Int. J. Metrol. Qual. Eng. **15**, 3 (2024)

https://doi.org/10.1051/ijmqe/2024001

International Journal of Metrology and Quality Engineering

Available online at:
www.metrology-journal.org

RESEARCH ARTICLE OPEN ACCESS

Low-voltage AC series arc fault detection based on Fisher-mutual information feature selection

Baichuan Qin[1], Wei Wang[1,*], Wei Hu[2], Lei Su[2], and Guofeng Zou[1]

[1] Shandong University of Technology, Zibo City 255000, Shandong, China
[2] Electric Power Research Institute of State Grid Hubei Electric Power Co., Ltd., Hubei Wuhan 430077, China

Received: 23 September 2023 / Accepted: 12 February 2024

Abstract. The detection of multi-feature fusion is a crucial approach to address the issue of series arc fault detection. Effective feature selection plays a vital role in enhancing the accuracy of the classifier and reducing system complexity. In this study, a feature selection algorithm based on Fisher-mutual information is proposed to tackle the problem of feature selection in multi-feature fusion detection. This algorithm utilizes the characteristics of arc fault voltage source to construct a feature pool. The Fisher-score algorithm and mutual information algorithm are employed to construct an optimal feature subset. The feature subset undergoes rough selection by retaining key features of the classifier and fine selection by eliminating redundant features. Experimental results and comparisons with related methods demonstrate that the proposed feature selection method significantly enhances the classifier's recognition accuracy, reduces classification and recognition time, diminishes the feature dimension, and outperforms other existing methods.

Keywords: Series arc fault / Fisher-score / mutual information / GA-BP neural network / feature selection

1 Introduction

When an arc fault occurs, the temperature at the fault point can reach thousands of degrees Celsius, which easily ignites the line and surrounding combustibles, making it a leading cause of electrical fires [1]. Among these faults, series arc faults pose a challenge in detection due to their lack of overcurrent and residual current characteristics, as well as difficulty in distinguishing their waveform characteristics from nonlinear loads [2]. Currently, research on series arc fault detection methods, both domestically and internationally, primarily relies on current-based approaches. These approaches can be categorized into two groups: detection using single features and detection using multi-feature fusion. Detection methods based on single features utilize parameters such as zero time of current [3,4], harmonic content [5], and similarity of adjacent wave current [6] to identify faults. However, due to the diverse nature of nonlinear load currents, single-feature based methods struggle to ensure applicability across various load types. Therefore, the mainstream approach for arc fault detection involves the fusion of multiple features. From the perspective of feature fusion, two methods are commonly employed: multi-feature decision fusion and feature fusion using artificial intelligence technology. In terms of decision fusion, for instance, reference [7] employs the ratio of current change rate to its effective value and the harmonic amplitude of the current signal in the range of 6–12 kHz as characteristics for arc fault detection. On the other hand, reference [8] calculates the cosine similarity of adjacent periodic current signals and the wavelet energy in the range of 3125–6250 Hz as characteristic quantities for arc fault detection. In reference [9], a study utilized four characteristic quantities − pulse width percentage, coefficient of variation, interharmonic mean, and wavelet singular entropy − to create an identification index for arc faults. However, the process of feature selection and threshold setting in the multi-feature decision fusion method is highly subjective. To address this issue, researchers have turned to feature fusion techniques using machine learning and deep learning, which have gained significant attention in recent years [10,11]. As described in [12], involves mathematical statistics, time domain analysis, time-frequency domain analysis of current data, and the synthesis of a fully connected neural network to propose an arc fault detection method. In reference [13] proposed a fusion method combining wavelet decomposition and empirical mode decomposition was proposed. This method extracts multi-dimensional features of current from the time domain, frequency domain, and wavelet domain, and employs

* Corresponding author: wwsdut@163.com

machine learning to construct an arc fault identification model. Similarly, in reference [14], a three-dimensional feature matrix was constructed based on current pulse width characteristics, interharmonic mean, and wavelet singular entropy, the parameters of the support vector machine were optimized using the particle swarm optimization algorithm to enable pattern recognition and arc fault detection. Additionally, references [15,16] employed a multi-layer convolutional neural network to extract high-dimensional features of the arc signal, enabling effective identification of arc faults.

For arc fault detection, it is not easy to find the most relevant and important features for pattern recognition classifiers. Considering that people always hope that the classifier contains more features, if the number of features is not limited, when the selected features are applied to the classifier recognition, not only the training speed of the classifier is slow, but also the accuracy of the classifier recognition is affected. Therefore, reducing the dimension of the selected arc features through feature selection is a necessary condition to improve the performance of the arc detection model.

Feature selection is a complex project that involves multiple transform domains and waveform morphologies. The ideal feature selection process should consider the complexity and real-time performance of the detection model, while also ensuring high detection accuracy. While there have been numerous studies on multi-feature fusion detection methods, there is a lack of research on the redundancy and optimality of multi-feature selection. Therefore, investigating feature selection methods that take into account feature redundancy and efficiency is of great engineering value. Such research can help improve the computational efficiency of multi-feature fusion detection methods and reduce the complexity of detection models.

Feature selection methods in academic research typically fall into three categories: filter, wrapper, and embedding. The filtering method sorting features based on an evaluation function and then training the classifier with the selected features. Various filtered feature selection methods are mentioned in reference [17]. Another method called maximum relevance minimum redundancy (mRMR) was proposed in [18], which utilizes the SVM classifier for pattern recognition. Reference [19] proposed the ReliefF feature selection method and selected SVM classifier for classification, the wraparound feature selection method is to use the final selected classifier as an evaluation function for feature selection, and the optimal feature subset is selected based on the evaluation criteria of the classifier. A feature selection method based on support vector machine recursive feature elimination (SVM-RFE) and binary particle swarm optimization (BPSO) was proposed in [20]. In this method, the BPSO fitness function incorporates a classifier for training and evaluation purposes. Lastly, the embedded method performs feature selection simultaneously with training and learning of the classifier. In reference [21], an average impure feature selection method was proposed and embedded into a single base learner of random forest.

When performing feature selection, for high-dimensional features, wraparound and embedded feature selection methods are often prone to overfitting phenomenon, while filtered feature selection methods are simpler and faster, so this paper chooses the filtered feature selection method for dimensionality reduction of arc features.

Based on the above background, this paper proposes a feature selection method based on Fisher-score-mutual information (Fisher-MI), which comprehensively considers the relevance of the selected features and the redundancy among the features to achieve the filtering and screening of the feature subset, and effectively achieves the reduction of the training time of the classifier and the enhancement of the recognition accuracy.

2 Feature selection method based on Fisher-MI algorithm

2.1 Principle of rough feature selection based on Fisher score

The Fisher-score [22] (also known as Fisher-score) aims to minimize the feature distance within the same category and maximize it between different categories. This approach helps identify features with high discriminative power.

Arc fault identification can be considered as a binary classification problem for which Fisher-score S_i for different features are calculated as shown below:

$$S_i = \frac{(\mu_k^r - \mu_k)^2 + (\mu_k^s - \mu_k)^2}{\frac{1}{n_r-1}\sum_{i=1}^{n_r}(x_{i,k}^r - \mu_k^r)^2 + \frac{1}{n_s-1}\sum_{i=1}^{n_s}(x_{i,k}^s - \mu_k^s)^2}, \quad (1)$$

where μ_k, μ_k^r, μ_k^s is the mean value of the kth feature over the whole dataset, over the rth class dataset and over the sth class dataset respectively; $x_{i,k}^r$ is the kth eigenvalue of the ith r-class sample and $x_{i,k}^s$ is the kth eigenvalue of the ith s-class sample [22].

The Fisher-score method is used to reflect the distance relationship between features by considering the variance. This approach helps to address the issue of large or small eigenvalues for a certain class. The numerator of the formula represents the inter-class variance of the sample, while the denominator represents the intra-class variance. The variance indicates the magnitude of the distance between selected features. From formula (1), it can be observed that a larger inter-class variance and a smaller intra-class variance lead to a higher calculated Fisher-score value. This suggests a stronger correlation between the feature and the category, the feature also has enhanced recognition capabilities.

In this paper, we propose a method that combines a classifier to achieve coarse feature selection. We sort the feature scores in descending order and input them into the classifier for learning and training. This allows us to evaluate the recognition accuracy under different feature dimensions and verify the recognition ability of the selected feature subset.

2.2 Arc recognition model based on GA-BP neural network

BP neural network is a multilayer feed-forward network model for error backward transfer, and its network structure consists of input layer neurons, hidden layer neurons, and output layer neurons. Each layer neuron is fully interconnected with the next layer neuron. The input layer neurons only receive external inputs and do not process the inputs. The neurons in the hidden layer and the output layer are functional neurons that process the inputs by means of an "activation function". The sigmoid function is used in this paper and its expression is:

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

Since BP neural network has good generalisation ability, nonlinear mapping ability and adaptive learning ability, this paper adopts BP neural network as the classification model for arc fault detection, and at the same time, adopts genetic algorithm to overcome the problems of slow convergence and easy to fall into the local minimum that exist in traditional BP neural network. It optimizes the initial weights and thresholds of the BP neural network using genetic algorithms to enhance their accuracy and convergence speed. The trained neural network is then used to classify and identify the arc data. The flow of the arc recognition method using the GA-BP neural network is illustrated in Figure 1.

2.3 Principle of feature selection based on mutual information

Mutual information (MI) is a widely used parameter for quantifying the degree of redundancy between information. When the MI algorithm is applied for feature selection, the magnitude of mutual information between features indicates the level of redundancy between them. The mutual information $I(X;Y)$ between features X and Y can be calculated as follows:

$$I(X;Y) = -\sum_{x\in X, y\in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}, \tag{3}$$

where $p(x)$ and $p(y)$ are the edge probability densities and $p(x,y)$ is the joint probability density.

The larger the value of mutual information $I(X,Y)$ between features X and Y, the higher the redundancy of features X and Y. When the value of $I(X;Y)$ is 0, it indicates that features X and Y are independent of each other, which can be used as a basis to achieve the selection of the optimal feature subset.

2.4 Fisher-MI based feature selection method

The flow of the Fisher-MI feature selection based method proposed in this paper is shown in Figure 2 with the following steps:

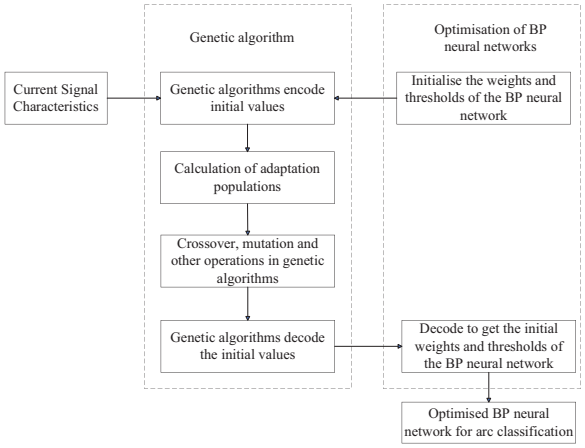

Fig. 1. Flowchart of BP neural network classification model optimised by genetic algorithm.

- The features in the feature pool are scored using the Fisher-score algorithm and ranked based on their scores to determine the importance of the selected features. Higher scores indicate higher importance.
- The selected features are sequentially input into the GA-BP neural network classification model in the order of their ranking for learning and training, the arc fault identification accuracy is obtained for different feature dimensions.
- By comparing the arc recognition accuracy under different feature dimensions, the feature set that achieves the highest recognition accuracy for the first time is selected as the roughly selected feature subset.
- The MI algorithm is used to calculate the mutual information of the roughly selected feature subset screened in (3). With the aim of removing redundant features in the roughly selected feature subset, the features corresponding to a mutual information value of 0 are selected to constitute the optimal feature subset for the Fisher-MI feature selection method.

3 Feature analysis and feature pool construction

3.1 Fault feature

There have been numerous studies on the feature extraction method of arc fault current. From a transform domain perspective, it can be divided into multi-domain features such as time domain, frequency domain, and wavelet domain. Examples of these features include zero rest time [3,4], current change rate [7], harmonic content [5,13], high-frequency content [7], and wavelet coefficient energy [8]. Additionally, features can be categorized as full-domain features, such as harmonic content [5,13] and waveform similarity [6,8], or local domain features, such as zero rest time [3,4] and current rate of change [7]. Furthermore, features can be described using statistical features [13] or waveform morphological features [13]. Examples of statistical features include the mean value,

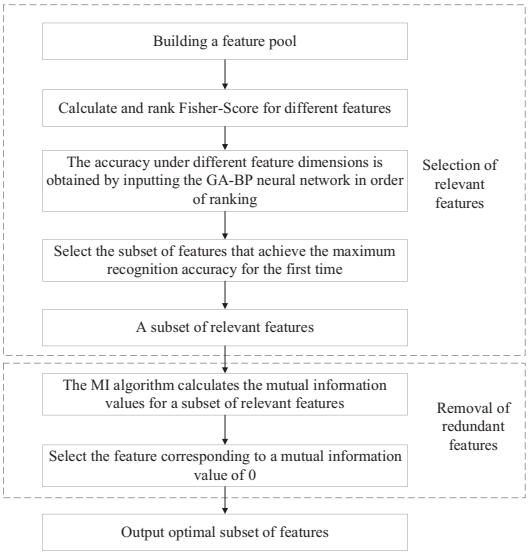

Fig. 2. Fisher-MI feature selection flowchart.

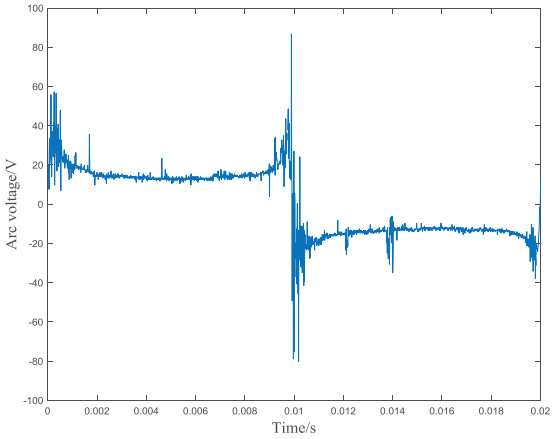

Fig. 3. Time domain waveform of arc fault voltage source.

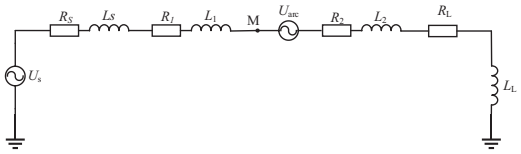

Fig. 4. Equivalent circuit diagram for series arc faults.

variance, and peak-to-peak values of the waveform, while morphological features include waveform cliffiness and skewness. Although there are numerous methods for describing arc current fault features, there is a lack of categorization methods for feature selection, and each feature has multiple description methods. Therefore, this paper aims to extract relevant fault features based on the principle of arc fault equivalent circuits to construct a feature pool.

Figures 3 and 4 depict the time domain waveform of the arc fault voltage source and the equivalent circuit diagram of the series arc fault, respectively. In Figure 4, point M represents the monitoring point in the line. According to the substitution theorem, a series arc fault in the line can be equivalently represented as a fault point in the series of an arc fault voltage source [2]. This arc fault voltage source can be approximated as a signal source composed of a square wave superimposed on a high frequency. Therefore, the randomness of the arc fault source waveform, caused by the variability of the carbonisation degree of the arc electrodes, results in the randomness of the fault current. Additionally, the odd harmonics (e.g., 1, 3, 5, 7) used to form the square wave in the arc fault source introduce additional odd harmonic characteristics to the fault current. The high-frequency component in the arc fault source introduces high-frequency characteristics to the fault current. Furthermore, the high resistance characteristic of the arc fault source near the current overcurrent is similar to the high resistance characteristic of the arc fault source at the fault point, leading to the zero-break characteristic of the fault current.

3.2 Current signal feature pool construction

Based on the analysis in the previous section, the references [3–8,13] are selected to describe the statistical, local and full domain features of fault currents, respectively, and then construct the feature pool. The current features described in the reference are shown in Table 1.

In Table 1, N is the number of current signal points in a cycle, I_k is the amplitude of the kth current signal, I_K is the vector consisting of all current signals in the Kth cycle, $I_k(0.002)$ is the amplitude of the current signal during the first 0.002 seconds of the current signal, a is the amplitude of the current signal during the first 0.002 seconds of the current signal, fc, 3rd hc, and 5th hc are the amplitudes of the fundamental, third harmonic, and fifth harmonic, respectively, after the FFT transform, I_{cd1}-I_{ca5} are the amplitudes of the signals in the cd1–ca5 layers of the current after doing the db4 wavelet transform, respectively. $I_{cd1(0.002)}$-$I_{ca5(0.002)}$ are the amplitude of the signal within the cd1–ca5 layer after doing the db4 wavelet transform for the signal within 0.002 seconds before the current.

4 Experimental validation

4.1 Experimental environment setup

As shown in Figure 5, reference GB 14187.4-2014 to build a series arc experimental platform and select five typical loads for testing, the load includes resistive loads (electric heater 800 W), resistive inductive loads (induction cooker 1800 W), non-linear loads (hoover 1300 W, microwave oven 1500 W, computer 180 W).

The experimental current signals were acquired using a Tek-MDO3014 oscilloscope and a 0.2 level low power current transformer with a sampling frequency of 100 kHz. Figure 6 shows the current waveforms before and after the above mentioned load failure.

Table 1. Description and physical significance of selected features.

Numbering	Current feature	Computing formula	Physical meaning
1	Mean value	$I_{mean} = \frac{1}{N}\sum_{k=1}^{N} I_k$	Randomness of arc current
2	Variance	$I_{var} = \frac{1}{N-1}\sum_{k=1}^{N}(I_k - I_{mean})^2$	Randomness of arc current
3	Skewness	$I_{skewness} = \frac{1}{N}\sum_{k=1}^{N}\frac{(I_k - I_{mean})^2}{I\ \mathrm{var}^{\frac{3}{2}}}$	Randomness of arc current
4	Kurtosis	$I_{kurtosis} = \frac{1}{N}\sum_{k=1}^{N}\frac{(I_k - I_{mean})^4}{I^2\mathrm{var}}$	High frequency characteristic
5	Peak-to-peak value	$I_{M-m} = I_{\max} - I_{min}$	High frequency characteristic
6	Maximum current change rate	$I_{\Delta\max} = \max\{I_{k+1} - I_k\}$	Zero-recess characteristic
7	Cosine similarity	$\cos\quad\theta = \frac{\sum I_K \times I_{K+1}}{\sqrt{\sum I_K^2} \times \sqrt{\sum I_{K+1}^2}}$	Current similarity
8	The zero rest time within 0.002 seconds before the current	$t_{(0.002)} = \frac{1}{N} * 0.02 \sum_{k-1}^{N}(\lvert I_{k(0.002)}\rvert < \sqrt{\frac{1}{N}\sum_{k-1}^{N} I_k^2 * 0.05})$	Zero-recess characteristic
9	The proportion of fundamental wave	$I_{fc} = \frac{fc}{fc+3\mathrm{rd}hc+5\mathrm{th}hc}$	Low frequency characteristics
10	The proportion of 3rd harmonic	$I_{3\mathrm{rd}hc} = \frac{3\mathrm{rd}hc}{fc+3\mathrm{rd}hc+5\mathrm{th}hc}$	Low frequency characteristics
11	Proportion of 5th harmonic	$I_{5\mathrm{th}hc} = \frac{5\mathrm{th}hc}{fc+3\mathrm{rd}hc+5\mathrm{th}hc}$	Low frequency characteristics
12	The energy of cd3 layer and	$E_{\mathrm{cd3}} = \sum_{k=1}^{N} I_{\mathrm{cd3}}^2$	High frequency characteristic
13	The energy of cd4 layer and	$E_{\mathrm{cd4}} = \sum_{k=1}^{N} I_{\mathrm{cd4}}^2$	High frequency characteristic
14	The maximum value of cd1 layer module	$M_{cd1} = \max\{cd1\}$	Zero-recess characteristic
15	The maximum value of cd2 layer module	$M_{cd2} = \max\{cd2\}$	Zero-recess characteristic
16	The maximum value of cd3 layer module	$M_{cd3} = \max\{cd3\}$	Zero-recess characteristic
17	The maximum value of cd4 layer module	$M_{cd4} = \max\{cd4\}$	Zero-recess characteristic
18	The maximum value of cd5 layer module	$M_{cd5} = \max\{cd5\}$	Zero-recess characteristic
19	The maximum value of ca5 layer module	$M_{ca5} = \max\{ca5\}$	Zero-recess characteristic
20	The energy of cd1 layer in 0.002 seconds before the current	$E_{cd1(0.002)} = \sum_{k=1}^{N} I_{cd1(0.002)}^2$	Zero-recess characteristic
21	The energy of cd2 layer in 0.002 seconds before the current	$E_{cd2(0.002)} = \sum_{k=1}^{N} I_{cd2(0.002)}^2$	Zero-recess characteristic
22	The energy of cd3 layer in 0.002 seconds before the current	$E_{cd3(0.002)} = \sum_{k=1}^{N} I_{cd3(0.002)}^2$	Zero-recess characteristic
23	The energy of cd4 layer in 0.002 seconds before the current	$E_{cd4(0.002)} = \sum_{k=1}^{N} I_{cd4(0.002)}^2$	Zero-recess characteristic
24	The energy of cd5 layer in 0.002 seconds before the current	$E_{cd5(0.002)} = \sum_{k=1}^{N} I_{cd5(0.002)}^2$	Zero-recess characteristic
25	The energy of ca5 layer in 0.002 seconds before the current	$E_{\mathrm{ca5}(0.002)} = \sum_{k=1}^{N} I_{\mathrm{ca5}(0.002)}^2$	Zero-recess characteristic

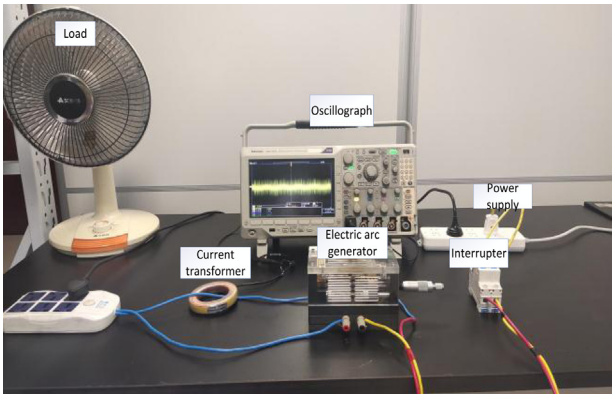

Fig. 5. Physical diagram of experimental circuit.

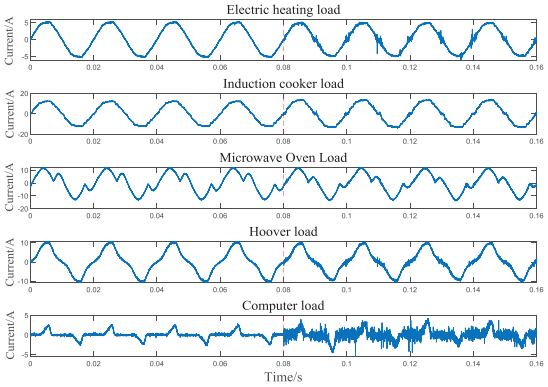

Fig. 6. Comparison of waveforms of different load currents.

4.2 Fisher-score based feature rough selection test

The Fisher-score algorithm is used to calculate the scores of the 25 features selected above respectively and sort them in the order of scores from high to low, Table 2 shows the normalised Fisher-score values of the selected features, the closer the Fisher-score value of a feature is to 1, the higher the importance of the feature.

Figure 7 presents the distribution of eigenvalues for certain features in Table 2 during normal and fault states. The first 4 cycles represent eigenvalues during normal current, while cycles 5–14 represent eigenvalues during fault current. The changes in the distribution of different eigenvalues before and after the fault indicate the impact of each feature on the classification performance of the classifier. It can be observed that the top-ranked feature shows a more distinct difference between normal and fault eigenvalues, whereas the lower-ranking features exhibits overlapping eigenvalues. Consequently the higher-ranked features demonstrate better recognition capabilities in identifying faulty arcs.

Based on the search strategy of sequential forward search [23], the 25 selected features are ranked according to the importance of the features calculated by Fisher-score, and a series of feature subsets $Feature(i) = \{F_1, F_2, ..., F_i\}$ are formulated as the input vectors of the GA-BP neural network. $Feature(i)$ is a subset of features consisting of the top i ($i = 1, 2, ..., 25$) features $F_1, F_2, ..., F_i$ in the ranking.

The parameters of the GA-BP neural network in this paper are shown as follows. The structure of the BP neural network is a total of three layers: input layer, hidden layer and output layer, the number of nodes in the input layer is i ($i = 1, 2, ..., 25$), the number of nodes in the hidden layer is 6, and the number of nodes in the output layer is 2. The parameters of the genetic algorithm are set as follows: the maximum number of iterations is 1000, the population size is 20, the number of evolutionary generations is 50, the learning rate is 0.1, the crossover probability is 0.3, and the mutation rate is 0.1. The number of training samples and test samples for the experiments in this paper are 1200 groups and 500 groups, respectively.

Based on the input features of different dimensions, the curved relationship between input feature dimensions and classification accuracy is plotted as shown in Figure 8:

Figure 8 shows that when the first-ranked feature is input into the GA-BP neural network for arc classification and identification, an accuracy of 96.4218% is achieved, and with the subsequent input of relevant features, the arc fault identification accuracy initially shows an upward trend, but when the input features reach a certain number, even if it continues to increase the number of input feature dimensions, the arc fault identification accuracy no longer changes and even shows a decreasing trend. In this experiment, when the first 15-dimensional features are input into the GA-BP neural network, the recognition accuracy of the system reaches the maximum value of 99.2467% for the first time, and when the feature input continues to be increased, the recognition accuracy of the system shows a decreasing or fluctuating state due to the effect of feature redundancy, and when all the features are input into the system, the system's recognition accuracy still does not reach the maximum value. This shows that Fisher-score feature selection method can effectively reduce the feature dimension and achieve the detection effect of all features by using only a small number of relevant features. According to the experimental results, this paper takes the first 15-dimensional features as the roughly selected feature subset.

4.3 MI-based feature selection test

To enhance the performance of the classifier and remove redundant relationships between features, we utilized the MI algorithm to select the features. We selected a 15-dimensional feature subset that achieved the highest recognition accuracy for the first time. The MI algorithm was then used to calculate the mutual information value between features, resulting in the creation of a mutual information matrix (Tab. 3). This matrix helped us identify mutually independent features and construct the optimal feature subset.

According to the results in Table 3, the features that remain after dimensionality reduction using the MI algorithm are: Maximum current change rate, The maximum value of cd2 layer module, The maximum value of cd1 layer module, The maximum value of cd3 layer module, The maximum value of cd4 layer module, The maximum value of cd5 layer module, The energy of ca5

Table 2. Fisher-score values corresponding to different features.

Ranking	Feature	Fisher-score value
1	Maximum current change rate	1.00
2	The maximum value of cd2 layer module	9.95×10^{-1}
3	The maximum value of cd1 layer module	9.45×10^{-1}
4	The maximum value of cd3 layer module	8.92×10^{-1}
5	The energy of cd1 layer in 0.002 seconds before the current	7.89×10^{-1}
6	The maximum value of cd4 layer module	5.73×10^{-1}
7	The energy of cd2 layer in 0.002 seconds before the current	5.69×10^{-1}
8	The energy of cd3 layer in 0.002 seconds before the current	3.62×10^{-1}
9	The maximum value of cd5 layer module	2.69×10^{-1}
10	The energy of ca5 layer in 0.002 seconds before the current	2.34×10^{-1}
11	The energy of cd3 layer	2.25×10^{-1}
12	The energy of cd4 layer	1.83×10^{-1}
13	The zero rest time within 0.002 seconds before the current	1.65×10^{-1}
14	The energy of cd4 layer in 0.002 seconds before the current	1.32×10^{-1}
15	Cosine similarity	1.08×10^{-1}
16	The energy of cd5 layer in 0.002 seconds before the current	1.05×10^{-1}
17	Peak-to-peak value	1.28×10^{-2}
18	The maximum value of ca5 layer module	2.80×10^{-3}
19	Skewness	2.40×10^{-3}
20	The proportion of 3rd harmonic	2.20×10^{-3}
21	The proportion of fundamental wave	8.38×10^{-4}
22	Variance	6.14×10^{-4}
23	Proportion of 5th harmonic	5.88×10^{-4}
24	Mean value	3.14×10^{-5}
25	Kurtosis	0

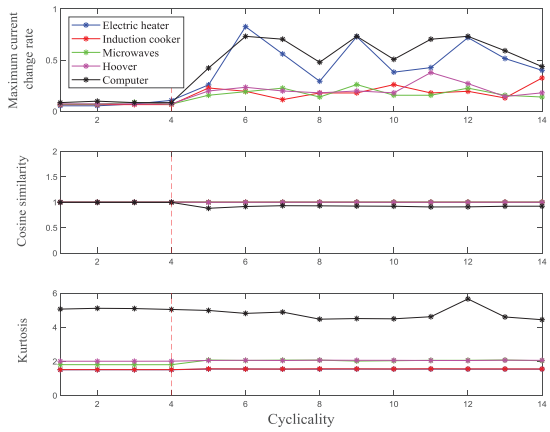

Fig. 7. Distribution of eigenvalues for different feature rankings.

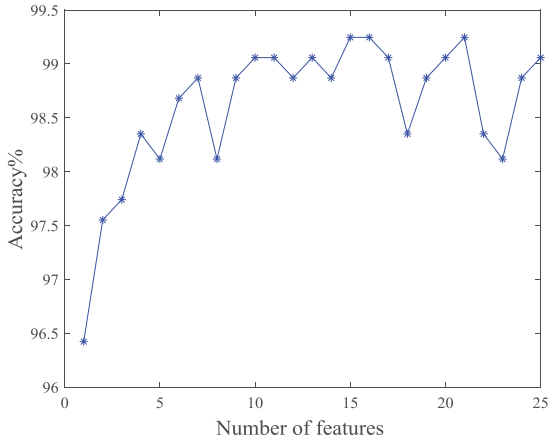

Fig. 8. Input feature vs. accuracy variation curve.

layer in 0.002 seconds before the current, The energy of cd3 layer, The zero rest time within 0.002 seconds before the current, and Cosine similarity. These features are considered the optimal subset for the Fisher-MI feature selection method. The recognition accuracy curve obtained by inputting this subset into the GA-BP neural network, following the aforementioned method in sequence, is shown in Figure 9. From Figure 9, it can be observed that the arc recognition accuracy reaches 99.6234% after removing the redundant features using the MI algorithm. This demonstrates that the dimensionality reduction process of the Fisher-MI feature selection method effectively improves the fault recognition accuracy while reducing the feature dimensionality.

Table 3. Mutual information matrix of roughly selected feature subsets.

Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1.0724	0	1.3194	1.1926	0	0	0	0.0199	0	0.9093	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1.3194	0	3.5603	1.4644	0	0	0	0.0325	0	1.0825	0
8	0	0	0	0	1.1926	0	1.4644	3.7457	0	0	0	0.0412	0	1.3170	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0.0199	0	0.0325	0.0412	0	0	0	0.1003	0	0.0486	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0.9093	0	1.0825	1.3170	0	0	0	0.0486	0	4.4433	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

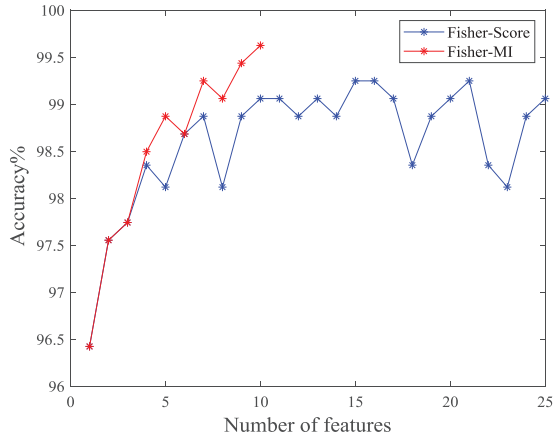

Fig. 9. Variation curve of recognition accuracy of Fisher-MI and Fisher-score feature selection.

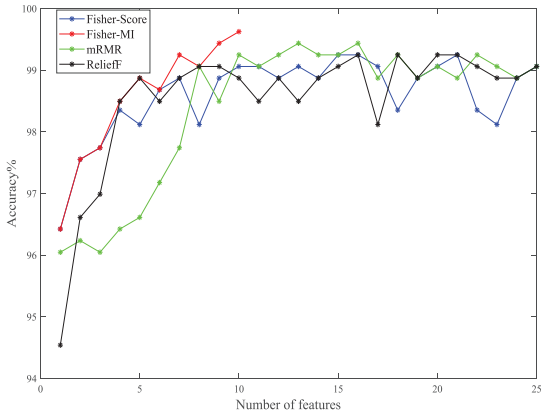

Fig. 10. Comparison of change in accuracy of different feature selection methods.

4.4 Performance comparison of different feature selection methods

In order to verify the advantages of the algorithm proposed in this paper, a comparison is made between other typical filter feature selection methods. The effects of the selected feature subsets from different feature selection methods on the performance of the classifier are evaluated from three perspectives: recognition accuracy, number of selected feature dimensions, and program running time.

Two filter feature selection methods, mRMR [18] and ReliefF [19], are chosen for comparative analysis with the feature selection methods proposed in this paper.

In this case, the mRMR feature selection method is to ensure the direct maximum correlation between features and categories while taking into account the redundancy between features, and when feature selection is performed, it ensures the high correlation between the feature set S and the category c. When two features x_i and x_j are highly correlated, then one of them is removed, which is calculated by the formula:

$$mRMR = \max\left[\frac{1}{|S|}\sum_{x_i} I(x_i, c) - \frac{1}{|S^2|}\sum_{x_i, x_j \in S} I(x_i, x_j)\right]. \tag{4}$$

The ReliefF feature selection method is based on the correlation between the selected features and categories to give different weights to the features, the larger the weight value of the features, the stronger the classification ability. The ReliefF algorithm firstly randomly draws a sample X, there are two nearest-neighbour samples of the sample: samples of the same category T, samples of the different categories F, according to the distance criterion to calculate the distance between the sample X and its nearest-neighbour samples to determine the correlation between sample X and feature A, which are denoted as $d(A,\ X,\ T)$ and $\mathrm{d}(A,\ X,\ F)$. The distance between the

Table 4. Indicators for different feature selection methods.

Feature selection method	Recognition accuracy rate/%	Feature dimension	Program running time/s
No feature selection	99.0584	25	0.1265
Fisher-score	99.2467	15	0.0684
mRMR	99.4350	13	0.0595
ReliefF	99.2467	16	0.0722
Fisher-MI	99.6234	10	0.0512

sample X and the nearest neighbour is calculated according to the distance metric to determine the correlation between the sample X and the feature A, which is denoted as $d(A, X, T)$ and d(A, X, F) respectively. If $d(A, X, T) < d(A, X, F)$, it means that the category is closer to feature A and its correlation with it is higher, then the weight will be increased; on the contrary, the weight will be decreased. Finally repeat n this above process to get the average of the results after n iterations for each feature. The formula for calculating the distance between two samples in feature A is as follows:

$$d(A,X,T) = \begin{cases} \dfrac{X_1[A] - X_2[A]}{\max(A) - min(A)} \\ A\text{continuous0}A\text{discreteand}X_1[A] \\ \quad = X_2[A]1A\text{discreteand}X_1[A] \neq X_2[A]. \end{cases} \tag{5}$$

The formula for its characteristic weights is as follows:

$$W[A] = W[A] - \frac{\sum_{i=1,j=1}^{k} d(A,X,T)}{(m \cdot k)} + \frac{\sum_{C \neq \text{class}(X_i)} \left[\frac{p(C)}{1-P(class(p_i))} \sum_{i=1,j=1}^{k} d(A, R_i, M_j(C)) \right]}{(m \cdot k)}, \tag{6}$$

where $M_j(C)$ represents the jth nearest-neighbour sample among the different classes, $p(C)$ represents the distribution probability of the class, $class(pi)$ denotes the class to which p_i belongs, m is the number of samples sampled, and k is the number of nearest-neighbour samples.

Figure 10 presents the comparison of the change curve of recognition accuracy between the feature selection method proposed in this paper and the two feature selection methods mentioned above. It can be observed that the Fisher-MI and mRMR feature selection methods, which consider both the correlation between features and categories and the redundancy between features, exhibit significantly higher accuracy in recognizing arc faults compared to the Fisher-score and ReliefF feature selection methods that only consider the correlation between features and categories. This implies that effectively reducing the negative impact of redundancy between features can improve the performance of the classifier.

In terms of the feature subset dimension, the mRMR and ReliefF feature selection methods yield feature subsets with dimensions of 13 and 16, respectively, in order to achieve maximum recognition accuracy for fault recognition of arcs. These subsets are considered as the optimal feature subsets selected by the feature selection methods. On the other hand, the feature selection method proposed in this paper results in a final feature dimension of 10. It is evident that this method not only achieves higher arc recognition accuracy but also effectively reduces network complexity, shortens procedure running time, and improves classifier performance.

The experimental environment of this paper uses Intel Core i5-8300H CPU to achieve the verification of the feasibility of the feature selection method in this paper, and the evaluation indexes of different feature selection methods are shown in Table 4.

5 Conclusion

In this paper, after comprehensively considering the relevance of the selected features and the redundancy between the features, a Fisher-MI feature selection method is proposed, in which the features with strong relevance to the classifier are firstly selected by the Fisher-Score algorithm to achieve the rough selection of the feature subset, and the MI algorithm is used to reject the features that are redundant with each other in the feature subset to achieve the selective selection of the feature subset.

The experimental results show that the feature subset filtered by the feature selection method proposed in this paper effectively improves the accuracy of arc recognition in arc fault identification and shortens the time of arc detection, and it is found by comparing with other filtered feature selection that selecting the feature subset with strong relevance and low redundancy can effectively improve the performance of the classifier.

Funding

This work is supported by National Science Foundation of China (Grant No. 52077221).

References

1. H. Gao, W. Wang, T. Shang, L. Zhai, Series arc fault detection method based on wavelet spectral entropy, New Technol. Electr. Eng. **41**, 63–71 (2022)
2. W. Wang, B. Xu, Z. Sun, D. Liang, Differential voltage detection method for arc faults in low voltage distribution networks, Chin. J. Electr. Eng. 1–14 (2023)
3. H. Zhao, H. Qin, K. Liu, L. Zhu, Series fault arc detection method based on correlation theory and zero rest feature fusion, J. Instrum. **41**, 218–228 (2020)
4. S. Ma, J. Bao, Z. Cai, C. Meng, Arc fault identification method based on information dimension and zero rest time, Chin. J. Electr. Eng. **36**, 2572–2579 (2016)
5. J. Cai, Research on fault arc identification method based on the probability distribution of 5th harmonic components[D]. Shenyang University of Technology, 2017
6. R. Ding, Y. Chen, L. Sun, Q. Cheng. Z. Liu, AC series arc fault identification based on neighbouring wave current difference and randomness, Power System Protection and Control. **51**, 169–178 (2023)
7. Y. Wang, Q. Wei, L. Ge, F. Niu, K. Li, L. Zhang, Series AC arc fault detection method based on high frequency component of arc current, Power Autom. Equip. **37**, 191–197 (2017)
8. Y. Wang, M. Tian, H. Xi, Y. Ban, L. Hou, K. Li, Series fault arc detection method based on current similarity and high frequency energy, Electr. Measur. Instrum. **59**, 158–165 (2022)
9. R. Cui, Y. Wang, C. Wang, Y. Li, F. Li, A multi-information fusion based arc identification method for series faults in aviation lines, J. Electrotechnol. **34**, 118–125 (2019)
10. Q. Yu, G. Huang, Y. Yang, Absolute value activated deep neural network for series fault arc detection, Comput. Appl. **39**, 54–59 (2019)
11. T. Zhang, C. Zhang, K. Yang, Arc fault detection system based on asymmetric convolutional neural network, J. Electr. Measur. Instrum. **36**, 116–125 (2022)
12. Y. Wang, Research on the dynamic characteristics and fault detection method of low-voltage AC arc [D]. Shanghai Jiao Tong University, 2020
13. Y. Yang, L. Huang, P. Li, L. Shen, Z. Lv, S. Yang, Arc fault detection method based on multi-dimensional feature extraction, J. Electr. Measur. Instrum. **35**, 107–115 (2021)
14. R. Cui, S. Wang, Aviation series fault arc detection based on multidimensional feature quantity, Sci. Technol. Eng. **17**, 38–43 (2017)
15. T. Shang, W. Wang, J Peng, B. Xu, H. Gao, G. Zhai, Series arc fault identification based on complete ensemble empirical mode decomposition with adaotive noise and convolution neural network, Int. J. Metrol. Qual. Eng. **13**, 11 (2022)
16. R. Chu, R. Zhang, K. Yang, C. Xiao, Series arc fault detection method based on multilayer convolutional neural network, Power Grid Technol. **44**, 4792–4798 (2020)
17. L. Wang, S. Jiang, S. Jiang, A feature selection method via analysis of relevance, redundancy, and interaction, Exp. Syst. Appl. **183**, 115365 (2021)
18. J. Liu, J. Liu, H. Zhao, Z. Wu, X. Liu, L. Wu, Short-term photovoltaic output prediction based on DKDE with improved m-RMR feature selection, Power Syst. Autom. **45**, 13–21 (2021)
19. B. Xue, K. Wen, W. Li, Z. Zhang, Y. Tang, Research on household load type identification method based on feature evaluation and screening by combining ReliefF and mutual information, Electr. Measur. Instrum. **57**, 38–45 (2020)
20. Y. Li, Research on multi-feature aviation arc fault detection algorithm based on pattern recognition [D]. Hebei University of Technology, 2020
21. J. Xu, X. Chen, Y. Dong, J. Yang, Random forest DDoS attack detection with fusion feature selection, Comput. Appl. **21**, 1–8 (2023)
22. P. Li, X. Dong, Q. Meng, J. Ynag, A transient stability assessment method for power systems based on Fisher score feature selection, Power Autom. Equip. **22**, 1–13 (2023)
23. H. Che, F. Lv, Z. Xiang, Defect recognition based on sequential forward floating search with time-frequency preferred features, J. Zhejiang Univ. **45**, 2235–2239 (2011)

Cite this article as: Baichuan Qin, Wei Wang, Wei Hu, Lei Su, Guofeng Zou, Low-voltage AC series arc fault detection based on Fisher-mutual information feature selection, Int. J. Metrol. Qual. Eng. **15**, 3 (2024)