This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3214628

A Novel Multiview Sampling-based Meta Self-Paced Learning Approach for Class-imbalanced Intelligent Fault Diagnosis

Pin Lyu, Pai Zheng, *Senior Member, IEEE*, Wenbing Yu, Chao Liu, *Member, IEEE*, and Min Xia, *Senior Member, IEEE*,

***Abstract*—In practical machine fault diagnosis, the obtained data samples under faulty conditions are usually far less than those under normal conditions, resulting in a class-imbalanced dataset issue. The existing solutions for class-imbalanced scenarios include data-level and model-level strategies which are either subject to over-generalization or time-consuming. To address it, this paper proposes a novel multiview sampling-based meta self-spaced learning approach. Firstly, the signal processing methods such as time-domain, frequency-domain and time-frequency domain are used to extract statistical features from the original data to form diverse views. Next, the meta self-paced learning technology is applied to select high-quality samples from multiview feature data to generate a class-balanced dataset. Finally, a fault diagnosis model is trained with the obtained class-balanced dataset. The main contribution of this research has twofold: 1) the introduced multiview sampling method adaptively learns the weight in the sampling process, and automatically delete the noise samples with large loss value, to improve the performance of the fault diagnosis model, and 2) the proposed meta self-spaced learning approach eliminates the error caused by setting parameters manually and ensures the quality of the extracted samples. To validate its performance, a comparative study is conducted on a public dataset and the one collected from an industrial motor test platform. Five baseline methods are compared with the proposed one based on the convolutional neural network model. Moreover, three traditional machine learning models are to verify the sample quality generated. The experimental results achieve above 90% diagnosis accuracy, which provides a new intelligent manner for the modular service application of class-imbalance fault diagnosis.**

***Index Terms*—Fault diagnosis, class-imbalanced data, multiview sampling, self-paced learning, meta self-paced learning.**

I. Introduction

WITH a prevailing development of the manufacturing industry towards digitization and servitisation [1], data-driven intelligence with advanced analytics has become one of the key enablers for the intelligent manufacturing process, such as quality prediction [2], and equipment fault diagnosis [3]. Specifically, data-driven fault diagnosis has been widely studied, which can be commonly divided into traditional machine learning methods (e.g., artificial neural network (ANN) [4], support vector machine (SVM) [5]) and deep learning methods (e.g., convolutional neural network (CNN) [6], denoising autoencoder [7], stacked auto encoder [8], adversarial transfer network [9] and dynamic joint distribution alignment network [10]). Most of them can perform fault diagnosis when the amount of data is balanced for the different health state categories of the machine.

However, in real industrial scenarios, the obtained data samples under faulty conditions are usually far less than those under normal conditions, resulting in a class-imbalanced dataset issue, of which the performance of the existing data-driven fault diagnosis models is relatively low [11]. This is because fault samples are prone to be ignored and flooded by the majority-class normal samples, which may lead to misclassification, and eventually lead to equipment failure, catastrophic accidents, and economic losses [12]. To address the class-imbalanced problem, many works have been done recently, of which the most representative methods can be roughly divided into three categories, i.e., data-level methods, model-level methods, and hybrid methods.

Data-level methods aim to change unbalanced datasets into balanced datasets by a resampling strategy. Synthetic minority over-sampling technique (SMOTE) [13] is one of the most famous technologies, which is an improved scheme based on random over-sampling. The purpose is to increase the sensitivity of the classifier to minority class. Other improved versions of SMOTE, such as adaptive synthetic sampling approach (ADASYN) [14], KMeans-SMOTE [15], Borderline-SMOTE [16], and so on, introduce more concepts to guide the process of rebalancing. Similarly, there are also many under-

This work was supported by the Mainland-Hong Kong Joint Funding Scheme (MHKJFS) of the Innovation and Technology Commission, HKSAR under Grant MHX/001/20, National Key R&D Programs of Cooperation on Science and Technology Innovation with Hong Kong, Macao and Taiwan under Grant SQ2020YFE020182 by the Ministry of Science and Technology of China, National Natural Science Foundation of China under Grant 52105534, and Shanghai Science and technology program under Grant 22010500900. (Corresponding author: Pai Zheng)

Pin Lyu is with the School of Electronic Information Engineering, Shanghai Dianji University, Shanghai 201306, China; and a senior visiting scholar in the Department of Industrial and Systems Engineering, at The Hong Kong Polytechnic University. (e-mail: lvp@sdju.edu.cn).

Pai Zheng is with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (e-mail: pai.zheng@polyu.edu.hk).

Wenbing Yu is with the School of Higher Vocational Technology, Shanghai Dianji University, Shanghai 201306, China (e-mail: yuwb@sdju.edu.cn).

Chao Liu is with the College of Engineering and Physical Sciences, Aston University, Birmingham B47ET, U.K. (e-mail: c.liu16@aston.ac.uk).

Min Xia is with Department of Engineering, Lancaster University, Lancaster LA1 4YW, U.K. (e-mail: m.xia3@lancaster.ac.uk).

sampling methods such as Easyensemble [17] and clustering analysis and instance selection [18]. At present, data-level methods have been widely applied to intelligent fault diagnosis. For example, the combination of SMOTE and Easyensemble was used to diagnose the wind turbine freezing fault [19]. Fan et al [20] utilized Borderline-SMOTE and ADASYN to process sampled data for chiller fault diagnosis. A dynamic modified SMOTE was proposed in chemical fault diagnosis [21]. However, these methods suffer from over-generalization when applied in practice. Over-generalization means the fault diagnosis model is applicable, but it cannot accurately classify minority-class samples in practice. Besides, recently, improved generative adversarial networks (GAN) are also used to expand the unbalanced dataset to the balanced dataset [22]. However, the improved GAN model needs to overcome the problem of mode collapse and vanishing gradient.

Model-level methods seek for a new model, which biases the recognition process to the minority classes during the training stage. For example, Duan et al. [11] used the combination of support vector data description and binary tree to study the problem of rotor unbalanced fault diagnosis. However, it is very difficult to select the appropriate kernel function for the support vector machine. He et al. [23] used focal loss function in the training of model spatio-temporal multiscale neural network to realize the imbalanced wind turbine fault diagnosis. To address the imbalanced rotating machinery fault diagnosis, Xu et al. [24] used cost matrix determination to adjust the classification boundary. Peng et al. [25] proposed maximum expected cost reduction and cost-weighted minimum margin to address imbalanced fault diagnosis for plasma etching. However, the cost matrix largely depends on expert knowledge, and there lacks an effective method to evaluate the performance of a cost-sensitive model.

Recently, hybrid methods which consider both data distribution and model improvements have attracted attention in the fault diagnosis domain. For example, Liu et al. [12] combined the bagging algorithm and SVM to reduce the imbalance rate of original data, and then used a new weighted cross entropy loss function when training the fault diagnosis model based on the bidirectional gated recurrent unit. Other studies of hybrid methods focus on the use of GAN. For example, Zareapoor et al. [26] presented a new adversarial network that not only generates faulty samples, but also implements fault classification. Kuang et al. [27] proposed a class-imbalanced adversarial transfer learning network, in which the class-imbalanced learning was embedded into the adversarial training process to solve the problem of cross-domain fault diagnosis with imbalanced data. Yang et al [28] investigated a data generation method based on GANs to solve the problem of data imbalance and utilize the multiscale CNN to realize the fault diagnosis of the harmonic drive. In addition, the distillation learning method is combined with a layer regeneration network under class-imbalanced samples to realize bearing fault diagnosis [29]. However, these hybrid methods are either too complex in process or too rigid in structure to be applied in various fault diagnosis tasks.

To overcome the above problems, this paper proposes a novel multiview sampling method, which first obtains multiview of the class-imbalanced data using different signal processing technologies, and then sample the multiview with meta self-paced learning to generate a class-balanced dataset. The main contributions of this work are summarized as follows:

1) Meta self-paced learning is applied to extract samples from multiview feature data to form a class-balanced dataset. It adaptively learns the sample weight in the sampling process, and automatically delete the noise samples with large loss value. To the best of our knowledge, it is the first attempt to adopt multiview sampling-based meta self-paced learning to solve the class-imbalanced problem in fault diagnosis. The class-imbalanced problem in fault diagnosis is addressed by considering the idea of multiview of industrial data.

2) A meta-learner is utilized to overcome the parameter manually setting problem of self-paced learning, which effectively obtain the best optimization value and decrease the computational cost. It helps to eliminates the error caused by setting parameters manually and ensures the quality of the extracted samples by meta self-spaced learning.

The proposed method is further verified by adopting a deep learning model and three traditional machine learning models on a benchmark industrial dataset and also one collected from an industrial motor test platform. The experimental results indicate that the proposed method can serve as a model-agnostic class-imbalanced diagnosis methods, which has a good generality to be adopted in many other scenarios. Moreover, compared with other baseline methods on the same dataset, the proposed method is proved to be superior.

The remaining content of this paper is organized as follows. Section II briefly introduces the theoretical background about Multiview learning, self-paced learning, and meta self-paced learning. Then, the proposed method for class-imbalanced fault diagnosis is presented in Section III. To demonstrate its effectiveness, Section IV presents a comparative study on a well-established open dataset and a practical industrial dataset of bearing. Finally, Section V summarizes this work and highlights the potential future research directions

II. Theoretical Premises

A. Multiview learning

Multiview learning is to learn the internal structure of data using multiple distinct feature sets from different perspectives [30]. It aims to improve the generalization performance of the model with multiview data, which refers to the data that the same object is multimodally described. Multiview learning has been successfully applied in many fields. Especially, recent works have shown that multiview features can be fused with its prior knowledge and high-order cross-view correlation to generate an accurate self-representation tensor [31], which indicates the powerful ability of information combination from different viewpoints. These works inspired us to introduce multiview idea to generate multiview samples for class-imbalanced fault diagnosis.

However, direct mining of useful information contained in multiview data will cause three problems. First, if all views are directly combined, the data containing noise will affect the performance of each view in the combined data. Second, high-dimensional data may lead to the construction of an invalid fault diagnosis model and overfitting due to the curse of dimensionality. Third, adjacent features may be similar characteristics, which will lead to overfitting during model training. To avoid these three problems, this work introduces a meta self-paced learning to analyze the importance of different samples in class-imbalanced fault diagnosis dataset in order to obtain a high-quality balanced dataset from multiview samples.

B. Self-paced learning

Self-paced learning [32], which can simulate the cognition of people, has the ability of step-by-step learning from simple to complex tasks to solve the over-generalization of existing data-level methods mentioned in the introduction. During self-paced learning, the metric criteria of samples are embedded into the optimization model. Hence, self-paced learning can achieve adaptive sample ranking.

The core idea of self-paced learning is as follows. Supposing there is a training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ denotes the *ith* *d*-dimension sample. $y_i \in \mathcal{Y}$ is the label of the *ith* sample. $\mathcal{Y} = \{1,2,\cdots,C\}$ denote a set with C classes. N is the number of samples in the training dataset. Let f_w is a supervised model, where w is the parameter of model. $\mathcal{L}(y_i, f_w(x_i))$ is the loss of the *ith* sample on D. The implementation mode of self-paced learning is to minimize the objective function shown in (1) and solve the weight parameters s_i that measure the complexity of each sample and model parameters w.

$$\min_{w,s} \sum_{i=1}^{N} \left(s_i \mathcal{L}\big(y_i, f_w(x_i)\big) + g(s_i, \lambda) \right) \quad (1)$$

where $s = [s_1, s_2, \cdots, s_N]$ is the sample weight, λ is the age parameter, $g(s_i, \lambda)$ is the self-paced regularization. The solution of (1) can be completed by the following three-step alternating iteration to update the parameters s and w [33].

Step 1: Fix parameter w^*, and compute weight s of the *ith* sample:

$$s_i^* = arg \min_{s\in[0,1]} s_i \mathcal{L}\big(y_i, f_w(x_i)\big) + g(s_i, \lambda) \quad (2)$$

Step 2: Fix parameter s^*, compute the weight w of model:

$$w^* = arg \min_{w} \sum_{i=1}^{N} s^* \mathcal{L}\big(y_i, f_w(x_i)\big) \quad (3)$$

Step 3: Increase the value λ after updating s and w so that more samples can be selected.

In the continuous iteration of the above three steps, the samples with large loss values are regarded as samples with low quality, and they are deleted from the training datasets by assigning them a small weight or 0 to reduce or avoid the negative impact of such samples on the performance of the classifier. Meanwhile, in the self-paced learning model shown in (1), age parameters λ are hyperparameters that need to be set for self-paced learning. However, the age parameter setting of self-paced learning usually requires priori knowledge of data. Reasonably setting the age parameter is a challenge of self-paced learning.

C. Meta self-paced learning

Recent studies have shown that meta learning can effectively avoid the parameter setting problem of self-paced learning [34]. Meta learning requires a small number of high-quality meta samples to be collected in advance. High quality means that the selected meta samples must reflect the real distribution of data to realize the meta guidance of training data. Supposing that $D_{meta} = \{(x_i, y_i)\}_{i=1}^{\mathcal{M}}$ represents the metadata set, where $\mathcal{M}$ represents the number of meta samples and $\mathcal{M} \ll N$, the learning of age parameters can be guided by the loss on the meta dataset, as shown in (4) and (5).

$$\hat{\lambda} = arg \min_{\lambda} \sum_{i=1}^{\mathcal{M}} \mathcal{L}\big(f_{w(\lambda)}(x_i, y_i)\big) \quad (4)$$

$$w(\lambda) = arg \min_{\lambda} \sum_{i=1}^{N} s_i \mathcal{L}\big(y_i, f_w(x_i)\big) + (s_i, \lambda) \quad (5)$$

Meta self-paced learning algorithm usually uses an approximate strategy to update the age parameters λ, sample weight s and model parameters w alternately [33]. Age parameters can be adjusted adaptively through meta data, which can be easily integrated into any machine learning model and has good generality.

III. The Proposed Method

A. Diagnostic flowchart

Fig.1 describes the flowchart of class-imbalanced fault diagnosis based on the proposed method. It consists of three steps. The first step is to segment the original vibration signal and generate multiple views using different signal processing technologies. The second step is to sample multiple views with meta self-paced learning to obtain a balanced dataset. The third step is to train a fault classification model using the new balanced dataset to realize the final fault diagnosis.

B. Multiview generation

For the time-domain (TD) view in Fig. 1, eight statistical features are used including mean value, root mean square, skewness, kurtosis, crest factor, pulse factor, margin factor and entropy. For the frequency-domain (FD) view in Fig. 1, Fast Fourier transform is performed on each segment and the spectrum associated with each segment is analyzed. Eight frequency domain features are extracted including frequency domain amplitude average, center of gravity frequency, mean square frequency, frequency variance, root mean square frequency, frequency amplitude variance, frequency amplitude skewness index and frequency amplitude kurtosis index. For the time-frequency domain (TFD) view in Fig. 1, wavelet packet transform is used to extract the features of the original vibration signal. Eight different statistical features are extracted according to the coefficient calculation formula of each wavelet packet. Finally, 256 statistical features are obtained.

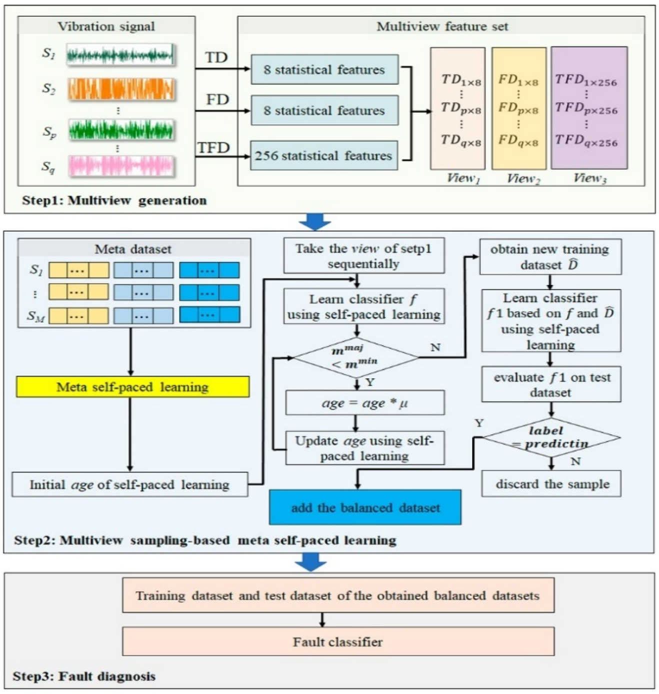

Fig. 1. Diagnostic flowchart based on the proposed method

C. Multiview sampling-based meta self-paced learning for class-imbalance fault samples

Algorithm 1 describes the process of generating balanced fault dataset by multiview sampling-based meta self-paced learning. First, the initial value of age parameter λ of self-paced learning is learned by meta self-paced learning. To the end, we construct a meta-sample set D_{meta} containing five clear meta samples using the method of selecting meta samples in section-IV, then optimize the objective function shown in (6).

$$\lambda^* = \underset{\lambda}{argmin} \sum_{i=1}^{5} \mathcal{L}\left(f_{w(\lambda)}(x_i, y_i)\right) \tag{6}$$

where $w(\lambda)$ is computed by formula (5). About the optimization of (6), random gradient descent is used to solve λ in the D_{meta}, using $\mathcal{M}'$ batch samples obtained from D_{meta} iterated in formula (7).

$$\lambda^t = \lambda^{t-1} - \eta \sum_{j=1}^{\mathcal{M}'} \nabla\left(f_{\widehat{w}^t(\lambda)}(x_j^{meta}), y_j^{meta}\right)\Big|_{\lambda^{t-1}} \tag{7}$$

where $\widehat{w}^t(\lambda)$ is computed in dataset D according to N batch samples in formula (8), η is learning rate.

$$\widehat{w}^t(\lambda) = w^{t-1} - \eta \sum_{j=1}^{N} \nabla_w s_j(\lambda)\mathcal{L}_D\left(f_w(x_j), y_j\right) \tag{8}$$

More samples are added in the next iteration with gradually increased values of λ and age growth factor μ, and low-quality samples are included in the training implementation increasingly, when the dataset does not achieve a balanced distribution.

Then, a classifier is iteratively learned on different views in turn. When the number of majority class samples m^{maj} is less than the number of minority class samples m^{min}, the age parameter is modified with the age growth factor μ, and then the self-paced learning is performed to obtain the updated age parameter to select more training samples. When the number of majority class samples is equal to the number of minority class samples, the under sampling of majority class terminates. At this time, a new training dataset is formed according to the sample weight. In this paper, the hard regularization is used, that is, all samples with weight 0 are deleted and all samples with weight 1 are formed into a new training dataset. The classifier is learned again on the new training dataset, and each sample in the test dataset is checked with the newly learned classifier. Real labels consistent with the prediction label are added to the fault balanced dataset.

Algorithm 1: Multiview sampling-based meta self-paced learning for class-imbalanced fault samples

Input: class-imbalanced training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, multiview $V = \{V_j\}_{j=1}^{P}$, age growth factor μ, initial sample weight s_0
Output: class-balanced dataset $\widetilde{D}$
1. Initialize λ and an empty $\widetilde{D}$
2. **For** $j = 1$ to P
3. Train a classifier f on V_j using self-paced learning
4. **While** $m^{maj} < m^{min}$ **do**
5. $\lambda *= \mu$
6. update λ based on self-paced learning
7. **End**
8. Obtain a new training dataset $\widehat{D}$ according to s_j
9. Train a new classifier f_1 on $\widehat{D}$ using self-paced learning
10. Predict the test dataset using f_1
11. Add samples in the test dataset whose real labels are consistent with the predicted results to the $\widetilde{D}_j$
12. $\widetilde{D} = \widetilde{D} \cup \widetilde{D}_j$
13. **End**
14. return $\widetilde{D}$

Algorithm 1 uses meta self-paced learning method to integrate age parameters, sample weight parameters and model parameters into the same optimization framework to realize the learning effect of dynamically adjusting age parameters from multiview data. This adaptive meta learning strategy enhance the robustness and accuracy of the original optimization problem for complex data deviations to ensure that high-quality samples can be sampled from the imbalanced fault dataset.

IV. CASE STUDY

A. Datasets description and preparation

1) **CWRU datasets**

CWRU bearing dataset [35] is a benchmark dataset for data-driven fault diagnosis, including vibration signals of various bearing conditions. In this experiment, the bearing data at the drive end and fan end are used. To better compare the fault diagnosis performance under different imbalance ratios, three new datasets D1, D2 and D3 with different imbalance ratios are constructed based on the vibration data files with load of 1hp, rotating speed of 1772r/min and sampling frequency of 48KHz.

The number of normal samples, the number of fault samples, different fault types and severities, and imbalance ratios are shown in Table I. According to the research conclusion of reference [36], the length of the signal segment is set to 1024. The imbalance ratios in Table I are 10:1, 50:1 and 100:1, respectively representing low, medium, and high imbalance case. The items in parentheses in Table I are the number of samples after signal segmentation.

2) **A bearing fault dataset from an industrial application**

In addition, to verify the effect of the proposed method in practical application, we use our developed intelligent motor test platform to collect the operation data of a three-phase asynchronous motor with bearing inner faults and outer faults and a healthy three-phase asynchronous motor under 0hp load. Based on the collected bearing datasets, three imbalanced datasets D4, D5 and D6 are constructed. The number of selected normal samples and fault samples, different fault types and imbalance ratio are shown in Table II.

B. Baseline methods

To verify the effectiveness of the proposed method, the following five baseline methods are used to make a comparative analysis with the proposed method.

SMOTE [13]: the number of nearest neighbors is set to 5.

Borderline-SMOTE [16]: the number of nearest neighbors is set to 5, and the number of nearest neighbors used to determine dangerous samples is set to 10.

ADASYN [14]: the number of nearest neighbors is set to 5.

Easyensemble [17]: CNN is used as the base classifier.

KMeans-SMOTE [15]: The number of nearest neighbors used to build a new sample is set to 4. The number of clusters is set to 10, 4 and 3 based on different experiment datasets in this paper, respectively.

TABLE I
NUMBER OF SAMPLES IN EACH CONSTRUCTED DATASET BASED ON CWRU

Sample category		Fault depth(inch)	Signal segmentation（number of samples）			class
			D1	D2	D3	
Majority	Normal		819 200 (800)	819 200 (800)	819 200 (800)	C0
Minority	Ball fault	0.007	81 920 (80)	16 384 (16)	8 192 (8)	C1
		0.014	81 920 (80)	16 384 (16)	8 192 (8)	C2
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C3
	Inner race fault	0.007	81 920 (80)	16 384 (16)	8 192 (8)	C4
		0.014	81 920 (80)	16 384 (16)	8 192 (8)	C5
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C6
	Outer race fault	0.007	81 920 (80)	16 384 (16)	8 192 (8)	C7
		0.014	81 920 (80)	16 384 (16)	8 192 (8)	C8
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C9
	Total		1 556 480 (1520)	966 656 (944)	892 928(872)	
	Imbalance ratio		10:1 Low	50:1 Medium	100:1 High	

TABLE II
NUMBER OF SAMPLES IN EACH CONSTRUCTED DATASET BASED ON AN INDUSTRIAL APPLICATION

Sample category	Fault category	Signal segmentation（number of samples）			class
		D4	D5	D6	
Majority	Normal	819 200 (800)	819 200 (800)	819 200 (800)	C0
Minority	Inner race fault	81 920 (80)	16 384 (16)	8 192 (8)	C1
	Outer race fault	81 920 (80)	16 384 (16)	8 192 (8)	C2
Total		983 040(960)	851 968 (832)	835 584(816)	
Imbalance ratio		10:1 Low	50:1 Medium	100:1 High	

TABLE III
DIFFERENT CNN STRUCTURE AND THEIR AVERAGE ACCURACY ON TRAINING DATASET

	Model_1	Model_2	Model_3	Model_4
CNN Structure	Conv(32,9)+Relu MaxPooling(2) Conv(32,9)+Relu MaxPooling(2) Dense(128)+Relu Dense(64)+Relu Dense(96)+Relu Dense(10)+Softmax	Conv(32,1)+Relu MaxPooling(2) Conv(32,5)+Relu MaxPooling(2) Dense(10)+Softmax	Conv(32,10)+Relu MaxPooling(2) Conv(32,10)+Relu MaxPooling(2) Dense(320)+Relu Dense(96)+Relu Dense(10)+ Softmax	Conv(32,10)+Relu MaxPooling(2) Conv(32,10)+Relu MaxPooling(2) Dense(10)+ Softmax
Average accuracy	0.9826±0.0041	0.9480±0.0063	0.9802±0.0052	0.9824±0.0034
Training time(s)	110.8	123.9	105.3	98.8
Prediction time(s)	1.6	1.7	1.6	1.9

C. Performance metrics

The essence of imbalanced bearing fault diagnosis is the classification of class-imbalanced datasets. Therefore, this paper selects three performance metrics G-mean, F1-score and area under curve (AUC) to evaluate the proposed method. The calculation method of performance metric is shown in the formula (9)-(14).

$$Sensitivity = \frac{TP}{TP + FN} \quad (9)$$

$$Specificity = \frac{TN}{TN + FP} \quad (10)$$

$$G - mean = \sqrt{Sensitivity \times Specificity} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$F1 - score = \frac{2 \times Sensitivity \times Specificity}{Sensivity + Specificity} \quad (13)$$

$$AUC = \frac{Sensitivity + Specificity}{2} \quad (14)$$

where *TP* represents the number of real minority class samples. *FN* is the number of real minority class samples misclassifying into majority class samples; *TN* represents the number of real majority class samples; *FP* is the number of real majority category samples misclassifying into minority class samples. In the multi-class classification problem in this paper, the *i*th fault category is regarded as a minority class, and all other faults are regarded as majority class. The above formula reveals that the higher the sensitivity, the higher the probability that minority class will be classified correctly. The higher the value of specificity, the higher the correct recognition rate of majority class samples. Only when the sensitivity and specificity are large, the values G-mean and F1-score are large. The larger the value of G-mean, the better the performance of the classifier on class-imbalanced datasets. The larger the value of F1-score, the higher the probability that minority samples are correctly classified. AUC can better reflect the classification accuracy of fault diagnosis model.

D. Comparative analysis of the proposed method and baseline methods

All experiments are run on TianKuo i620-g30 server, and the experimental environment is Python 3.7.0, Tensorflow, Keras and Pytorch, and machine learning toolkits include imblearn and scikit-learn.

1) **The structure selection of CNN**

CNN is selected as the fault diagnosis model to evaluate the proposed method and five baseline methods in this paper. To determine a simple and optimal CNN structure, we use the class-balanced dataset used to construct Table I. Sampling data point of each class is divided into segments length 1024 and resized to a two-dimensional matrix of size 32 by 32. There is no overlap between segments. For each category 460 such segments are taken. Thus, total size of the data becomes (4600, 32, 32). Out of 460 samples were randomly selected as the test data and the rest are used for training data.

Ten trials are conducted on the training dataset in this experiment. The hyperparameters of epoch and batch size are 100 and 128 during the training process, respectively. Therefore, the average accuracy of four different CNN structures is the average of ten trials in Table III. We record the training time of ten times and prediction time of the four different CNN structures in Table III to demonstrate that using CNN designed in this work will not cause expensive computing cost. Overall, the single training time of four CNN models does not exceed eighteen seconds. From the last two rows in Table III, although the training time of Model_1 is two seconds longer than Model_4, Model_1 has a faster response for the prediction. In addition, the four confusion matrices on test dataset based on different CNN models are also given. Integrating the experimental result of Table III and Fig.2, it is feasible to select Model_1 as the fault diagnosis model.

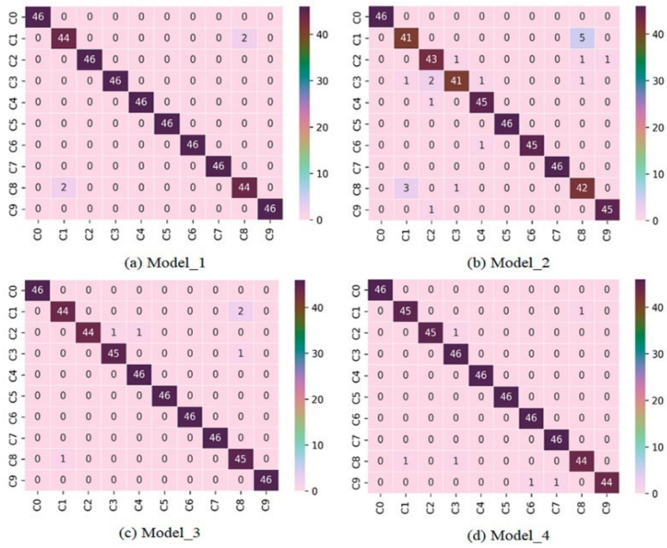

Fig. 2. Confusion matrix on test dataset based on different CNN structure.

2) **Selection of the optimal age growth factor**

The selection of the optimal age growth factor μ in Algorithm 1 is studied based on the balanced dataset for picking the optimal CNN structure. In our experiment, the range of μ is set to the interval [0.05,2.84], and the step is 0.31. The accuracy of the fault diagnosis in test set is shown in Fig.3, which generally follows the trend of increasing first and then decreasing, with the increase of μ. It also can be observed that when μ is less than 1, the accuracy increases continuously, and the maximum accuracy reaches 86.88% when μ is 0.98. As the value of the age growth factor exceeds 1, the accuracy of the model decreases from slow to sharp, and finally converges. This indicates that in the iterative process of Algorithm 1, if μ is too large, some noise samples are selected to the process of training. Therefore, in the follow-up experiments, the age growth factor μ is set to be 0.98.

3) **Performance analysis on CWRU dataset**

This section gives the comparison between the proposed method and five baseline methods in terms of three performance metrics. Logistic regression model is selected as the classifier in self-paced learning. First, the original class-imbalance dataset is randomly divided into a training dataset (70%) and a test dataset (30%). The 90% training dataset is

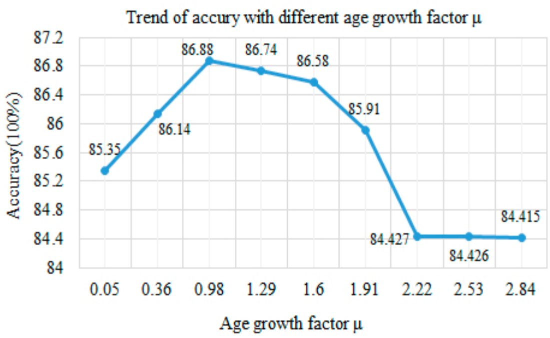

Fig.3. Selection of the optimal age growth factor

meta support set to learn the initial age parameter λ of self-paced learning.

The meta samples in this paper are selected from the validation set, because the high-quality validation set can reflect the real distribution of data and guide the training data in a good state. The selection process of validation set is as follows: the balanced fault dataset composed of (4600, 32, 32) samples in the part of the structure selection of CNN in section IV-D is divided randomly for 10 times. At each division, the whole dataset is split into training set, validation set and test set with a ratio of 8:1:1. Model_1 from the part of the structure selection of CNN in section IV-D is trained using each divided training set, adjust the hyperparameters of Model_1 the with the corresponding validation set, and use the Model_1 with the optimal parameter to predict the test set. The accuracy of these 10 trials is 94.9%, 97.4%, 94.7%, 97.3%, 91.7%, 95.7%, 94.3%, 93.5%, 92.1% and 93.8%, respectively.

It can be seen from the experimental results of 10 trails that the test performance of the second trial is the best. Therefore, the second randomly divided validation set is taken as the meta dataset, in which five samples are randomly selected as meta samples. When randomly selecting the meta samples, it is necessary to ensure that there are both normal samples and fault samples in the meta samples in order to correctly learn the age parameter of self-paced learning.

Each fault classification model is trained for 50 epochs. Each performance metric takes the mean and standard deviation of the results of 10 test runs. The performance of the fault diagnosis model on the balanced dataset obtained by six different methods is shown in Table IV, V and VI.

Table IV shows the performance of five different baseline methods and the proposed method in terms of F1-score. It can be seen from Table V that the proposed method is superior to all baseline methods in all datasets. It reached the highest value of 91.69% on the D1 dataset, about 3.95% higher than the worst SMOTE method. On the other hand, the average standard deviation of the proposed method is lower than that of the baseline methods, which indicates that the proposed method has high robustness. In addition, the average F1-score of the proposed method is about 2.73% higher than that of five baseline methods. This is because the proposed method uses meta self-paced learning to assign weights to samples. The adaptive learning of sample weight makes full use of the prior distribution information of the original data, which better distinguish the differences between samples. Therefore, the proposed method can minimize the impact of noise samples on training datasets and improve the over-generalization of the diagnostic model. It can be seen from Table V that the G-mean of the proposed method is also higher than that of the baseline methods, where the G-mean is the highest on the D3 dataset. For example, the G-mean on D3 dataset of SMOTE method is 87.32%, the proposed method is 95.67%, and the improvement range is 8.35%. At the same time, we find that he standard deviation of G-mean is higher than that of F1-score, which means that the proposed method focuses on minority class samples. In addition, the G-mean of the proposed method is about 4.22% higher than that average of the baseline methods. This shows that Multiview idea can enhance the discrimination ability of the diagnostic model and is effective for fault datasets with different imbalance ratios.

Table VI shows that the AUC of the proposed method is higher than that of the baseline methods, and the average standard deviation is lower than that of other methods, which indicates the proposed method is robust. In addition, the average AUC of the proposed method is about 3.7% higher than that of five baseline methods, which shows that the proposed method is conducive to improving the accuracy of fault diagnosis model.

In short, the proposed method is superior to the five baseline methods on three datasets in terms of the three metrics, which shows that the proposed method is effective and robust.

4) **Performance analysis on a bearing fault dataset from an industrial application**

Except for the verification of the proposed method on public datasets, the proposed method is also evaluated with a bearing dataset from practical industrial applications. The number of neurons in the last layer of fault classification model CNN is 3. Table VII, VIII and IX show the performance of the fault diagnosis model on the balanced dataset obtained by different methods, of which it is found that the average performance of the proposed method on the bearing data set of practical industrial application is better than that of the public datasets, especially the average value of F1-score. This shows that the proposed method can correctly classify fault samples in practical industrial applications with high probability even though the dataset is imbalanced. In addition, it can be seen from Table XIII that the G-mean of the proposed method is lower than SMOTE on dataset D6. This is because SMOTE first increases the number of minority samples according to the nearest internal distance in the minority samples. However, some noise data are not contained in the generated balanced dataset with our proposed method. The test dataset may still contain these noises. The incorrect prediction causes the change of G-mean accordingly.

5) **Running time of the proposed method and baseline methods**

To verify the efficiency of the proposed method, we calculated the running time for the proposed method and five

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2022.3214628

TABLE IV
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE F1-SCORE ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D1	0.8774±0.052	0.8983±0.046	0.8994±0.034	0.8993±0.045	0.8783±0.054	**0.9169±0.024**
D2	0.8676±0.061	0.8765±0.017	0.8876±0.033	0.8743±0.022	0.8786±0.034	**0.8934±0.011**
D3	0.7856±0.032	0.8577±0.023	0.8669±0.032	0.8963±0.046	0.8675±0.058	**0.8889±0.005**
Average F1-score	0.844±0.048	0.8775±0.029	0.8846±0.032	0.8809±0.038	0.8748±0.049	**0.8997±0.013**

TABLE V
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE G-MEAN ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D1	0.8788±0.048	0.8929±0.031	0.8979±0.031	0.9218±0.028	0.9076±0.011	**0.9299±0.032**
D2	0.8654±0.082	0.8678±0.001	0.8854±0.047	0.8965±0.110	0.8958±0.063	**0.9189±0.024**
D3	0.8732±0.013	0.9043±0.003	0.8988±0.032	0.9365±0.032	0.9233±0.058	**0.9567±0.049**
Average G-mean	0.8725±0.048	0.8866±0.012	0.8940±0.037	0.9031±0.057	0.9089±0.021	**0.9352±0.035**

TABLE VI
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE AUC ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D1	0.9066±0.015	0.9123±0.024	0.9242±0.067	0.9365±0.032	0.8778±0.040	**0.9432±0.012**
D2	0.9032±0.023	0.9365±0.041	0.9301±0.074	0.8978±0.023	0.9335±0.016	**0.9524±0.015**
D3	0.8465±0.030	0.8836±0.030	0.9002±0.056	0.9285±0.032	0.9149±0.030	**0.9432±0.029**
Average AUC	0.8854±0.023	0.9111±0.032	0.9182±0.066	0.9209±0.029	0.9087±0.029	**0.9463±0.019**

TABLE VII
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE F1-SCORE ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D4	0.9480±0.028	0.9220±0.031	0.8318±0.084	0.9232±0.051	0.8260±0.019	**0.9590±0.028**
D5	0.9149±0.022	0.8970±0.017	0.8971±0.014	0.9012±0.013	0.9035±0.012	**0.9639±0.007**
D6	0.9231±0.008	0.9310±0.041	0.7238±0.034	0.7648±0.028	0.9319±0.011	**0.9371±0.006**
Average F1-score	0.9287±0.019	0.9167±0.029	0.8176±0.044	0.8631±0.031	0.8871±0.014	**0.9533±0.014**

TABLE VIII
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE G-MEAN ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D4	0.9479±0.027	0.8453±0.071	0.9062±0.036	0.7594±0.037	0.9179±0.010	**0.9589±0.027**
D5	0.9004±0.022	0.9048±0.011	0.9019±0.016	0.9063±0.012	0.9026±0.014	**0.9634±0.007**
D6	**0.9532±0.007**	0.9257±0.028	0.7899±0.024	0.9341±0.009	0.9345±0.034	**0.9389±0.006**
Average G-mean	0.9338±0.019	0.8919±0.037	0.8660±0.025	0.8666±0.019	0.9183±0.018	**0.9537±0.013**

TABLE IX
COMPARISION WITH THE DIAGNOSTIC MODELS IN TERMS OF THE AUC ON DIFFERENT DATASET

Dataset	SMOTE	Borderline-SMOTE	ADASYN	EasyEnsemble	KMeans-SMOTE	Proposed method
D4	0.9923±0.008	0.9813±0.012	0.9649±0.026	0.9859±0.017	0.9345±0.057	**0.9500±0.111**
D5	0.9229±0.010	0.9540±0.005	0.9380±0.004	0.9655±0.003	0.9254±0.014	**0.9859±0.006**
D6	0.9311±0.009	0.8001±0.011	0.8236±0.009	0.9348±0.012	0.9368±0.008	**0.9408±0.006**
Average AUC	0.9488±0.009	0.9118±0.009	0.9087±0.013	0.9621±0.011	0.9322±0.026	**0.9589±0.041**

baseline models to obtain a balanced dataset. The experimental results are shown in Fig. 4 and Fig. 5. It should be noted that the running time presented here is the average for obtaining ten balanced datasets. From Fig.4 and Fig.5, it can be seen that the running time decreases with the reduction of the number of processed samples. For example, the running time of the proposed method on D1 (1520 samples) is 231.8 seconds, which is 131.4 seconds longer than that on D3 (872 samples). The running time of the proposed method on D4 (960 samples) is 100.3 seconds, which is 75.6 seconds longer than that on D6 (816 samples). It indicates: 1) the size of imbalanced dataset is one factor influencing the efficiency of the proposed method, and 2) the less sample categories in the imbalanced dataset, the higher efficiency of the proposed method. This is because the fewer sample categories are, the easier implementation of learning mechanism of meta self-paced learning is, thereby reducing the running time to obtain the initial value of age parameters.

Moreover, according to Fig. 4 and Fig.5, one can also observe that: 1) Among all the baseline methods, over-sampling strategies take less running time than under-sampling strategies. For example, compared to its improved version, SMOTE takes the least running time and EasyEnsemble takes the most on all datasets. This is because EasyEnsemble uses the ensemble method in the process of obtaining balanced datasets. It not only needs to sample multiple subsets from majority classes, but also needs to train multiple base classifiers and learn their results. 2) Except EasyEnsemble, the proposed method generally takes more running time than the baseline methods. For example, on D1, the proposed method takes 77.5 seconds more than SMOTE which takes the least running time. On D4, the proposed method is 43.5 seconds longer than EasyEnsemble which takes the most time. This is because the proposed method includes the running time to extract three different kinds of statistical features from the unbalanced dataset, which is not included in the baseline methods. 3) Running time of the proposed method is lower than EasyEnsemble on D2, D3, D5 and D6. The above analysis shows that the proposed method is more effective than baseline ones in the class-imbalanced fault diagnosis.

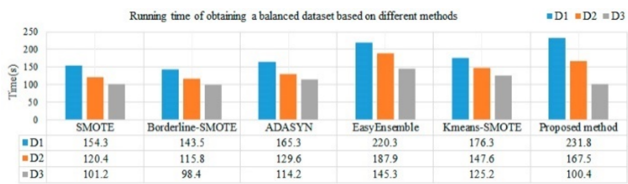

Fig.4. Running time of different methods on D1, D2 and D3

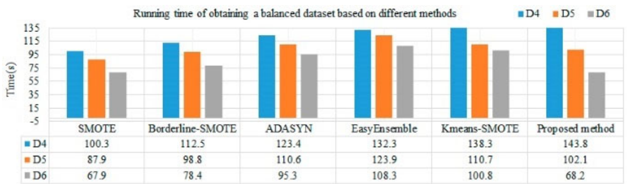

Fig.5. Running time of different methods on D4, D5 and D6

E. Influence of number of majority-class samples for the proposed method

To evaluate the performance of the proposed method when the number of majority class samples in the dataset is small and close to the number of fault samples, the sample numbers of majority class each dataset are changed to 100, 20 and 10 in Table I and II, respectively. The new constructed dataset is randomly divided into a training dataset (70%) and a test dataset (30%). The 90% training dataset is meta support set to learn the initial age parameter of self-paced learning. Model_1 is still used as a diagnostic model. The performance of the proposed method on the reconstructed dataset is shown in Table X and XI. Since the reconstructed dataset is class-balanced, accuracy is used as performance metric.

From Table X and XI, we can observe that 1) the performance of the proposed method is directly related to the amount of data. When the amount of original data is large, the more high-quality samples that meta self-paced learning can learn from multiple views, the better the performance of the model. For example, in Table X, when the number of majority class samples is reduced by 5 times from 100, the average accuracy of the diagnostic model is reduced by 29.41%. When the number of normal samples is reduced by 10 times from 100, the average accuracy of the diagnostic model is reduced by about 50%; 2) with the reduction of majority class samples, the fewer fault classes are, the better the performance of the proposed method. For example, when the fault class changes from 10 and 3, the performance on is 39.89% in D3_new and 81.54% on D6_new. This means that fault features are effectively mined using the multiview idea; 3) when the number of majority class samples becomes small, for example, from 100 to 10, the optimal average accuracy of the proposed method is 81.54%, which shows that the proposed method is more suitable to class-imbalanced scenario.

F. Portability of the proposed method

Based on the experimental results in Subsection D, one can know that the combination of multiview sampling-based meta self-paced learning and the fault diagnosis model based on CNN has a good effect on the industrial application scenario dataset. To further verify the balanced dataset generated by the proposed method is effective for traditional machine learning methods, we select SVM, RandomForest (RF) and ANN as diagnosis model for the practical industrial application dataset. The baseline methods KMeans-SMOTE and SMOTE are selected, which have the best G-mean in the industrial application scenarios to compare with the proposed method.

These three diagnostic models based on traditional machine methods are trained on the datasets D4, D5 and D6. The experimental results are shown in Table XII, XIII, and XIV. In this experiment, the features of TD, FD and TFD views are selected as the features of the three fault diagnosis models.

It can be seen from Table XII, XIII, and XIV that no matter which features is used in the three fault diagnosis models, the value of G-mean obtained by the proposed method is higher than that of the other two methods. The highest G-mean value of the proposed method is 96.72% of SVM on D5, followed by 95.27 % of RF on D4 dataset, and the lowest G-mean value is 89.43% of ANN on D4 dataset. These values are slightly inferior to those in the last column in Table VIII. But the difference between lowest G-mean of traditional machine learning method and average G-mean of deep learning modelCNN in Table VIII is up to 5.94%. This shows that the method proposed in this paper is independent of the model and has good portability.

Therefore, it is an effective way to convert the imbalanced dataset into a balanced dataset composed of high-quality samples through sampling, and then different fault diagnosis models are selected according to the actual configuration of computing resources in industrial scenarios. For example, in the fault diagnosis system based on cloud edge architecture, if

TABLE X
NUMBER OF SAMPLES IN EACH CONSTURCTED DATASET BASED ON CWRU

Sample category		Fault depth(inch)	Signal segmentation（number of samples）			class
			D1_new	D2_new	D3_new	
Majority	Normal		102 400 (100)	20 480 (20)	10 240 (10)	C0
		0.007	81 920 (80)	16 384 (16)	8 192 (8)	C1
	Ball fault	0.014	81 920 (80)	16 384 (16)	8 192 (8)	C2
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C3
		0.007	81 920 (80)	16 384 (16)	8 192 (8)	C4
Minority	Inner race fault	0.014	81 920 (80)	16 384 (16)	8 192 (8)	C5
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C6
		0.007	81 920 (80)	16 384 (16)	8 192 (8)	C7
	Outer race fault	0.014	81 920 (80)	16 384 (16)	8 192 (8)	C8
		0.021	81 920 (80)	16 384 (16)	8 192 (8)	C9
Total			839 680 (820)	36 864 (164)	83968(82)	
Imbalance ratio			1:1	1:1	1:1	
Average accuracy			0.8965±0.019	0.6024±0.043	0.3989±0.062	

TABLE XI
NUMBER OF SAMPLES IN EACH CONSTRUCTED DATASET BASED ON AN INDUSTRIAL APPLICATION

Sample category	Fault category	Signal segmentation（number of samples）			class
		D4_new	D5_new	D6_new	
Majority	Normal	102 400 (100)	20 480 (20)	10 240 (10)	C0
Minority	Inner race fault	81 920 (80)	16 384 (16)	8 192 (8)	C1
	Outer race fault	81 920 (80)	16 384 (16)	8 192 (8)	C2
Total		266 240 (260)	53 248 (52)	835 584(26)	
Imbalance ratio		1:1	1:1	1:1	
Average accuracy		0.9895±0.011	0.9268±0.038	0.8154±0.164	

TABLE XII
G-MEAN OF DIAGNOSTIC MODEL BASED ON SVM

Method	TD			FD			TFD		
	D4	D5	D6	D4	D5	D6	D4	D5	D6
SMOTE+SVM	0.8641	0.8452	0.8502	0.8016	0.8531	0.8609	0.8609	0.8248	0.8966
KMeans-SMOTE+SVM	0.8873	0.8271	0.8993	0.8236	0.8946	0.8527	0.8527	0.8524	0.8819
Proposed method+SVM	0.9389	0.9164	0.9436	0.9257	**0.9672**	0.9393	0.9393	0.9027	0.9348

TABLE XIII
G-MEAN OF DIAGNOSTIC MODEL BASED ON RF

Method	TD			FD			TFD		
	D4	D5	D6	D4	D5	D6	D4	D5	D6
SMOTE+RF	0.8654	0.8623	0.8482	0.8346	0.8631	0.8529	0.8529	0.8328	0.8852
KMeans-SMOTE+RF	0.8783	0.8771	0.8865	0.8443	0.8734	0.8765	0.8765	0.8564	0.8942
Proposed method+RF	0.9376	0.9034	0.9266	0.9527	0.9298	0.9003	0.9003	0.9174	0.9178

TABLE XIV
G-MEAN OF DIAGNOSTIC MODEL BASED ON ANN

Method	TD			FD			TFD		
	D4	D5	D6	D4	D5	D6	D4	D5	D6
SMOTE+ANN	0.9001	0.9001	0.8652	0.8554	0.8554	0.813	0.813	0.813	0.8451
KMeans-SMOTE+ANN	0.8965	0.8965	0.8883	0.8632	0.8632	0.8789	0.8789	0.8789	0.8693
Proposed method+ANN	0.9042	0.9042	0.9232	0.8943	0.8952	0.9104	0.9133	0.9027	0.8948

the computing resources at the edge are limited, the traditional machine learning model can be selected instead. Therefore, the proposed method in this paper provides a new idea for the modular service application of category imbalance fault diagnosis. For example, in the fault diagnosis system based on cloud edge architecture, if the computing resources at the edge are limited, the traditional machine learning model can be selected instead. Therefore, the proposed method in this paper provides a new idea for the modular service application of category imbalance fault diagnosis.

V. CONCLUSION

This paper proposed a multiview sampling-based meta self-paced learning for class-imbalanced fault diagnosis. It can automatically remove some noise samples from imbalanced dataset and identify a suitable subset to form a balanced dataset, thus improving fault diagnosis performance. By comparing with five baseline models, when a CNN model is used as a diagnosis model, the performance of the proposed method is on average the best among the bearing fault datasets from two different application scenarios including the public dataset CRWU and an industrial dataset collected by our own test platform. More importantly, even if the imbalance ratio is 100, the proposed method still shows better performance than the five baseline methods. Subsequently, the influenced number of majority-class samples for the proposed method is observed. The experimental results show that the proposed method is more suitable to class-imbalanced scenarios.

In addition, the proposed method can also achieve high fault diagnosis performance when the traditional machine learning algorithm is used as the fault diagnosis model. It further indicates that the generality of the proposed method, as a model-agnostic class-imbalanced diagnosis method. However, this research work still has some limitations, for instances:1) multiview generation. The proposed method only samples high-quality instances from three conventional views such as time domain, frequency domain and time-frequency domain. In fact, there are other kinds of statistical feature views. Hence, in future work, we will further explore the number of multivew and different combination of multiview in order to produce the high-quality balanced dataset for class-imbalance fault diagnosis. 2) meta samples. The quality of meta dataset directly affects the initial value of age parameter. In this work, meta samples are selected using validation set in advance, which not necessarily guarantee the quality of the meta sample. Our future research will develop a method that can automatically extract meta sample. Taking the above factors into consideration, potential future work will focus on the combination strategy analysis of multiview for vibration signal data and automatic selection of meta data for class-imbalanced fault diagnosis, which can further take full advantages of multimodal data and cognitive mechanism.

References

[1] C. Liu, P. Zheng, and X. Xu, “Digitalisation and servitisation of machine tools in the era of Industry 4.0: a review,” *Int. J. Prod. Res.*, pp. 1-31, Aug. 2021, doi: 10.1080/00207543.2021.1969462.

[2] L. Xia, P. Zheng, X. Huang, and C. Liu, “A novel hypergraph convolution network-based approach for predicting the material removal rate in chemical mechanical planarization,” J. Intell. Manuf., pp. 1-21, May. 2021, doi: 10.1007/s10845-021-01784-1.

[3] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, and C. W. de Silva, “Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning,” Reliab. Eng. Syst. Saf., vol. 215, Nov. 2021, doi: 10.1016/j.ress.2021.107938.

[4] D. S. Huang, “A constructive approach for finding arbitrary roots of polynomials by neural networks,” IEEE Trans. Neural Net., vol. 15, no. 2, pp. 477-491, Mar. 2004, doi: 10.1109/TNN.2004.824424.

[5] Y. Li, M. Xu, Y. Wei, and W. Huang, “A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree,” Measurement, vol. 77, pp. 80-94, Jan. 2016, doi: 10.1016/j.measurement.2015.08.034.

[6] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, and M. Xia, “Adversarial domain-invariant generalization: a generic domain-regressive framework for bearing fault diagnosis under unseen conditions,” IEEE Trans. Ind. Informat., vol. 18, no. 3, pp. 1790–1800, Mar. 2022, doi: 10.1109/TII.2021.3078712.

[7] M. Xia, T. Li, T. Shu, J. Wan, C. W. de Silva, and Z. Wang, “A two-stage approach for the remaining useful life prediction of bearings using deep neural networks,” IEEE Trans. Ind. Informat., vol. 15, no. 6, pp. 3703–3711, Jun. 2019, doi: 10.1109/TII.2018.2868687.

[8] H. Shao, M. Xia, J. Wan, and C. de Silva, “Modified stacked auto-encoder using adaptive morlet wavelet for intelligent fault diagnosis of rotating machinery,” IEEE/ASME Trans. Mechatron., vol. 27, no. 1, pp. 24-33, Feb. 2021, doi: 10.1109/TMECH.2021.3058061.

[9] Z. Chen, G. He, J. Li, Y. Liao, K. Gryllias, and W. Li, “Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery,” IEEE Trans. Instrum. Meas., vol. 69, no. 11, pp. 8702–8712, Nov. 2020, doi: 10.1109/TIM.2020.2995441.

[10] C. Shen, X. Wang, D. Wang, Y. Li, J. Zhu, and M. Gong, “Dynamic joint distribution alignment network for bearing fault diagnosis under variable working conditions,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1-13, Feb. 2021, doi: 10.1109/TIM.2021.3055786.

[11] L. Duan, M. Xie, T. Bai, and J. Wang, “A new support vector data description method for machinery fault diagnosis with unbalanced datasets,” Expert Syst. Appl., vol. 64, pp. 239–246, Dec. 2016, doi: 10.1016/j.eswa.2016.07.039.

[12] H. Liu, Z. Liu, W. Jia, D. Zhang, and J. Tan, “A novel imbalanced data classification method based on weakly supervised learning for fault diagnosis,” IEEE Trans. Ind. Informat., vol. 18, no. 3, pp. 1583-1593, Mar. 2021, doi: 10.1109/TII.2021.3084132.

[13] N. v. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, pp.321-357, Jun. 2002, doi: 10.1613/jair.953.

[14] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: adaptive synthetic sampling approach for imbalanced learning,” in Proc. IJCNN, June. 2008, pp. 1322-1328.

[15] G. Douzas, F. Bacao, and F. Last, “Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE,” Inf. Sci., vol. 465, pp. 1-20, Oct. 2018, doi: 10.1016/j.ins.2018.06.056.

[16] H. Han, W. Y. Wang, and B. H. Mao, “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning,” ICIC 2005, Aug. 2005, pp.878-887.

[17] X. Y. Liu, J. Wu, and Z. H. Zhou, “Exploratory under-sampling for class-imbalance learning,” IEEE Trans. Syst. Man Cybern, Part B (Cybernetics), vol. 39, no. 2, pp. 539-550, Apr. 2009, doi: 10.1109/TSMCB.2008.2007853.

[18] C. F. Tsai, W. C. Lin, Y. H. Hu, and G. T. Yao, “Under-sampling class imbalanced datasets by combining clustering analysis and instance selection,” Inf. Sci., vol. 477, pp. 47–54, Mar. 2019, doi: 10.1016/j.ins.2018.10.029.

[19] Z. Wu, W. Lin, and Y. Ji, “An integrated ensemble learning model for imbalanced fault diagnostics and prognostics,” IEEE Access, vol. 6, pp. 8394-8402, Feb. 2018, doi: 10.1109/ACCESS.2018.2807121.

[20] Y. Fan, X. Cui, H. Han, and H. Lu, “Chiller fault detection and diagnosis by knowledge transfer based on adaptive imbalanced processing,” Sci. Technol. Built. Environ., vol. 26, no. 8, pp.1082-1099, Apr. 2020, doi: 10.1080/23744731.2020.1757327.

[21] X. Gu, Y. Zhao, G. Yang, and L. Li, “An imbalance modified convolutional neural network with incremental learning for chemical fault diagnosis,” IEEE Trans. Ind. Informat., vol. 18, no. 6, pp. 3630 – 3639, Sep. 2021, doi: 10.1109/TII.2021.3112988.

[22] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, “A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks,” IEEE Trans. Instrum. Meas., vol. 70, pp.1-17, July. 2021, doi: 10.1109/TIM.2020.3009343.

[23] Q. He, Y. Pang, G. Jiang, and P. Xie, “A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data,” IEEE Trans. Ind. Informat., vol. 17, no. 10, pp. 6875 – 6884, Nov. 2021, doi: 10.1109/TII.2020.3041114.

[24] Q. Xu, S. Lu, W. Jia, and C. Jiang, “Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning,” J. Intell. Manuf., vol. 31, no. 6, pp. 1467-1481, Dec. 2020, doi: 10.1007/s10845-019-01522-8.

[25] P. Peng, W. Zhang, Y. Zhang, Y. Xu, H. Wang, and H. Zhang, “Cost sensitive active learning using bidirectional gated recurrent neural networks for imbalanced fault diagnosis,” Neurocomputing, vol. 407, pp. 232–245, Sep. 2020, doi: 10.1016/j.neucom.2020.04.075.

[26] M. Zareapoor, P. Shamsolmoali, and J. Yang, “Oversampling adversarial network for class-imbalanced fault diagnosis,” Mechanical Systems and Signal Process., vol. 149, Feb. 2021, doi: 10.1016/j.ymssp.2020.107175.

[27] J. Kuang, G. Xu, T. Tao, and Q. Wu, “Class-imbalance adversarial transfer learning network for cross-domain fault diagnosis with imbalanced data,” IEEE Trans. Instrum. Meas., vol. 71, pp.1-11, Dec. 2021, doi: 10.1109/TIM.2021.3136175.

[28] G. Yang, Y. Zhong, L. Yang, H. Tao, J. Li, and R. Du, “Fault diagnosis of harmonic drive with imbalanced data using generative adversarial network,” IEEE Trans. Instrum. Meas., vol. 70, pp.1-11, June. 2021, doi: 10.1109/TIM.2021.3089240.

[29] F. Li, J. Chen, S. He, and Z. Zhou, “Layer regeneration network with parameter transfer and knowledge distillation for intelligent fault diagnosis of bearing using class unbalanced sample,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1-10, July. 2021, doi: 10.1109/TIM.2021.3097408.

[30] Q. Lin, Z. Wang, Y. Chen, and P. Zhong, “Supervised multi-view classification via the sparse learning joint the weighted elastic loss,” Signal Proces., vol. 191, Feb. 2022, doi: 10.1016/j.sigpro.2021.108362.

[31] Xiao, X., Chen, Y., Gong, Y. J., & Zhou, Y, “Prior knowledge regularized multiview self-representation and its applications,” IEEE Trans. Neural Netw. Learn Syst., vol.32, no.3, pp.1325-1338, Mar.2021, DOI: 10.1109/TNNLS.2020.2984625.

[32] M. P. Kumar, B. Packer, and D. Koller, “Self-paced learning for latent variable models.” NIPS 2010, Dec. 2010, pp.1-9.

[33] J. Shu, D. Meng, and Z. Xu, “Meta self-paced learning,” Scientia Sinica Informationis, vol. 50, no. 6, pp.781-793, June. 2020, doi: 10.1360/SSI-2020-0005.

[34] Q. Wang, Y. Zhou, Z. Cao, and W. Zhang, “M2SPL: generative multiview features with adaptive meta-self-paced sampling for class-imbalance learning,” Expert Syst. Appl., vol. 189, Mar. 2022, doi: 10.1016/j.eswa.2021.115999.

[35] Bearing Data Center, “Bearing data center,” [Online]. Available: http://csegroups.case.edu/ bearingdatacenter/home.

[36] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, “An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors,” IEEE Trans. Ind. Informat., vol. 13, no. 6, pp.2758-2769, Dec. 2017, doi: 10.1109/TII.2017.2755064.

Pin Lyu received the M.S. and Ph.D. degrees in the School of Computer Science and Technology from Wuhan University of Technology, Wuhan, China, in 2001 and 2014, respectively. She is currently a professor with the School of Electronics Information Engineering, Shanghai Dianji University, and a senior visiting scholar in the Department of Industrial and Systems Engineering, at The Hong Kong Polytechnic University. Her current research interests include machine learning, and big data analytics in smart manufacturing.

Pai Zheng (Senior Member, IEEE) is currently an Assistant Professor in the Department of Industrial and Systems Engineering, at The Hong Kong Polytechnic University. He has been a Research Fellow in the Delta-NTU Corporate Laboratory for Cyber-Physical Systems, in the School of Electrical and Electronic Engineering, at Nanyang Technological University (NTU), Singapore from Jan. 2018 to Sep. 2019. Dr. Zheng received the Dual Bachelor’s Degrees in Mechanical Engineering (Major) and Computer Science and Engineering (Minor) from Huazhong University of Science and Technology, Wuhan, China, in 2010, the Master’s Degree in Mechanical Engineering from Beihang University, Beijing, China in 2013, and the Ph.D. degree in Mechanical Engineering at The University of Auckland, Auckland, New Zealand in 2017. His research interest includes smart product-service systems, human-robot collaboration, and intelligent manufacturing systems. He serves as the Associate Editor of Journal of Intelligent Manufacturing and Journal of Cleaner Production, Editorial Board Member of the Journal of Manufacturing Systems and Advanced Engineering Informatics, and Guest Editor/Reviewer for several high impact international journals in the manufacturing and industrial engineering field.

Wenbing Yu received the M.S. degree in the Department of Photoelectric Engineering from Huazhong University of Science and Technology, China, in 2004. He is currently a senior engineer with the school of Higher Vocational Technology, Shanghai Dianji University. He has presided and participated in four scientific research projects including NSFC general project and Hubei Natural Science Foundation. He published more than twenty academic papers including SCI or EI. His current research interests include intelligent computing, photonic crystal fiber sensors and periodically poled lithium niobate all-optical communication devices

Chao Liu (Member, IEEE) is a Lecturer in the College of Engineering and Physical Sciences at Aston University. He received his Ph.D. in Mechanical Engineering at the University of Auckland, New Zealand. He obtained his Bachelor’s degree and Master’s degree in Mechanical Engineering both from Northeastern University, China. Prior to joining Aston University, Dr. Liu has worked as a Research Assistant Professor in the Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University in Hong Kong, and as a Postdoctoral Research Associate in the School of Engineering at Cardiff University, UK. He is a member of SME and a Managing Guest Editor for Robotics and Computer-Integrated Manufacturing. His research interests include Industry 4.0, smart manufacturing, machine tools, digital twin, CPS, IoT, big data analytics, AR, machine learning, metal additive manufacturing, etc.

Min Xia (Senior Member, IEEE) received the B.S. degree from Southeast University, Nanjing, China, in 2009, the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada, in 2017. He is currently a Lecturer with the Department of Engineering, Lancaster University, Lancaster, U.K. His research interests include smart manufacturing, machine diagnosis and prognosis, deep neutral networks, autonomous vehicle, and sensor fusion. He serves as the Associate Editor of IEEE Transactions on Instrumentation and Measurement (2021-), IET Intelligent Collaborative Manufacturing (2021-), and Mechatronic Systems and Control (2020 -), Topic Editor of Electronics (2020 -) and Junior Associate Editor of Unmanned Systems Technology (2020 -).