Measurement 188 (2022) 110531

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Fault anomaly detection of synchronous machine winding based on isolation forest and impulse frequency response analysis

Yu Chen [a], Zhongyong Zhao [a,*], Hanzhi Wu [a], Xi Chen [a], Qianbo Xiao [b], Yueqiang Yu [a]

[a] College of Engineering and Technology, Southwest University, Chongqing 400716, China
[b] State Grid Chongqing Electric Power Company, Chongqing 400014, China

ARTICLE INFO

Keywords:
Isolation forest
Unsupervised learning
Impulse frequency response
Anomaly detection
Synchronous machine
Winding fault

ABSTRACT

Synchronous machine is one of the critical power generation parts in the power system. Its stable operation ensures people's normal economic activities. Winding is an essential component of a synchronous machine, and the winding fault is a common fault type. The reliable and efficient fault diagnosis of synchronous machine winding is of great significance to ensure the stability of the power system. Therefore, this paper proposes an anomaly detection method of synchronous machine winding fault based on isolation forest (IF) and impulse frequency response analysis (IFRA). Firstly, the basic principle of the anomaly detection method is introduced, and mathematical-statistical indicators of IFRA signatures used are then explained. Besides, the experimental verification is carried out on a 5 kW synchronous machine, and the performance of the anomaly detection method for winding fault is compared with other conventional methods. The experimental results show that the proposed method is feasible and effective, and the generalization ability of the data is strong. The comparative experimental results show that the proposed method is superior to the existing conventional supervised learning method. It has a shorter calculation time and higher accuracy, with stronger robustness, which is more suitable for the actual data structure.

1. Introduction

A synchronous machine or a generator is one of the crucial parts of a power system, which converts mechanical energy to electrical energy. At present, in all kinds of power generation operation modes, 90% of the power is provided by the synchronous machine [1]. Various studies show that the normal operation of a synchronous machine may be affected by many factors, for instance, thermal stress, electromagnetic stress, mechanical stress, electromagnetic noise, vibration, rotor centrifugal force, insulation material aging, *etc* [2]. These failure factors will cause serious faults to the active or static parts of the generator [3], leading to the shutdown of the entire synchronous machine. The abrupt shutdown of synchronous machines in operation will affect the power supply quota of power plants, which will destroy the power supply–demand relationship [4-7], having a large negative impact on the economy, politics, activities of human beings [5,6]. For example, in 2011, a 300 MW steam turbine generator in Dalong Guizhou power plant suffered an inter-turn short circuit accident of rotor winding, which resulted in an unplanned shutdown for 60 days, resulting in an economic loss of about 400 million dollars [8]. In high voltage machines, the probability of stator failure is 2/3, greater than that of bearings, rotors, and other parts [4]. As an important part of stator, relevant researches [3,6,8] show that winding fault is not only the most common fault in stator but also one of the crucial causes of a machine fault. Therefore, it is necessary to detect and diagnose the winding faults of synchronous machines.

Relevant researchers proposed the electric or current analysis, for instance, the DC resistance measurement method, AC impedance/dielectric dissipation method [10], the recurrent surge oscilloscope (RSO) method [11], open-end transformer method, detection coil method, and traveling wave method [12]. In recent years, some scholars worldwide start to use frequency response analysis (FRA), which has already been widely used in the winding fault diagnosis of power transformers, to detect the winding fault of synchronous machines [3,9,13-25]. The FRA technique can be divided into sweep FRA (SFRA) and impulse FRA (IFRA), according to the nature of the input excitation signal. IFRA, as a derivative branch of FRA, has the advantages of high detection sensitivity, fast diagnosis speed, low cost, and non-destructive process [3,9,13-22]. Its basic principle is that the winding can be equivalent to a broadband equivalent circuit model composed of

* Corresponding author.
E-mail address: zhaozy1988@swu.edu.cn (Z. Zhao).

https://doi.org/10.1016/j.measurement.2021.110531
Received 13 August 2021; Received in revised form 17 November 2021; Accepted 28 November 2021
Available online 2 December 2021

capacitance, resistance, and inductance under the excitation of the high-frequency impulse signal. The faults such as an inter-turn short circuit of windings will inevitably change the structure and parameters of the equivalent circuit model, and the frequency response characteristics of windings will change. Based on this, the fault state of windings can be inferred. Besides, it is noted that there are some other methods in fault detection of machines, such as thermal analysis [26,27], vibration analysis [28], acoustic analysis [29], etc. These methods are significant for the condition maintenance of machines, which can provide guidance to the detection and diagnosis of the fault. However, these methods study the bearing, electric drills, or other mechanical components of the device, which are not focused on diagnosing synchronous machine windings.

Moreover, for the fault diagnosis of synchronous machines based on the FRA method, few reports exist regarding diagnosing the presence of a fault. Some similar research methods can be found in fault diagnosis of transformer winding mechanical deformation. In conclusion, two categories can be summarized as follows: 1. Calculate various mathematical-statistical indicators of FRA curves in the sub-frequency band, including the resonant point-based statistical indicator and entire frequency point-based statistical indicator. The calculated indicator is compared with the threshold value [3,9,15,21,27,30]. 2. The threshold comparison of the first method is replaced with the advanced recognition method based on machine learning-based classifiers to diagnose the fault types [31,32].

In the first method, the winding fault characteristics can be identified based on the resonant point of the FRA signature in Ref. [3]. The influence of various faults on the amplitude and frequency of resonant points of the FRA curve is explored. Nevertheless, this method is vulnerable to noise because the amplitude and frequency of the resonance point may be sensitive to noise, which will have an impact on the accuracy of the method. In Ref. [30], using the correlation coefficient indicator, a method for diagnosing the fault degree of the inter-turn short circuit of the synchronous machine is proposed. This method is simple; however, the used mathematical indicator is a single indicator. The threshold value is fuzzy. It can not be used to judge the winding fault extent of other types of synchronous machines. In the second method, a classification method of winding deformation fault based on support vector machine (SVM) optimized by particle swarm optimization (PSO) is presented in Ref. [31]. This method is fast and accurate; however, a large number of winding fault data will be collected from simulation or experiment before performing the fault diagnosis. The experiment data is obtained from the destructive test, which would cause damage to the experimental winding; nevertheless, the simulated data may not emulate the actual situation of the winding fault. Besides, a neural network fault diagnosis method based on genetic algorithm (GA) optimization is proposed in Ref. [32]. This method is accurate and can extract the mapping relationship between various mathematical-statistical indicators and fault conditions. However, the training process of a network is time-consuming, and the well-trained network may not be suitable for winding of different fault types; thus, the generalization performance of the network is not strong.

In conclusion, to make the fault detection method more convenient and accurate, IF is used to process the IFRA data of synchronous machine windings for the first time to diagnose whether the winding has a fault. The main contributions of this study are as follows:

1) Based on a large number of healthy data and a small amount of faulty data, a new fault diagnosis method of synchronous machine winding which corresponds to this data structure is proposed and introduced. This method is an anomaly detection method, which automatically diagnose a large amount of health data as normal, and a small amount of fault data as abnormal. In addition, this method belongs to unsupervised learning. It does not need to label all data in advance. It can significantly save time in the data collection stage.
2) Compared with other traditional machine learning methods, the proposed method is more suitable for processing data structure in real situations. This method is accurate and rapid.
3) The proposed method can be applied to the periodic detection of synchronous machines. The fault anomaly detection of synchronous machine winding can be detected offline. Preliminary screening is made for subsequent manual fault classification, which saves many resources. Moreover, this method uses the mathematical-statistical indicators to analyze the degree of difference between the faulty IFRA curve and its reference, thus it can provide guidance for unknown faults with indicators deviating from the normal range.

It can be seen that there exist some issues regarding the fault diagnosis of synchronous machines based on the existing FRA method, for instance, time-consuming, low accuracy, demand for massive fault experimental data. Therefore, from another perspective of solving the problem, the fault anomaly detection technique, for the first time, is introduced to fault diagnosis of synchronous machine winding in this study. This paper proposes a fault anomaly detection of synchronous machine winding based on IF and IFRA. The basic principle of anomaly detection based on IF and IFRA is introduced in section II. The mathematical-statistics indicators used in winding fault detection are introduced and selected in section III. Section IV then carries out the experimental verification on a 5 kW, three-phase, synchronous machine. The comparison of the proposed method with other conventional supervised learning methods is presented in section V. Finally, section VI summarizes and concludes the result.

2. Basic principles

2.1. Introduction of anomaly detection theory based on isolation forest

Anomaly detection is targeted at a few unpredictable or uncertain, rare events. It has unique complexity, making general machine learning and deep learning techniques ineffective [33-38].

It is meaningful to use anomaly detection in synchronous machines of different sizes. For small and medium-sized synchronous machines, anomaly detection can replace artificial subjective judgment to detect winding faults. After that, related staff can choose to replace the synchronous machine according to the fault or not, without detecting the fault type, which can significantly improve the accuracy and speed of fault detection. For large synchronous generators, the current related research can not unify the method to identify all kinds of winding faults [3,9,24,25,30]. However, the short time required for anomaly detection can provide information on whether to determine the winding faults before performing recognition, reducing the amount of data calculation in the power system.

Moreover, for the currently unknown types of faults, anomaly detection can make some guidance recommendations. There is a large amount of healthy winding data and a small amount of fault winding data during the research process. The anomaly detection algorithm can show better classification performance than other classifiers in this data structure. A flowchart for the application of anomaly detection is shown in Fig. 1 when the synchronous machines are regularly inspected.

IF is an anomaly detection algorithm proposed by Professor Zhihua Zhou and others at the 8th IEEE Data Mining International Conference in 2008 [35]. It generally targets anomaly points in continuous structured data. The algorithm is an ensemble-based anomaly detection method and therefore has linear time complexity. Compared with other anomaly detection methods, such as the Gaussian mixture model [39], it does not need to calculate the inverse of the covariance matrix and covariance matrix between the indicators. If too many dimensions of the evaluation indicators lead to the matrix inversion into singular values and calculation is slow. The IF algorithm does not need complicated matrix calculation, so it has higher accuracy. Compared with other anomaly detection algorithms, the IF algorithm has a smaller computational load

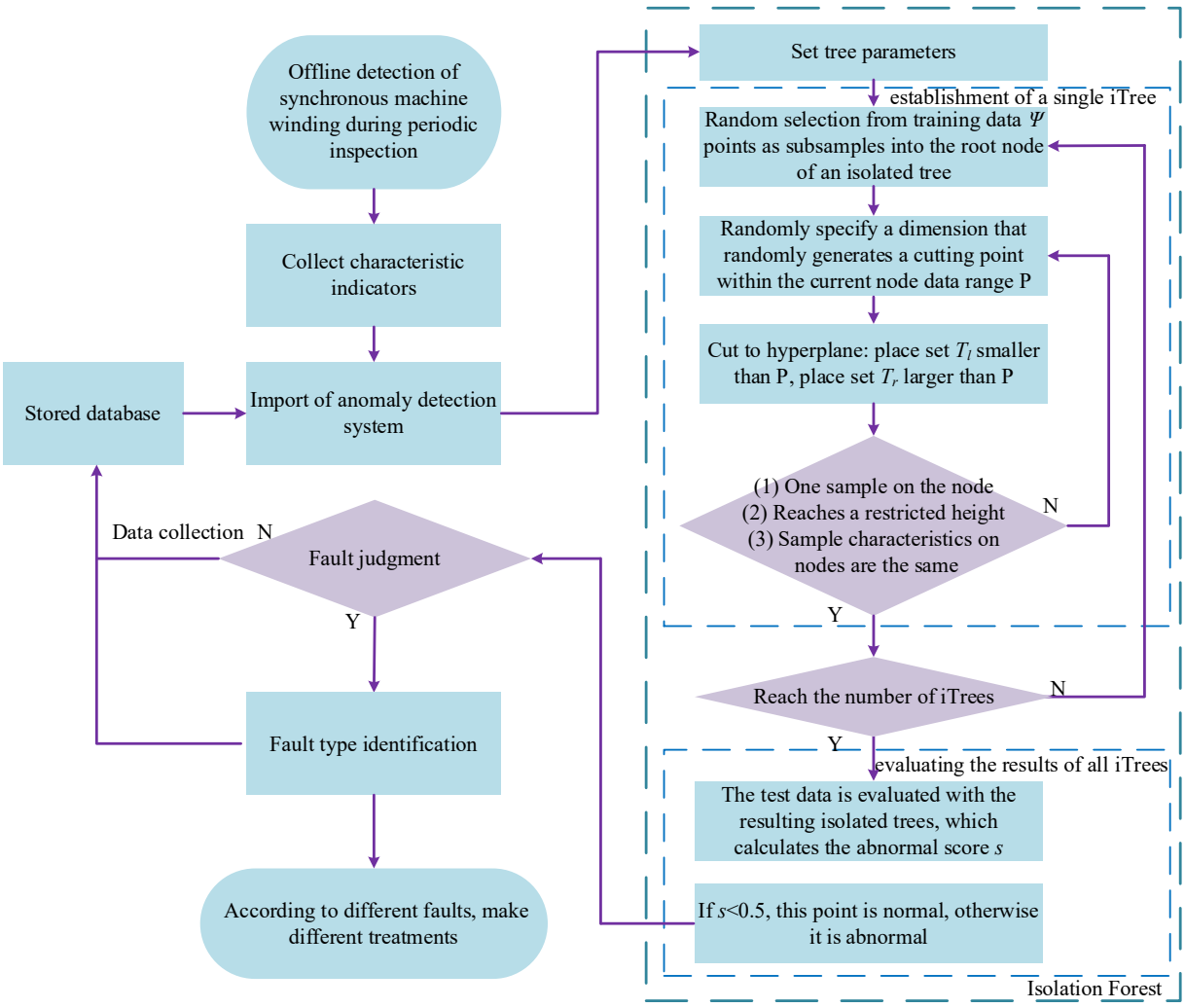

Fig. 1. Flow chart of synchronous machine winding detection system combined with IF.

when dealing with large data, allowing more evaluation indicators, so it is widely used in the industry [36-38].

The IF algorithm is divided into two processes: (1) training of a single iTree (2) evaluating the score of all iTrees. The flow chart of the IF algorithm is shown in Fig. 1.

The most important part in Fig. 1 is the construction of a single iTree. The iTree realizes the separation of data through a random hyperplane to cycle the separation until there is only one data point in the subspace. The normal value is in the high-density area, which requires multiple data isolation and complete separation. The abnormal data is in the low-density area, which can be completely separated after a few isolations. The core problem of IF is the method of separating data. The construction process is as follows:

1) Randomly select Ψ points from training data as subsamples into the root node of an iTree.
2) Randomly specify a dimension that randomly generates a cutting point within the current node data range P.
3) Cut to hyperplane: place set T_l smaller than P, place set T_r larger than P.
4) Go back to step 2) and 3) to form a new node. When the data cannot be segmented or the number of segmentation times reaches $\log_2\Psi$, stop segmentation.

Besides, the algorithm needs to set a few parameters, including the number of iTrees and the number of subsamples. Then repeat the steps in Fig. 1 to establish an isolated forest. It can also be found that the generation of each iTree is independent of each other, indicating that the growth of an isolated forest can grow multiple iTrees at the same time, that is, the isolated forest can support parallel operation. When faced with a large amount of data, the parallel operation algorithm can greatly improve operation efficiency.

Its random process is the idea of Monte Carlo simulation. If the number of iTrees tends to be infinite, according to the law of large numbers, the final score of all samples tends to be a stable value. The abnormal score of each collected sample can be calculated and evaluated by the constructed iFroest. That is, traverse all the iTrees in iForest one by one, extract the node depth of the sample in each iTree, and then calculate the average depth of the sample in iForest, as shown in Fig. 2. Through the further transformation of the depth of iTree, the anomaly score of the sample can be obtained.

Calculate the abnormal score of sample x, as shown in formula (1):

$$s(x,\psi) = 2^{-\frac{E(h(x))}{c(\psi)}} \tag{1}$$

where $h(x)$ is the path length of the sampling point x from the root node of the iTree to the leaf node. $E(h(x))$ is the average path length of the same sample in all iTrees. $c(\psi)$ represents the average depth of iTrees in the constructed iForest. ψ is the number of subsamples.

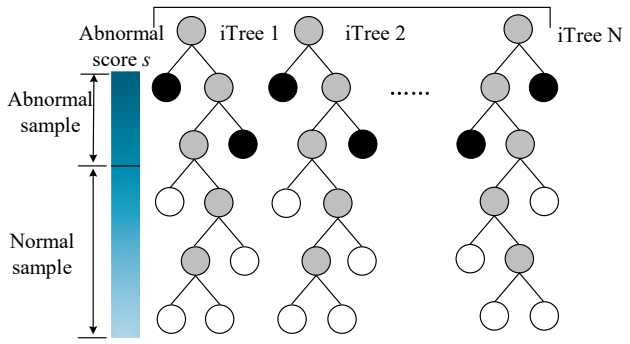

Fig. 2. Detection process of IF.

$$c(\psi) = \begin{cases} 2H(\psi-1) - 2(\psi-1)/\psi \ , \psi > 2 \\ 1 \ , \psi = 2 \\ 0 \ , otherwise \end{cases} \quad (2)$$

where $H(i)$ is a harmonic number, and it can be estimated as $\ln(i) + 0.5772156649$.

The closer the $s(x)$ score is to 1, the higher the probability that the sample is an abnormal point. The closer it is to 0, the greater the probability that it is a normal sample. Furthermore, set 0.5 as the classification threshold regularly.

2.2. Principle of impulse frequency response analysis

In recent years, many scholars have used the IFRA to extract the winding fault characteristics of transformers and achieved good results [13-23]. In this paper, according to the application method of impulse frequency response in transformer windings, it is applied to the diagnosis extraction of the synchronous generator winding faults [3,9,24,25,30]. In Ref. [30], the related researchers have constructed the winding model of the synchronous machine in Fig. 3. Moreover, the frequency characteristic curve of the circuit model is in good agreement with the actual frequency characteristic curve. Changing the parameter value of the basic circuit can simulate the corresponding fault type, which also confirmed the feasibility of IFRA to judge the winding fault of synchronous machines.

First, the impulse signal V_{in} is injected into the first section of the winding by the pulse generator. The current response signal I_{out} is collected at the end of the winding. Then the fast Fourier transform (FFT) is applied to the input impulse signal, and the current response signal and their frequency spectrum values are obtained:

$$V_{in}(k) = \sum_{n=0}^{N-1} V_{in}(n) e^{-j\frac{2\pi}{N}kn} \quad (3)$$

$$I_{out}(k) = \sum_{n=0}^{N-1} I_{out}(n) e^{-j\frac{2\pi}{N}kn} \quad (4)$$

where $V_{in}(n)$ is the N points sampling signal of excitation voltage, $I_{out}(n)$ is the N points sampling signal of response current; $V_{in}(k)$ and $I_{out}(k)$ are fast Fourier transforms of $V_{in}(n)$ and $I_{out}(n)$.

Finally, the transfer function $H(dB)$ expressed as the gain is obtained:

$$H(dB) = 20log_{10}\frac{|I_{out}(k)|}{|V_{in}(k)|} \quad (5)$$

This transfer function can be used as the characteristic information of the winding to characterize the healthy state of the winding at this time. Researchers often refer to this transfer function as the "fingerprint" of the winding's gain image at different frequencies [31,40,41].

3. Indicators selection

Compared with the use of full-band frequency information, the use of mathematical-statistics indicators can simplify the amount of data, accelerate the calculation speed and reduce the dimension of indicators [31,40]. Moreover, the abnormal winding fingerprint can be transformed into outlier different from the healthy winding fingerprint in multi-dimensional space. The impulse frequency response of faulty winding becomes the abnormal point in anomaly detection. According to the reference [31], this paper takes Eq. (5) of healthy winding at different frequencies as the benchmark to construct the mathematical-statistical indicators between each sample and healthy sample.

According to Ref. [41], all selected indicators should have a linear relationship with fault degree. The flow chart of selecting indicators combined with IFRA and IF is shown in Fig. 4. According to Ref. [31,41], the indicators included are shown in Table 1.

In Table 1, n is the number of samples, Y_i is the value of health IFRA, and X_i is the value of other winding's IFRA.

Correlation coefficient and Euclidean distance are usually the two most common indicators to judge winding fault based on the frequency response curve. The correlation coefficient can describe whether the relationship between healthy data and measured data is close or not and can also describe the similarity between them to a certain extent. Euclidean distance describes the actual distance between two points in multidimensional space. The maximum of difference represents the maximum deviation between the measured data and the healthy data, which is often used to reflect the resonance point offset caused by the fault of the winding. The integral of absolute difference and the sum squared ratio error reflects the "gap" between the measured and healthy data. They can reflect the up and downshift of frequency response amplitude caused by a fault. Sum squared and root mean square errors can reflect the dispersion degree between fault data and normal data and measure their deviation. The sum squared max–min ratio error can highlight the influence of relatively large errors and reduce relatively

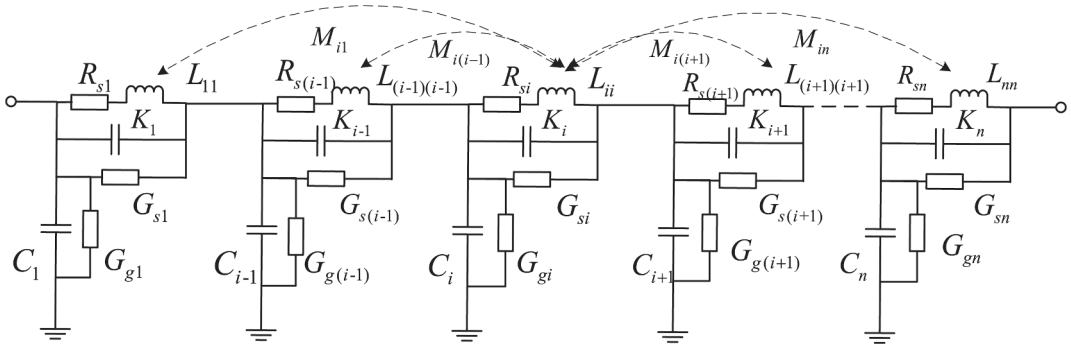

Fig. 3. Circuit model structure of synchronous machine [30].

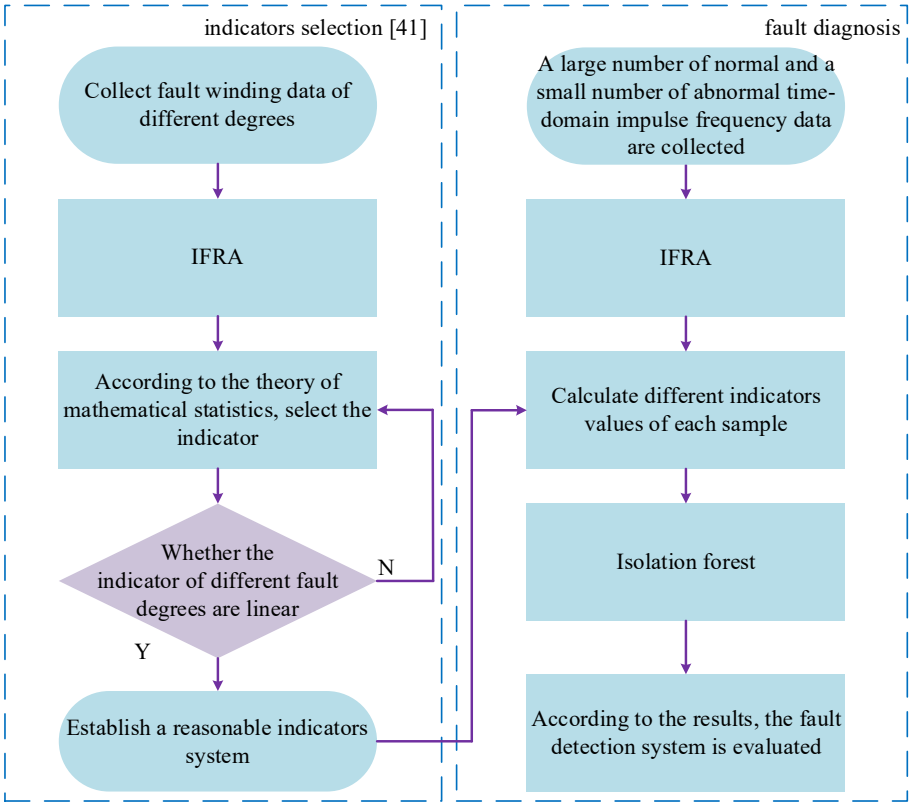

Fig. 4. Establish the flow chart of the fault detection system.

Table 1
Mathematical-statistical indicators.

Indicator name	Calculation formula
Correlation coefficient (CC)	$\frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n}[X_i]^2 \sum_{i=1}^{n}[Y_i]^2}}$
Euclidean distance (ED)	$\sqrt{\sum_{i=1}^{n}(Y_i - X_i)^2}$
Maximum of difference (MAX)	$max\|(Y_i - X_i)\|$
Integral of absolute difference (IA)	$\frac{1}{n}\sum_{i=1}^{n}(Y_i - X_i)^2$
Sum squared error (SSE)	$\frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i}{X_i} - 1)^2$
Sum squared ratio error (SSRE)	$\int \|Y_i - X_i\| df$
Sum squared max–min ratio error (SSMMRE)	$\frac{1}{n}\sum_{i=1}^{n}(\frac{max(Y_i, X_i)}{min(Y_i, X_i)} - 1)^2$
Root mean square error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{\|Y_i\| - \|X_i\|}{\frac{1}{n}\sum_{i=1}^{n}\|X_i\|})^2}$

small errors. In Table 1, the calculation of these indicators is closely related to the health samples. To a certain extent, these indicators extracted from the impulse frequency response can effectively increase the difference between the fault winding and the health winding and ensure the possibility of anomaly detection.

References [3,9,23,30,31,40,41] show that the impulse frequency response (1 kHz–1 MHz) can effectively reflect the characteristic information of the difference between the healthy fingerprint and the fault fingerprint of transformer windings, so this paper only calculates the mathematical-statistical indicators within this frequency interval.

4. Experimental validation

IFRA measurement is carried out in this experiment on a 5 kW, three-phase, salient pole synchronous machine winding. The nameplate of the synchronous machine is shown in Table 2. Fig. 5(a) shows the measurement wiring diagram, and the actual experimental site diagram is shown in Fig. 5(b). The measuring system comprises a homemade pulse generator, a current sensor, a voltage probe, and an oscilloscope. The pulse generator is mainly composed of the Marx circuit, power module, drive circuit, etc., as shown in Fig. 6. The pulse generator can continuously adjust the pulse amplitude of 0–4 kV, the leading edge time is within 40 ns, and the pulse width is adjusted in 10–1000 ns. In Fig. 6, the excitation voltage and response current are measured through a broadband voltage probe (model: P5100a, Tektronix, Oregon, America) and a current sensor (model:150, Pearson, California, America), respectively. The waveform is recorded by an oscilloscope (model: MDO4104C, Tektronix, Oregon, America). Many IFRA tests are carried out under the same machine state. The excitation and response signals of 64 measurements are averaged to reduce the impact of white noise on the measurement. Then the average signal is used for analysis as a time-domain signal. Here is the setup of pulse parameter and sampling, the pulse amplitude is 480 V, the pulse width is less than 700 ns, the rising time is less than 50 ns, as shown in Fig. 7. And the sampling rate is 25 MHz, and the sampling point is 10 k.

Table 2
Nameplate values of synchronous machine

Characteristics	Parameter value
Rated power	5 kW
Rated voltage	380 V
Frequency	50 Hz
Pole pairs	1
Number of slots	36
Rated speed	1500 rpm

In addition, the angular placement of the rotor will have a certain impact on the IFRA curve of the stator. Therefore, in experimental validation, only the stator winding of the synchronous machine is tested

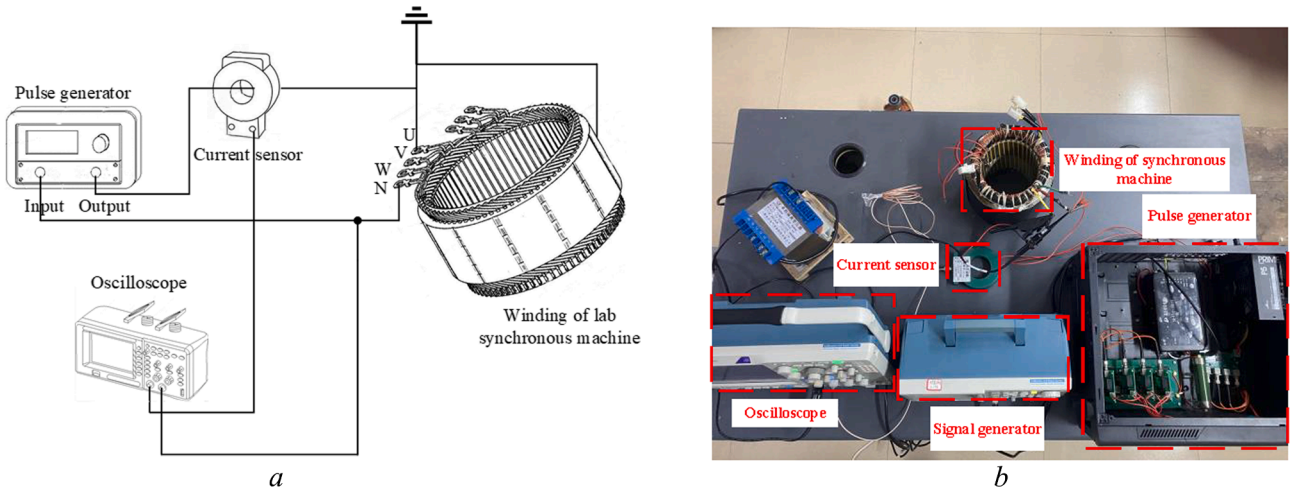

Fig. 5. Measurement experiment diagram **(a)** measurement wiring diagram, **(b)** actual wiring diagram.

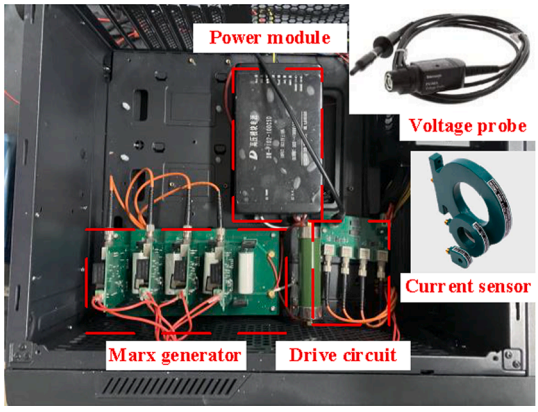

Fig. 6. Structure of pulse generator, current sensor and voltage probe.

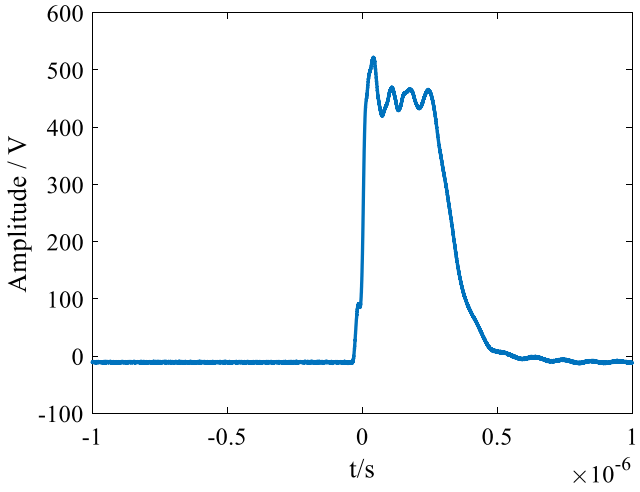

Fig. 7. Pulse waveform.

without any rotor installed.

The measurement wiring diagram without any simulated fault is shown in Fig. 5. After that, the inter-turn and ground faults of the motor are simulated according to the literature [3,9]. The experiment simulates different degrees of inter-turn faults by the parallel inductance of 0.1/10/46/100/330uH between turns. Because the inter-turn short circuit fault will reduce the effective number of turns, inducing the variation of inductance [42]. Therefore, turn to turn parallel inductance is selected to simulate inductance reduction [30]. The simulated wiring diagram of inter-turn faults is shown in Fig. 8. Experiment by paralleling 210/140/70/35/21 Ω resistance at a point in the winding simulates varying degrees of grounding failure. The value of parallel inductance and resistance are selected by order of magnitude of inductance and resistance in the circuit model established in Fig. 3 [3,9,17,30,43,44]. The connection diagram for the ground fault is shown in Fig. 9.

To generate the data structure with a large amount of healthy winding information and a small amount of fault winding information, a total of 320 impulse frequency response data are measured, including the impulse frequency response of 300 healthy windings, two groups (2 × 5) of inter turn faults with different degrees, and two groups (2 × 5) of grounding faults with different degrees. The measured time-domain waveforms are then changed by FFT. The IFRA of 15 healthy windings is shown in Fig. 10, 5 examples of inter-turn faults of different degrees are offered in Fig. 11, and 5 samples of grounding faults of different degrees are shown in Fig. 12.

As shown in Fig. 10, noise interference during the measurement process can cause jitter in the impulse frequency curves of healthy windings. These result in slight differences between the mathematic-statistical indicators of healthy windings, which is consistent with the anomaly detection algorithm's requirements of normal data.

As shown in Figs. 11 and 12, all kinds of faults result in an impulse frequency response curve that is different from the healthy windings, and winding faults translate into farther outliers in multidimensional space. Faults can cause large fluctuations in the mathematical-statistical

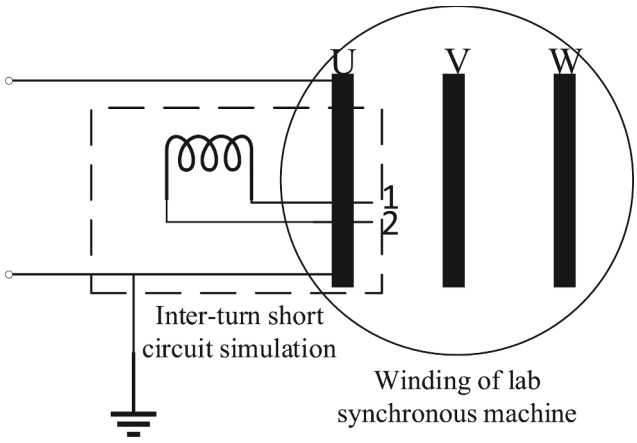

Fig. 8. Simulated wiring diagram of inter-turn faults.

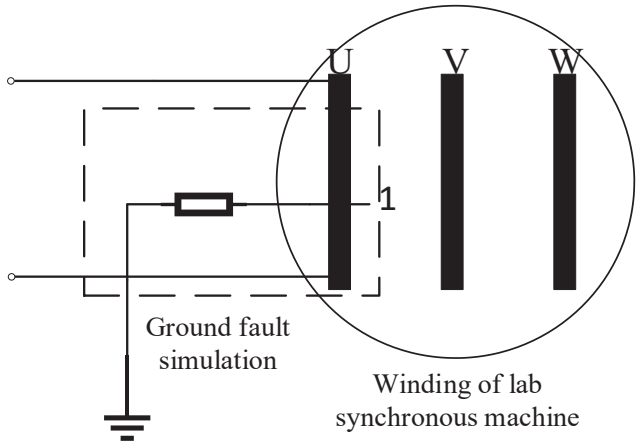

Fig. 9. Simulated wiring diagram of ground faults.

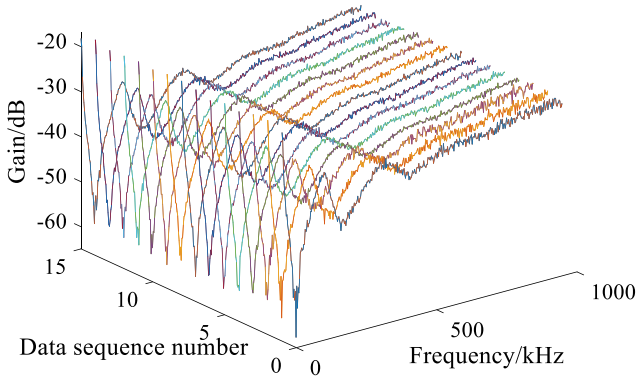

Fig. 10. IFRA of the healthy winding.

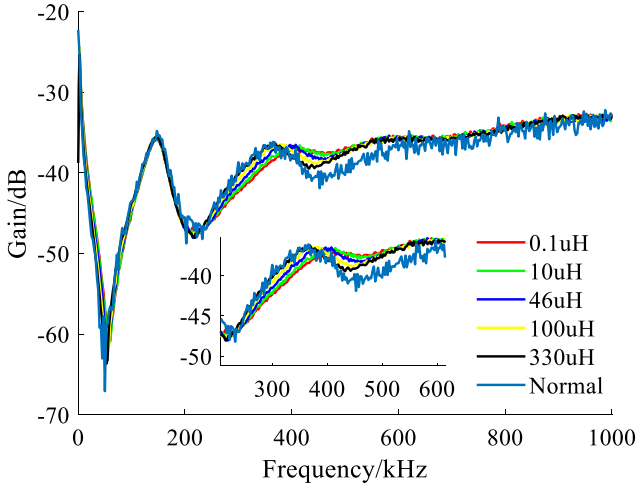

Fig. 11. IFRA of winding inter-turn short circuit faults in different degrees.

indicators compared with the healthy windings, which can be detected as outliers in multidimensional space by the anomaly algorithm through anomaly indicators.

In the anomaly detection algorithm, because of its particular data structure, the experiment can not use the accuracy of the test set to measure the ability of the algorithm to identify anomalies, namely, the quality of fault detection [35]. Therefore, a new evaluation indicator,

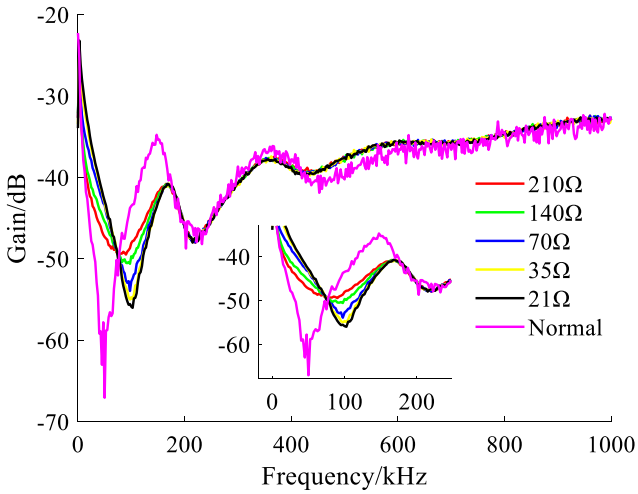

Fig. 12. IFRA of winding grounding faults in different degrees.

the area under the receiver operating characteristic curve (AUC), is introduced [38,45]. The closer the AUC is to 1, the better the performance of anomaly detection is. According to a series of dichotomous ways, the receiver operating characteristic curve (ROC) is drawn, the true-positive rate is the vertical axis, and the false-positive rate is the horizontal axis:

$$TRRate = \frac{TP}{TP + FN} \tag{6}$$

$$FPRate = \frac{FP}{FP + TN} \tag{7}$$

where *TP* is the number of true positives, *FP* is the number of false positives, *FN* is the number of false negatives, and *TN* is the number of true negatives. The meaning of *TPrate* is the proportion of all samples whose real category is 1 and whose prediction category is 1. The meaning of *FPrate* is the proportion of all samples with 0 real class and 1 predicted class.

AUC is calculated as follows:

$$AUC = \frac{\sum_{i \in positiveClass} rank_i - \frac{P(1+P)}{2}}{P \times N} \tag{8}$$

where, *rank* is the positive sample starting from 1 to the end, *P* is the number of positive samples, and *N* is the number of negative samples.

The IF algorithm is used to detect 320 samples of data. The number of iTrees is 100, the number of iterations is 30, and the subsample size is 64. The algorithm itself is unsupervised learning and does not need any labels on the data. By default, it judges a large number of similar data as normal and a small number of data with large differences as abnormal. However, to evaluate this algorithm's performance, the label is brought into the calculation of AUC, and the average AUC size of each generation is 0.9947, and the training time is 6.32 s. It does not need to train the system to identify the various characteristic information of the impulse frequency response, and it only determines whether the winding is faulty. In addition, the parallel calculation is carried out when calculating different iTrees. Those are two reasons for the fast calculation speed of the proposed method. The ROC curve is shown in Fig. 13, and the AUC value of 30 iterations is shown in Fig. 14.

It can be seen from Fig. 13 that the value of the AUC of the IF algorithm is large. From the value of AUC, this algorithm is suitable for this data structure, namely, ideal for fault diagnosis of synchronous machine winding. This algorithm can accurately identify the abnormal in mathematical-statistics indicators, namely, the impulse frequency response of fault winding.

It can be seen from Fig. 14 that the AUC value of the IF algorithm in

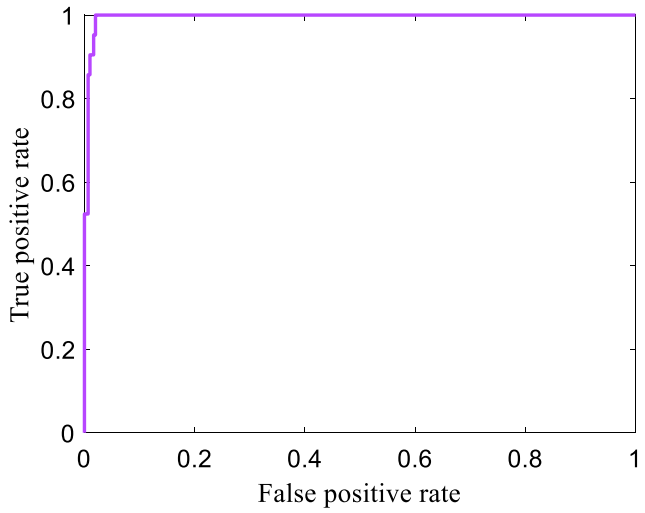

Fig. 13. Receiver operating characteristic curve.

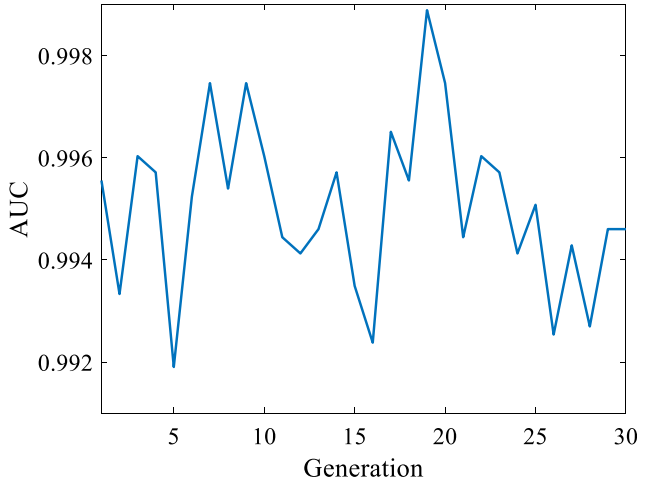

Fig. 14. AUC values at different generations.

repeated generations is above 0.99, indicating that this algorithm has certain advantages in accuracy. There is no test set in the sense of theory in the IF algorithm, but the performance of each generation can be regarded as the test set verification of supervised learning. AUC value is larger and more stable in each generation, which shows that this proposed method has a strong generalization ability in winding fault detection.

5. Comparative experiment and analysis

In order to explore the effect of different parameters on the algorithm's performance, comparative experiments are analyzed. Other setting parameters are the same, and the performance of the IF algorithm is shown in Table 3 under the different number of subsamples. "Time" in the tables indicates the time taken for the algorithm to run once.

Table 3
Performance comparison table of isolation forest algorithm under different number of subsamples.

Number of subsamples	32	64	128	256
AUC	0.9845	0.9965	0.9865	0.9847
Time	0.98 s	1.32 s	2.44 s	6.56 s

Before sampling, normal and abnormal samples may be overlap, so it isn't easy to separate them. However, after sampling, abnormal and normal samples can easily be separated. In addition, sampling can reduce the waste of computing time and space. An appropriate subsample size can also get a good AUC value in this experiment, closely related to the experiment's sample size. However, when the subsample size is more than 320, the AUC value tends to be stable because the subsample size is larger than the sample size, and resampling is also a duplicate sample. Therefore, in this experiment, the relatively appropriate amount of sampling can significantly reduce the operation time, and the AUC value is also high. Consequently, it is suggested that 20% of the sample size as a subsample size can provide a higher AUC value and reduce the calculation amount.

The different number of iTrees will also have a certain impact on the experiment. When other setting parameters are the same, the performance comparison table of the IF algorithm under the different number of iTrees is shown in Table 4.

It can be seen from Table 4 that if the number of iTrees is relatively large, the AUC value will tend to be stable, but the calculation speed will slow down. When the number of iTrees is relatively small, the AUC value fluctuates to a certain extent, which is related to the fact that Equation (1) can not approach stability when it is less than 100 but can approach stability after 100. This experiment gets the same conclusion as Ref. [35]: the number of iTrees should be more than 100.

In order to illustrate the superiority of this algorithm, supervised learning methods such as SVM, k-nearest neighbor (KNN), random decision forests (RDF), and BP neural network are compared with the proposed method in this paper. The results are shown in Table 5. "Accuracy" in the tables represents the probability of correct prediction in the test set. In this paper, the accuracy of the last generation of the proposed method is taken as the accuracy of the IF test set.

It can be seen from Table 5 that compared with other algorithms, and the IF algorithm has more strong adaptability to the data structure provided in this paper, stronger robustness, and less time. The AUC of other algorithms is small because the accuracy of the test set is usually used as the objective function in the process of generation, which is not suitable for this kind of data structure.

Take the KNN algorithm as an example, and its confusion matrix is shown in Fig. 15. Although its accuracy is 90.9% in the test set, its AUC is 0.70. It can be seen from Fig. 15 that KNN has lost the ability to judge because it considers all the data as normal data. If KNN is used, it can not effectively carry out unsupervised learning on the fault winding because the outliers in the data can not be clustered. The accuracy of RDF on the test set is 93.2%, but its AUC is only 0.74. It also shows that the accuracy of the test set can not reflect the classification ability of the algorithm in this data structure. In addition, although the experimental results of BP and SVM are relatively good, they only train the model for this experimental data, and their generalization ability is weak. The trained model is not suitable for other windings' data.

It can be seen from the result analysis in Tables 3–5:

1. Reasonable selection of hyperparameters values can improve the performance of the classification algorithm. However, in the case of fault anomaly detection of synchronous machine windings, the change of hyperparameters does not induce a significant variation in performance (all AUC values are larger than 0.98). It verifies that the classifier used in this study is suitable for the provided data structure: a small sample size with a large amount of normal data and a small amount of failure data [35,45].

Table 4
Comparison table of different methods

Number of iTree	50	100	150	200
AUC	0.9843	0.9947	0.9948	0.9948
Time	1.32 s	6.56 s	8.40 s	12.73 s

Table 5
Comparison table of different methods

Method	IF	SVM	KNN	RDF	BP
AUC	0.99	0.74	0.70	0.74	0.95
Time	6.32 s	5.44 s	6.56 s	7.36 s	148.56 s
Accuracy	100%	90.9%	90.9%	93.2%	97.7%

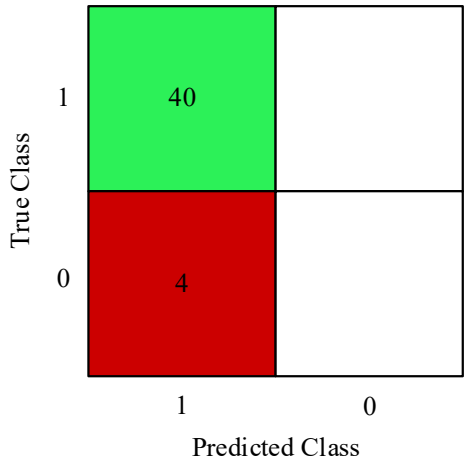

Fig. 15. Confusion matrix of KNN algorithm.

2. If the data structure is not uniformly distributed, the classifier tends to classify all samples into the category of a large number of samples [45]. Therefore, it is unreasonable to use the accuracy of the test set to judge the algorithm's performance. Nevertheless, the AUC can avoid this problem, and the value of AUC can be used to evaluate the performance of binary classification algorithms for unevenly distributed data [38]. From Table 5, the poor classification performance of other traditional methods is due to the unique robustness of these algorithms, which ignores the characteristics of this data structure.

In addition, this study also uses other unsupervised learning, such as Gaussian mixture model [39]. However, when using the same data and indicators for classification, the covariance matrix falls into singular values. Hence it will not be involved here, and it also proves that the multivariate Gaussian model is not suitable for high-dimensional data.

6. Conclusion

This study proposes a fault anomaly detection method of synchronous machine winding based on IF and IFRA. Firstly, we introduce the principle of the IF algorithm and IFRA. Secondly, the mathematical-statistical indicators of IFRA curves are selected and analyzed. Then, the experimental verification of fault anomaly detection is carried out on a 5 kW synchronous machine by artificially simulating winding short circuit fault and fault classification. Finally, the performance of other traditional machine learning classification methods is compared to that of the proposed method. The following conclusions are drawn:

1) As an unsupervised learning method, this method does not need to diagnose the machine fault type in advance but only collects a large amount of normal data and a small amount of abnormal data. This method automatically analyzes a large amount of health data as normal. It considers a small amount of failure data abnormal, which is more suitable for the actual working state. It also conforms to the fact that there are many normal data and a small amount of fault data in practice. Because of the adaptability of this data structure, the classification ability of the anomaly detection algorithm is stronger than other classifiers.
2) The proposed method establishes a fault evaluation indicator system for synchronous machine windings. The value of each indicator can be linearly related to the severity of the fault, so this indicator system has a more comprehensive ability to evaluate the fault than a single evaluation indicator.
3) The proposed method has a strong ability to distinguish faults and a fast calculation speed for periodic fault detection of synchronous machine winding. At the same time, this offline detection can also be carried out before the synchronous machine is cold standby to ensure that the synchronous machine put into the power system has no winding failures. The methods mentioned in this paper have strong industrial application ability.
4) Traditional identification methods use many artificially simulated faults, which are different from the real fault information in fact. However, the method proposed in this paper can overcome this difficulty. It has certain guidelines for unknown winding fault types, ensuring that no faulty machines with doubts are used.

In the future, the proposed method will be applied to the detection of large synchronous machines in power plants, the corresponding improvements will also be made according to the results, and finally, we hope to make available offline detection equipment for practical application.

CRediT authorship contribution statement

Yu Chen: Writing – original draft, Methodology, Validation. **Zhongyong Zhao:** Supervision. **Hanzhi Wu:** Software. **Xi Chen:** Investigation. **Qianbo Xiao:** Conceptualization. **Yueqiang Yu:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 51807166, the Natural Science Foundation of Chongqing under Grant cstc2019jcyj-msxmX0236, and the Venture and Innovation Support Program for Chongqing Overseas Returnees under Grant cx2019123.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.measurement.2021.110531.

References

[1] H.A. Toliyat, S. Nandi, S. Choi, et al., Electric machines: modeling, condition monitoring and fault diagnosis, CRC Press, 2012.
[2] H. Henao, et al., Trends in fault diagnosis for electrical machines: a review of diagnostic techniques, IEEE Ind. Electron. 8 (2) (2014) 31–42.
[3] F.R. Blánquez, C.A. Platero, E. Rebollo, F. Blázquez, Field-winding fault detection in synchronous machines with static excitation through frequency response analysis, Int. J. Electr. Power Energy Syst. 73 (2015) 229–239.
[4] K.N. Gyftakis, A.J. Marques-Cardoso, Reliable Detection of Very Low Severity Level Stator Inter-Turn Faults in Induction Motors, in: IECON 2019–45th Annual Conference of the IEEE Industrial Electronics Society, 2019, pp. 1290–1295.
[5] A. Mugarra, C.A. Platero, J.A. Martínez, U. Albizuri-Txurruka, Validity of Frequency Response Analysis (FRA) for Diagnosing Large Salient Poles of Synchronouos Machines, IEEE Trans. Ind. Appl. 56 (1) (2020) 226–234.
[6] N. Jordan Jameson, M.H. Azarian, M. Pecht, Improved electromagnetic coil insulation health monitoring using equivalent circuit model analysis, Int. J. Electr. Power Energy Syst. 119 (2020), 105829.

[7] M. Cuevas, R. Romary, J. Lecointe, F. Morganti, T. Jacq, Noninvasive detection of winding short-circuit faults in salient pole synchronous machine with squirrel-cage damper, IEEE Trans. Ind. Appl. 54 (6) (2018) 5988–5997.
[8] J. Zhang, Study on electromagnetic characteristic analysis and fault diagnosis method of generator under rotor winding inter-turn short circuit. Dissertation for the master degree in engineering, North China Electric Power University, 2016 (in Chinese).
[9] F.R. Blánquez, C.A. Platero, E. Rebollo F. Blánquez, Evaluation of the applicability of FRA for inter-turn fault detection in stator windings, 2013 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), Valencia, 2013, pp. 177-182.
[10] Z.D. Jia, H. Lu, Z.P. Zhang, et al., Modeling simulation and diagnosis analysis of inter-turn short circuits fault in generator rotor, High Voltage Eng. 45 (12) (2019) 3932–3940 (in Chinese).
[11] T. Li, Simulation study on interturn short circuit of rotor windings in generator by RSO method, ICHVE, ATHENS, Greece, 2018, pp. 1-4.
[12] G. Capolino, A. Cavagnino, et al., New trends in electrical machines technology—Part II, IEEE Trans. Ind. Electron. 61 (9) (2014) 4931–4936.
[13] R.K. Senobari, J. Sadeh, H. Borsi, Frequency response analysis (FRA) of transformers as a tool for fault detection and location: a review, Electr. Pow. Syst. Res. 155 (2018) 172–183.
[14] D. Smugala, M. Bonk, R. Ziemski, Single-phase magnetic cores' faults diagnosis using FRA approach, Measurement 114 (2018) 428–435.
[15] E.G. Luna, G.A. Mayor, C.G. Garcia, et al., Current status and future trends in frequency-response analysis with a transformer in service, IEEE Trans. Power Del. 28 (2) (2013) 1024–1031.
[16] M. Wang, A.J. Vandermaar, K.D. Srivastava, Improved detection of power transformer winding movement by extending the fra high frequency range, IEEE Trans. Power Del. 20 (3) (2005) 1930–1938.
[17] V.A. Lavrinovich, A.V. Mytnikov, Development of pulsed method for diagnostics of transformer windings based on short probe impulse, IEEE Trans. Dielectr. Electr. Insul. 22 (4) (2015) 2041–2045.
[18] E.G. Luna, G.A. Mayor, J.P. Guerra, et al., Application of wavelet transform to obtain the frequency response of a transformer from transient signals—part 1: theoretical analysis, IEEE Trans. Power Del. 28 (3) (2013) 1709–1714.
[19] C.G. Yao, Z.Y. Zhao, Y. Chen, et al., Transformer winding deformation diagnostic system using online high frequency signal injection by capacitive coupling, IEEE Trans. Dielectr. Electr. Insul. 21 (4) (2014) 1486–1492.
[20] E.G. Luna, G.A. Mayor, J.P. Guerra, et al., Application of wavelet transform to obtain the frequency response of a transformer from transient signals—part ii: practical assessment and validation, IEEE Trans. Power Del. 29 (5) (2014) 2231–2238.
[21] A. Abu-Siada, N. Hashemnia, S. Islam, et al., Understanding power transformer frequency response analysis signatures, IEEE Electr. Insul. Mag. 29 (3) (2013) 48–56.
[22] A. Shintemirov, W.H. Tang, Q.H. Wu, Hybrid winding model of disc-type power transformers for frequency response analysis, IEEE Trans. Power Del. 24 (2) (2009) 730–739.
[23] N. Abeywickrama, Y.V. Serdyuk, S.M. Gubanski, High-frequency modeling of power transformers for use in frequency response analysis (FRA), IEEE Trans. Power Del. 21 (3) (2006) 1375–1382.
[24] C.A.P. Gaona, F. Blázquez, P. Frías, M. Redondo, A novel rotor ground-fault-detection technique for synchronous machines with static excitation, IEEE Trans. Energy Convers. 25 (4) (2010) 965–973.
[25] C.A. Platero, F. Blázquez, P. Frías, M. Pardo, New on-line rotor ground fault location method for synchronous machines with static excitation, IEEE Trans. Energy Convers., 26, (2), pp. 572-580.
[26] A. Glowacz, Ventilation diagnosis of angle grinder using thermal imaging, Sensors 21 (8) (2021) 2853.
[27] A. Glowacz, Fault diagnosis of electric impact drills using thermal imaging, Measurement 171 (2021) 0263–2241.
[28] A. Kumar, G. Vashishtha, C.P. Gandhi, Y. Zhou, A. Glowacz, J. Xiang, Novel Convolutional Neural Network (NCNN) for the Diagnosis of Bearing Defects in Rotary Machinery, IEEE Trans. Instrum. Meas. 70 (2021) 1–10.
[29] A. Glowacz, R. Tadeusiewicz, S. Legutko, et al., Fault diagnosis of angle grinders and electric impact drills using acoustic signals, Appl. Acoust. 179 (2021) 0003–682X.
[30] Z. Zhao, Y. Chen, Y. Yu, M. Han, C. Tang, C. Yao, Equivalent broadband electrical circuit of synchronous machine winding for frequency response analysis based on gray box model, IEEE Trans. Energy Convers., doi: 10.1109/TEC.2021.3081933 (early access).
[31] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, S. Islam, Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine, IEEE Access 7 (2019) 112494–112504.
[32] W. Shuting, H. Peng, Rotor winding inter-turn short circuit fault diaginosis system based on artificial neural network, 2007 8th International Conference on Electronic Measurement and Instruments, 2007, pp. 3-581-3-585.
[33] Lingqiang Xie, Dechang Pi, Xiangyan Zhang, Junfu Chen, Yi Luo, Wen Yu, Graph neural network approach for anomaly detection, Measurement, 2021, pp. 0263-2241.
[34] Longmei Li, Ruifeng Yang, Chenxia Guo, Shuangchao Ge, Binglu Chang, The data learning and anomaly detection based on the rudder system testing facility, Measurement 152 (2020) 0263–2241.
[35] F.T. Liu, K.M. Ting, Z. Zhou, Isolation Forest, 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 413-422.
[36] K. Sadaf, J. Sultana, Intrusion detection based on autoencoder and isolation forest in fog computing, IEEE Access 8 (2020) 167059–167068.
[37] E. Khaledian, S. Pandey, P. Kundu, A.K. Srivastava, Real-time synchrophasor data anomaly detection and classification using isolation forest, KMeans, and LoOP, IEEE Trans. Smart Grid 12 (3) (2021) 2378–2388.
[38] S. Ahmed, Y. Lee, S. Hyun, I. Koo, Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest, IEEE Trans. Inform. Forensics Security 14 (10) (2019) 2765–2777.
[39] X. Qin, B. Li, J. Huang, A new spatial steganographic scheme by modeling image residuals with multivariate gaussian model, in: ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2617–2621.
[40] J. Liu, Z. Zhao, K. Pang, D. Wang, C. Tang, C. Yao, Improved winding mechanical fault type classification methods based on polar plots and multiple support vector machines, IEEE Access 8 (2020) 216271–216282.
[41] J. Ni, Z. Zhao, S. Tan, Y. Chen, C. Yao, C. Tang, The actual measurement and analysis of transformer winding deformation fault degrees by FRA using mathematical indicators, Electr. Power Syst. Res. 184 (2020) 0378–7796.
[42] Y. Qi, E. Bostanci, M. Zafarani, B. Akin, Severity estimation of interturn short circuit fault for PMSM, IEEE Trans. Ind. Electron. 66 (2019) 7260–7269.
[43] Carolina María Martín, José Manuel Guerrero, Carlos A. Platero, Pablo Gómez Mourelo, Ground Faults Location for Synchronous Machine Poles through Frequency Response Analysis, in: Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe) 2020 IEEE International Conference on, 2020, pp. 1–6.
[44] F. Blánquez Delgado, Development and improvement of protection system algorithms for ground-fault detection and location in synchronous machines, Ph. D. dissertation, E.T.S.I.I., U.P.M, Madrid, 2015.
[45] A.B. Nassif, M.A. Talib, Q. Nasir, F.M. Dakalbab, Machine learning for anomaly detection: a systematic review, IEEE Access 9 (2021) 78658–78700.

Yu Chen was born in Wenzhou, Zhejiang, China, in 2000. He is currently pursuing the bachelor's degree with the Department of Electrical Engineering, College of Engineering and Technology, Southwest University, Chongqing, China. His areas of research include condition monitoring and fault diagnosing for power transformer, and application of artificial intelligence.

Zhongyong Zhao was born in Guangyuan, Sichuan, China. He received the B.S. degree and the Ph.D. degree from Chongqing University, Chongqing, China, in 2011 and 2017, respectively, all in electrical engineering. He received a scholarship from China Scholarship Council to enable him to attend a joint-training Ph.D. program in Curtin University, Perth, Western Australia, in 2015-2016. He is currently an associate professor in the College of Engineering and Technology, Southwest University, Chongqing, China. His areas of research include condition monitoring and fault diagnosing for HV apparatus, and artificial intelligence.

Hanzhi Wu was born in Qindao, Shandong, China, in 1999. He is currently pursuing thebachelor's degree with the Department of Electrical Engineering, College of Engineering and Technology, Southwest University, Chongqing, China. His areas of research include condition monitoring and fault diagnosing for synchronous machine.

Xi Chen was born in Wenzhou, Zhejiang, China, in 2001. He is currently pursuing the bachelor's degree with the Department of Electrical Engineering, College of Engineering and Technology, Southwest University, Chongqing, China. His areas of research include condition monitoring and fault diagnosing for synchronous machine.

Qianbo Xiao was born in Chongqing, China, in 1986. He received the B.S. degree from Shandong University, Shandong, China, in 2008, and the master's degree from Chongqing University, Chongqing, China, in 2011. He currently works as a senior engineer in State Grid Chongqing electric power company. His areas of research include condition monitoring and fault diagnosing for power equipment.

Yueqiang Yu was born in Nanchong, Sichuan, China, in 1994. He is currently pursuing the master's degree with the Department of Electrical Engineering, College of Engineering and Technology, Southwest University, Chongqing, China. His areas of research include condition monitoring and fault diagnosing for synchronous machine.