issue in different types of electrical networks. The arc fault characteristics of different electrical networks are varied due to diverse equipment and topologies. The method proposed in this paper can provide a basis for arc fault diagnosis in different electrical networks, and researchers can select the appropriate features and adjust the model parameters according to the actual situation to apply the method in different electrical networks.

Author Contributions: Conceptualization, X.W.; methodology, X.W.; software, X.W.; validation, Y.Z.; data curation, Y.Z.; writing—original draft preparation, Y.Z.; writing—review and editing, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by State Grid Corporation Headquarters Science and Technology Project under funding number 5700-20215204A-0-0-00.

Data Availability Statement: The data presented in this study are available from the corresponding author upon request. The data are not publicly available due to privacy reasons.

Acknowledgments: The authors of this article appreciate the referees for their valuable suggestions, which contributed to improving this paper.

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

References

- Zhang, B.; Wang, S.; Chen, L.; Li, X.; Tang, N. Influence of Oxygen on Solid Carbon Formation during Arcing of Eco-Friendly SF₆-Alternative Gases. *J. Phys. Appl. Phys.* **2023**, *56*, 365502. [[CrossRef](#)]
- Luan, W.; Lin, J.; Liu, B.; Zhao, B. Arc Fault Detection and Identification via Non-Intrusive Current Disaggregation. *Electr. Power Syst. Res.* **2022**, *210*, 108113. [[CrossRef](#)]
- Meng, Y.; Yang, Q.; Chen, S.; Wang, Q.; Li, X. Multi-Branch AC Arc Fault Detection Based on ICEEMDAN and LightGBM Algorithm. *Electr. Power Syst. Res.* **2023**, *220*, 109286. [[CrossRef](#)]
- Wang, Z.; Han, C.; Gao, H.; Guo, F. Identification of Series Arc Fault Occurred in the Three-Phase Motor With Frequency Converter Load Circuit via VMD and Entropy-Based Features. *IEEE Sens. J.* **2022**, *22*, 24320–24332. [[CrossRef](#)]

5. Calvo, J.L.; Schweitzer, P.; Weber, S.; Tisserand, E.; Joyeux, P. Arcing Detection at Home System Using Correlation Analysis. In Proceedings of the 2014 27th International Conference on Electrical Contacts, Dresden, Germany, 22–26 June 2014.
6. Ming, Z.; Tian, Y.; Zhang, F. Design of Arc Fault Detection System Based on CAN Bus. In Proceedings of the 2009 International Conference on Applied Superconductivity and Electromagnetic Devices, IEEE, Chengdu, China, 25–27 September 2009; pp. 308–311.
7. Hadziefendic, N.; Kostic, M.; Radakovic, Z. Detection of Series Arcing in Low-voltage Electrical Installations. *Eur. Trans. Electr. Power* **2009**, *19*, 423–432. [[CrossRef](#)]
8. Mo, J.; Yang, H. Sampled Value Attack Detection for Busbar Differential Protection Based on a Negative Selection Immune System. *J. Mod. Power Syst. Clean Energy* **2023**, *11*, 421–433. [[CrossRef](#)]
9. Du, L.; Xu, Z.; Chen, H.; Chen, D. Feature Selection-Based Low-Voltage AC Arc Fault Diagnosis Method. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [[CrossRef](#)]
10. Zou, G.; Fu, G.; Han, B.; Wang, W.; Liu, C. Series Arc Fault Detection Based on Dual Filtering Feature Selection and Improved Hierarchical Clustering Sensitive Component Selection. *IEEE Sens. J.* **2023**, *23*, 6050–6060. [[CrossRef](#)]
11. Gao, X.; Zhou, G.; Zhang, J.; Zeng, Y.; Feng, Y.; Liu, Y. Fault Arc Detection Based on Channel Attention Mechanism and Lightweight Residual Network. *Energies* **2023**, *16*, 4954. [[CrossRef](#)]
12. Zhang, S.; Qu, N.; Zheng, T.; Hu, C. Series Arc Fault Detection Based on Wavelet Compression Reconstruction Data Enhancement and Deep Residual Network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–9. [[CrossRef](#)]
13. Li, J.; Zou, G.; Wang, W.; Shao, N.; Han, B.; Wei, L. Low-Voltage Series Arc Fault Detection Based on ECMC and VB-SCN. *Electr. Power Syst. Res.* **2023**, *218*, 109222. [[CrossRef](#)]
14. Wang, Y.; Zhang, F.; Zhang, S. A New Methodology for Identifying Arc Fault by Sparse Representation and Neural Network. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2526–2537. [[CrossRef](#)]
15. Jing, T.; Huang, D.; Mi, Z.; Yao, L.; Liu, X. An Intelligent Recognition Method of a Short-Gap Arc in Aviation Cables Based on Feature Weight Enhancement. *IEEE Sens. J.* **2023**, *23*, 3825–3836. [[CrossRef](#)]
16. Guo, M.-F.; Liu, W.-L.; Gao, J.-H.; Chen, D.-Y. A Data-Enhanced High Impedance Fault Detection Method Under Imbalanced Sample Scenarios in Distribution Networks. *IEEE Trans. Ind. Appl.* **2023**, *59*, 1–14. [[CrossRef](#)]
17. Jiang, W.; Liu, B.; Yang, Z.; Cai, H.; Lin, X.; Xu, D. Non-Intrusive Arc Fault Detection and Localization Method Based on the Mann–Kendall Test and Current Decomposition. *Energies* **2023**, *16*, 3988. [[CrossRef](#)]
18. Bao, G.; Jiang, R.; Gao, X. Novel Series Arc Fault Detector Using High-Frequency Coupling Analysis and Multi-Indicator Algorithm. *IEEE Access* **2019**, *7*, 92161–92170. [[CrossRef](#)]
19. Bao, G.; Jiang, R.; Liu, D. Research on Series Arc Fault Detection Based on Higher-Order Cumulants. *IEEE Access* **2019**, *7*, 4586–4597. [[CrossRef](#)]
20. Lu, Q.W.; Wang, T.; Li, Z.R.; Wang, C. Detection Method of Series Arcing Fault Based on Wavelet Transform and Singular Value Decomposition. *Trans. China Electrotech. Soc.* **2017**, *32*, 208–217. (In Chinese) [[CrossRef](#)]
21. Jimenez-Guarneros, M.; Morales-Perez, C.; Rangel-Magdaleno, J.D.J. Diagnostic of Combined Mechanical and Electrical Faults in ASD-Powered Induction Motor Using MODWT and a Lightweight 1-D CNN. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4688–4697. [[CrossRef](#)]
22. Bagheri, S.; Moravej, Z.; Gharehpetian, G.B. Classification and Discrimination Among Winding Mechanical Defects, Internal and External Electrical Faults, and Inrush Current of Transformer. *IEEE Trans. Ind. Inform.* **2018**, *14*, 484–493. [[CrossRef](#)]
23. Zhou, R.; Wang, X.; Wan, J.; Xiong, N. EDM-Fuzzy: An Euclidean Distance Based Multiscale Fuzzy Entropy Technology for Diagnosing Faults of Industrial Systems. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4046–4054. [[CrossRef](#)]
24. Chen, Z.; Yang, Y.; He, C.; Liu, Y.; Liu, X.; Cao, Z. Feature Extraction Based on Hierarchical Improved Envelope Spectrum Entropy for Rolling Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [[CrossRef](#)]
25. Guo, J.; He, Q.; Zhen, D.; Gu, F. Intelligent Fault Detection for Rotating Machinery Using Cyclic Morphological Modulation Spectrum and Hierarchical Teager Permutation Entropy. *IEEE Trans. Ind. Inform.* **2023**, *19*, 6196–6207. [[CrossRef](#)]
26. Shi, E.; Sun, L.; Xu, J.; Zhang, S. Multilabel Feature Selection Using Mutual Information and ML-ReliefF for Multilabel Classification. *IEEE Access* **2020**, *8*, 145381–145400. [[CrossRef](#)]
27. Sun, P.; Reid, M.D.; Zhou, J. An Improved Multiclass LogitBoost Using Adaptive-One-vs-One. *Mach. Learn.* **2014**, *97*, 295–326. [[CrossRef](#)]
28. Li, P. ABC-Boost: Adaptive Base Class Boost for Multi-Class Classification. In Proceedings of the 26th Annual International Conference on Machine Learning, ACM, Montreal, QC, Canada, 14–18 June 2009; pp. 625–632.
29. Li, P. Robust LogitBoost and Adaptive Base Class (ABC) LogitBoost. In Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI'10), Catalina Island, CA, USA, 8–11 July 2010; AUAI Press: Arlington, VA, USA, 2010; pp. 302–311.
30. Cao, B.; Zhang, W.; Wang, X.; Zhao, J.; Gu, Y.; Zhang, Y. A Memetic Algorithm Based on two_Arch2 for Multi-Depot Heterogeneous-Vehicle Capacitated Arc Routing Problem. *Swarm Evol. Comput.* **2021**, *63*, 100864. [[CrossRef](#)]
31. Dehghani, M.; Hubalovsky, S.; Trojovský, P. Northern Goshawk Optimization: A New Swarm-Based Algorithm for Solving Optimization Problems. *IEEE Access* **2021**, *9*, 162059–162080. [[CrossRef](#)]
32. Jiang, Y.; Li-an, C. Improved Impedance Arc Model Based on Cassie Theory. In Proceedings of the 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Chongqing, China, 7–9 July 2022; IEEE: Shanghai, China, 2022; pp. 97–102.
33. Zeng, H.; Wang, L.N.; Lin, W.X.; Zhao, S. The Application of a New Dynamic Arc Length Model in Grounding Arc. *Comput. Simul.* **2018**, *32*, 36–41. (In Chinese)

34. He, D.; Du, L.; Yang, Y.; Harley, R.; Habetler, T. Front-End Electronic Circuit Topology Analysis for Model-Driven Classification and Monitoring of Appliance Loads in Smart Buildings. *IEEE Trans. Smart Grid* **2012**, *3*, 2286–2293. [[CrossRef](#)]
35. QU, N.; Wang, J.H.; Liu, J.H.; Wu, F.C. A Series Arc Fault Detection Method Based on Cassie Model and L3/4 Norm. *Power Syst. Technol.* **2018**, *42*, 3992–3997. (In Chinese) [[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.