Improved Interpretation of Impulse Frequency Response Analysis for Synchronous Machine Using Life Long Learning Based on iCaRL

Yu Chen, Zhongyong Zhao, *Member*, *IEEE*, Yueqiang Yu, Yunlong Guo, and Chao Tang, *Member*, *IEEE*

***Abstract*—Winding inter-turn and ground short circuit (SC) faults are common fault types of synchronous machines. The sweep frequency response analysis (SFRA) has been recently introduced to detect winding SC faults. Additionally, impulse FRA (IFRA) provides an alternative that serves the same objective. Whether SFRA or IFRA, so far, there is still no standard and reliable interpretation code. The present interpretation of frequency response still calls for experienced personnel, which could be subjective. Thus, many researchers have used machine or deep learning-based models to detect winding SC faults from the frequency response automatically. However, most proposed models cannot provide real-time feedback and update. Therefore, this study proposes an improved model using life long learning strategy based on incremental classifier and representation learning (iCaRL) to interpret and analyze the IFRA curves. This study artificially simulates and records winding SC faults on a 5 kW synchronous machine. The proposed method is then verified on the test set and compared with other traditional life long learning strategies. The experimental results show that the accuracy of the proposed model is higher than 90% under all types of fault data streams in real-time. The comparative experimental results show that the proposed model performs better than other life long learning strategies. (Github code: https://github.com/cy1034429432/Improved-Interpretation-of-IFRA-based-on-iCaRL/tree/main)**

Index Terms*—winding; synchronous machine; iCaRL; life long learning; fault detection.

I. Introduction

As one of the most significant power generation equipment, synchronous machines are used in thermal and hydraulic power plants [1-3]. With the more extensive application of synchronous machines, people pay much attention to their operational safety and stability. However, according to Ref. [4], 2/3 of the faults in large synchronous machines are owing to winding faults. If synchronous machines are used in critical production processes or equipment, winding SC faults may lead to downtime and loss of production capacity, further increasing economic losses [5]. Therefore, detecting synchronous machine winding SC faults is of great significance in ensuring the safe and reliable operation of the machine.

At present, the winding SC fault detection methods mainly include repetitive surge oscilloscope (RSO) [6], motor current signature analysis technique (MCSA) [7], impedance dielectric dissipation method [8], etc. These methods are significant, but new methods have also been proposed and studied. In recent years, some researchers [2, 3, 8-10] have introduced frequency response analysis (FRA), which has demonstrated promising results in detecting faults in power transformer windings, to detect winding SC faults of synchronous machines offline, and it can be used for the regular inspection of synchronous machines. FRA can be divided into sweep FRA (SFRA) and impulse FRA (IFRA). SFRA obtains the frequency response curve of machines through sweep frequency signal, and frequency response is the amplitude-frequency characteristic of the winding transfer function [11]. IFRA provides an alternative way to get the frequency response by injecting high-frequency impulse signals. However, it gets rapidity by losing certain stability because of the nature of high-frequency impulse signals [3].

Regardless of whether SFRA or IFRA is used, there are three primary methods for interpreting the frequency response curve to diagnose the presence, extent, and location of winding SC faults: 1. Mathematical-statistical indicators are constructed and computed, and a threshold value is established and compared to detect winding faults. 2. Based on the machine's equivalent high-frequency electrical circuit model, the frequency response curves of winding under various faults are analyzed, and the amplitude and frequency information of resonant points are used to detect faults. 3. Advanced artificial intelligence (AI) is used to process the frequency response, and many machine learning and deep learning-based models are used as classifiers.

In the first method, various mathematical-statistical indicators are used to analyze the relationship between winding faults and frequency response curves. Ref. [12] provides a series of mathematical-statistical indicators linearly correlated with winding SC faults' degrees. Besides, the correlation coefficient (CC), the absolute sum of logarithmic

* This work was supported in part by the Sichuan Science and Technology Program under Grant 2023NSFSC0829, in part by the Fundamental Research Funds for the Central Universities under Grant SWU-KT22027, and in part by the National Natural Science Foundation of China under Grant 51807166 (Corresponding author: Zhongyong Zhao).

Yu Chen is with the College of Engineering and Technology, Southwest University, Chongqing 400716, China, and also with the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yu_chen2000@hust.edu.cn).

Zhongyong Zhao is with the College of Engineering and Technology, Southwest University, Chongqing 400716, China, and also with Yibin Academy of Southwest University, Sichuan 64400, China (e-mail: zhaozy1988@swu.edu.cn).

Yueqiang Yu and Chao Tang are with the College of Engineering and Technology, Southwest University, Chongqing 400716, China (e-mail: yyq14011@email.swu.edu.cn; swutc@swu.edu.cn).

Yunlong Guo is with the School of Automation, Beijing Institute of Technology, Beijing, 100081, China (yunlong.guo2022@bit.edu.cn).

error (ASLE), standard deviation (SD), and other indicators are also often used [13-15]. The indicators under each sub-frequency band of frequency response are compared to the threshold values to quantitatively and qualitatively analyze the winding fault. However, there is no standard and defined threshold values for the present method, and even if a suggested threshold value is given, it does not apply to all machines. Besides, there is no basis for sub-frequency band division [2].

In the second method, Refs. [9-11, 16] analyze the influence of various winding faults on the frequency response curve by modeling the equivalent broadband or high-frequency circuits of machines, then the information of resonant points is often analyzed to make decisions on diagnosis results regarding the winding faults. However, machines' equivalent circuits should be first established, which could be complex and time-consuming. Meanwhile, this method is not entirely objective or intelligent, as it heavily relies on manual analysis.

Therefore, there is still no standard and reliable interpretation code, and the present interpretation of frequency response still calls for experienced personnel, which would be subjective. Recent studies use advanced AI techniques to process the frequency response to solve this problem, elaborated in the third method.

In the third method, many researchers use various machine learning-based classifiers to detect winding faults. For example, Ref. [17] uses the fault diagram for supervised learning fault detection; Ref. [8] uses isolation forest (IF) for unsupervised learning fault detection. Machine learning and deep learning-based models seem to overcome some drawbacks in the first two methods.

However, for the existing AI methods, there are some problems in winding faults detection:

1. The generalization performance of some machine learning and deep learning-based detection models is inadequate. Several supervised learning models can only learn feature mapping from their own winding fault data and are incapable of detecting other fault types that are not included in the dataset. For example, some traditional machine learning models mentioned in Ref. [18] only concentrate on winding inter-turn SC faults and exhibit excellent performance in analyzing inter-turn SC faults. Nevertheless, the trained model is unsuitable for predicting ground SC faults, as this fault is an entirely new fault type for the model.
2. The current models do not have any real-time feedback capability. Specifically, the trained models cannot be updated in real-time, so researchers need to train fault detection models from scratch when a new fault needs to be added. Due to the lack of self-adaptability to new data, most current methods are not applicable in engineering and are only suitable for laboratory settings.

To solve the above problems, this study proposes an improved detection method of winding SC faults for the synchronous machine using life long learning based on iCaRL.

The main contributions of this study are as follows:

1. Life long learning strategy is the first to be introduced into fault diagnosis in this study, and the proposed method has strong engineering practicality according to the characteristics of life long learning by considering data input, data storage, *etc*.
2. An improved winding SC fault diagnosis method is proposed in conjunction with IFRA, which can automatically identify the type and degree of winding SC faults with high accuracy and improve the detection model's generalization performance in interpreting the frequency response.
3. Compared to all previous works, the proposed method's biggest highlight is the model's real-time update capability. This capability allows the model's parameters to be updated based on users' real-time feedback, enabling adaptation to new winding fault types and closing the gap between the trained fault detection model and its updating ability.

The remainder of this study is organized as follows. The basic principles of IFRA and iCaRL are introduced in section II. The IFRA dataset obtained from the artificially simulated winding SC fault experiment is introduced in section III. The experimental results of verifying the proposed method are analyzed in section IV. The comparison of the proposed method with other traditional life long learning strategies is presented in section V. The discussion and limitations are given in section VI. The conclusions are given in section VII.

II. Theoretical Basis of Proposed Methods

A. The basic principle of IFRA

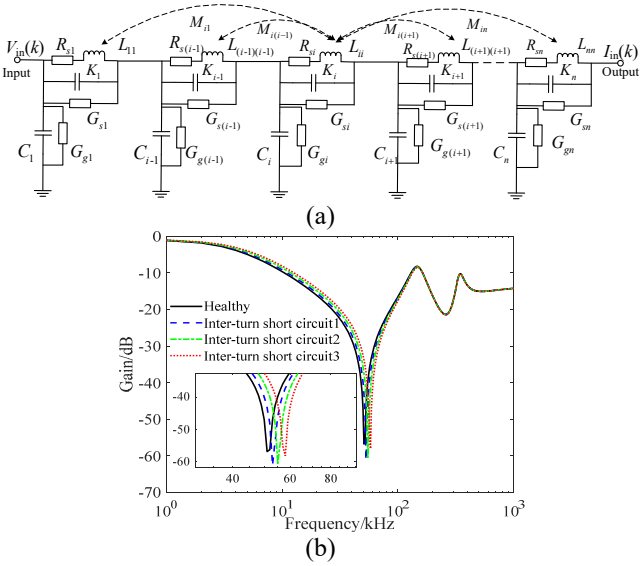


Fig. 1. Synchronous machine winding's equivalent circuit model structure and simulated winding fault. (a) equivalent circuit model structure. (b) inter-turn SC fault simulated by the built equivalent circuit[11].

The application of IFRA in detecting synchronous machine winding SC faults has become increasingly popular among researchers due to its high sensitivity, rapid diagnostic speed,

cost-effectiveness, and non-destructive detection process [2, 8]. The winding of the synchronous machine can be modeled as an equivalent two-port circuit consisting of inductance, capacitance, and resistance, as shown in Fig.1(a) [11]. When a high-frequency impulse signal $v_{in}(n)$ is injected at one terminal of winding, the response current $i_{out}(n)$ at the other terminal can be recorded, and both $v_{in}(n)$ and $i_{out}(n)$ are used to construct the transfer function $H(k)$ based on Fourier transform of time domain signal to the frequency domain, as shown in Equation (1)-(3). The amplitude-frequency characteristic of transfer function $H(k)$ is also called the frequency response curve. After manufacturing the synchronous machine winding, its transfer function $H(k)$ is determined accordingly. The presence of ground or inter-turn SC faults in the synchronous machine winding will alter the parameters of its equivalent circuit model, resulting in changing the frequency response curve ($H(k)$). For instance, Fig.1(b) shows $H(k)$ in different degrees of inter-turn SC winding faults simulated by short-circuiting equivalent units [11]. The discrepancy in $H(k)$ between the faulty and normal windings, such as shifts in resonance frequency, is employed for detecting any defects in the winding [2, 8].

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

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

$$H(k)=20\log_{10}\frac{|I_{out}(k)|}{|V_{in}(k)|} \tag{3}$$

where $v_{in}(n)$ *and* $i_{out}(n)$ are the N points sampling signal of high-frequency impulse signal and response current; $V_{in}(k)$ and $I_{out}(k)$ are fast Fourier transforms (FFT) of $v_{in}(n)$ and $i_{out}(n)$; $H(k)$ is the transfer function.

B. The basic principle of iCaRL

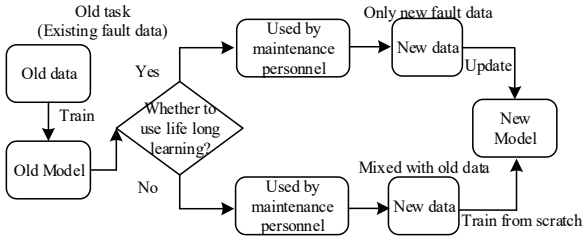


Fig. 2. Differences in the training process between with or without life long learning strategy.

Fig.2 shows differences in the training process between with or without learning life long learning [19-22]. As can be seen from Fig.2, the fault detection model (old model) is trained by the existing fault data (old data) at the beginning. After exiting new faults, there are two ways to build a new model: (1) Without life long learning, it should train a new model from scratch with all data, including new and old data. (2) With life long learning, it could train a new model by updating the old model only with new data. With the new data feedback, the old model with life long learning can learn new faults when in use. A fault detection model with a life long learning strategy automatically updates the old model, overcoming the limitation of weak self-adaptability to new fault data.

Fig.3 shows the training data input way of life long learning during the training process, usually called the training data stream. Each input data stream is a task that can contain one or more classes. Fig.3 is an example, which shows that in task 1, the model learns four different degrees of ground SC faults, and in task 2, the model learns four different inter-turn SC faults. With the new training data streams, the model can be continuously updated. Therefore, the model with life long learning strategy is also called a class-incremental learner. In addition, life long learning also includes another type called instance-incremental learning [19]. In the research scenario of this study, instance-incremental learning means learning the same fault type with different types of synchronous machines. Because of the limitation of experimental equipment, this study only adopts the former.

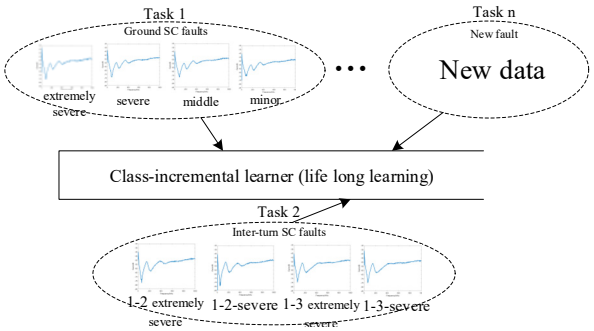


Fig. 3. The training data stream of life long learning.

Generally, when the model's classification classes (fault types) increase, it is necessary to modify the model's last layer structure first, then mix new data with old data, and finally train a new model from scratch. It will require many memory resources to store new and old training data and need more computing power and time to build a new model. This problem can be avoided using the data stream in Fig.3 [19]. In addition, the collecting data method in real life is more similar to the data stream, and it is impossible to collect all winding fault data simultaneously. Therefore, the model should only detect fewer winding faults initially, and with the users' real-time feedback, the model will be updated to detect more winding faults.

In life long learning, there are many methods, such as elastic weight consolidation (EWC) [22] and learning without forgetting (LWF) [23], which show promising results on some simple datasets. The iCaRL proposed in the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) shows much better performance than the above two methods on some complex datasets, and iCaRL combined with IFRA has the following advantages: (1) For the IFRA data of new winding faults at different times, the fault detection model could be trainable. (2) At any time, the model can present relatively high performance for the winding faults that have occurred. (3) The model updating time and storage resources of winding faults detection equipment with the diagnosis model are limited, or the growth of resources is slow

with the growth of new winding faults.

The overall framework diagram of iCaRL is shown in Fig.4. The pseudo-code of the algorithms is shown in Appendix. The main characteristics of its algorithms are as follows:

(1) iCaRL Classify: Unlike traditional *softmax*, iCaRL uses the *nearest-mean-of-exemplars* classifier. In addition, the feature map uses iCaRLNet, and the iCaRLNet structure refers to Avalanche [21]. This algorithm is responsible for classifying the input.

(2) iCaRL IncrementalTrain: This algorithm is responsible for learning new classes.

(3) iCaRL UpdataRepresentation: Knowledge distillation is added to the loss function compared to traditional cross-entropy. This algorithm is responsible for updating the feature map.

(4) iCaRL ConstructExemplarSet: This algorithm is responsible for constructing its exemplar set for the current winding faults.

(5) iCaRL ReduceExemplarSet: This algorithm is responsible for determining the number of IFRA pictures restored in each new task and deleting the IFRA pictures of the old task.

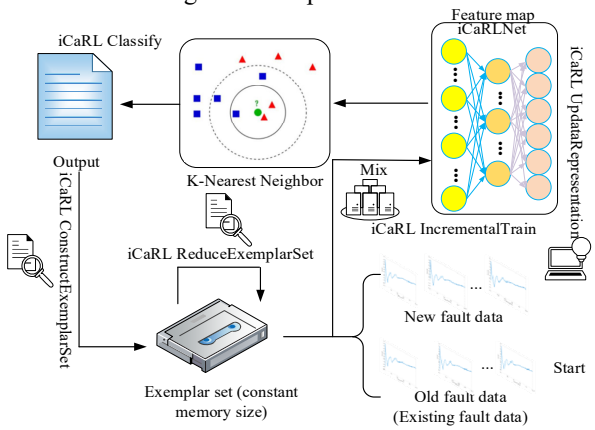


Fig. 4. Overall framework diagram of iCaRL. This figure shows the relationship between the algorithms. The initial input data is the existing fault data, and the subsequent input data is the new fault data.

III. Winding SC Fault Experiment Setup and IFRA Datasets

The synchronous machine winding SC faults artificial simulation platform is built, and the winding SC faults simulation experiments are carried out to verify the effectiveness of the proposed method. The experimental platform is shown in Fig.5. In Fig.5, the experimental platform includes a pulse generator, a current sensor, a voltage probe, an oscilloscope, and a 5kW synchronous machine without parallel branches and rotor (Therefore, there is no need to consider IFRA measurement for rotating machines with unstable repeatability [3, 8].). These devices' key nameplate parameters are shown in Tab.I.

In Fig.5, the voltage probe and current sensor measure impulse signal $v_{in}(n)$ and response current $i_{out}(n)$, and the waveform is recorded by oscilloscope. Set the sampling rate on the oscilloscope as 25 MHz and the sampling point as 10k. Then, $v_{in}(n)$ and $i_{out}(n)$ (measured 64 times) are averaged on the oscilloscope to reduce the impact of white noise (The above measurement parameters are referenced from our previous work [3].). Finally, IFRA is calculated on the average time-domain signal.

TABLE I
NAMEPLATE VALUES OF SYNCHRONOUS MACHINE

Characteristics		Parameter value
Rated power		5kW
Rated voltage		380V
Frequency		50Hz
Pole pairs		1
Number of slots		36
Rated speed		1500rpm
Equipment	**Model**	**Key parameter value**
Pulse generator	Homemade	amplitude: 0-4kV, leading-edge: 40ns, pulse width: 10-1000ns
Current sensor	Pearson 150	bandwidth: 40Hz-20MHz, sensitivity: 0.5V/A
Oscilloscope	Tektronix MDO4104C	bandwidth: 1GHz
Voltage probe	Tektronix P5100A	bandwidth: 500 MHz

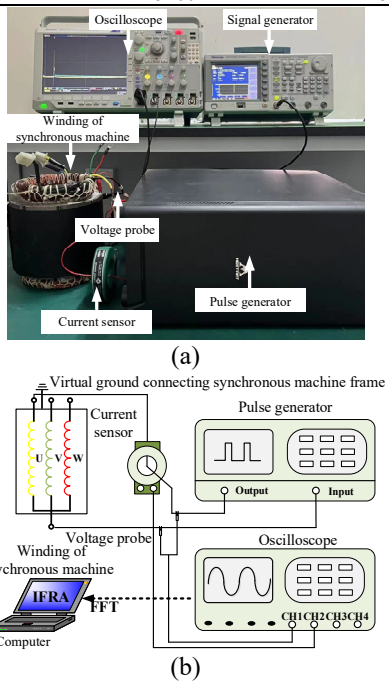


Fig. 5. Measurement experiment diagram. (a) actual wiring diagram. (b) measurement wiring diagram.

According to Refs. [2, 3, 8, 9, 14, 24-26], the ground and inter-turn SC faults are simulated artificially. In the experiment, ground SC faults of varying degrees are simulated by connecting 40/20/10/0 Ω resistance in parallel with winding slot 1 of the U-phase, called ground SC fault (1-G-minor, middle, severe, and extremely severe), respectively. Windings in slots 1, 2, and 3 of the U-phase are short-circuited to simulate different degrees of inter-turn SC faults. In addition, to obtain more data on inter-turn SC fault in different degrees, this research parallels the 10 Ω resistance to slots 1, 2, and 3, called inter-turn SC fault (1-2-severe, extremely severe, 1-3-severe, and extremely severe) respectively. The wiring diagram is shown in Fig.6, and the several IFRA curves are shown in Fig.7. Then, repeat the experiment and construct

the IFRA dataset containing nine classes, whose details are shown in Tab.II.

For the above, it must be noted that this study only treats different connection resistance values as degrees of winding SC fault, which only verifies the proposed method. For different synchronous machine windings, the simulated fault resistance values (According to the relevant Refs. [3, 9, 27, 28], resistance values are mostly around 0-100 Ω.) may be different, which can be determined according to the actual situation. This study focuses on winding SC faults detection methodology, and the selection of connection resistance values can be referred to [3].

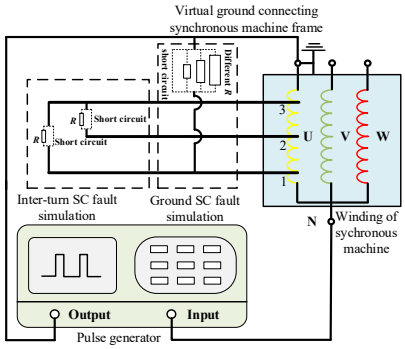
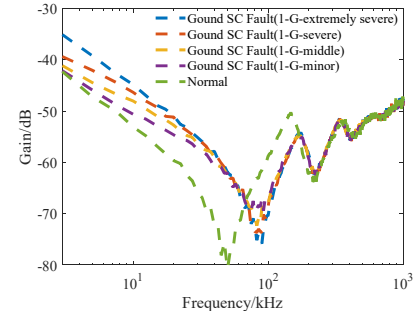
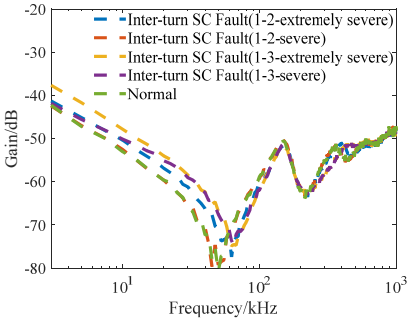


Fig. 6. Wiring diagram of artificially simulated synchronous machine winding SC faults.



(a)



(b)

Fig. 7. IFRA curves of the synchronous machine in various winding fault states. (a) IFRA curves of ground SC faults in different degrees. (b) IFRA curves of inter-turn SC faults in different degrees.

IV. Validation Experiments of Proposed Method

A. The comparison of different feature maps

The feature map uses the image classification model for the following reasons: (1) The critical rule that IFRA curves can detect winding faults is to use the normal winding IFRA curve to make a horizontal comparison with the faulty winding IFRA curves. Then relevant researchers get the changing trend to detect winding faults. The IFRA curves' change is usually the deviation of resonance frequency and resonance point, and the image classification model can excavate the relationship between IFRA curves' differences and various winding faults. (2) If a sequence model is used as the feature map, the input sequence dimension will be very long, and the different selection of sampling frequency will lead to difficult training and poor generalization of the feature map. Additionally, we explained how the image classification models understand IFRA for detecting synchronous machine winding SC faults based on Smooth Grad-CAM++ in previous work [2].

TABLE II

IFRA Datasets of Synchronous Machine Winding

Abbreviation		Type	Number of samples (training set)	Number of samples (test set)
Fault 0 or 0	Ground SC fault	1-G-extremely severe	202	30
Fault 1 or 1		1-G-severe	202	30
Fault 2 or 2		1-G-middle	201	30
Fault 3 or 3		1-G-minor	201	30
Fault 4 or 4	Inter-turn SC fault	1-2-extremely severe	201	30
Fault 5 or 5		1-2-severe	201	30
Fault 6 or 6		1-3-extremely severe	200	30
Fault 7 or 7		1-3-severe	200	30
Normal or 8		Normal	200	30

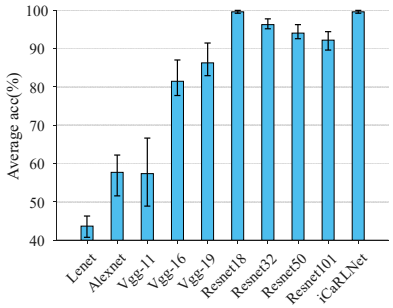


Fig. 8. The comparison of different feature maps.

The more epochs a model undergoes training, the more severe the problem of catastrophic forgetting becomes. Therefore, the accuracy obtained by using a feature map alone represents the upper bound of the feature map when employing a life long learning strategy [22, 23]. The first step in adopting a life long learning strategy is to select a high-accuracy and appropriate feature map. The loss function uses *softmax* to fairly compare the feature maps' winding SC faults detection performance. The results are shown in Fig.8.

As shown in Fig.8, the accuracy of the feature maps is more than 90% based on Resnet, while the low accuracy of large-scale Resnet is related to the IFRA dataset being too small [2]. The accuracy of other simpler models is low because the

extraction ability is weak. The iCaRLNet is a model built based on Resnet's basic structure residual block [21], and the average accuracy has reached 99.63%. The Resnet18 and iCaRLNet demonstrate similar performance, but iCaRLNet has proven to exhibit high performance on numerous datasets [21], and this study chooses to utilize iCaRLNet.

B. The selection of training hyperparameters

TABLE III

SERVER HARDWARE AND SOFTWARE CONFIGURATION

Device	Model
CPU	Inter(R) Xeon(R) Gold 6268CL × 2
GPU	NVIDIA RTX A4000
RAM	128G
SOFTWARE	Python
PACKAGE	Numpy, PyTorch, and Avalanche[21]

All experiments in this research are trained on the server, and the server configuration is shown in Tab.III. In addition, all codes are based on Avalanche [21] and PyTorch. The details of the iCaRL and the training hyperparameters settings are shown in the code on GitHub.

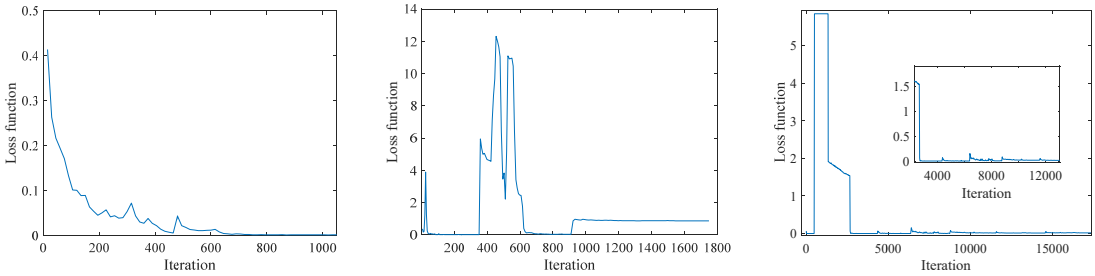
C. Experimental result

As shown in Tab.II, the IFRA dataset has 9 classes. Therefore, this research simulates three IFRA training data streams (a single task containing 1, 3, and 9 classes). When the model is trained, the combination and arrangement order of classes will also affect the accuracy of the model [21-23]. For comparison, Fig.9, 10, and 11 show the loss function curve, confusion matrix, and accuracy curve under the same arrangement order training data stream. The average accuracy of the proposed model on the test set under different arrangement order training data streams is shown in Tab.IV, where the values in brackets are the maximum and minimum values of the accuracy.

TABLE IV

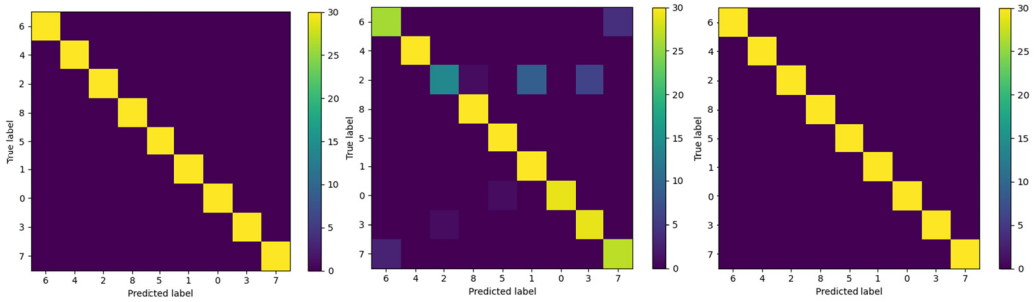
AVERAGE ACCURACY UNDER DIFFERENT ARRANGEMENT ORDERS OF THE TRAINING DATA STREAM

Type	9 classes per task	3 classes per task	1 class per task
Average accuracy(%)	99.25(97.78-100)	94.44(90.70-100)	92.59(90.70-100)



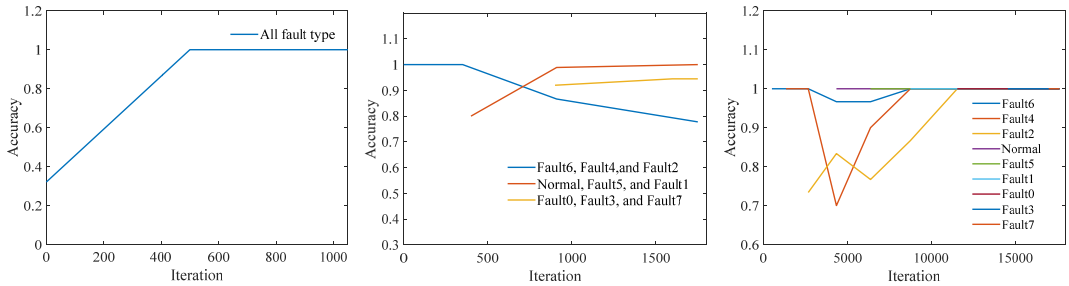
(a) (b) (c)

Fig. 9. Loss function under different IFRA data streams. (a) 1 task. (b) 3 tasks. (c) 9 tasks.



(a) (b) (c)

Fig. 10. Confusion matrix under different IFRA data streams. (a) 1 task. (b) 3 tasks. (c) 9 tasks.



(a) (b) (c)

Fig. 11. Accuracy curve under different IFRA data streams. (a) 1 task. (b) 3 tasks. (c) 9 tasks. The training data stream in Fig.9, 10, and 11 is in a specific order, but the results will be different under different training data streams in other arrangement orders.

As shown in Fig.9, the value of the loss function suddenly increases during training because a new task is added. In addition, the mutation interval of the loss curve becomes larger and larger because the input training data increases with more tasks, while the batch size and epoch do not change, so the interval becomes larger. It can be seen from Fig.10(b) that

task 1 has obvious catastrophic forgetting in this case. It can also be concluded from the accuracy curve in Fig.11(b) that task 1 did not get satisfactory accuracy at the end, which is related to the fact that this order is not conducive to learning in this case, but catastrophic forgetting may be reduced in other cases according to Tab.IV. It can be seen from Fig.11(c) that although there will be some catastrophic forgetting in the training process, it has been overcome in training, and the accuracy of any task has reached 100% at the end, which is closely related to the use of knowledge distillation in the loss function.

According to the statistical results in Tab.IV, the average accuracy of using iCaRL in different cases is more than 90%, and the robustness of this algorithm is high, which is feasible for synchronous machine winding SC faults detection equipment. The average accuracy of 9 classes per task reaches 99.25%, which shows that the classification efficiency using the *nearest-mean-of-exemplars* classifier is the same as that using *softmax*. In addition, when the parameters of the feature map change, the *nearest-mean-of-exemplars* classifier will change and modify self-adaptively, which can also reduce catastrophic forgetting. The average accuracy change trend in Tab.IV is also consistent with life long learning [19-21, 23]. The model for training all data must be the upper bound of life long learning performance because the more tasks, the more severe and unavoidable catastrophic forgetting.

V. Ablation Experiment and Comparison of the Proposed Method with Other Life Long Learning Strategies

A. Ablation experiment of the critical hyperparameter

The iCaRL has a critical hyperparameter K related to iCaRL ConstructExemplarSet and ReduceExemplarSet, as introduced in Appendix. Hyperparameter K (memory size) determines the resource size of storing IFRA data in the exemplar set. The K also simulates the internal memory of detection equipment, whose value significantly impacts the model to overcome catastrophic forgetting.

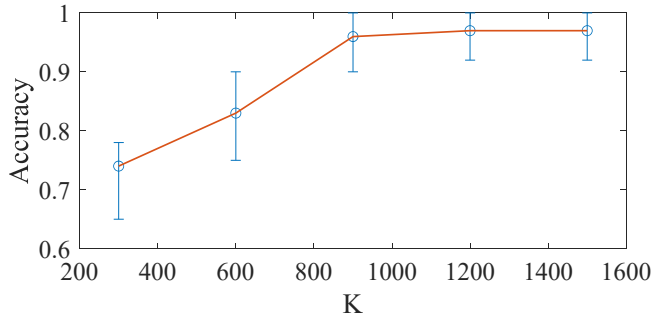


Fig.12. Average accuracy of different IFRA data streams under different K. The above results are obtained from 20 repeated experiments. The solid line is the average accuracy, and the length of the lower and upper parts represent the minimum accuracy and the maximum accuracy.

Fig.12 shows the average accuracy under different K. For the experimental synchronous machine in this study, it can be seen from Fig.12 that choosing 900 as K has certain advantages. After being greater than 900, the average accuracy does not increase significantly, while less than 900 will significantly impact the average accuracy.

In model training, there exist numerous hyperparameters that necessitate prior knowledge of training high-performance models and appropriate adjustments for diverse physical scenarios. However, these conventional hyperparameter settings concerning the training model are not the critical parameters of the proposed method, so these hyperparameter settings are presented on the GitHub code.

B. Comparison results of the proposed method with other life long learning strategies

To demonstrate the indispensability of the fault diagnosis model incorporating life long learning strategy and highlight the exceptional performance of iCaRL, this study undertakes experiments including the fault diagnosis model without life long learning and with conventional life long strategies. For the fairness of comparison, all experiments are conducted using one class per task, consistent with the same training data stream (same training order) in Fig.11(c).

TABLE V
ACCURACY WITHOUT LIFE LONG LEARNING STRATEGY

Accuracy		Tested on								
		6	4	2	8	5	1	0	3	7
After training	6	1.00	0	0	0	0	0	0	0	0
	4	0.13	1.00	0	0	0	0	0	0.90	0
	2	0	0.73	1.00	0	0	0	0	0	0
	8	0	0	0.07	1.00	0	0	0	0	0
	5	0	0	0	0	1.00	0	0	0	0
	1	0	0	0	0	0.10	1.00	0	0	0
	0	0	0	0	0	0.13	0.73	0.40	0	0
	3	0	0	0	0	0	0	0	1.00	0
	7	0	0	0	0	0	0	0	0	1.00

Tab.V shows the training results of iCaRLNet when the life long learning strategy is not implemented (*softmax* is used in the last layer). The training order of each class is represented by the vertical and horizontal axes, with the abbreviations of winding fault types from Tab.II displayed on both axes (For example, “6” stands for “inter-turn SC fault (1-3-extremely severe)” in Tab.II.). The values in the table indicate the accuracy of each task (calculate the accuracy on the test set of a single class) in the test set obtained after completing each task, i.e., the model will be “tested on” test sets of all classes each time a task is “after training”. The above description also applies to Tab.VI-VII. It can be seen from Tab.V that without life long learning, the model only cares about the current task and has a tremendous catastrophic forgetting of the previous tasks. It finally regards all winding faults as fault 7, which is related to the fact that the last constructed loss function only contains fault 7.

In the field of life long learning, Deepmind proposes EWC to overcome catastrophic forgetting [22]. Compared with the traditional multi-class classification, this method adds a regularization-like term to the loss function. In addition, LWF is used for the IFRA dataset [23]. Like iCaRL, LWF also uses knowledge distillation technology to solve catastrophic forgetting. Tab.VI-VII show the training results of iCaRLNet with EWC and LWF. It can be seen from Tab.VI that although EWC is better than not implementing life long learning

strategy and can only accurately identify the last two winding fault types at the end, it is still much worse than iCaRL. It can be seen from Tab.VII that LWF has less catastrophic forgetting in fewer training classes (<3), while LWF is challenging to overcome catastrophic forgetting when the classes are greater than 4. In addition, when using LWF, because the classes are 9, it is necessary to set 9 penalty hyperparameters. Compared with iCaRL, there are too many critical hyperparameters, but the performance is much worse than iCaRL.

TABLE VI

ACCURACY UNDER EWC

Accuracy		Tested on								
		6	4	2	8	5	1	0	3	7
After training	6	1.00	0	0	0	0	0	0	0.83	0
	4	1.00	0.90	0	0	0	0	0	0.90	0
	2	0	0	1.00	0	0	0	0	0	0
	8	0	0.30	0.00	1.00	0	0	0	0	0
	5	0	0	0	0	1.00	0	0	0	0
	1	0	0	0	0	0.10	0.96	0	0	0
	0	0	0	0	0	0.13	0.73	0.40	0	0
	3	0	0	0	0.07	0.03	0	0	1.00	0
	7	0	0	0	0	0	0	0	1.00	1.00

TABLE VII

ACCURACY UNDER LWF

Accuracy		Tested on								
		6	4	2	8	5	1	0	3	7
After training	6	1.00	0	0	0	0	0	0	0	0
	4	0.73	1.00	0	0	0	0	0	0	0
	2	0.3	0.73	1.00	0	0	0	0	0	0
	8	0	0.3	0.50	1.00	0	0	0	0	0
	5	0	0	0	0.1	1.00	0	0	0	0
	1	0	0	0	0	0.96	1.00	0	0	0
	0	0	0	0	0	0.13	0.73	1.00	0	0
	3	0	0	0	0	0	0	0.73	1.00	0
	7	0	0	0	0	0	0	0.3	0.73	1.00

VI. DISSCUSION AND LIMITATION

Currently, many AI-based works are related to detecting synchronous machines' winding SC faults [2, 3, 8, 11, 24, 28]. However, their work only focuses on the performance of fault diagnosis models and does not discuss the practical application of the fault detection models outside of the laboratory. As the proposed method does not solely target the performance of fault models, it cannot be fairly compared with previous works. This study offers a practical and suitable model for making fault-detection equipment. Using fault-detection equipment embedded with life long learning strategy makes it possible to continuously learn different types of winding faults throughout the lifetime of the synchronous machine.

Although the model with life long learning strategy exhibits excellent performance, there are still many limitations to the proposed method:

1. This study focused on the same machine and did not conduct experiments on other types of synchronous machines. This is due to limitations in the authors' experimental equipment and sample dataset. Future research could consider expanding the sample dataset and experimental equipment to explore the application and effectiveness of this method on different types of machines. In addition, it is difficult to effectively define winding faults of the same degree for different machines. A minor ground SC fault resistance may be in the megaohm level for large synchronous machines but only in the 10-ohm level for small machines.
2. The fault diagnosis of synchronous machines should consider diverse signals, such as vibration and sound signals. But this study only focuses on electrical signals, which may lead to incomplete diagnosis results. Additionally, the selection of feature maps should be diverse. Choosing a multi-modal model can simultaneously consider diverse signals, expand the discernible fault types, and further improve fault diagnosis performance.
3. In fact, many of the effects in use are not considered at all. Further research is needed to explore the feasibility and effectiveness of implementing the proposed method in real-world scenarios.

VII. CONCLUSION

This study presents an improved detection method of winding SC faults for synchronous machines using life long learning strategy based on IFRA and iCaRL. According to the experimental results and comparative analysis, the following conclusions are obtained:

1. The average accuracy of the proposed detection model exceeds 90% in all cases. Furthermore, the generalization performance of the proposed method can be continuously enhanced by inputting newly diverse data.
2. The iCaRL has a better ability to overcome the problem of catastrophic forgetting than the EWC and LWF. Besides, without life long learning strategy, the winding SC faults detection accuracy of iCaRLNet reaches 99.63%, which is also better than most other traditional image classification models.
3. The proposed fault detection model can guide the practical application of the deep learning model, and the detection equipment embedded with the proposed method can have the ability for life long learning. After getting users' feedback, the equipment can update the model in real-time, and there is no need to train the model from scratch.
4. With the unification of image input and sequence input models in recent years, a multi-modal model with the ability to detect any fault of the synchronous machines can be established with life long learning strategy in the future.

APPENDIX

The following is the pseudo-algorithm code for iCaRL. X^s represents the IFRA training image set of s-type winding fault, Θ represents the feature map's parameters, K represents the maximum number of IFRA pictures that can be restored in the exemplar set, and P represents the current exemplar set.

Algorithm 1 iCaRL Classify

Input x //IFRA image to be classified

require $p = (P_1, ..., P_t)$ //class exemplar sets
require $\varphi: \chi \to R^d$ //feature map
for $y = 1, ..., t$ **do**
$\mu_y \leftarrow \frac{1}{|P_y|} \sum_{p \in P_y} \varphi(p)$ //mean-of-exemplars
end for
$y^* \leftarrow \arg\min_{y=1,...,t} \| \varphi(x) - \mu_y \|$
output class label y^*

Algorithm 2 iCaRL IncrementalTrain

Input $X^s, \ldots, X^t$ // training examples in per-class sets
input K // memory size
require Θ // current model parameters
require $p = (P_1, ..., P_{s-1})$ //current exemplar sets
$\Theta \leftarrow \text{UpdateRepresentation}(X^s, ..., X^t; p, \Theta)$
$m \leftarrow K/t$ // number of exemplars per class
for $y = 1, \ldots, s-1$ **do**
$P_y \leftarrow \text{ReduceExemplarSet}(P_y, m)$
end for
$p \leftarrow (P_1, ..., P_t)$ // **new exemplar sets**

Algorithm 3 iCaRL UpdataRepresentation

Input $X^s, \ldots, X^t$ // training images of classes $s, \ldots, t$
require $p = (P_1, ..., P_{s-1})$ // exemplar sets
require Θ // current model parameters
// form combined training set:
$D \leftarrow \bigcup_{y=s,...,t} \{(x, y) : x \in X^y\} \cup \bigcup_{y=1,...,s-1} \{(x, y) : x \in P^y\}$
//store network outputs with pre-update parameters:
for $y = 1, \ldots, s-1$ **do**
$q_i^y \leftarrow g_y(x_i)$ *for* all $(x_i) \in D$
end for
run network training with loss function

$$\ell(\Theta) = -\sum_{(x_i, y_i)} [\sum_{y=s}^{t} \delta_{y=y_i} \log g_y(x_i) + \delta_{y \neq y_i} \log(1 - g_y(x_i)) + \sum_{y=1}^{s-1} q_i^y \log g_y(x_i) + (1 - q_i^y) \log(1 - g_y(x_i))]$$

that consists of classification and distillation terms

Algorithm 4 iCaRL ConstructExemplarSet

Input image set $X = \{x_1, ..., x_n\}$ *of* class y
input m target number of exemplars
require current feature function $\varphi: \chi \to R^d$
$\mu \leftarrow \frac{1}{n} \sum_{x \in X} \varphi(x)$ //current class mean
for $k = 1, \ldots, m$ **do**
$p_k \leftarrow \arg\min_{x \in X} \| \mu - \frac{1}{k} [\varphi(x) + \sum_{j=1}^{k-1} \varphi(p_j)] \|$
end for
$p \leftarrow (P_1, ..., P_m)$
output exemplar set P

Algorithm 5 iCaRL ReduceExemplarSet

Input m //target number of exemplars
input $P = (p_1, ..., p_{|P|})$ //current exemplar set
$P = (p_1, ..., p_{|m|})$ //current class mean
output exemplar set P

REFERENCES

[1] H. Toliyat, S. Nandi, S. Choi, and H. Meshgin-Kelk, Electric Machines: Modeling, Condition Monitoring, and Fault Diagnosis. 2017.
[2] Y. Chen, Z. Zhao, Y. Yu, W. Wang, and C. Tang, "Understanding IFRA for Detecting Synchronous Machine Winding Short Circuit Faults Based on Image Classification and Smooth Grad-CAM++," IEEE Sensors Journal, vol. 23, no. 3, pp. 2422-2432, 2023.
[3] Y. Yu, Z. Zhao, Y. Chen, H. Wu, C. Tang, and W. Gu, "Evaluation of the Applicability of IFRA for Short Circuit Fault Detection of Stator Windings in Synchronous Machines," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-12, 2022.
[4] K. N. Gyftakis and 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, vol. 1, pp. 1290-1295.
[5] L. Hao, J. Wu, Z. Chen, and H. Song, "Mechanism of Effects of Inter-turn Short Circuits in Field Windings on Large Turbo generator Vibration," Power system automation (in Chinese), vol. 38(04), pp. 25-31+50, 2014.
[6] T. Li et al., "Simulation Study on Interturn Short Circuit of Rotor Windings in Generator by RSO Method," in 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), 2018, pp. 1-4.
[7] N. Bessous, S. E. Zouzou, S. Sbaa, and A. Khelil, "New vision about the overlap frequencies in the MCSA-FFT technique to diagnose the eccentricity fault in the induction motors," in 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), 2017, pp. 1-6.
[8] Y. Chen, Z. Zhao, H. Wu, X. Chen, Q. Xiao, and Y. Yu, "Fault anomaly detection of synchronous machine winding based on isolation forest and impulse frequency response analysis," Measurement, vol. 188, p. 110531, 2022.
[9] F. R. Blánquez, C. A. Platero, E. Rebollo, and F. Blánquez, "Evaluation of the applicability of FRA for inter-turn fault detection in stator windings," in 2013 9th IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED), 2013, pp. 177-182.
[10] F. R. Blánquez, C. A. Platero, E. Rebollo, and F. Blázquez, "Field-winding fault detection in synchronous machines with static excitation through frequency response analysis," International Journal of Electrical Power & Energy Systems, vol. 73, pp. 229-239, 2015.
[11] Z. Zhao, Y. Chen, Y. Yu, M. Han, C. Tang, and C. Yao, "Equivalent Broadband Electrical Circuit of Synchronous Machine Winding for Frequency Response Analysis Based on Gray Box Model," IEEE Transactions on Energy Conversion, vol. 36, no. 4, pp. 3512-3521, 2021.
[12] J. Ni, Z. Zhao, S. Tan, Y. Chen, C. Yao, and C. Tang, "The actual measurement and analysis of transformer winding deformation fault

degrees by FRA using mathematical indicators," Electric Power Systems Research, vol. 184, p. 106324, 2020.
[13]M. F. M. Yousof, A. A. Alawady, S. M. Al-Ameri, N. Azis, and H. A. Illias, "FRA Indicator Limit for Faulty Winding Assessment in Rotating Machine," in 2021 IEEE International Conference on the Properties and Applications of Dielectric Materials (ICPADM), 2021, pp. 346-349.
[14]A. Mugarra, C. A. Platero, J. A. Martínez, and U. Albizuri-Txurruka, "Validity of Frequency Response Analysis (FRA) for Diagnosing Large Salient Poles of Synchronouos Machines," IEEE Transactions on Industry Applications, vol. 56, no. 1, pp. 226-234, 2020.
[15]H. Mayora, R. Alvarez, G. Bossio, and E. Calo, "Condition Assessment of Rotating Electrical Machines using SFRA - A Survey," in 2021 IEEE Electrical Insulation Conference (EIC), 2021, pp. 26-29.
[16]Y. Chen, X. Ji, and Z. Zhao, "Synchronous Machine Winding Modeling Method Based on Broadband Characteristics," Applied Sciences, vol. 11, p. 4631, 2021.
[17]A. Mugarra, H. Mayora, J. M. Guerrero, and C. A. Platero, "Frequency Response Analysis (FRA) Fault Diagram Assessment Method," IEEE Transactions on Industry Applications, vol. 58, no. 1, pp. 336-344, 2022.
[18]A. Moradzadeh, H. Moayyed, B. Mohammadi-ivatloo, G. B. Gharehpetian, and A. P. Aguiar, "Turn-to-Turn Short Circuit Fault Localization in Transformer Winding via Image Processing and Deep Learning Method," IEEE Transactions on Industrial Informatics, vol. 18, no. 7, pp. 4417-4426, 8 2021.
[19]S. A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental Classifier and Representation Learning," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5533-5542.
[20]B. Maschler, H. Vietz, N. Jazdi, and M. Weyrich, "Continual Learning of Fault Prediction for Turbofan Engines using Deep Learning with Elastic Weight Consolidation," in 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 2020, pp. 959-966.
[21]V. Lomonaco et al., "Avalanche: an End-to-End Library for Continual Learning," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2021, pp. 3595-3605.
[22]J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 114, no. 13, pp. 3521-3526, 2017.
[23]Z. Li and D. Hoiem, "Learning without Forgetting," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 2935-2947, 2018.
[24]J. Zhang, Y. Wang, K. Zhu, Y. Zhang, and Y. Li, "Diagnosis of Interturn Short-Circuit Faults in Permanent Magnet Synchronous Motors Based on Few-Shot Learning Under a Federated Learning Framework," IEEE Transactions on Industrial Informatics, vol. 17, no. 12, pp. 8495-8504, 2021.
[25]H. Ehya, T. N. Skreien, and A. Nysveen, "Intelligent Data-Driven Diagnosis of Incipient Interturn Short Circuit Fault in Field Winding of Salient Pole Synchronous Generators," IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3286-3294, 2022.
[26]Y. Qi, E. Bostanci, M. Zafarani, and B. Akin, "Severity Estimation of Interturn Short Circuit Fault for PMSM," IEEE Transactions on Industrial Electronics, vol. 66, no. 9, pp. 7260-7269, 2019.
[27]S. Uhrig, F. Öttl, R. Hinterholzer, and N. Augeneder, "Reliable Diagnostics on Rotating Machines using FRA," in 2020 International Conference on Diagnostics in Electrical Engineering (Diagnostika), 2020, pp. 1-6.
[28]C. Platero, F. Blazquez, P. Frías, and D. Ramirez, "Influence of Rotor Position in FRA Response for Detection of Insulation Failures in Salient Pole Synchronous Machines," IEEE Transactions on Energy Conversion, vol. 26, p. 671, 2011.

Yu Chen was born in Wenzhou, Zhejiang, China, in 2000. He received the B.S. degree from the College of Engineering and Technology, Southwest University, Chongqing, China, in 2022. He is currently pursuing a master's degree in the School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan, Hubei. His research areas include condition monitoring, fault diagnosing for power equipment, and the application of artificial intelligence.

ZHONGYONG ZHAO (M'19) 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. His areas of research include condition monitoring and fault diagnosing for HV apparatus, and artificial intelligence.

Yueqiang Yu was born in Nanchong, Sichuan, China, in 1994. He is currently pursuing the master 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 machine.

Yunlong Guo was born in Handan, Hebei, China, in 2000. He is currently pursuing the master degree in the Department of Automation, Beijing Institute of Technology, Beijing, China. His research interests include machine learning, networked systems and swarm intelligence.

Chao Tang was born in Sichuan, China, in 1981. He received the M.S. and Ph.D. degrees in electrical engineering from Chongqing University, China, in 2007 and 2010, respectively. As a Ph.D.student (2008-2009), and as a Visiting Scholar (2013-2013) and (2015-2016), he studied with the Tony Davies High Voltage Laboratory, University of Southampton, U.K., doing some research on the dielectric response characteristics and space charge behaviors of oil-paper insulation. He is currently a Professor with the College of Engineering Technology, Southwest University, China. His research interests include mainly in the field of on-line monitoring of insulation conditions and fault diagnosis for high-voltage equipment.