# rgfc-Forest: An enhanced deep forest method towards small-sample fault diagnosis of electromechanical system

Yuhang Ming[1], Haidong Shao[1*], Baoping Cai[2], Bin Liu[3]

1 College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

2 College of Mechanical and Electronic Engineering, China University of Petroleum, Qingdao 266580 China

3 Department of Management Science, University of Strathclyde, Glasgow G1 1XQ, UK

**Corresponding author: Haidong Shao (hdshao@hnu.edu.cn)**

Email:

Yuhang Ming: mingyuhang@hnu.edu.cn

Haidong Shao: hdshao@hnu.edu.cn

Baoping Cai: caibaoping@upc.edu.cn

Bin Liu: b.liu@strath.ac.uk

**Abstract:** Deep forest models offer a promising alternative to traditional deep neural networks by demanding fewer training samples and hyperparameters. However, existing deep forest fault diagnosis models encounter persistent challenges such as insufficient representation of multi-grained spatial information and redundancy of cascaded forest features. To address the above challenges, an enhanced deep forest method called random multi-grained fusion cascade forest (rgfc-Forest) is presented for fault diagnosis of electromechanical systems with limited training samples. First, a random multi-grained scanning module is designed to improve feature information learning. Subsequently, a feature fusion cascade forest module is constructed to improve the representativeness of features in multi-grained scanning and cascade forest delivery while ensuring data diversity. Finally, a decision tree self-growth strategy is combined to refine the classification capability of the high-level forest. To evaluate the effectiveness of our proposed method, we applied it to experimental data related to motor system and gearbox faults. Our results demonstrate significant improvements over existing methods: With just 20 samples per class, our method achieved an average accuracy of 84.41% for motor System Diagnosis. Similarly, for the gearbox system, we attained an impressive accuracy of up to 92.72% with the same limited dataset. These outcomes underscore the superior feature representation and fault classification capabilities of our approach compared to both benchmark deep forest models and mainstream deep learning methods when confronted with small training datasets.

**Keywords:** Fault diagnosis; Small training samples; rgfc-Forest; Random multi-grained scanning; Feature fusion cascade forest

## 1. Introduction

Electromechanical systems are integral to various industries, underpinning the reliable and efficient operation of devices and machinery (Xu et al., 2023; Wen and Xu, 2021; Li et al., 2022; Liao et al., 2023). However, components of machinery operating under complex conditions are susceptible to various complex loads, inevitably leading to various fault modes in the mechanical system. Such faults, when they occur during operation, can lead to substantial maintenance expenses and even pose significant safety risks. Thus, exploring advanced fault diagnosis techniques is of paramount importance, offering the potential to reduce economic losses and enhance production safety (Yan et al., 2023; Liu et al., 2023; Xu et al., 2022; Peng et al., 2023).

Traditional fault diagnosis process relies on a priori knowledge and expert experience, which come with inherent limitations. Deep learning-based fault diagnosis overcomes this shortcoming by mining valuable information from raw data, thus enabling end-to-end intelligence for signal feature extraction and fault pattern recognition (Liu et al., 2022; Su et al., 2022; Qian et al., 2022; Chen et al., 2023; Cheng et al., 2020). Consequently, deep learning methods such as deep belief networks (Zhong et al., 2021), sparse auto-encoders (Liu et al., 2021), recurrent neural networks (Kang, 2020), and convolutional neural networks (Wu et al., 2023; Liang et al., 2019; Han et al., 2021) are extensively researched and used in fault diagnostics. Jiao et al. (2018) presented a convolutional neural network according to data from a multivariate encoder and used this model for defect diagnostics of planetary gearboxes. Their findings revealed remarkable accuracy, with the test set achieving up to 97.32% accuracy when the training dataset contains500 samples for each class. Shao et al. (2019) merged wavelet transform and convolutional neural networks in order to diagnose inductive motor issues, achieving an impressive accuracy of 98.72% on the test set when each class in the training set had 800 samples. Zhang et al. (2021) used recurrent neural networks for identifying defects in whirling machinery and applied this method to the CWRU dataset, achieving an accuracy of 99.96% when each class in the training set had 800 samples. While intelligent diagnostic technologies based on deep learning have demonstrated promising results, their effectiveness heavily relies on the availability of abundant high-quality training samples. In real-world engineering scenarios, practical challenges emerge, such as high labeling costs and difficulty in collecting fault data. Consequently, these methods tend to overfit when dealing with small training samples, leading to suboptimal performance (Hu et al., 2018). Therefore, the pursuit of fault diagnosis with small sample sizes holds practical significance, as it closely mirrors real-world engineering applications while also mitigating model computation and data collection expenses (Lin et al., 2023; Liang et al., 2023).

To resolve the aforementioned issues, Zhou and Feng (2019) presented a model named deep forest. This model builds upon the foundation of a random forest while incorporating a deep structure akin to deep learning. It presents an effective solution with tunable parameters, capable of adapting to varying

data sizes (Zhu et al., 2018). However, the model still has some challenging issues to be addressed to further improve its diagnostic performance:

1) Multi-grained scanning structures: Existing multi-grained scanning structures rely on the interception of features using sliding windows, resulting in inadequate learning of representations for other uncorrelated spatial features. Moreover, there is an imbalance in the number of feature scans, with the middle position experiencing significantly more scans than the two ends, leading to the loss of feature elements at the extremities.

2) Feature redundancy in multi-grained scanning: When processing high-dimensional data samples, current multi-grained scanning encounters problems of feature redundancy. Furthermore, the feature dimension of multi-grained scanning often surpasses the dimension of enhanced features generated by the cascade forest. This significant dimension difference leads to feature submersion during concatenation.

3) Fixed decision tree classifiers: In the existing cascade forest framework, the number of decision tree classifiers remains fixed at every level. This rigidity poses challenges in adapting to the growing complexity of the model as the number of cascade forest layers increases.

Addressing these challenges is pivotal for advancing the diagnostic capabilities of deep forest models and further enhancing their utility in practical applications. To this end, this article presents a random multi-grained fusion cascade forest model, called the rgfc-Forest model, for small-sample fault diagnosis of electromechanical system. The following encapsulates the primary contributions of this article:

(1) A random multi-grained scanning module is designed to enrich the expression of the initial vibration signals, enhancing the learning of spatially uncorrelated features.

(2) A feature fusion cascade forest module is constructed to address the issues of feature redundancy and feature submergence in the input cascade forest, further improving the quality of feature learning.

(3) A decision tree self-growing strategy is combined bolster the feature discrimination capabilities of higher-level cascade forests.

The structure of the article is organized as below. The related literature on deep learning and deep forest are elaborated in Section 2. The fundamental method is presented in Section 3. Section 4 provides the specifics and overall framework of this proposed method. Section 5 substantiate the effectiveness of our proposed method through an array of experimental comparisons. The concluding remarks and key takeaways of the article are outlined in Section 6.

## 2. Related works

### *2.1 Deep learning*

In real-world industrial scenarios, the obtained vibration data of the machine is complex, and since

the machine usually works in a healthy condition, the measured abnormal data of the machine is very limited. This presents a formidable challenge for conventional deep learning techniques, rendering them ineffective in addressing such issues Consequently, novel small sample training strategies based on deep learning have emerged to tackle these challenges, thereby promoting the application of intelligent diagnostic methods in real-life scenarios.

At present, the small sample training strategy of deep learning includes strategies such as parameter optimization and data expansion (Wang et al., 2023). Kumagai (2016) proposed a method of parameter transfer and deduced the learning boundary of the parameter transfer algorithm. Bertinetto et al. (2016) proposed a method for learning another network parameter with a network. In offline mode, the learner's network is trained through a training set; and in online mode, the learner's network generates the parameters of the pupil network through samples ultimately serving for classification tasks. Ma et al. (2021) applied the parameter transfer approach to ascertain the initial parameters of the DNN, followed by fine-tuning these parameters on an existing dataset, resulting in significant model improvement. Li et al. (2020) proposed an adversarial feature illusion network. The model is based on conditional Wasserstein Generative Adversarial networks, which can generate various and distinguishable features in the case of a few labeled samples. Zhang et al. (2018) proposed a MetaGAN method for the image classification task of feed-shot learning, which generates indistinguishable fake samples from real sample data, thereby providing more training signals for classifiers and thereby improving the accuracy of classification. Although the above-mentioned small sample training strategy methods for deep learning have good classification effects, they do exhibit certain drawbacks. For example, due to the training of numerous hyperparameters, these methods tend to have relatively lengthy training times. Furthermore, their training processes can be intricate; some approaches necessitate pre-trained models as a prerequisite.

### *2.2 Deep forest*

The deep forest algorithm addresses the challenges associated with a high number of hyperparameters and complex training processes. At present, deep forest has been widely used in the fields of tobacco drying condition recognition, image recognition, and power system steady state research and has achieved promising research results, while electromechanical fault research is in the research and exploration stage (Bi et al., 2022; Xia et al., 2021). Li et al. (2021) used the deep forest model for the evaluation of the transient stability of power systems, revealing that the proposed method is simpler in parameter setting and faster in training in comparison to the deep neural networks. Ding et al. (2021) used the deep forest model for fault diagnosis of bearing data and matched it up against the deep neural networks model and random forest model, demonstrating the strong generalization capability.

As the research on the deep forest model advances, scholars have identified certain structural

shortcomings in the original model. Fan et al. (2019) optimized the structure of cascade forest. Every forest's standard deviation for various of its most essential characteristics is integrated into a new feature for subsequent cascade tier. Additionally, to increase the likelihood of training, each level of the cascade structure is expanded to two sublayers. Testing this approach on five datasets from the UCI machine learning library, it exhibited superior classification performance, particularly on small-scale datasets. Li et al. (2022) proposed a deep forest model for predicting the occurrence of rock explosions, which adjusts the hyper-parameters of the whole model through Bayesian optimization, and the results show the capability to predict the occurrence of rock explosions with a high degree of accuracy. Wu et al. (2022) applied the deep forest method to IoT, employing it for the detection of botnet traffic on the Internet of Things. They selected dominant features using the Fisher score and input them into the deep forest model, demonstrating that the deep forest model remains robust even in the face of vast high-dimensional data. Li et al. (2022) combined the methods of wavelet analysis and deep forest for the diagnosis of rolling bearing faults and finally verified that the method has good diagnostic accuracy when dealing with small samples. Su et al. (2023) proposed an improved gc-Forest algorithm for defect detection of flip-chips, which added the Kernel principal component analysis (KPCA) strategy to the cascade forest structure and modified the classifiers in the cascade forest to improve the generalization ability of the model.

The studies discussed above reveal that existing enhancements to the deep forest model have typically focused on either multi-grained scanning or cascade forest structures, but not both simultaneously. Motivated by these approaches, this paper proposes a random multi-grained fusion cascade forest model for small sample fault diagnosis of electromechanical systems.

## 3. Benchmark deep forest

Deep forest is an ensemble learning approach for decision trees that largely consists of two components: multi-grained scanning and cascade forest (Rodriguez-Galiano et al., 2012). It enhances the feature representation capability by transforming the input original features using the multi-grained scanning method, followed by a cascade structure that facilitates layer-by-layer feature learning.

### *3.1 Multi-grained scanning*

The multi-grained scanning framework is shown in **Fig. 1**, where the original features are scanned through a sliding window. If a sample contains $K$- dimensional (dim) features and the sliding window of length is $L$ with a step size of S, then after sliding sampling, we obtain $N$ $L$-dim feature vectors.

$$N = \frac{(K - L)}{S} + 1 \tag{1}$$

Subsequently, each of the sliding window scanning feature vectors is put into both ordinary random forest (ORF) and completely random forest (CRF). As a result, every type of the forest will

produce *N* instances of dimension *P*. Finally, these probability vectors are concatenated to obtain the output of the multi-grained scanning structure.

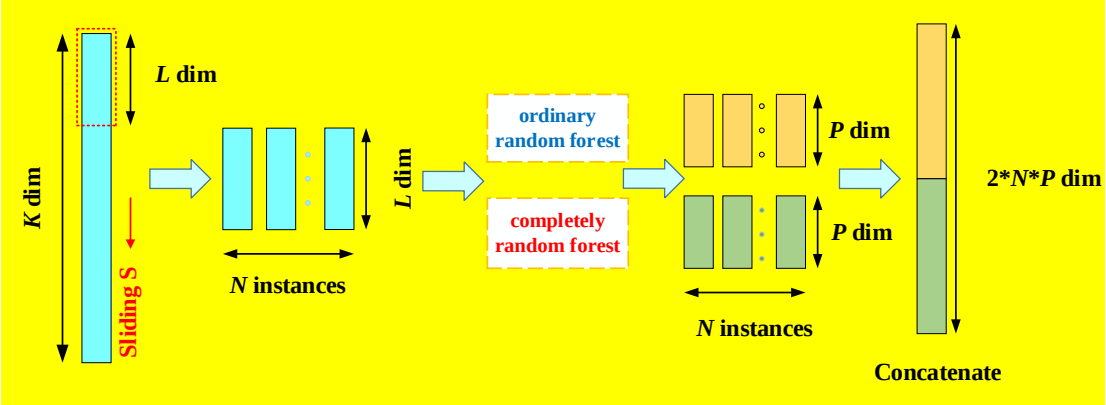



**Fig. 1.** Multi-grained scanning structure diagram.

### *3.2 Cascade forest*

The cascade forest framework structure is illustrated in **Fig. 2**. Every layer of the cascade forest consists of several ORF and CRF (Breiman, 2001). At every node of the decision tree, the CRF randomly picks a feature to divide until each node includes only instances of the same class. On the other hand, the ORF randomly selects $\sqrt{r}$ features as options (where *r* is the number of input features), computes the information gain for every feature, and selects the feature with the highest information gain to grow the tree.

The specific method is as follows: For the *i*-th feature, $x_i$, its information gain is calculated as the difference between the entropy of the feature set $D_{FS}$ and the empirical conditional entropy of $D_{FS}$ given $x_i$, $H(D_{FS}|x_i)$. The expression for the information gain function $Gain(\cdot|\cdot)$ is:

$$Gain(D_{FS}|x_i) = H(D_{FS}) - H(D_{FS}|x_i) \tag{2}$$

with

$$H(D_{FS}|x_i) = \sum_{a \in P_i(D_{FS})} \frac{|D_{FSx_i=a}|}{|D_{FS}|} H(D_{FSx_i=a}) \tag{3}$$

$$H(D_{FS}) = -\sum_{c \in C} \frac{|D_{FS_y=c}|}{|D_{FS}|} \log_2 \frac{|D_{FS_{y=c}}|}{|D_{FS}|} \tag{4}$$

$$D_{FSx_i=a} = \{(x,y) \in D_{FS} | x_i = a\} \tag{5}$$

$$D_{FS_y=c} = \{(x,y) \in D_{FS} | y = c\} \tag{6}$$

where $P_i(D_{FS})$ is the proportion of the feature $x_i$ in the total features of the dataset $D_{FS}$ and *c* represents the possible class labels of the element $x_i$.

Each forest has a number of decision trees, each of which produces a class vector as a result. The mean of the output class vector is taken as the final output *Q* of the forest.

$$Q = \{Q_1, Q_2, \cdots Q_P\} \tag{7}$$

$$Q_j = \frac{1}{t}\sum_{l=1}^{t} Q_{jl},\ j = 1, 2, \cdots, P; l = 1, 2, \cdots, t \tag{8}$$

where $P$ stands for a $P$-categorization problem and $t$ stands for the number of decision trees in the forest. At the final level of the cascade, all the forest decisions are averaged at the same level, and the category corresponding to the maximum value is taken as the prediction result for the sample at that level of the cascade.

In **Fig. 2**, the output of the multi-grained scanning is used as the input for the cascade forest but is also concatenated with the output of each layer in the cascade forest to create enhanced features for the next level input. Additionally, it is possible to independently determine how many levels there are in the cascade forest, allowing the deep forest model to adaptively determine the model's depth based on the size of the dataset, reducing the manual effort required for parameter tuning.

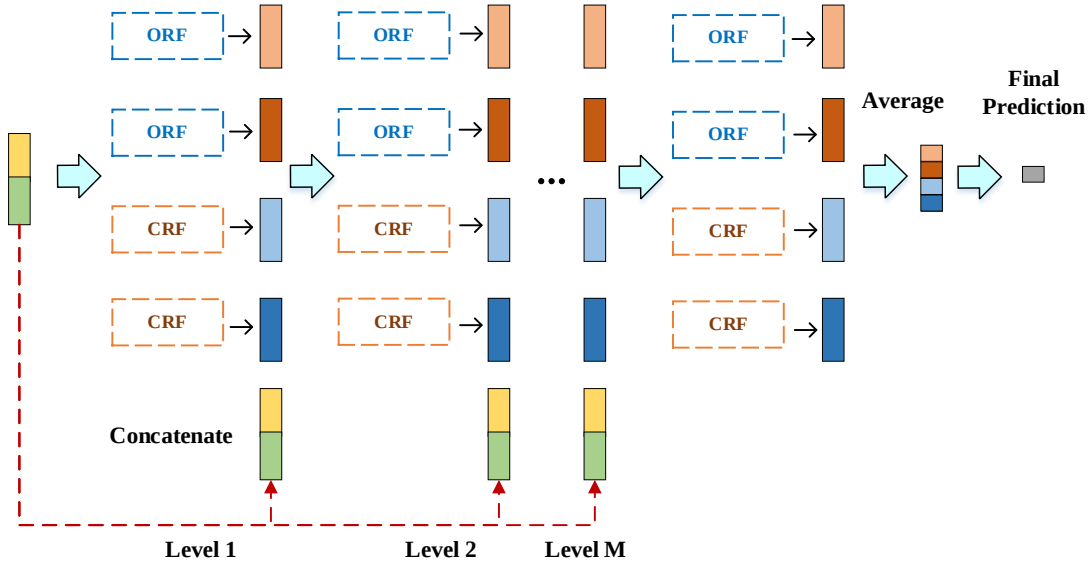



**Fig. 2.** Cascade forest structure diagram. (In the figure ORF stands for ordinary random forest and CRF stands for completely random forest.)

## 4. Proposed method

### *4.1 Overall framework*

The proposed method comprises two main components: the design of a random multi-grained scanning module and the construction of a feature fusion cascade forest module, complemented by a decision tree self-growing strategy. In addition to these two parts, **Fig. 3** shows the overall fault diagnosis process and describes the five main steps in detail.

**Step 1: Data collecting.** The one-dimensional vibration signal of the electromechanical system is firstly acquired, and then the data set is normalized to obtain the original feature data.

**Step 2: Random multi-grained scanning.** The original feature dataset is split up into two portions, the training set and the test set, with the training set be input into the random multi-grained scanning

module. The original training fault features are scanned through the multi-dimensional window and random feature extraction, and then input into the random forest for feature transformation, and finally the obtained features are stitched together to obtain the high-dimensional transformed features.

**Step 3: PCA feature extraction.** The high-dimensional transformed features output by the random multi-grained scanning module are normalized, the optimal number of PCA principal components is selected, then the optimal principal components are selected according to the singular value decomposition (SVD), and finally the feature space is transformed to obtain the reduced-dimensional transformed features before being input into the cascade forest.

**Step 4: Feature fusion cascade forest.** The PCA-extracted features from step 3 are first fed directly into the first layer of the cascade forest for training, and then the enhanced features from each layer are spliced with the PCA-extracted transformed features from step 3 and transported into the subsequent level of the cascade forest for training, and so on.

**Step 5: Fault diagnosis performance testing.** Every level of the cascade forest is tested for diagnostic performance using Z-fold cross-validation, and if the performance does not converge then return to step 3 to start a new level of cascade forest growth until the diagnostic performance converges and the cascade forest stops growing. Then the test dataset is transported into the trained model to make classification predictions and output diagnostic results.

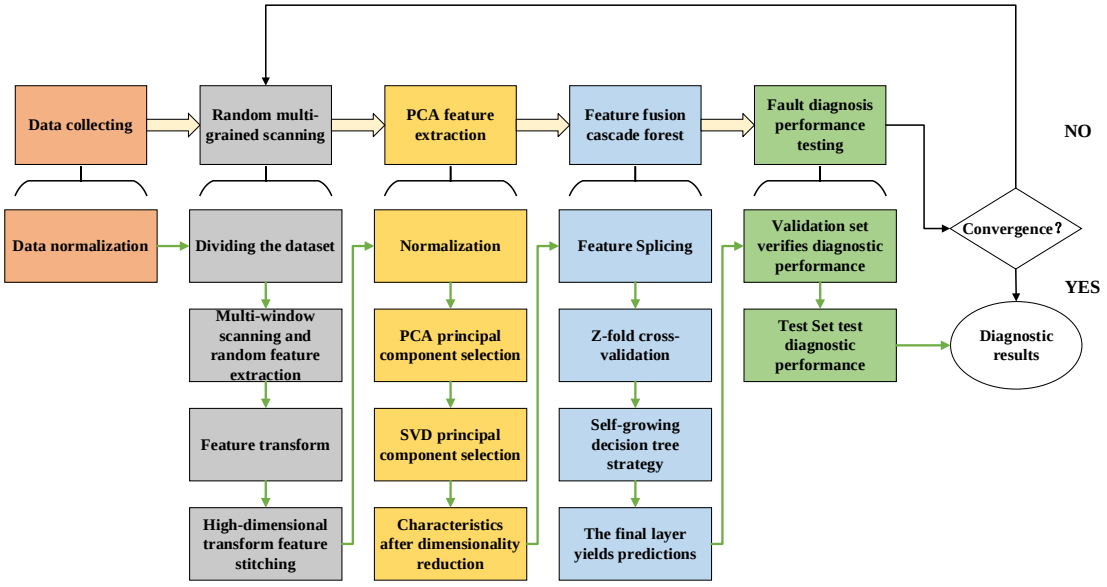



**Fig. 3.** Overall diagnostic framework diagram.

### *4.2 Design of a random multi-grained scanning module*

Due to the insufficient representation learning capability of the benchmark deep forest's multi-grained scanning structure for spatially unrelated features, a random multi-grained scanning module is designed. In addition to preserving various sliding window extracted features, this module introduces a random feature extraction component. Specifically, it randomly selects several features from each feature category to compose feature fragments, enhancing the representation capability of

the original features. After extracting features using both sliding window and random feature extraction, the resulting feature fragments are fed into ORF and CRF for training. Every feature fragment generates a corresponding classification probability vector. These probability vectors are then concatenated to obtain transformed features. The random multi-grained scanning module's final output is created by concatenating the modified features from each feature fragment. In **Fig. 4**, the overall structure is depicted.

To increase the diversity of features obtained by the model, multiple different sliding windows can be used for multi-scale feature sampling, resulting in feature segments of varying scales. Assuming the original input features are $d$-dim, we can utilize sliding windows of sizes $d/4$, $d/8$, and $d/16$ for scanning. Under a stride of 1, each sliding window will generate $(3d/4+1)$, $(7d/8+1)$, and $(15d/16+1)$ features, respectively. Then, we randomly extract $a$-dim features from the $d$-dim features $b$ times ($a$ and $b$ are the new parameters.) These feature vectors, generated through window scanning and random extraction, are separately input into a ORF and a CRF to create transformed features. If it is an $n$ classification problem and the above scanning window is used, the $d/4$ window input forest will produce $(3d/4+1)$ $n$-dim features, and finally the features produced by the two forests will be concatenated together to produce $2n(3d/4+1)$-dim transformed features corresponding to the original $d$-dim features. Similarly, the $d/8$ window generates $2n(7d/8+1)$-dim features, the $d/16$ window generates $2n(15d/16+1)$-dim transformed features and the randomly selected window generates $2bn$-dim transformed features, and finally the four sets of transformed features are concatenated together to obtain the output of the random multi-grained scanning module.

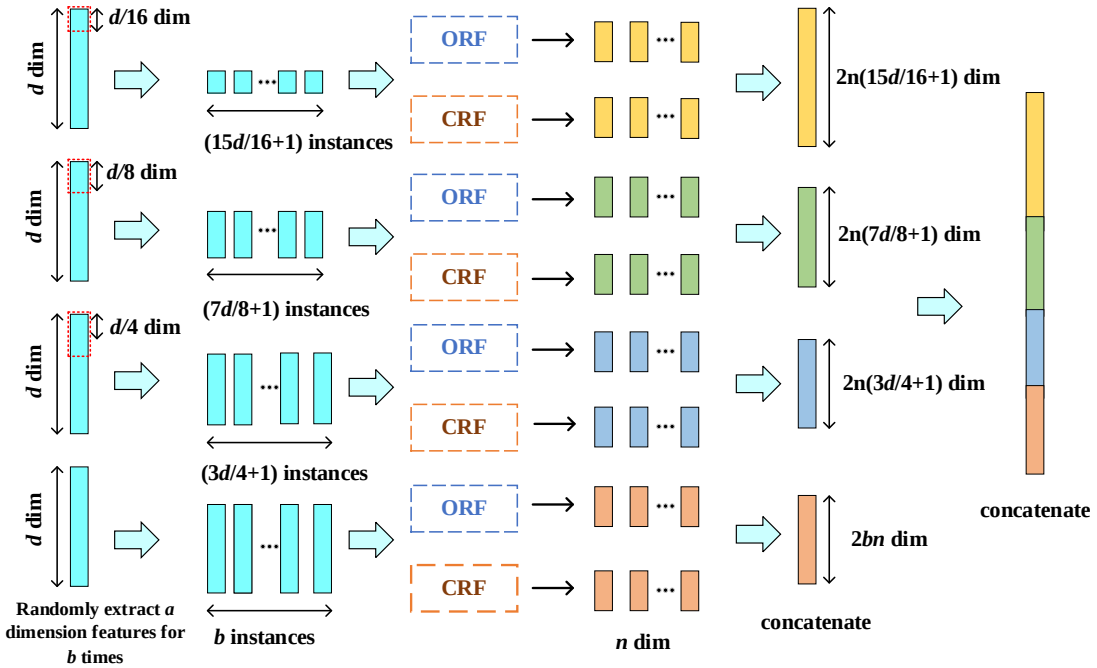



**Fig. 4.** Random multi-grained scanning module.

### *4.3 Feature fusion cascade forest module construction and decision tree self-growing strategy*

***combination***

In the benchmark deep forest model, the input of the cascade forest structure is the high-dimensional features output by the original features through the multi-grained scanning structure. After one layer of training, the generated enhanced features from the cascade forest are directly concatenated with the high-dimensional transformed features. This may lead to feature redundancy, decreasing the effectiveness of the method as a result. Moreover, since the dimension of the original transformed features is much larger than that of the enhanced features generated by the cascade forest, it can, to some extent, overshadow the enhanced features, resulting in insufficient learning of the new enhanced features and hence lowering the accuracy of the model.

To address the aforementioned issues, a feature fusion cascade forest module is constructed. A principal component analysis (PCA) (Ebied, 2012) module is added between the multi-grained scanning structure and the cascade forest structure. By extracting the high-dimensional features from the multi-grained scanning structure with principal component analysis and then inputting them into the cascade forest structure, the quantity of faulty features might be decreased, thus solving the problem of feature redundancy caused by long characteristic single-sample data such as vibration data, and the low efficiency of the algorithm operation. The variation features after dimensionality reduction are then spliced with the enhanced features of the cascade forest to increase the validity of this method by improving the masking of the enhanced features by the original features to some extent.

As the complexity of the deep forest increases with the number of cascade layers, a decision tree self-growing strategy is combined. Deep forest is an ensemble learning modelling built on decision trees, whose accuracy is influenced by the decision trees comprised. Since the number of trees in a decision tree is a crucial parameter in the model, its size affects the diagnostic results. As the quantity of cascade layers rises the complexity of this entire model also grows. Therefore, different levels of cascade forest require different demands for the complexity of decision trees: (1) At the first layer of the cascade forest, a smaller number of decision trees can be set, which not only allows for better learning of features, but also increases the training rate of this method. (2) As the quantity of layers rises the number of decision trees per layer can be increased to achieve better learning of higher level features.

The framework diagram of the feature fusion cascade forest module and the decision tree self-growing strategy is shown in **Fig. 5**. After the original features undergo the random multi-grained scanning module, high-dimensional transformed features are generated. Given such high-dimensional features, the following steps are required for processing.

(1) The high-dimensional features are first extracted by PCA features and then fed into the first layer of the cascade forest for training. Assuming two ORF and two CRF in every layer, the first layer will output 4$n$-dimensional enhanced features.

(2) The enhanced features of 4$n$ dimensions are concatenated with the transformed features

obtained after PCA feature extraction. They will then feed into the cascade forest's second level. This process is repeated for the subsequent layers of the cascade forest. As the quantity of layers rises the amount of decision trees in every layer will also increase.

(3) The general efficacy to the complete cascade forest for each additional cascade layer will be calculated on each validation set of Z-fold cross-validation. Subsequently, the average diagnostic accuracy of the entire cascade forest on all validation sets of Z-fold cross-validation will be taken as the overall diagnostic accuracy. If the overall diagnostic accuracy does not show improvement compared to the previous level, the validation performance converges, and the training process will terminate.

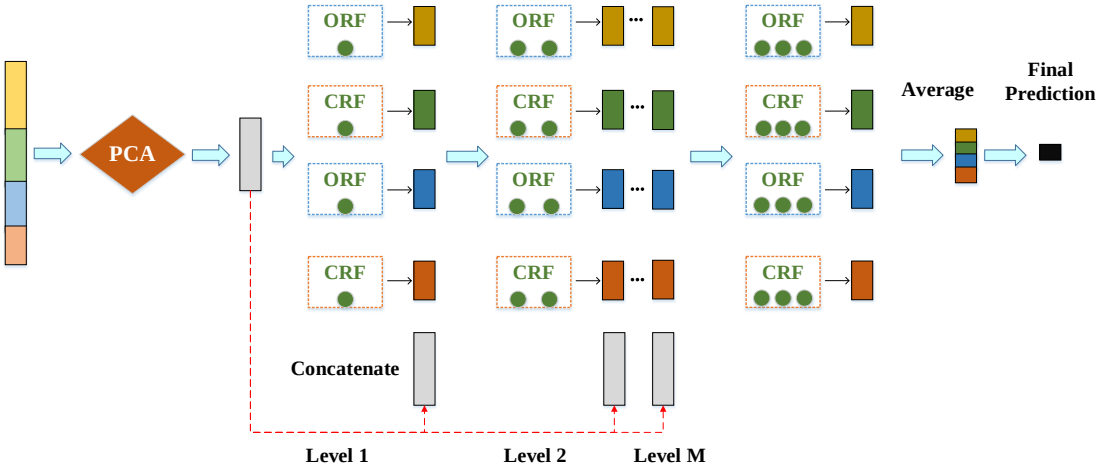



**Fig. 5.** Feature fusion cascade forest module with decision tree self-growing strategy framework diagram. (Green balls represent decision trees.)

## 5. Case study

### *5.1 Introduction of the data set*

The superiority and effectiveness of the method for fault diagnosis was verified on two data sets.

**Case 1:**

The data set is collected on an asynchronous motor test bench (Liang et al., 2021). The bench consists of a test asynchronous motor and vibration sensor; a vibration signal monitor; and a data acquisition device, as shown in **Fig. 6**. The model of the motor used is YE2-100L2-4, the model of the vibration signal monitor is PCH1028, the data acquisition equipment is IPC-610L, the sampling frequency of 250K/s PCI-1716 data acquisition card is selected, the sensor is selected CT1020L, the voltage sensitivity is 200mv/g, there are three sensors, respectively test the motor three directions of vibration. Damage to the motor is man-made and there are five common types of faults, as shown in **Fig. 7**. The signal sampling frequency was 10KHz and the experiment was collected for 10s for each fault type, with data collected for a total of eight health states. Each sample in this experiment had a length of 1024, for a total of 8000 samples and the dataset is shown in **Table 1**.

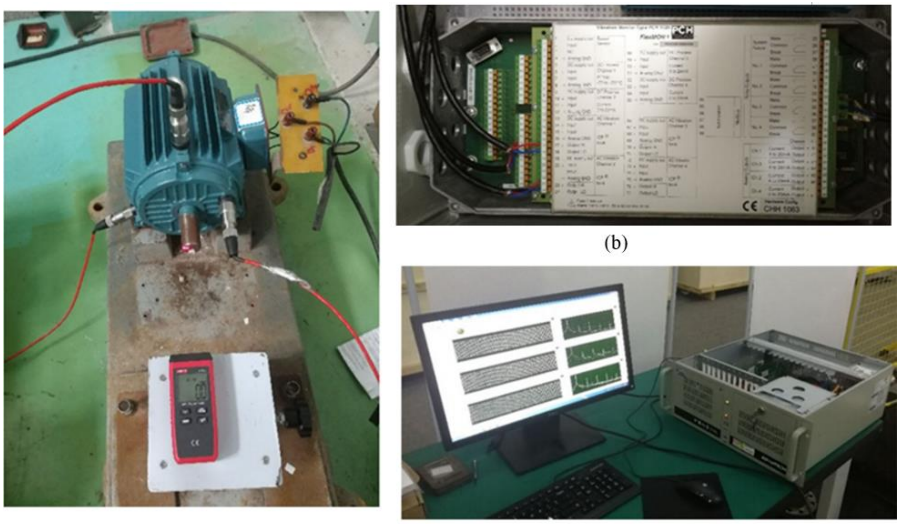



**Fig. 6.** Asynchronous motor test bench (a) Test asynchronous motors and horizontal, vertical and axial vibration sensors; (b) Vibration sensor; (c) Advantech data acquisition equipment.

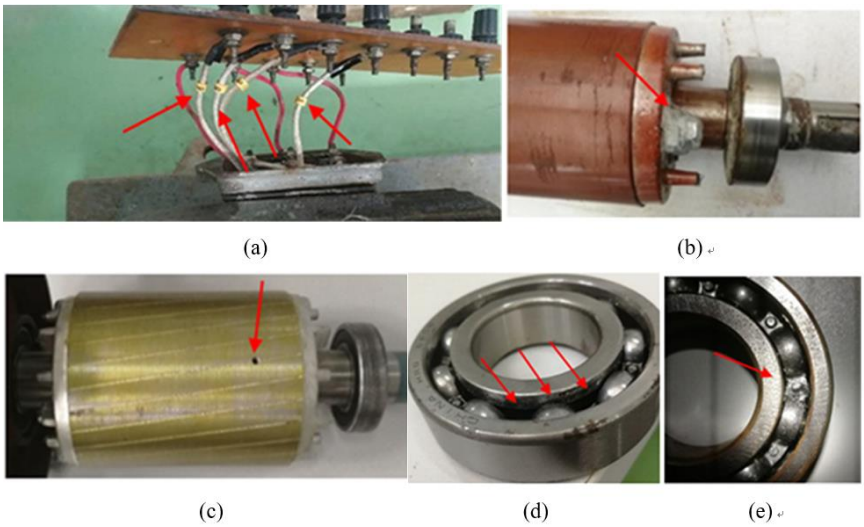



**Fig. 7.** Common forms of fault in asynchronous motors. (a) Short circuit; (b) air-gap eccentricity; (c) Rotor bar broken; (d) Bearing abrasion fault; (e) Bearing cage broken.

**Table 1**

Introduction of the motor data set in Case 1.

| Labels of conditions | Health conditions of motor system |
|---|---|
| 0 | Normal |
| 1 | 2 turns short circuit |
| 2 | 4 turns short circuit |
| 3 | 8 turns short circuit |
| 4 | Air gap eccentricity |
| 5 | Rotor rod fracture |
| 6 | Bearing cage fracture |
| 7 | Bearing wear |

**Case 2:**

The data set was obtained from a drivetrain diagnostics simulator at Southeast University (Shao et al., 2018), as shown in **Fig. 8**. The speed and load of this simulator were configured at 20Hz-0V, and there were 8 sensor channels measuring the vibration and torque of the motor, planetary gearbox and parallel gearbox in x, y and z directions respectively, with a sampling frequency of 5KHz. 4 fault states for each of the gears and bearings in the experiment are shown in **Fig. 9**, plus two normal states for a total of 10 fault states. This experiment gathered a total of 10,230 samples, each of which was 1024 in length, and the data set is introduced as shown in **Table 2** below.

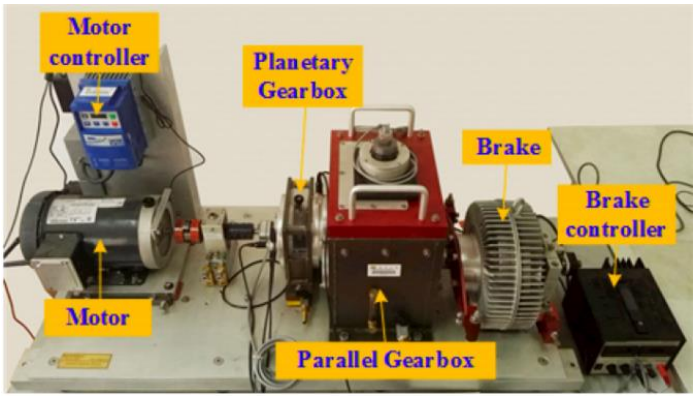



**Fig. 8.** Southeastern University gearbox experimental bench.


**Fig. 9.** Common types of gearbox faults.

**Table 2**

Introduction of the gearbox data set in Case 2.

| Labels of conditions | Health conditions of gearbox system |
|---|---|
| 0 | Health Gear |
| 1 | Chipped Tooth |
| 2 | Missing Tooth |
| 3 | Root Fault |
| 4 | Surface Fault |
| 5 | Health Bearing |
| 6 | Inner ring |
| 7 | Outer ring |
| 8 | Combo |
| 9 | Ball |

### *5.2 Comparison with the improved deep forest model*

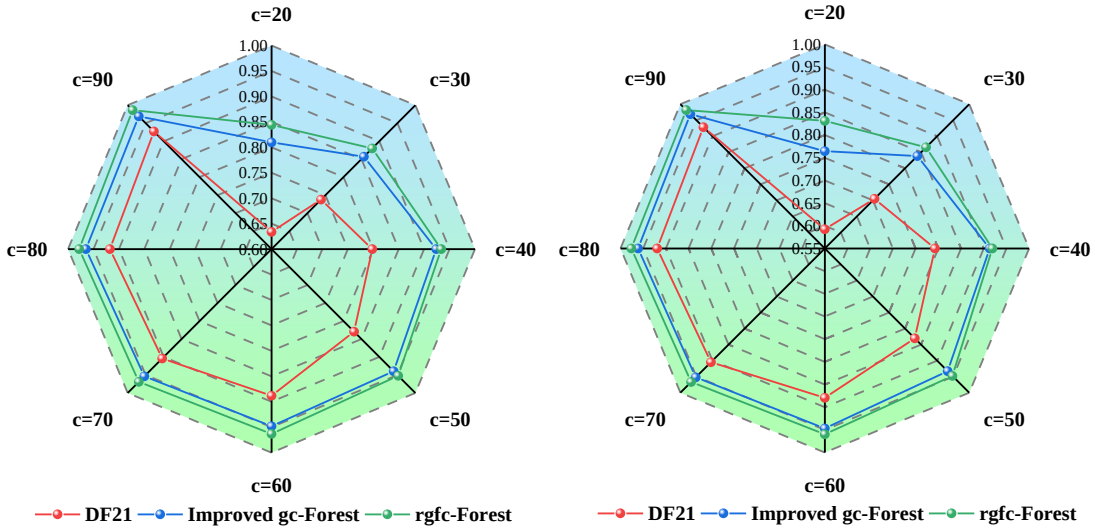
The rgfc-Forest model was contrasted with the Deep Forest 21 (DF21) model and the improved gc-Forest (Su et al., 2023) for trials in order to demonstrate its superiority. The length of every fault sample is configured at 1024, so the selected multi-grained scanning window sizes are 64-dimensional, 128-dimensional and 256-dimensional respectively. The number of random forests for the multi-grained scanning structure was set to 2 and the decision tree for each random forest was set to 1000. There are 4 random forests in the cascade forest structure, and each of them has 600 decision trees. The structural parameters of the benchmark and modified deep forest in this experiment are set to the above values.

In Case 1, the experiments were conducted with the number of training samples of 20, 30, 40, 50, 60, 70, 80 and 90 in each class, and the test set was 200 samples in every class, for a total of five rounds of experiments, and the average diagnostic accuracy and F1 scores for these two models were obtained respectively as shown in **Table 3** below, and the radar plot for the third experiment can be seen in **Fig. 10**.

**Table 3**
Diagnostic results for each method in Case 1.

| Quantity per category | Metric | **rgfc-Forest** | DF21 | Improved gc-Forest |
|---|---|---|---|---|
| 20 | Accuracy | **84.41** | 65.96 | 80.97 |
| | F-score | **82.16** | 61.13 | 76.44 |
| 30 | Accuracy | **87.99** | 75.13 | 85.69 |
| | F-score | **86.55** | 71.90 | 83.83 |
| 40 | Accuracy | **93.37** | 80.16 | 92.37 |
| | F-score | **91.90** | 79.27 | 91.31 |
| 50 | Accuracy | **95.2** | 83.93 | 93.94 |
| | F-score | **94.74** | 83.02 | 93.26 |
| 60 | Accuracy | **96.32** | 88.48 | 94.81 |
| | F-score | **95.97** | 87.92 | 94.69 |
| 70 | Accuracy | **96.63** | 90.41 | 95.37 |
| | F-score | **96.65** | 90.48 | 95.18 |
| 80 | Accuracy | **97.55** | 91.76 | 96.50 |
| | F-score | **97.62** | 91.94 | 96.12 |
| 90 | Accuracy | **98.34** | 92.69 | 96.88 |
| | F-score | **98.15** | 92.83 | 96.89 |

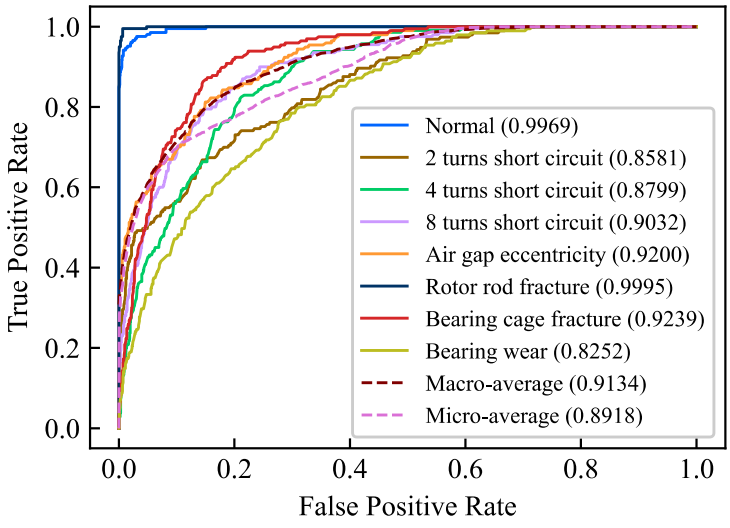
From the experimental results in **Fig. 10** it can be seen that the diagnostic accuracy and F1 score of the rgfc-Forest model are comprehensively better than those of the two improved deep forest models. When there are 20 training samples in each class, the gap between the diagnostic accuracy of the proposed model and the DF21 model is the largest. The average diagnostic accuracy of the DF21 model is only 65.96%, while the average diagnostic accuracy of the rgfc-Forest model can reach 84.41%, which is also larger than that of the other improved deep forest method, thus proving that the rgfc-Forest model has a better diagnostic accuracy even when the number of samples is very small. With the increase in the number of training samples in each class, the diagnostic accuracy of all three models is improving, and the gap between the diagnostic accuracies of the three models is gradually decreasing, especially the gap between the proposed model and another improved deep forest model, which is very small. When the number of samples per class is 90, the rgfc-Forest model still has a performance improvement of 1.46% compared to the other improved deep forest model, thus indicating that the proposed model has better test performance.



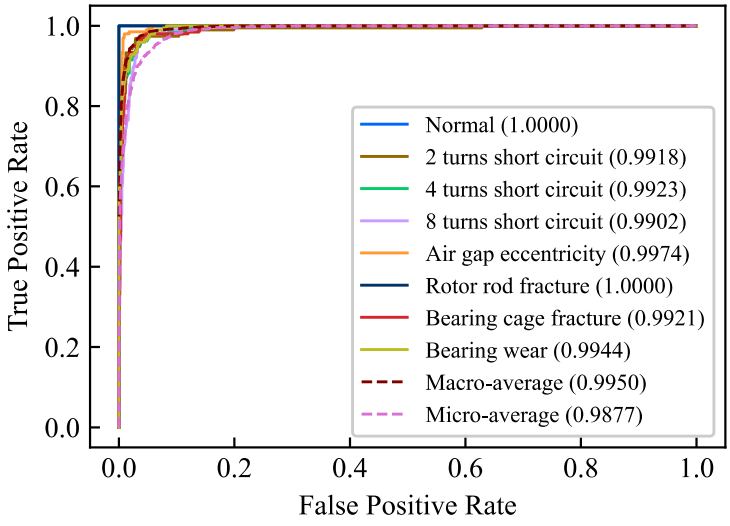

(a) (b)

**Fig. 10.** Diagnostic radar plot for each method with sample size c in each category in Case 1 (a) Accuracy; (b) F1.

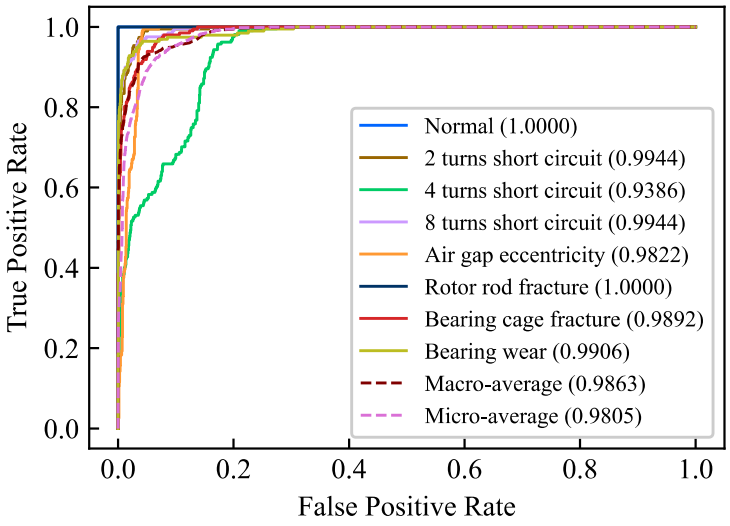
To further compare the diagnostic effectiveness of these three methods for different states with small datasets, the receiver operating characteristic (ROC) curves for the three models with 20 training samples in each class were obtained as shown in **Fig. 11** below. Analysis of the results shows that the proposed method has the highest area under curve (AUC) value, with the proposed method having an AUC value higher than 0.98. Comparing the AUC values of the proposed model with the DF21 model shows that the AUC values are close to each other in the detection of normal category and rotor rod fracture, but there is a large difference in the remaining six fault states. The AUC values of the proposed method and another improved deep forest model are almost the same for several fault states, except for a large gap in the 4-turn short circuit.




(a)




(b)




(c)

**Fig. 11.** ROC curves for three experimental methods in Case 1 (a) DF21; (b) rgfc-Forest; (c) Improved

gc-Forest.

### *5.3 Ablation experiments*

Ablation tests are used to verify the efficiency of the rgfc-Forest model's upgraded module in this section.

In Case 2, the experiments were conducted on four models, namely, the benchmark deep forest model, the benchmark deep forest with random grained module, the benchmark deep forest with PCA module, and the rgfc-Forest. The training set has 20, 30, 40, 50, 60, and 70 samples per class, whereas the test set contains 160 samples each class. The average accuracy, precision and recall for each class obtained from five rounds of experiments for each class are shown in **Table 4** below, and the accuracy plots obtained for each class are shown in **Fig. 12**. A box plot with 20 training samples in each category and 60 training samples in each category can be seen in **Fig. 13**.

**Table 4**

Diagnostic results by method in Case 2.

| Quantity per category | Metric | **rgfc-Forest** | gc-Forest | gc-Forest + PCA | gc-Forest + Random Grained |
|---|---|---|---|---|---|
| 20 | Accuracy | **92.72** | 83.13 | 88.11 | 90.27 |
| | Precision | **93.33** | 78.96 | 91.40 | 92.26 |
| | Recall | **92.52** | 82.91 | 88.99 | 90.08 |
| 30 | Accuracy | **95.11** | 86.27 | 91.40 | 92.32 |
| | Precision | **95.16** | 89.52 | 91.93 | 93.28 |
| | Recall | **94.94** | 86.08 | 91.43 | 92.11 |
| 40 | Accuracy | **96.02** | 91.63 | 92.56 | 93.26 |
| | Precision | **96.13** | 92.67 | 92.32 | 93.55 |
| | Recall | **95.93** | 91.48 | 92.21 | 93.05 |
| 50 | Accuracy | **96.31** | 93.22 | 94.25 | 93.71 |
| | Precision | **96.52** | 93.29 | 93.81 | 93.74 |
| | Recall | **96.21** | 92.97 | 93.75 | 93.4 |
| 60 | Accuracy | **96.71** | 93.57 | 95.13 | 94.16 |
| | Precision | **96.67** | 93.34 | 94.87 | 94.20 |
| | Recall | **96.61** | 92.93 | 94.62 | 93.95 |
| 70 | Accuracy | **97.13** | 93.76 | 95.54 | 94.38 |
| | Precision | **97.13** | 93.20 | 95.19 | 94.29 |
| | Recall | **97.07** | 93.06 | 95.14 | 94.14 |

The following results can be obtained from **Table 4**:

(1) Comparing the benchmark deep forest with the benchmark deep forest with the addition of random grained module, when there are 20 training samples for every class, the average test accuracy for the former is only 83.13%, whereas the average test accuracy of the latter might be as high as 90.27%. The addition of the random grained module to the multi-grained scanning has a positive impact when there are few training examples available, but this module does not significantly improve the model accuracy when the quantity of training samples is larger than 50, as can be seen from **Fig. 12**. As the quantity of samples per class is greater than 50, the difference in diagnostic accuracy between the two types of models becomes less obvious. This may be due to the fact that the module enhances the learning of features when the training samples are small and causes feature redundancy when the training samples are large.

(2) Comparing the benchmark deep forest with the benchmark deep forest with PCA module, the average test accuracy of adding the PCA module at 20 samples per class is 88.11%, which is lower than the accuracy of adding the random grained module. The average test accuracy of the former model was higher than that of the latter when an amount of specimens per class was greater than 50, demonstrating that the addition of the PCA module has a positive impact on the improvement of model accuracy when the number of training data per class is relatively high, which may be caused by the fact that the addition of PCA module can solve the problem of feature redundancy when there are a huge amount of samples.

(3) Finally, comparing the rgfc-Forest model with all other models, it is found that the diagnostic performance of this model is the highest. When there are 20 training samples each class, the average test accuracy is 92.72%, and when there are 70 training samples per class, it is 97.13%, indicating that the method has a great improvement on the original model when the quantity of samples is small and large, indicating that these two added modules are effective and necessary. According to **Fig. 13**, it might be found that the fluctuation of the presented method's test metrics is low, which proves the stability of the presented method's diagnostic performance.

To better show the feature extraction performance of all the methods, the test set results of the above four classes of models at a training sample size of 20 were visualized in two dimensions by the t-distributed neighborhood embedding algorithm (Van et al., 2008), and the results are shown in **Fig. 14**. It can be clearly observed that the features learned by the proposed method are the most representative, with each class basically having clearer classification boundaries, while some of the classes of the other three models show more class mixing.

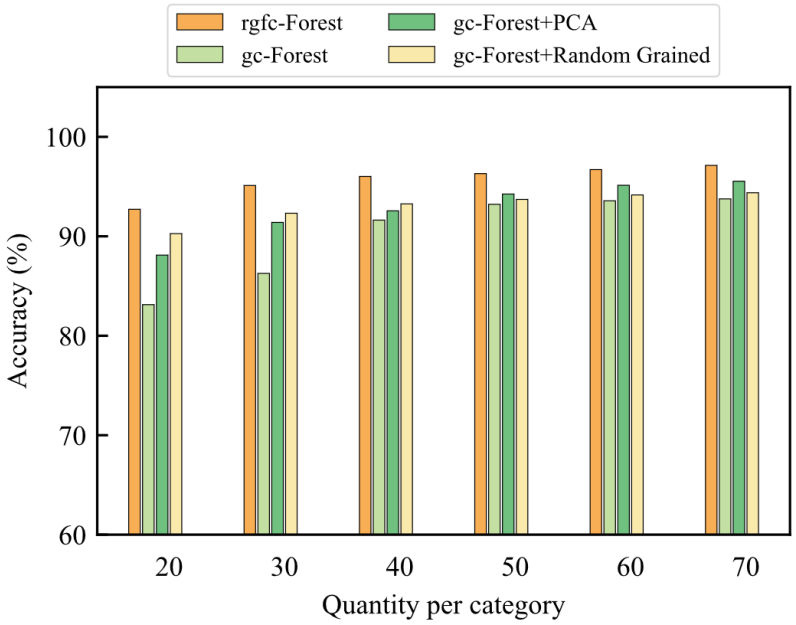



**Fig. 12.** Accuracy graphs for each comparison model for each type of experiment in Case 2.

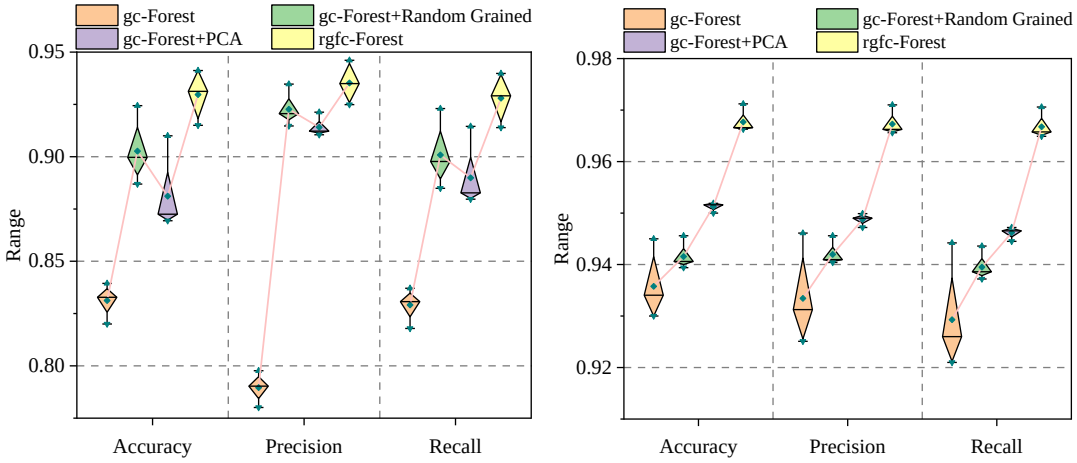
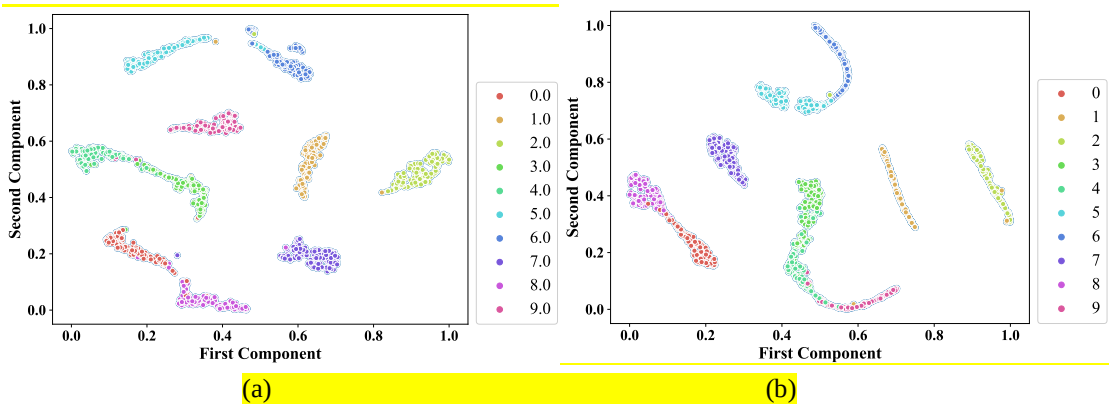
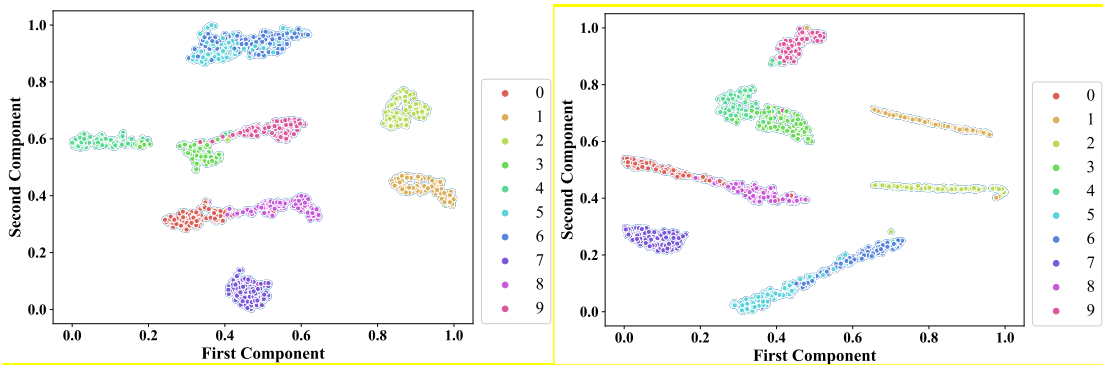



**Fig. 13.** Comparative model box plots in Case 2 (a) Box plots for 20 training samples per class; (b) Box plots for 60 training samples per class.






(c) (d)

**Fig. 14.** 2-dimensional visualization of test set features at 20 training samples per class in Case 2 (a) rgfc-Forest; (b) gc-Forest + Random Grained; (c) gc-Forest + PCA; (d) gc-Forest.

### *5.4 Comparison with deep learning methods*

To verify the superiority of this method under small training samples, comparative studies are conducted with the current mainstream deep learning fault diagnosis methods, including 1DCNN (Zhao et al., 2020), SAE (Zhao et al., 2020), AlexNet (Zhao et al., 2020), ResNet18 (Zhao et al., 2020), Convformer (Han et al., 2022), 1DCNN + parametric transfer learning (1DCNN + TL) and AlxeNet +parametric transfer learning (AlexNet + TL). Among them, the depth of the 1DCNN model is 5 layers, the optimizer for all models is Adam, the initial learning rate is 0.001, the epoch is 60 and the batch size is 40. In the model with the addition of a migration learning strategy, the first round is pre-trained for 20 rounds in noisy data with a signal-to-noise ratio of 8, and then the predictions are obtained by training 40 rounds in the original data. In Case 2, the trials on the aforementioned models were run using 160 test samples each category and 15, 20, 25, 30, 35, and 40 training samples each category. These average accuracy line graphs obtained from five rounds of experiments for each class are shown in **Fig. 15**.

From the **Fig.15**, it might be observed that the diagnostic performance of this proposed method is better than other deep learning models, and the diagnostic performance of the SAE model is worse. When the amount of training samples in each class is less than 25, the proposed method has a large gap in diagnostic accuracy compared with other original deep learning models, which is in line with the theoretical characteristics of deep neural networks that rely on a large amount of training data. The method of the 1DCNN model plus the parameter migration learning strategy has the smallest gap with the proposed method at this time, which, side by side, reflects the effectiveness of the parameter migration strategy for small samples. However, as the quantity of samples per class rises, the gap between the diagnostic accuracy of the proposed method and other deep learning models reduces, especially when the amount of training samples per class is 40, which is almost the same as the average diagnostic accuracy of the ResNet18 model. The proposed method has better model stability and better diagnostic accuracy when the amount of training samples each category is small, which verifies the stability and superiority of the proposed method in diagnosing with small samples. Overall, the

degradation of the diagnostic performance of the presented method is much lower than that of other deep learning models when the amount of training samples is decreased.

To further demonstrate the diagnostic effectiveness of all the methods for different states, the confusion matrix for this experiment with the amount of training samples of 20 for every class is visualized as shown in **Fig. 16**, where 0-9 correspond to the fault states in **Table 2**. It can be found that each model has different advantages and shortcomings for handling different state classes of faults, but the proposed method in this paper compares to other deep learning models for each class of faults. The superiority of the proposed method is verified by the better diagnosis results for each class of faults compared with other deep learning models.

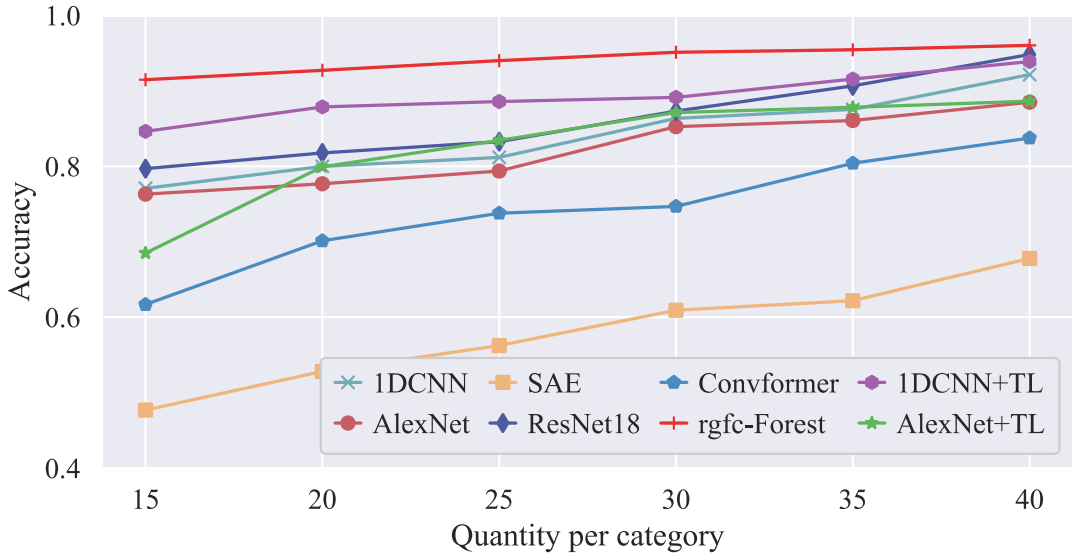
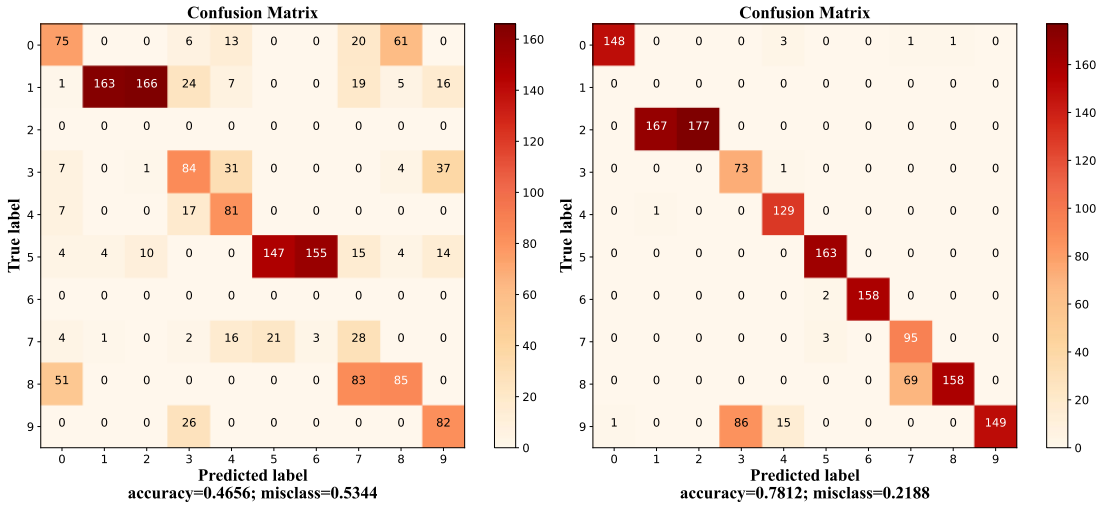
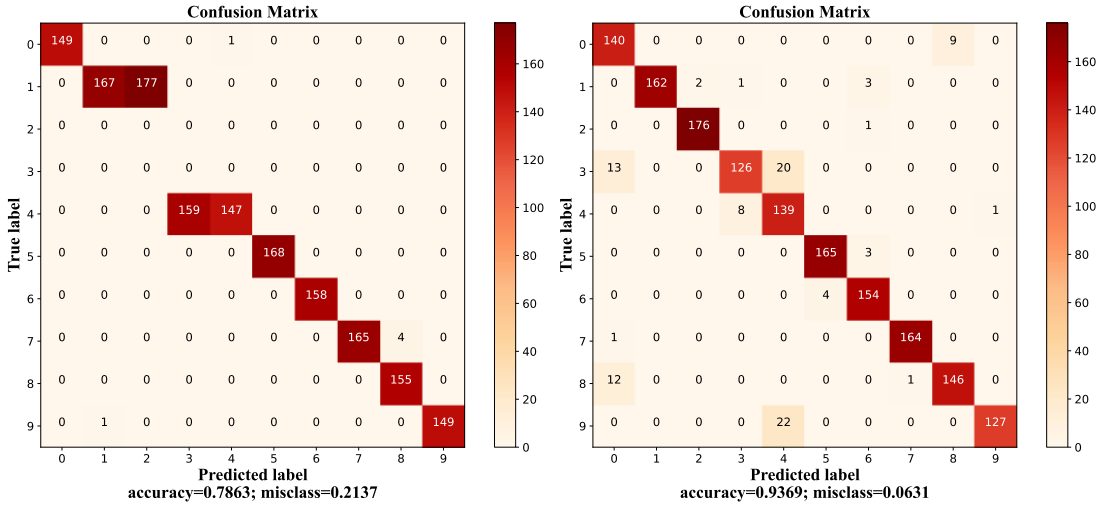



**Fig. 15.** Line graphs for each comparison model in Case 2.



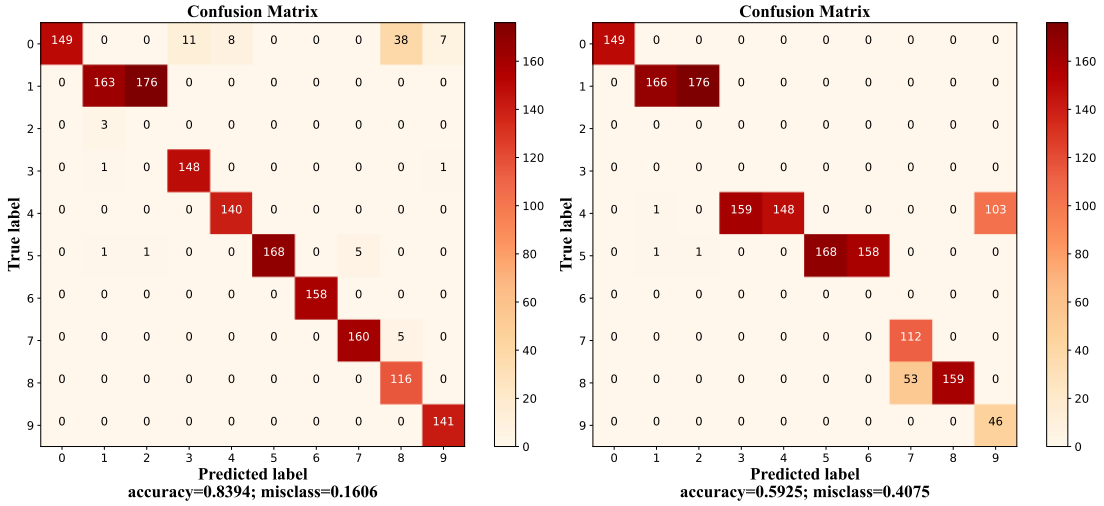

(a) (b)




(c)



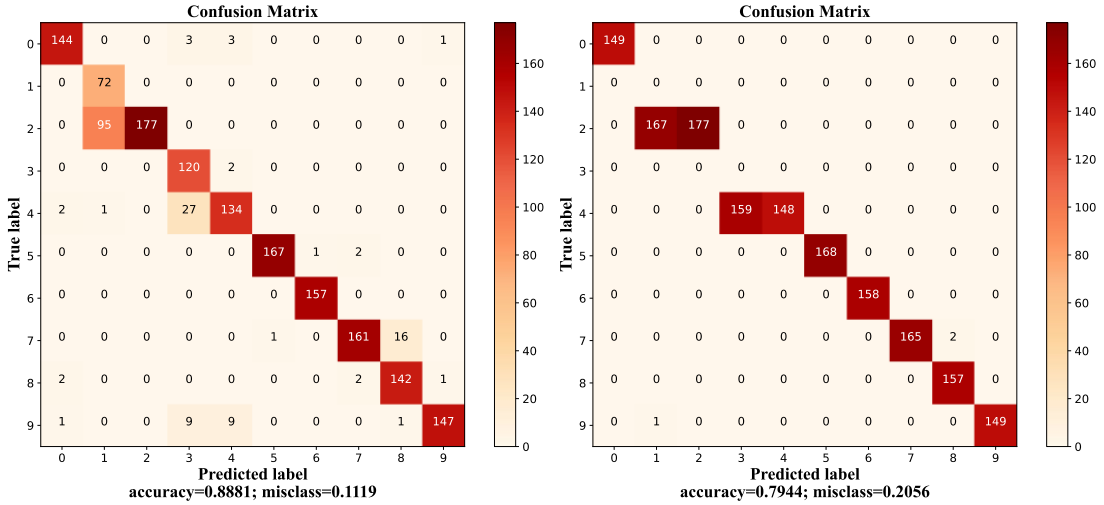

(d)




(e)




(f)




(g)




(h)

**Fig. 16.** Confusion matrix for each experiment at 20 training samples per class in Case 2 (a) SAE; (b) 1DCNN; (c) AlexNet; (d) rgfc-Forest (The proposed method); (e) ResNet18; (f) Convformer; (g) 1DCNN+TL; (h) AlexNet+TL.

## 6. Conclusions

In order to improve the diagnostic performance of the model with small training samples, this paper develops a random multi-grained fusion cascade forest model based on the benchmark deep forest. Specifically, a random grained module is integrated into the benchmark multi-grained scanning to fully exploit the information in the features. A PCA module is added between the multi-grained scanning and the cascade forest to reduce the redundant features in the transfer process between the two. The diagnostic quality of the algorithm is increased in the cascade forest by adjusting the number of decision trees at different stages.

The superiority and effectiveness of the method is verified through case studies from asynchronous motors and gearboxes under small training samples. The diagnostic results show: (1) The performance of the proposed method is superior and more robust compared with the original base method and mainstream deep learning methods under small training samples. (2) The designed random multi-grained scanning module and the constructed feature fusion cascade forest module are able to increase the model's capacity for accurate diagnosis. Although the proposed model has a greater improvement in diagnostic accuracy compared with the base model, when the structure of the deep forest is analyzed, it is found that the selection of base classifiers in the model affects the diagnostic accuracy of the model, and the structure of the cascade forest can be further optimized to improve the efficiency of its operation, which suggests that there is still room for improvement in the model. In future research, the rgfc-Forest model will be further optimized from these two aspects to improve the efficiency of the model for fault diagnosis.

## Declaration of Competing Interest

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

## Acknowledgements

This research is supported by the National Natural Science Foundation of China (No. 52275104), the Science and Technology Innovation Program of Hunan Province (No. 2023RC3097), and the Natural Science Fund for Excellent Young Scholars of Hunan Province (No. 2021JJ20017).

## References

Xu, J., Liang, S., Ding, X., & Yan, R. (2023). A zero-shot fault semantics learning model for compound fault diagnosis. Expert Systems with Applications, 221, 119642.

Wen, X., & Xu, Z. (2021). Wind turbine fault diagnosis based on ReliefF-PCA and DNN. Expert Systems with Applications, 178, 115016.

Li, H., Lv, Y., Yuan, R., Dang, Z., Cai, Z., & An, B. (2022). Fault diagnosis of planetary gears based on intrinsic feature extraction and deep transfer learning. Measurement Science and Technology, 34(1), 014009.

Liao, J. X., Dong, H. C., Sun, Z. Q., Sun, J., Zhang, S., & Fan, F. L. (2023). Attention-embedded quadratic network (qttention) for effective and interpretable bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 72, 1-13.

Yan, S., Shao, H., Min, Z., Peng, J., Cai, B., & Liu, B. (2023). FGDAE: A new machinery anomaly detection method towards complex operating conditions. Reliability Engineering & System Safety, 236, 109319.

Liu, L., Liu, J., Zhou, Q., & Huang, D. (2023). Influence of sample attributes on generalization performance of machine learning models for windage alteration fault diagnosis of the mine ventilation system. Expert Systems with Applications, 213, 119320.

Xu, J., Zhou, L., Zhao, W., Fan, Y., Ding, X., & Yuan, X. (2022). Zero-shot learning for compound fault diagnosis of bearings. Expert Systems with Applications, 190, 116197.

Peng, D., Wang, H., Desmet, W., & Gryllias, K. (2023). RMA-CNN: A residual mixed-domain attention CNN for bearings fault diagnosis and its time-frequency domain interpretability. Journal of Dynamics, Monitoring and Diagnostics, 2(2), 115-132.

Liu, Y., Li, J., Li, Q., & Wang, Q. (2022). Transfer learning with inception ResNet-based model for rolling bearing fault diagnosis. Journal of Advanced Mechanical Design, Systems, and Manufacturing, 16(2), JAMDSM0023-JAMDSM0023.

Su, H., Xiang, L., Hu, A., Xu, Y., & Yang, X. (2022). A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions. Mechanical Systems and Signal Processing, 169, 108765.

Qian, C., Zhu, J., Shen, Y., Jiang, Q., & Zhang, Q. (2022). Deep transfer learning in mechanical intelligent fault diagnosis: application and challenge. Neural Processing Letters, 54(3), 2509-2531.

Chen, X., Shao, H., **ao, Y., Yan, S., Cai, B., & Liu, B. (2023). Collaborative fault diagnosis of rotating machinery via dual adversarial guided unsupervised multi-domain adaptation network. Mechanical Systems and Signal Processing, 198, 110427.

Cheng, C., Zhou, B., Ma, G., Wu, D., & Yuan, Y. (2020). Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data. Neurocomputing, 409, 35-45.

Zhong, T., Qu, J., Fang, X., Li, H., & Wang, Z. (2021). The intermittent fault diagnosis of analog circuits based on EEMD-DBN. Neurocomputing, 436, 74-91.

Liu, H., Yan, G., Duan, Z., & Chen, C. (2021). Intelligent modeling strategies for forecasting air quality

time series: A review. Applied Soft Computing, 102, 106957.

Kang, J. L. (2020). Visualization analysis for fault diagnosis in chemical processes using recurrent neural networks. Journal of the Taiwan Institute of Chemical Engineers, 112, 137-151.

Wu, H., Xin, C., Liu, Y., Rao, R., Li, Z., Zhang, D., ... & Han, S. (2023). Intelligent fault diagnosis for triboelectric nanogenerators via a novel deep learning framework. Expert Systems with Applications, 226, 120244.

Liang, P., Deng, C., Wu, J., Yang, Z., Zhu, J., & Zhang, Z. (2019). Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform. Computers in Industry, 113, 103132.

Han, S., Zhong, X., Shao, H., Zhao, R., & Cheng, J. (2021). Novel multi-scale dilated CNN-LSTM for fault diagnosis of planetary gearbox with unbalanced samples under noisy environment. Measurement Science and Technology, 32(12), 124002.

Jiao, J., Zhao, M., Lin, J., & Zhao, J. (2018). A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes. Knowledge-Based Systems, 160, 237-250.

Shao, S., Yan, R., Lu, Y., Wang, P., & Gao, R. X. (2019). DCNN-based multi-signal induction motor fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 69(6), 2658-2669.

Zhang, Y., Zhou, T., Huang, X., Cao, L., & Zhou, Q. (2021). Fault diagnosis of rotating machinery based on recurrent neural networks. Measurement, 171, 108774.

Hu, G., Li, H., Xia, Y., & Luo, L. (2018). A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis. Computers in Industry, 100, 287-296.

Lin, J., Shao, H., Zhou, X., Cai, B., & Liu, B. (2023). Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals. Expert Systems with Applications, 120696.

Liang, X., Zhang, M., Feng, G., Xu, Y., Zhen, D., & Gu, F. (2023). A Novel Deep Model with Meta-learning for Rolling Bearing Few-shot Fault Diagnosis. Journal of Dynamics, Monitoring and Diagnostics, 1-22.

Zhou, Z. H., & Feng, J. (2019). Deep forest. National science review, 6(1), 74-86.

Zhu, Q., Zhu, Q., Pan, M., Jiang, X., Hu, X., & He, T. (2018, December). The phylogenetic tree based deep forest for metagenomic data classification. In 2018 IEEE international conference on bioinformatics and biomedicine (BIBM) (pp. 279-282). IEEE.

Wang, J., Liu, K., Zhang, Y., Leng, B., & Lu, J. (2023). Recent advances of few-shot learning methods and applications. Science China Technological Sciences, 66(4), 920-944.

Kumagai, W. (2016). Learning bound for parameter transfer learning. Advances in neural information processing systems, 29.

Bertinetto, L., Henriques, J. F., Valmadre, J., Torr, P., & Vedaldi, A. (2016). Learning feed-forward one-shot learners. Advances in neural information processing systems, 29.

Ma, C., Mu, X., Zhao, P., & Yan, X. (2021). Meta-learning based on parameter transfer for few-shot classification of remote sensing scenes. Remote Sensing Letters, 12(6), 531-541.

Li, K., Zhang, Y., Li, K., & Fu, Y. (2020). Adversarial feature hallucination networks for few-shot learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13470-13479).

Zhang, R., Che, T., Ghahramani, Z., Bengio, Y., & Song, Y. (2018). Metagan: An adversarial approach to few-shot learning. Advances in neural information processing systems, 31.

Bi, S., Mu, L., & Liu, X. (2022). Deep multi-sequence multi-grained cascade forest for tobacco drying condition identification. Drying Technology, 40(9), 1832-1844.

Xia, M., Wang, Z., Han, F., & Kang, Y. (2021). Enhanced Multi-Dimensional and Multi-Grained Cascade Forest for Cloud/Snow Recognition Using Multispectral Satellite Remote Sensing Imagery. IEEE Access, 9, 131072-131086.

Li, M., Lei, M., & Zhou, T. (2021). Transient stability assessment method for power system based on deep forest. Electrical Measurement and Instrumentation (In Chinese), 58(02), 53-58.

DING, J, M., Wu, Y, H., LUO, Q, B., & Du, Y. (2021). A fault diagnosis method of mechanical bearing based on the deep forest. Journal of Vibration and Shock (In Chinese), 40(12), 107-113.

Fan, Y., Qi, L., & Tie, Y. (2019, October). The cascade improved model based deep forest for small-scale datasets classification. In 2019 8th international symposium on next generation electronics (ISNE) (pp. 1-3). IEEE.

Li, D., Liu, Z., Armaghani, D. J., **ao, P., & Zhou, J. (2022). Novel ensemble tree solution for rockburst prediction using deep forest. Mathematics, 10(5), 787.

Wu, Y., He, X., & Chen, X. (2022, November). IoT-Botnet Traffic Detection Based on Deep Forest. In 2022 IEEE 22nd International Conference on Communication Technology (ICCT) (pp. 1388-1393). IEEE.

Li, X., Zhang, Y., Wang, F., & Sun, S. (2022). A fault diagnosis method of rolling bearing based on wavelet packet analysis and deep forest. Symmetry, 14(2), 267.

Su, L., Hu, X., Gu, J., Ji, Y., Wang, G., Ming, X., ... & Pecht, M. (2023). Intelligent defect inspection of flip chip based on vibration signals and improved gcForest. Measurement, 214, 112782.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS

journal of photogrammetry and remote sensing, 67, 93-104.

Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

Ebied, H. M. (2012, May). Feature extraction using PCA and Kernel-PCA for face recognition. In 2012 8th International Conference on Informatics and Systems (INFOS) (pp. MM-72). IEEE.

Liang, Y., Li, B., & Jiao, B. (2021). A deep learning method for motor fault diagnosis based on a capsule network with gate-structure dilated convolutions. Neural Computing and Applications, 33, 1401-1418.

Shao, S., McAleer, S., Yan, R., & Baldi, P. (2018). Highly accurate machine fault diagnosis using deep transfer learning. IEEE Transactions on Industrial Informatics, 15(4), 2446-2455.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen, X. (2020). Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. ISA transactions, 107, 224-255.

Han, S., Shao, H., Cheng, J., Yang, X., & Cai, B. (2022). Convformer-NSE: A novel end-to-end gearbox fault diagnosis framework under heavy noise using joint global and local information. IEEE/ASME Transactions on Mechatronics, 28(1), 340-349.