ISA Transactions 136 (2023) 442–454

Contents lists available at ScienceDirect

ISA Transactions

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

Practice article

Intelligent fault monitoring and diagnosis of tunnel fans using a hierarchical cascade forest☆

Zhi-Xin Yang [a], Chao-Shun Li [b], Xian-Bo Wang [c,a], Hao Chen [a,*]

[a] State Key Laboratory of Internet of Things for Smart City, University of Macau, 999078, Macao Special Administrative Region of China
[b] The School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[c] Hainan Institute of Zhejiang University, SanyaChina 572000, China

ARTICLE INFO

Article history:
Received 9 August 2021
Received in revised form 10 August 2022
Accepted 29 October 2022
Available online 8 November 2022

Keywords:
Tunnel fans
Intelligent fault diagnosis
Deep forest
Random forest
Hierarchical cascade structure
Confidence estimation

ABSTRACT

Tunnel fan is critical fire-fighting equipment, and its safe and stable operation is very important for the efficiency and safety of tunnel traffic. Existing studies commonly train the fault diagnosis methods with the goal of minimizing mean error which ignores the difference between classes in feature distribution. To solve the problem of inaccurate prediction caused by mean error evaluation, this paper presents a non-neural deep learning model, namely hierarchical cascade forest, which has three characteristics: (1) A hierarchical cascade structure is constructed, of which the output comes from each layer; (2) Each fault class is evaluated and recognized independently, the result of fault classes that are easy to distinguish is output earlier; (3) A confidence-based threshold estimate method is proposed in HCF and used to improve the training method to increase the reliability of HCF. Based on these, HCF improves the cascade forest structure and implements the proper matching of different depth of feature and fault patterns. The effect of HCF is verified through experiments based on the tunnel fans testing rig. Experimented results show that, compared to Deep Forest, the accuracy of HCF increases by 0.6% to 10.8%, and the training time of HCF is reduced 33.24%.

1. Introduction

In recent years, the development of tunnel traffic has improved the traffic conditions in areas that are difficult to reach by conventional traffic. Since the tunnel is a relatively closed environment and the traffic density is high, the atmosphere of the tunnel is polluted by various harmful small particles and toxic exhaust gas and difficult to dissipate. Unplanned downtime of tunnel fans will incur higher costs and losses than ever before [1]. Therefore, tunnel fans, as shown in Fig. 1, play a very important role in tunnel traffic for maintaining fresh air underground, lower gas concentration, and ensuring safe and reliable traffic. In this context, intelligent fault monitoring and diagnosis (IFMD) for tunnel fans are getting more and more attention [2].

Data-driven approaches for IFMD have seen widespread adoption and outstanding outcomes [3–5]. These methods can perform adaptive analysis on the real-time monitoring data, extract sensitive fault information, and distinguish the health status of the machine. Data-driven IFMD approaches typically include two principal procedures: extract fault-sensitive indicators from vibration signals based on signal processing techniques and deep learning methods [6,7]; classifiers are utilized to the recognition of fault modes [8].

Traditional shallow fault diagnosis methods use artificial features as input, and focus the optimization of the classification algorithm [9–12]. Zheng et al. [13] used multi-scale fuzzy entropy as an artificial feature, designed an intelligent fault diagnosis method of rolling bearing based on support vector machines. Martin-Diaz et al. [14] investigated the current monitoring for effective motor fault diagnosis by using random forest algorithm. Liang et al. [15] proposed a modified K-Nearest Neighbor algorithm employing Shared Nearest Neighbor based Distance for unmanned aerial vehicles fault diagnosis. Wei et al. [16] presented a random forest based fault diagnosis method for planetary gearboxes employing a novel signal processing scheme by combining refined composite hierarchical fuzzy entropy. However, due to the limited artificial features and simple model structure, shallow machine learning has gradually been unable to meet the needs of complex fault diagnosis.

Nowadays, deep learning is one of the most popular learning algorithms, and most of the current deep learning methods are neural networks [17]. The outstanding achievements of deep

☆ This work was funded in part by the Science and Technology Development Fund, Macau SAR (No. 0018/2019/AKP, 0008/2019/AGJ, FDCT/194/2017/A3, and SKL-IOTSC-2021-2023), in part by the Ministry of Science and Technology of China (Grant no. 2019YFB1600700), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant no. 2020B1515130001).

* Corresponding author.
E-mail addresses: zxyang@um.edu.mo (Z.-X. Yang), csli@hust.edu.cn (C.-S. Li), xianbowang@um.edu.mo (X.-B. Wang), chen.hao@connect.um.edu.mo (H. Chen).

https://doi.org/10.1016/j.isatra.2022.10.037
0019-0578/

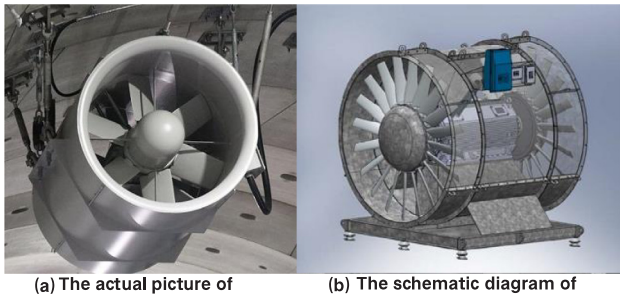

Fig. 1. The tunnel fans. (a) The actual picture of the tunnel fan. (b) Schematic diagram of tunnel fan.

learning on classification [18] demonstrated the potential of deep models in fault pattern recognition. Zhang et al. [19] introduced a deep residual learning-based fault diagnosis method for rotating machinery. Shao et al. [20] proposed a novel method called adaptive deep belief network with dual-tree complex wavelet packet and applied to the fault diagnosis of rolling bearings. Xu et al. [21] developed an improved multi-scale convolutional neural network integrated with a feature attention mechanism model to address the bearing fault diagnosis under unsteady and complex working environments. The advantages of deep learning are reflected in its deep expression of fault characteristics for precise classification, which benefits from its deep structure.

Fault monitoring and diagnosis is ultimately a decision-making problem for equipment, and IFMD methods are tools to help decision-making, so its reliability needs to be studied and improved [22]. It has been proven to be invaluable for permitting engineers to have confidence in the fault diagnosis [23]. Therefore, high-precision fault diagnosis and accurate evaluation of diagnosis results are two important issues of IFMD. Existing methods, such as stacked auto-encoder (SAE) in [24], BP neural network in [25], and convolutional neural network (CNN) in [26], commonly calculate test accuracy as the confidence. In [27], each neuron is used to represent the classification confidence of each health condition respectively. Karabadji et al. [28] calculated the trust index based on the classification accuracy taking the size of the training set and testing set into consideration. The confidence proposed by existing methods describes the probability that the samples are classified correctly.

The structural characteristics of deep learning are analyzed separately and expanded to a broader perspective in [29], and a novel deep learning method based on decision tree-based ensemble method is proposed, namely Deep Forest (DF) [30]. DF is an ensemble learning method which leads the research of non-neuron networks. Existing research found that a set of classifiers with an ensemble consensus mechanism demonstrates superior classification performance than one single classifier [31]. Complex meta-learners in DF, namely random forests, enable the complexity of the model structure to be appropriately reduced. So, fewer parameters need to be set and more efficient training is achieved.

Although DF inherits the advantages of shallow machine learning and deep learning well [32], there are still some issues worth exploring: (1) The classification accuracy of DF has the potential to be further improved. DF generates the optimal depth by continuously verifying the average accuracy of each layer, however, the average accuracy cannot accurately evaluate the separability of the feature distribution of each class, which leads to failure to reach the optimal performance; (2) The purpose of IFMD is to provide reliable equipment condition prediction, the evaluation of reliability is lacking in current IFDM methods; (3) The fault monitoring system commonly works online, which has higher requirements for the efficiency of the fault recognition algorithm. However, The decision tree searches for the optimal split point by traversing the attributes, so it is inefficient when facing high-dimensional data.

In order to solve the problems mentioned above, the structure of the cascading forest is improved by evaluating the performance of each layer so that the data can be shunted and complex multi-classification problems are simplified. In this improved framework, difficult sub-classifications are solved by extracting deep features, and simple sub-classifications are completed based on shallow features. In this way, as the data flowing to deeper layers, the number of samples decreases so as to increase the training efficiency, and the sample space is purified making classification easier. The middle layer features of DF is interpretable, so it is more reasonable to implement the hierarchical output structure based on DF than neural networks. The output accuracy threshold (OAT) of each layer is predicted by preset confidence to provide the optimal hierarchical output. For improving the computational efficiency of random forests, signal process methods are employed for constructing low-dimension artificial features [33].

Based on the above discussion, a hierarchical cascade forest (HCF) model is proposed in this paper. The HCF employs a cascade structure to integrate the random forests into a deep model and performs evaluation for each category in each layer. HCF assigns the classification tasks of different classes to different layers of HCF and integrates the output of each layer as the final result. The research on HCF aims to improve class accuracy and training efficiency, and achieve accurate confidence evaluation. The contributions of this research can be summarized as follows:

1. The cascaded forest method is improved for accurate mechanical fault diagnosis. A cascaded forest-based hierarchical framework is proposed, in which the feature space is continuously simplified, so that complex classification problems are disassembled and simplified to improve accuracy;
2. A novel confidence is defined and a confidence evaluate algorithm is proposed to predict the OAT for each layer, the OAT is employed to implement appropriate hierarchical output thereby improving accuracy and efficiency.
3. The hierarchical cascade structure improves the training efficiency compared with the deep forest model in [30] and deep neural networks in [34];
4. The effect of the OAT of HCF on classification accuracy is discussed, and the proposed confidence evaluation method is verified on public and experimental platform datasets;

The rest of this paper is organized as follows. In Section 2, the structure of the DF is briefly introduced. In Section 3, the HCF method is proposed, and the theory of the confidence evaluation algorithm is provided. In Section 4, based on bearing and gearbox data, the improvement of accuracy and training efficiency of HCF are verified. The influence of artificial features and the OAT on fault diagnosis will be also discussed in this section. The paper is finally concluded in Section 5.

2. A brief review of deep forest

Deep learning has stronger feature mining and knowledge representation capabilities which are produced by the layer-by-layer structure. The DF model [30] has a non-neural network deep model framework that employs a cascade structure to process feature layer-by-layer. DF replaces the neurons in the neural network with random forests and preserves deep structure, forming a deep model with shallow model as the unit.

Each layer of DF consists of several random forests which are basic units, similar to neurons in neural networks, called meta-classifiers. The probability evaluated class vector is employed to

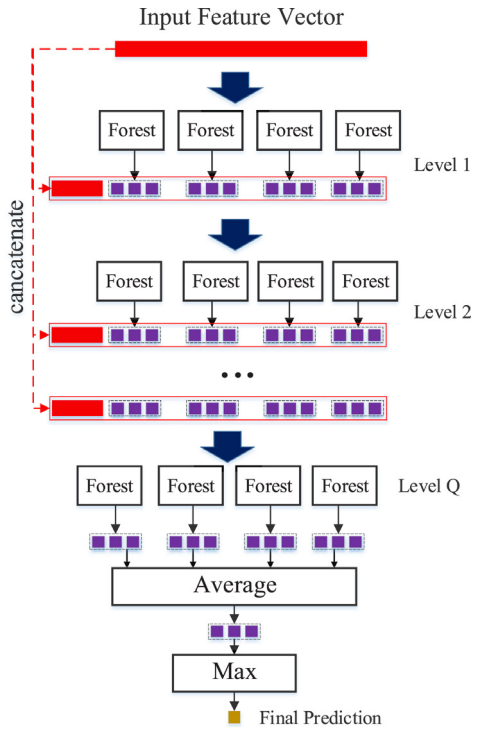

Fig. 2. Illustration of the cascade forest structure.

represent the class distribution predicted by the random forests. And then the output vectors of RFs are concatenated with the original data as the input of the next layer. In the last layer, the probability distributions of all random forests outputs are averaged, and the class with the highest probability is taken as the classification result. An example of the DF is illustrated in Fig. 2, each layer contains four random forests that output 3-dimensional feature vectors.

3. The hierarchical cascade forest

3.1. Structure of hierarchical cascade forest

HCF is a deep learning model with a cascade structure which consists of a series of forest layers, as shown in Fig. 3. Each forest layer only outputs the samples that are predicted to belong to certain classes as a partial final result and these classes are called the output classes of this forest layer. For a multi-classification task, each forest layer solves part of the problem. The predicted labels of all samples can be obtained by integrating the output of all forest layers. As shown in Fig. 4, assume that each forest layer of HCF contains 4 random forest classifiers. The input feature vector is assigned to each random forest through the Bagging strategy [35]. The class distribution vectors are averaged and then the maximum value is the label predicted by the forest layer. Samples that are predicted to belong to the output classes are output in this forest layer, and the class distribution of the remaining samples is concatenated with raw input and output to the next forest layer.

Given a training dataset $D = \{(X_1, C_1), \ldots, (X_N, C_N)\}$ with N samples, $X_i \in \mathbb{R}^K$ is a K-dimensional attribute vector of sample i and C_i is the true label of sample i. Decision trees of HCF are trained based on training dataset D along with bootstrap sampling. In each decision tree, the training dataset keeps splitting and divided into subtrees until the instances of each node are pure or the tree meats the maximum depth, the splitting attributes are chosen by minimizing information entropy.

For the tth decision tree in HCF, the proportion for the ith class at the terminal node h can be represented as:

$$p_{i,h}^{t} = \frac{1}{m_h} \sum_{j \in h} \mathring{1}\left(Y_j = i\right) \tag{1}$$

where, m_h is the number of instances that are divided into the node h. $\mathring{1}(\cdot)$ is an indicator function. For the test case X_{test}, the tth decision tree outputs the predicted class distribution, namely the class vector, by the following equation:

$$\hat{Y}\left(X_{test}\right) = \left\{\hat{Y}_1\left(X_{test}\right), \hat{Y}_2\left(X_{test}\right), \ldots, \hat{Y}_c\left(X_{test}\right)\right\} \tag{2}$$

$$\hat{Y}_i\left(X\right) = \frac{1}{T} \sum_{t=1, X \in h_t}^{T} p_{i,h_t}^{t} \quad i = 1, 2, \ldots, C \tag{3}$$

$\hat{Y}_i(X)$ is the proportion in the class vector of class i, and T is the number of decision trees of each forest. For each layer in HCF, the class vectors are concatenated with raw input to form an augmented feature vector and fed to the next layer. The augmented feature vector can be represented as:

$$\mathscr{F}_{aug} = \left\{ \hat{Y}^{(1)}\left(X_{test}\right) \;\vdots\; \cdots \;\vdots\; \hat{Y}^{(S)}\left(X_{test}\right) \;\vdots\; X_{test} \right\} \tag{4}$$

The dimension of $\mathscr{F}_{aug}$ is $C' \times S + dim\left(X_{test}\right)$, where C' is the number of remaining classes, S is the number of forests in each layer, and $dim\left(X_{test}\right)$ is the dimension of raw input. Samples predicted to be N_l classes are output in layer l. Therefore dimension of augmented feature vector decreases layer by layer. N_l is determined by a threshold discrimination function, by the following equation:

$$g_l\left(X\right) = \begin{cases} 1 & \frac{1}{D}\sum_{i}^{D} Pr\left(\max\limits_{j \in C}\left(\bar{Y}\left(X\right)\right) \neq Y\right) \leq \theta_l \\ 0 & otherwise \end{cases} \tag{5}$$

$Pr\left(\max\limits_{j \in C}\left(\bar{Y}\left(X\right)\right) \neq Y\right)$ denotes the probability of misclassification on dataset X. $\bar{Y}(X)$ is average of class vectors of S forests and $\max\limits_{j \in C}(\cdot)$ is the class label with the largest proportion in the class vector. Note that the threshold discrimination function evaluates a class of instances rather than one, $g_l(\cdot) = 1$ means samples of the class can be output at layer l and vice versa.

For concise expression, each layer of the HCF model is defined as a transfer function. The output of this function is the input of the next layer. The transfer function can be represented as Eq. (6).

$$f_l = \begin{cases} y_l\left(\mathbf{x}\right) & l = 1 \\ y_l\left(\{f_{l-1}\left(\mathbf{x} \cdot \mathcal{G}\left(\mathbf{x}, l\right)\right)\}\right) & l > 1 \end{cases} \tag{6}$$

Where $y_l(\mathbf{x})$ is the combination of all forests in layer l and raw input, that can be represented as Eq. (4), and $\mathcal{G}(\mathbf{x}, l)$ is an indicator function defined as Eq. (7). Where $\acute{\mathbf{x}}$ is a subset of $\mathbf{x}$ and $\mathcal{G}(\mathbf{x}, l)$ indicates the instances that does not belong to the classes output at layer $l-1$.

$$\mathbf{x} \cdot \mathcal{G}\left(\mathbf{x}, l\right) = \acute{\mathbf{x}} \; where \; \acute{\mathbf{x}} \in \mathbf{x}, g_{l-1}\left(\acute{\mathbf{x}}\right) = 0 \tag{7}$$

So far, the predictive function of HCF can be defines as Eq. (8), where $arg\max\limits_{c}\left[\bar{f}_l\left(\mathbf{x}\right)\right]$ denotes the class c with the largest proportion in the output of layer l and $\bar{f}_l(\mathbf{x})$ is average of $f_l(\mathbf{x})$ for S forests. L is the number of layers of HCF.

$$\mathbb{H}\left(\mathbf{x}\right) = \sum_{l=1}^{L}\left[c \cdot \mathring{1}\left(arg\max_{c}\left[\bar{f}_l\left(\mathbf{x}\right)\right] \in \left\{Y_i \cdot g_l\left(\mathbf{x}\right) | i = 1, 2, \cdots, C\right\}\right)\right] \tag{8}$$

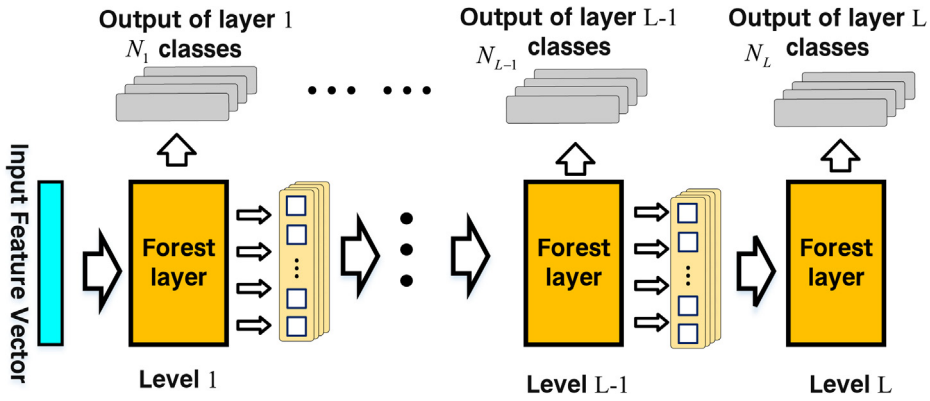

Fig. 3. Structure of the hierarchical cascade forest.

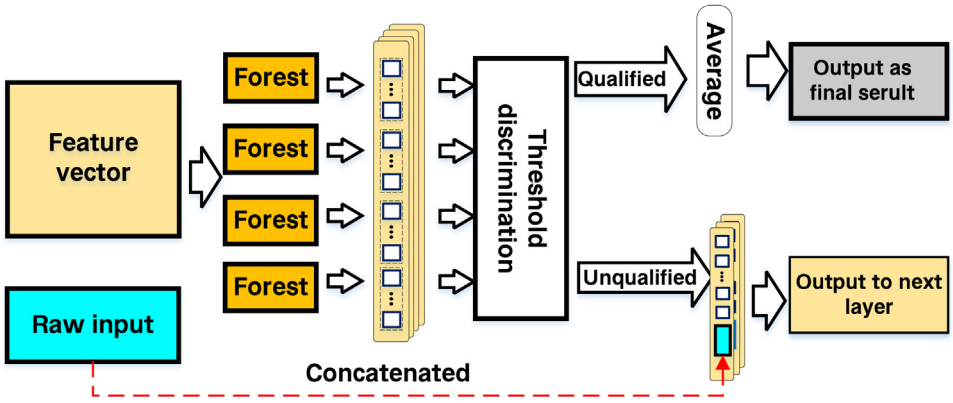

Fig. 4. Illustration of the forest layer of hierarchical cascade forest.

3.2. Confidence evaluation algorithm

In this paper, a confidence evaluation algorithm is proposed to improve and evaluate the stability of the model over whole sample space. The existing confidence calculation methods for fault diagnosis usually face three problems: (1) For multi-classification problems, the confidence is used to evaluate the performance of the model on all classes, however, the performance on various classes is not comparable; (2) The existing evaluation of confidence is sensitive to the quality of the test set and unrelated to the training procedure, which makes it ineffective for improving the performance; (3) The validity of the confidence evaluation has not been sufficiently validated in the existing studies. The confidence evaluation algorithm of HCF is proposed to solve these problems. The confidence is calculated for each forest layer independently and employed to evaluate the result of each forest layer. Besides, the confidence of HCF is a preset parameter used to predict the OAT of each forest layer to adjust training to get the most effect.

Generalization error is an effective means to evaluate the generalization performance of a model. The proposed method provides threshold parameters for model training by predicting the Ge, in order to promote the training effect and accurately quantify the confidence. Vapnik Chervonenkis (VC) dimension [36] is wildly used as a measurement for qualitative analyzing the Ge of classifiers. However, the VC dimensions of complex models are very difficult to calculate, especially for deep learning models. Since there is a lot of parameters related to a deep learning model, which makes the hypothetical space too huge to calculate exactly. Gaussian complexity [37] is a specific quantization of the hypothetical space complexity of a model, which can be obtained by calculating the fitness of the model to random noise. In this paper, Gaussian complexity is employed to design a confidence calculating method.

Assuming that $\hat{f}$ is the trained model, the error of this model for predicted unknown instances is defined as Ge. The expected risk of a model can be quantized by generalization error, the models with small Ge have high confidence. The Ge can be calculated by Eq. (9).

$$R_{gen}\left(\hat{f}\right) = \mathbf{E}\left[L\left(Y, \hat{f}\,(x)\right)\right] = \int_{x\times y} L\left(Y, \hat{f}\,(x)\right) P\,(x, y)\,dxdy \tag{9}$$

Where $R_{gen}\left(\hat{f}\right)$ denotes the generalization error of $\hat{f}$; $L\left(Y, \hat{f}\,(x)\right)$ is the loss function of learning algorithm, $R_{gen}\left(\hat{f}\right)$ is the expectation of $L\left(Y, \hat{f}\,(x)\right)$ in sample spaces $\{X, Y\}$. The whole sample space distribution is unknown, thereby it is difficult to calculate $R_{gen}\left(\hat{f}\right)$. Although the accurate value of $R_{gen}\left(\hat{f}\right)$ cannot be obtained, evaluating generalization error by calculating an upper bound of generalization error is a common approach for analyzing generalization performance.

Existing research has investigated the complexity of a model that can be represented with a class of functions. In [37], a probability evaluation of the error bound of a decision tree model based on the Gaussian complexities is proposed. Then the error can be evaluated by formula (10).

$$Pr\,(y \neq f\,(x)) \leq \hat{P}_N\,(y \neq f\,(x)) + \sum_{l}^{L} \min\left(\tilde{P}_N\,(l)\,, cdG_N\,(H)\right) + \sqrt{\frac{\ln\left(\frac{L}{\delta}\right)}{2N}} \tag{10}$$

Where $Pr\left(y \neq f\left(x\right)\right)$ denotes the expectation of the probability of misclassification on unknown datasets, $\hat{P}_N\left(y \neq f\left(x\right)\right)$ denotes the expectation of the probability of misclassification on the training data. $\tilde{P}_N\left(l\right)$ denotes the proportion of all training examples which reach leaf l and are correctly classified. L is the number of leaf nodes, N is the scale of the training dataset, and c is an absolute constant and d is the depth of the decision tree. $G_N\left(H\right)$ is the Gaussian complexity of a function family H, defined by Eq. (11).

$$G_N\left(H\right) = \mathbf{E}\left[\hat{G}_N\left(H\right)\right] = \mathbf{E}\left[\mathbf{E}\left[\sup_{f\in F}\left|\frac{2}{N}\sum_{i=1}^{N} g_i f\left(X_i\right)\right| \left|X_1, X_2, \cdots, X_N\right.\right]\right] \tag{11}$$

Where $g_1, g_2, \ldots, g_N$ are independent Gaussian $N\left(0, 1\right)$ random variables, and $X_1, \ldots, X_N$ are independent samples. Assume that the model is trained in a reasonable hypothesis space, that is, each f in F has a well classification performance. Then $\sum_l \tilde{P}_N\left(l\right)$ is close to 1 for each f in F and $G_N\left(F\right)$ is less than $\frac{1}{N}$. Obviously, formula (12) holds.

$$\sum_l \min\left(\tilde{P}_N\left(l\right), cdG_N\left(H\right)\right) \leq \sum_l cdG_N\left(H\right) \tag{12}$$

To simplify the presentation, the second term in formula (10) is simplified, the case where $\tilde{P}_N\left(l\right)$ is less than $cdG_N\left(H\right)$ is ignored. And let the constant c be the reciprocal of $N \times d$ the simpler form of formula(10) is provided as formula (13).

$$Pr\left(y \neq f\left(x\right)\right) \leq \hat{P}_N\left(y \neq f\left(x\right)\right) + \frac{1}{N} + \sqrt{\frac{ln\left(\frac{L}{\delta}\right)}{2N}} \tag{13}$$

According to formula (13), the generalization error of random forests can be obtained with a parameter δ, δ means the probability that the error exceeds the estimated bound. The probability $1 - \delta$ is defined as the confidence of the result of HCF. The confidence is corresponded with a bound of error (the right side of formula (13)), which is employed to set the OAT of each layer of HCF.

3.3. The procedure of HCF

Based on the confidence defined previously, a training mechanism is proposed to improve the training efficiency and accuracy. HCF adopts a layer-by-layer training method, and the depth of HCF is generated during training, which reduces the labor cost of optimizing structural parameters. In the procedure, some parameters need to be initialized first, including the maximum depth, the number of random forests included in each forest layer, and the number of decision trees included in the random forest.

In the training phase, the training data is assigned to each random forest in the forest layer via Bagging. The OAT of the current forest layer can be calculated according to the preset confidence. Then the training data is divided into k parts, of which k−1 parts are used as the training set to train the random forest, and one part is used as the verification set to verify the accuracy of the model. According to the validation set results, the categories whose accuracy is higher than the OAT can be filtered out and assigned as the output class of the forest layer. The samples corresponding to the output class are output in this layer. The next forest layer receives the remaining samples and performs the same training until all samples are output. The detail confidence-based training algorithm is summarized in Algorithm 1.

In the test phase, let all samples pass through the forest layers. Each forest layer outputs samples that are predicted as the output class, and passes the remaining samples to the next layer. Repeat the foregoing process until all categories are output. The output results of all forest layers are integrated to get the final classification result. To intuitively understand the proposed model, the flow chart of the proposed framework is shown in Fig. 5

In the proposed training mechanism, confidence is preset parameter rather than the evaluation of the trained model. Therefore, in the testing process, the effectiveness of this training mechanism and confidence evaluation algorithm can be verified by analyzing the statistical indicators of multiple tests.

Algorithm 1 Confidence-based training mechanism

Input: maximum of the depth of HCF M; the confidence of HCF *conf*; the training data D_{Train}
Output: output results

Initialize the model: the set of rest calsses $C_{rest} = \{c_1, c_2, \ldots, c_n\}$; n is the number of classes; the layer id $l = 0$
while C_{rest} is not $\emptyset$ & $L \leq M$ **do**
  Generate a new forest layer
  $l = l + 1$
  Calculate the generalization error $R_{gen}(l)$
  Devide D_{Train} into k sets: $D_{Train} = \{S_1, S_2, \ldots, S_k\}$
  for *epoch* $= 1$ to k **do**
    $\complement_{D_{Train}} S_{epoch}$ is train set, S_{epoch} is verification set
    Train HCF on S_{train} and record
    Test HCF on S_{ver} and record
  for $i = 1$ to number of rest classes **do**
    Calculate the target accuracy of each class
    $\theta_{tar}(c_i) = \theta_{train}(c_i) - R_{gen}(l)$
    Calculate the average test accuracy θ_{ave}
    if $\theta_{ave}(c_i) \geq \theta_{tar}$ **then**
      Output the samples which belong to c_i as final result
      Record c_i as the output class of forest layer l
    else
      Output the samples which belong to c_i as input of next layer
Integrate the output of each forest layer
return result

4. Experimental verification

4.1. Preprocessing: Feature extraction

To obtain enough information for the diagnosis, the input length commonly covers several periods of fault emergence. However, HCF is a random forest-based model, and random forests suffer efficiency and training challenges when fed with high-dimensional features. In this paper, we refer to a previous research method of fault diagnosis based on deep forest to deal with this problem(see Fig. 6). [32].

Variational mode decomposition (VMD) [38] is a signal decomposition method that is less sensitive to noise and sampling. Three types of mode features can be calculated based on the intrinsic mode functions (IMFs) obtained by VMD, including energy entropy, the center frequency, and singular value features. Besides, 23 time–frequency domain features are selected as suggested in [39]. The energy entropy features used to express the global feature of VMD components are obtained as suggested in [40] and the singular value features used to reduce the dimension and express decorrelation characteristics are obtained as suggested in [41].

As presented in Fig. 4, the feature vector consists of 12 time-domain features, 11 frequency-domain features, and $2K + 1$ VMD

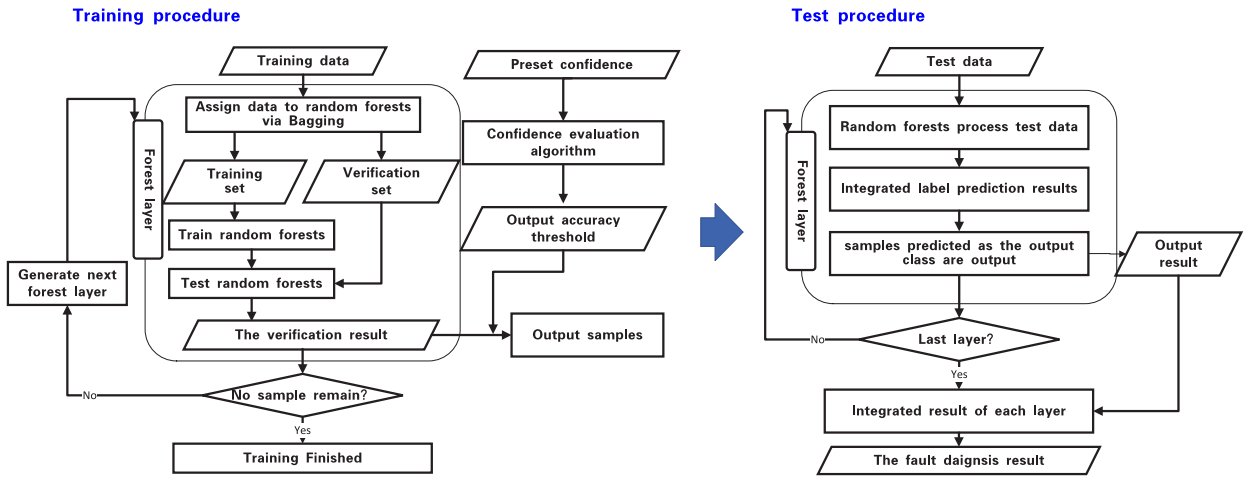

Fig. 5. The procedure of HCF.

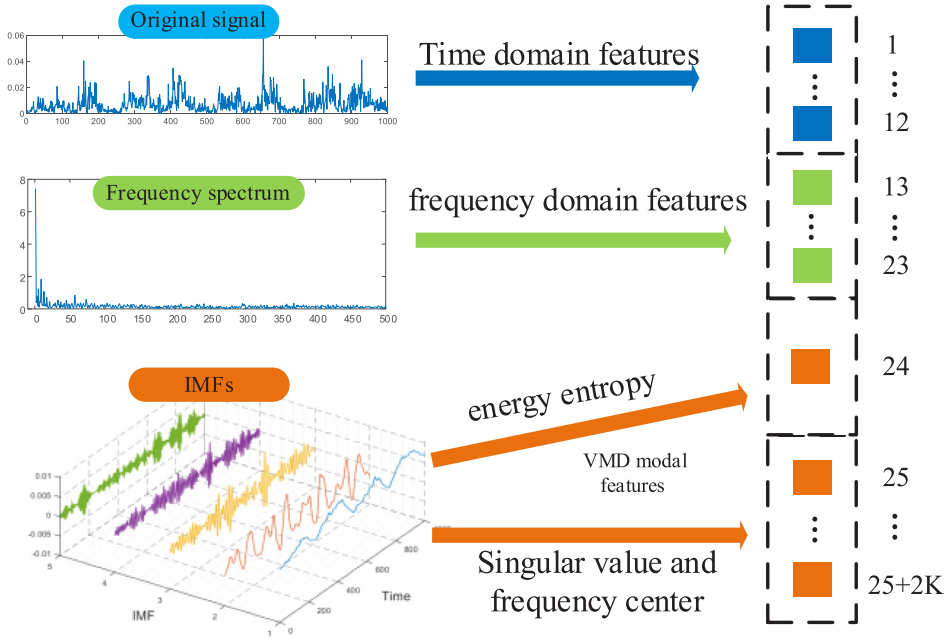

Fig. 6. The artificial features consist of 12 time domain features, 11 frequency domain features and 2K+1 VMD modal features.

modal features. K is the number of IMFs, K is set as 5 by trial and error in this paper.

4.2. Experimental data description

The vibration signal recording of Case Western Reserve University is an authoritative dataset for bearing fault diagnosis that has been widely used in IFD methods. This dataset is sampled at 12 KHz and contains four health states: normal state (N) and the fault states of ball (B), inner race (IR), and outer race (OR). For fault state B and IR, the dataset recorded the four damage diameters of 0.007, 0.014, 0.021, and 0.027 inches, and the recording of OR contains damage diameters of 0.007, 0.014, 0.021 inches. Each state consists of samples under four different motor loads of 0, 1, 2 and 3 horsepower(hp) (motor speeds of 1730, 1750, 1772, and 1797 rpm).

CWRU dataset is divided into two datasets A and B based on different loads and different fault diameters. Dataset A and B contain 12 × 80 and 13 × 80 samples respectively, and 75% of them are training data and 25% are test data. Each sample contains 1024 data points. The details are shown in Table 1.

Another dataset is collected from the gearbox dynamics simulator (GDS) of Spectra Quest company in our laboratory, as shown in Fig. 7. In the experiment, the vibration signals under different health conditions are obtained by changing the bearing and gear closest to the sensor. Fig. 8 shows the location of the faulty gear, faulty bearing, and sensors. The input speed of the device is set to 1400 rpm, and the speed of fault gear and the gear meshed with fault gear is 1184 rpm and 840 rpm respectively. The acquired vibration signals contain 8 conditions, that is, a normal condition(C0), and 7 fault condition: chipped tooth (C1), missing tooth (C2), gear crack (C3), bearing outer ring fault (C4), bearing inner ring fault (C5), loose fixed(C6) and gear unbalance(C7). The photos of the fault components are shown in Fig. 9. Datasets C, D, and E are constructed by combining these single states. The details of C, D, and E are shown in Table 2. The vibration signals of the gearbox with different conditions are shown in Fig. 10.

Table 1
Component fault scenarios and health condition information.

State	A		B	
	Load(hp)	Fault diameter(inch)	Load(hp)	Fault diameter(inch)
N	3	\	0	0
B	3	0.007/0.014/0.021/0.027	0/1/2/3	0.07
IR	3	0.007/0.014/0.021/0.027	0/1/2/3	0.07
OR	3	0.007/0.014/0.021	0/1/2/3	0.07

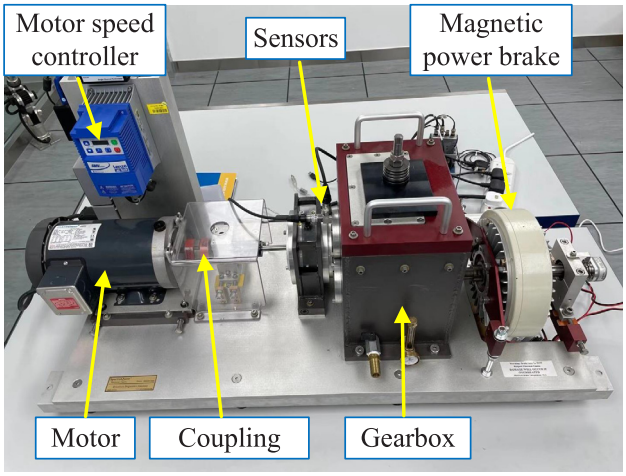

Fig. 7. The Spectra Quest gearbox dynamics simulator.

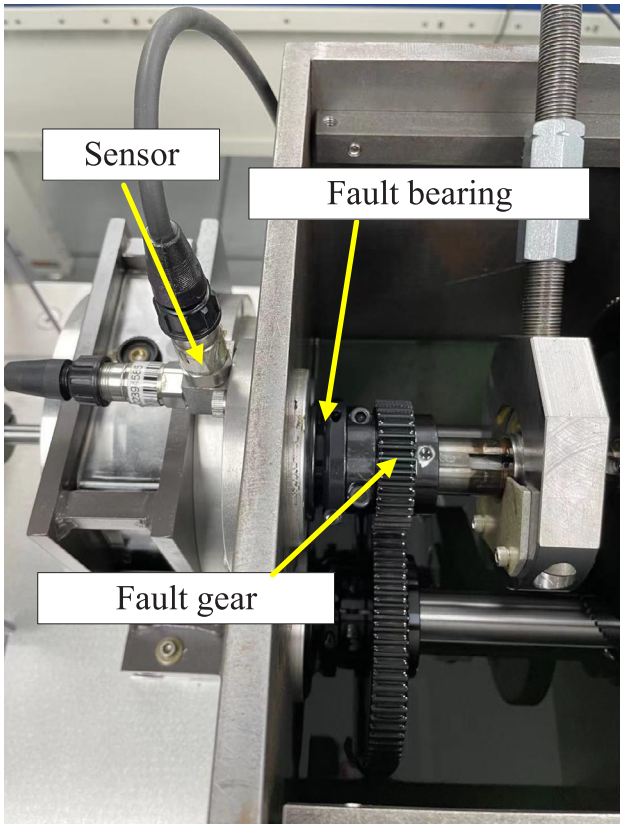

Fig. 8. The location of the faulty gear, faulty bearing and sensors.

4.3. Case I: Analysis on the confidence of HCF

By predicting the OAT of each layer of the HCF, the proposed confidence evaluation algorithm promotes model training, ensuring that the training results satisfy the preset confidence criteria. In Case I, the effect of the OAT on the diagnosis accuracy is analyzed and the robustness evaluation ability of confidence evaluation algorithm is verified.

Each layer of HCF consists of four random forests, and each random forest contains 50 C4.5 decision trees. The depth of HCF is determined adaptively and the maximum is 10. 200 HCF models with OAT settings from 0.8 to 0.99 are tested in case I, confidence settings as 0.9, and each model has performed 50 repeated experiments. Fig. 11 reflects the ratio that the test accuracy is higher than the OAT in 50 repeated experiments, called control rate, and the overall accuracy of the model changes with different OATs. The test accuracy higher than the OAT can be understood as the test effect of the model has reached the expected. The trend reflected in Fig. 11 is that the lower the OAT is set, the higher the control rate is, but the overall accuracy is lower. The OATs predicted by the confidence evaluation algorithm are at the inflection point of the rising curves, indicating that the confidence evaluation algorithm gives a reasonable and effective forecast. In Fig. 11, the result where the OAT predicted by the proposed algorithm reaches 100%, 90%, and 74%, respectively, and the average control rate reaches 92.45%, exceeding the preset confidence.

4.4. Case II: Diagnostic performance with artificial features

As mentioned in Section 4.1, the signal process method is employed for obtaining artificial features. In case II, several signal process methods are compared on dataset B, the results are shown in Table 3. Four combinations of artificial features are tested, "time-fre" is a combination of 12 time-domain features and 11 frequency-domain features explained in Section 4.1; empirical mode decomposition (EMD) and empirical wavelet transform (EWT) are compared with VMD, and the feature extraction method is the same as explained in Section 4.1. EMD decomposes the signal according to the time scale characteristics, without any basis function set in advance, and continuously extracts the components of various scales that make up the original signal from high frequency to low frequency. Our experiments use a PC with Intel Core i7-9700 CPU (8 cores), and all the methods do not use parallel computing or GPU.

The results in Table 3 contain the average accuracy of fault diagnosis, the signal process time cost, and the leaf number of the first layer of HCF. The parameters of HCF are the same as that in case I. The leaf number can indicate the complexity of the feature space, less leaf number is beneficial for accurate and stable classification. Compared with the other three features, the accuracy of the Time-fre+VMD feature is improved by 16.31%, 3.34%, and 9.76%, and the model complexity is reduced by 32.33%, 12.48%, and 13.94%. To clearly show the performance of class recognition, the details of 13 classes are shown in Fig. 12. In summary, the performance of Time-fre+VMD is the best.

4.5. Case III: Comparisons with previous work

Case III aims to compare the fault diagnosis performance of HCF and other existing fault diagnosis methods. The experiment

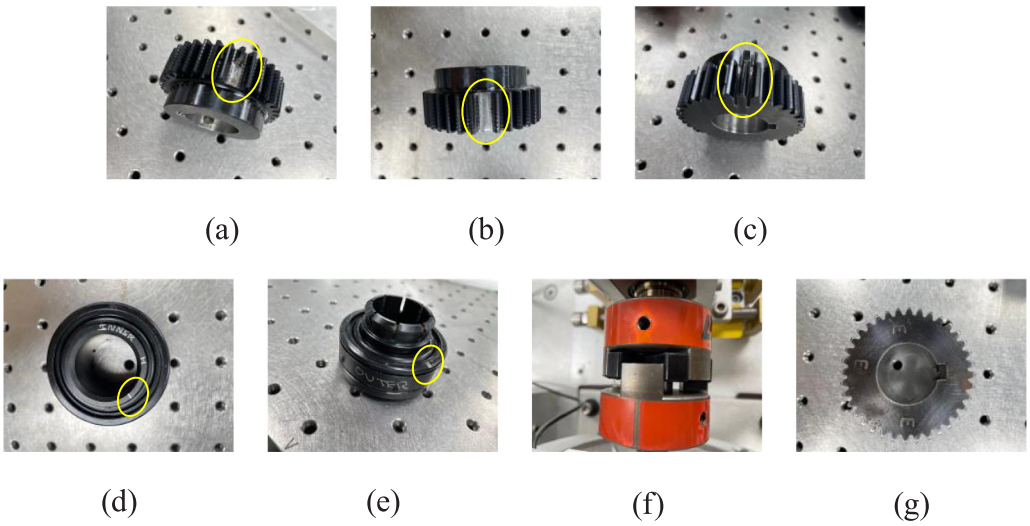

Fig. 9. Photos of the faults. (a) Chipped tooth. (b) Missing tooth. (c) Gear crack. (d) Broken ball bearing. (e)Wear of outer race of bearing. (f) Loose fixed. (g) Gear unbalance.

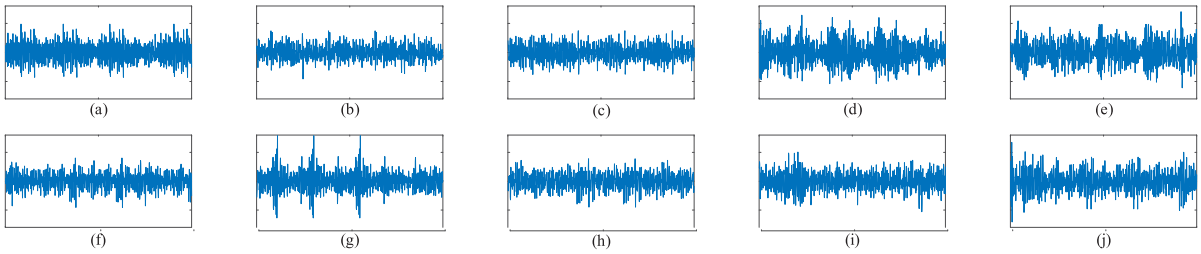

Fig. 10. Vibration signals of gearbox with different conditions. (a) N. (b) F1. (c) F2. (d) F3. (e) F4. (f) F5. (g) F6. (h) F7. (i) F8. (j) F9.

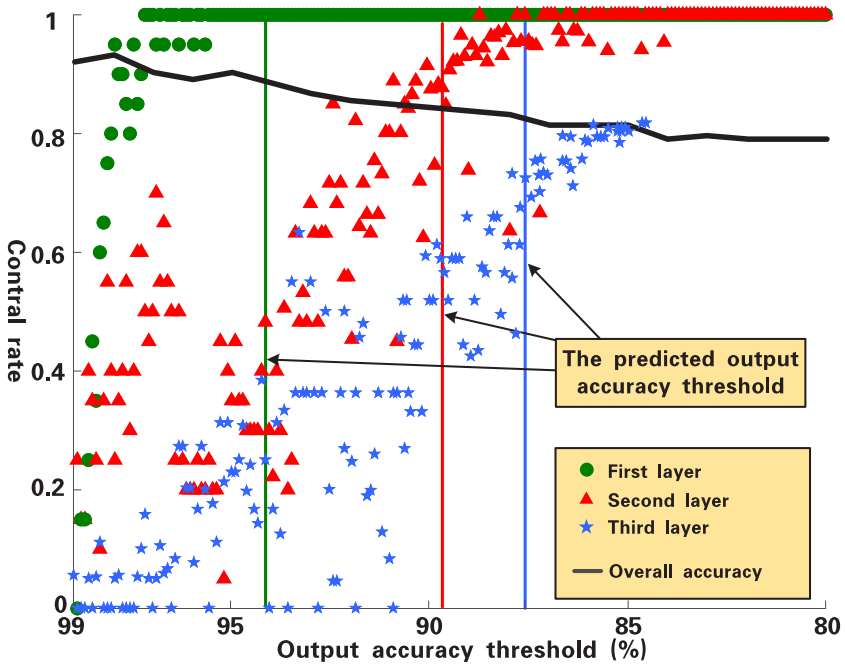

Fig. 11. The control rate of the first three layers of HCF with the OAT parameter from 80% to 99%.

is executed on datasets C, D, and E. DF, VMD-CNN [34], and standard SAE are taken to compare methods. VMD-CNN is a bearing fault diagnosis method that merges multi-channel features based on VMD and employs CNN for feature extraction. Note that there are two signal channels of datasets C, D, and E, in addition to this, the parameters of VMD-CNN are the same as those in [34]. The parameters of HCF are set based on the proposed confidence evaluate algorithm and the compared methods are implemented with the optimal parameters. The details of the parameters of experimental methods are shown in Table 4. Each item in the

Table 2
Gearbox dynamics simulator dataset description.

Dataset	State	Train	Test	C0	C1	C2	C3	C4	C5	C6	C7
C	N	105	35	✓							
	F1	105	35		✓		✓				
	F2	105	35		✓			✓			
	F3	105	35			✓		✓			
D	N	90	30	✓							
	F1	90	30		✓	✓					
	F2	90	30		✓		✓				
	F3	90	30		✓			✓			
	F4	90	30		✓				✓		
	F5	90	30		✓					✓	
	F6	90	30			✓		✓			
	F7	90	30			✓			✓		
E	N	75	25	✓							
	F1	75	25		✓	✓					
	F2	75	25		✓		✓				
	F3	75	25		✓			✓			
	F4	75	25		✓				✓		
	F5	75	25		✓					✓	
	F6	75	25			✓		✓			
	F7	75	25			✓			✓		
	F8	75	25			✓				✓	
	F9	75	25							✓	✓

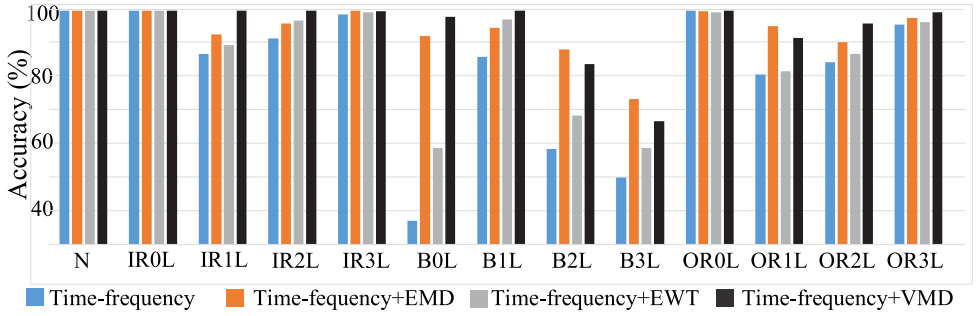

Fig. 12. Accuracy of HCF based on different artificial features.

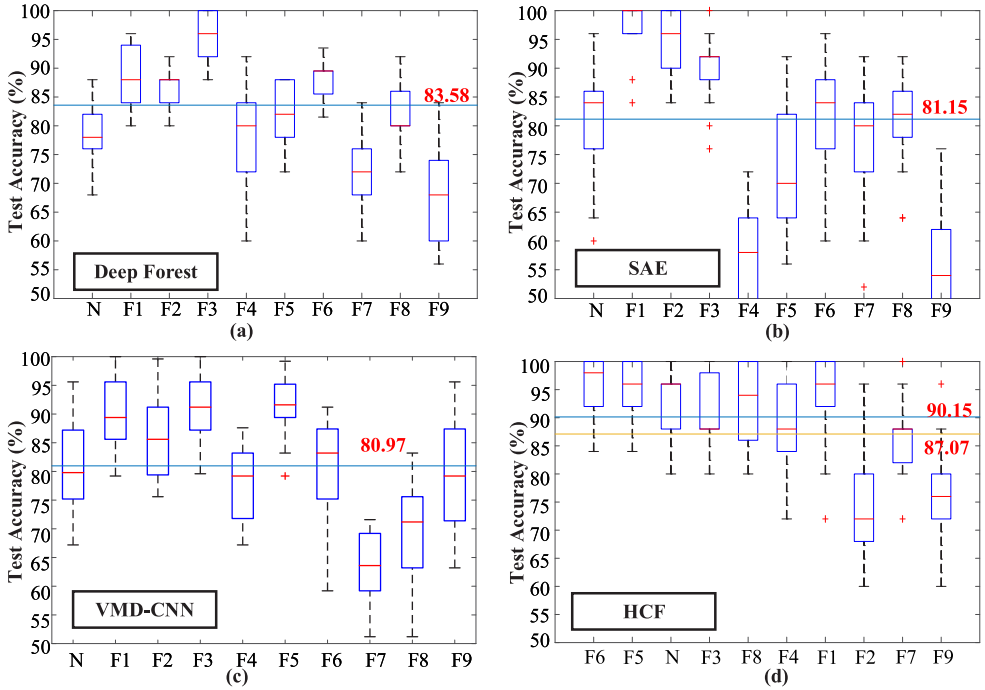

Fig. 13. The test accuracy of each class. (a) is the result of DF; (b) is the result of SAE; (c) is the result of VMD-CNN and (d) is the result of HCF.

Table 3
Comparison of different artificial features.

Features	Accuracy (%)	Training time (s)	Leaf num
time-fre	81.79	13.63	56.40
time-fre+EMD	92.05	55.88	47.94
time-fre+EWT	86.67	47.52	48.56
time-fre+VMD	95.13	66.27	42.62

Case III is repeated 20 times under the same conditions to obtain the result that can reflect the stable performance.

Table 5 shows the accuracy of each class, the mean accuracy, and the training time cost. The optimal results of each dataset are shown in bold. HCF stands out in complex problems since HCF can hierarchically simplify the sample space. From the result on dataset E, HCF increases by 0.6% compared to DF in the four easiest categories (Selected according to the results of DF), and the four most difficult categories (Selected according to the results of DF) increased by 10.8%. This can demonstrate that the performance improvement of HCF is attributed to improve difficult classification sub-problems. Compared with DF, the training time of HCF is reduced by 27.35%, 33.24%, and 17.23% respectively corresponding to the dataset C, D, and E. This shows that HCF effectively improves the training efficiency of the cascade structure by outputting layer by layer. Fig. 14 shows the confusion matrix obtained from the proposed method in Case III.

To visually show the advantages of HCF in confidence evaluation, the statistical result of 20 repeated experiments are analyzed and displayed in Fig. 13. The subfigure (a), (b), (c), and (d) are corresponding to the results of DF, SAE, VMD-CNN, and HCF, where (d) is arranged in order of output of each class. The horizontal lines in subfigures (a), (b), and (c) are the mean accuracy of DF, SAE, and VMD-CNN, which are normally used as the confidence of the corresponding methods and called the model accuracy. The 2 horizontal lines in (d) are the OATs of the first 2 layers of HCF with the confidence is 95%. The OAT of the first layer is the same for each repeated case and that of the second layer is an average because the OAT parameter is related to the number of samples. From Fig. 8, 60%, 40%, and 50% of the test results of DF, SAE, and VMD-CNN is lower than the model accuracy, and it should be emphasized that which classes have poor results are unknown in compare methods. In contrast, the confidence proposed in this paper can be set in advance, and HCF can calculate a reasonable model accuracy index based on the set confidence. Fig. 13 not only shows the good model accuracy of HCF but also verifies the effectiveness of the confidence evaluation algorithm of HCF. In 20 repeated experiments on dataset E, the first and second layers of HCF output 2.6 and 5.2 classes on average. From Fig. 13, the mean accuracy of the classes in the first two layers is over the layer accuracy and the lower quartile of class accuracy in the first layer is over the layer accuracy.

4.6. Deep feature evaluation

In order to further analyze the effectiveness of the hierarchical output schema, the quality of the deep features is evaluated and compared, and the deep features of each layer are displayed. Taking the Case III as an example, considering that the deep features are all high-dimensional data, thus, t-distributed Stochastic Neighbor Embedding (t-SNE) is used for data dimension reduction. The deep features are converted to two-dimensional visualization data by t-SNE. The feature distributions of different layers of HCF, gcForest, SAE, and VMD-CNN are shown in Fig. 15. Since the four methods are inconsistent in-depth, the distributions of the deep feature of the first layer, a middle layer, and the output layer are shown here, corresponding to the first column, the second column, and the third column in Fig. 15 respectively.

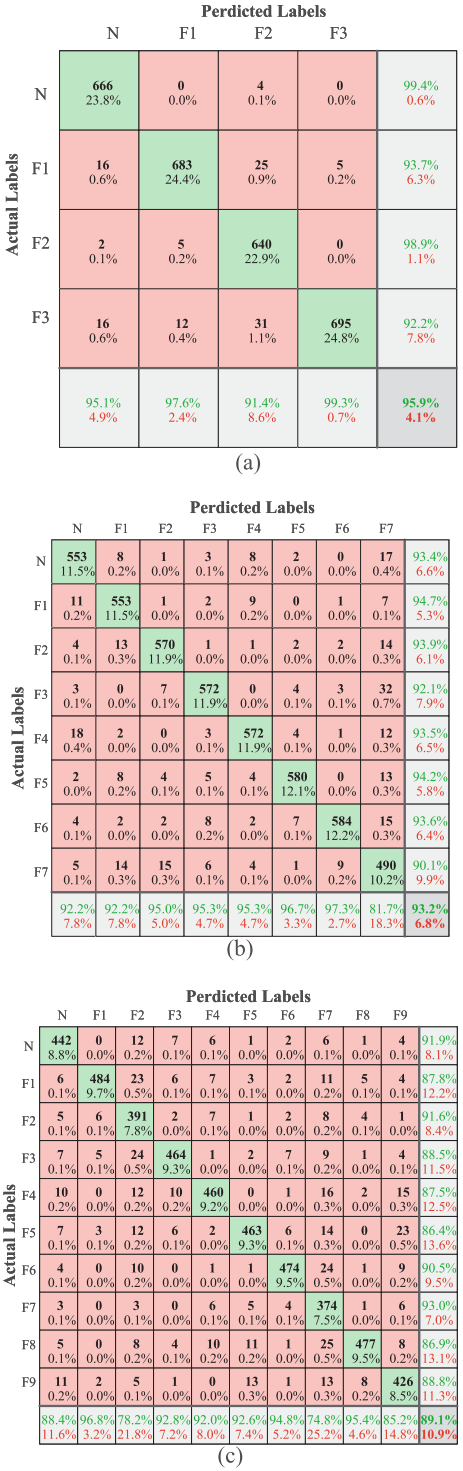

Fig. 14. The confusion matrix using the proposed method on dataset C, D and E. (a) experimental result on dataset C; (b) experimental result on dataset D; (c) experimental result on dataset E.

It can be clearly seen from Fig. 15 that as the depth increases, the highly overlapping area of the deep feature learned by HCF gradually shrinks, and the classes that were originally in the overlapping area will be separated. Comparing (a), (b), and (c) in

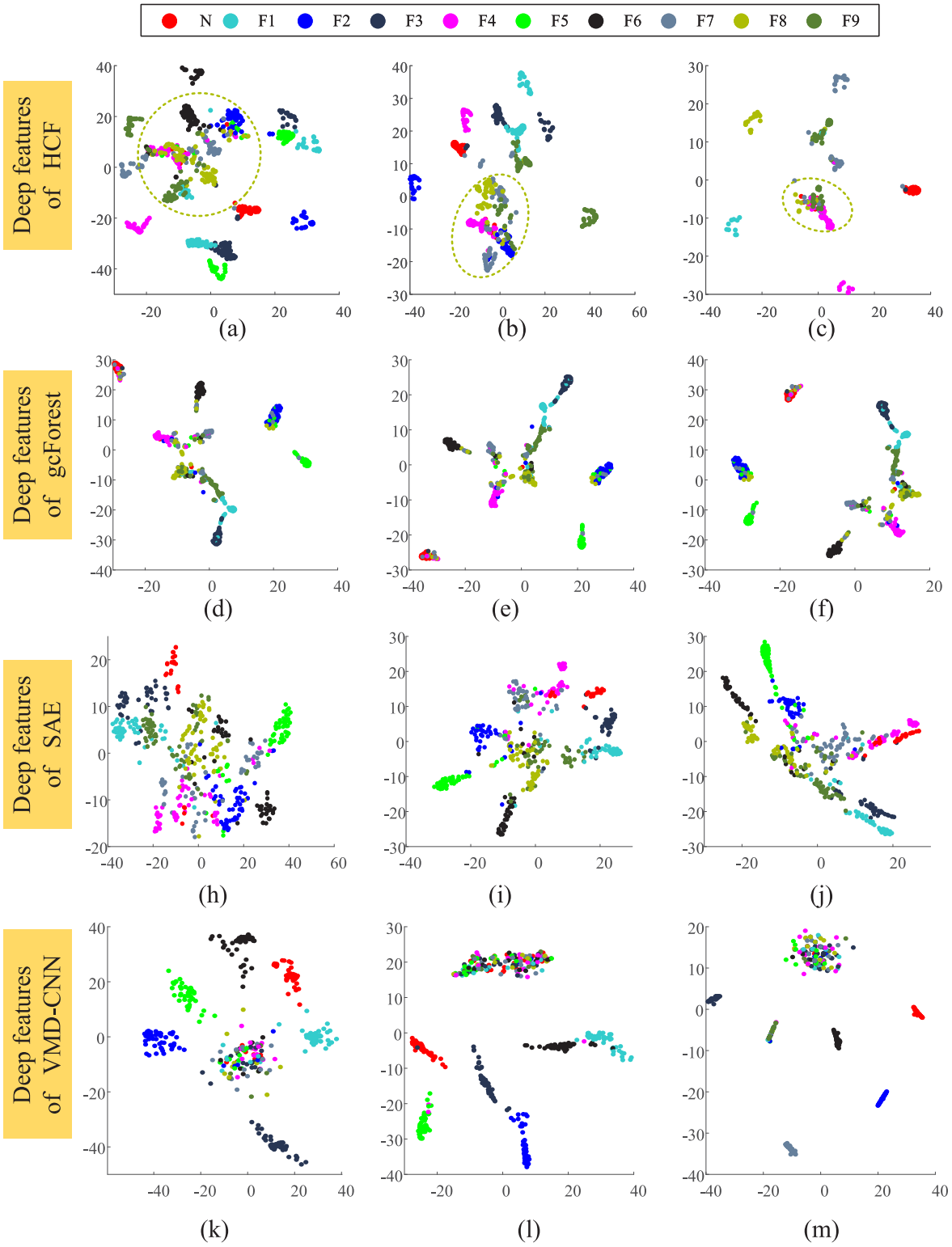

Fig. 15. Two-dimensional visualizations of the different features: (a),(b),(c): Deep features learned by the proposed HCF; (d),(e),(f): Deep features learned by the gcForest; (h),(i),(j): Deep features learned by the SAE; (k),(l),(m): Deep features learned by the VMD-CNN.

Fig. 15, it can be found that each layer of HCF outputs samples outside the highly overlapping area, and the feature distribution of the results of the next layer is significantly optimized. That is, each layer of HCF can focus on solving the problem that is too challenging for the previous layer to handle. It is the advantage of the hierarchical output. In contrast, although the deep features of gcForest, SAE, and VMD-CNN are also obviously optimized, the highly overlapping area has not been resolved. This also shows that there are differences in the classification difficulty between different classes.

Table 4
Description of experimental methods.

Methods	Architecture
DF	2 forests per layer, each forest consists of 50 decision trees. Max depth is 7.
VMD-CNN	the LeNet-5 classic network model with two convolution layers, two Max-Pooling layers and a Fully-connected layer. The hyper parameters of the VMD-CNN are consistent with that in [34]
SAE	(input dimension)-256-128-64-32-16-(output dimension).
HCF	2 forests per layer, each forest consists of 50 decision trees. Max depth is 7.

Table 5
Comparison of test accuracy and training time cost.

Data	Methods	N	F1	F2	F3	F4	F5	F6	F7	F8	F9	Mean	Traing time(s)
C	DF	85.29	98.43	99.29	**99.86**	\	\	\	\	\	\	**95.72**	17.01
	SAE	81.71	99.00	**100.00**	92.71	\	\	\	\	\	\	93.36	36.87
	VMD-CNN	94.29	**100.00**	97.14	82.86	\	\	\	\	\	\	93.57	41.62
	HCF	**95.14**	97.57	91.43	99.29	\	\	\	\	\	\	95.85	14.08
D	DF	86.67	81.67	84.50	**96.83**	80.83	93.50	89.33	**87.33**	\	\	87.58	62.91
	SAE	80.83	83.67	86.17	95.33	64.00	85.83	85.00	70.67	\	\	81.44	63.53
	VMD-CNN	83.54	93.23	91.12	93.42	64.42	81.41	66.66	73.33	\	\	80.89	79.55
	HCF	**92.17**	**92.17**	**95.00**	95.33	**95.33**	**96.67**	**97.33**	81.67	\	\	**93.21**	42.86
E	DF	78.20	91.20	86.80	**95.80**	78.60	81.80	86.40	71.80	82.40	68.60	83.83	87.29
	SAE	80.40	**97.40**	**94.40**	89.80	56.00	72.20	81.80	**77.20**	80.60	55.80	81.15	144.83
	VMD-CNN	81.36	90.50	86.02	91.64	78.14	91.46	79.62	63.52	68.26	79.20	80.97	120.55
	HCF	**88.40**	96.80	78.20	92.80	**92.00**	**92.60**	**94.80**	74.80	**95.40**	**85.20**	**89.10**	63.42

5. Conclusion

To achieve reliable and accurate fault diagnosis for mechanical systems, a hierarchical cascade deep learning method is proposed in this paper. In the existing fault diagnosis methods, the classification of all classes is completed simultaneously, and test accuracy is used to evaluate the confidence of the model. However, the accuracy of each class cannot be accurately predicted through an average test accuracy, which influences the fault location. And the efficiency of the deep learning models needs further improvement. To solve this problem, this paper proposes the HCF model which can output hierarchically according to a preset confidence. This preset confidence is used to calculate an OAT for each layer based on the Gaussian complexity theory. The OAT filters the results of each layer so that the output of each layer meets the confidence. The research shows that the proposed HCF model can effectively evaluate the fault recognition performance for each class. The sample space is gradually simplified by hierarchical output mechanism, and the accuracy of difficult classes is improved. By outputting part of samples in advance, HCF is significantly more efficient than other deep learning methods.

Although this method has made progress in accuracy and efficiency, there are still some problems that need to be further explored and optimized. In the confidence evaluation algorithm, all samples are treated equally, however, the length of the sample for fault diagnosis based on vibration data is important. In addition, it is worth studying to improve the recognition rate of concurrent faults by combining outputs based on HCF.

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.

References

[1] Abid A, Khan MT, Lang H, De Silva CW. Adaptive system identification and severity index-based fault diagnosis in motors. IEEE/ASME Trans Mechatronics 2020;24(4):1628–39. http://dx.doi.org/10.1109/TMECH.2019.2917749.

[2] Zhang Y, Ji J. Intelligent fault diagnosis of a reciprocating compressor using mode isolation-convolutional deep belief networks. IEEE/ASME Trans Mechatronics 2020;1. http://dx.doi.org/10.1109/TMECH.2020.3027912.

[3] Zhong JH, Wong PK, Yang ZX. Fault diagnosis of rotating machinery based on multiple probabilistic classifiers. Mech Syst Signal Process 2018. http://dx.doi.org/10.1016/j.ymssp.2018.02.009.

[4] Chen J, Wang J, Zhu J, Lee TH, De Silva C. Unsupervised cross-domain fault diagnosis using feature representation alignment networks for rotating machinery. IEEE/ASME Trans Mechatronics 2020;1. http://dx.doi.org/10.1109/TMECH.2020.3046277.

[5] Wang J, Fu P, Zhang L, Gao RX, Zhao R. Multilevel information fusion for induction motor fault diagnosis. IEEE/ASME Trans Mechatronics 2019;24(5):2139–50. http://dx.doi.org/10.1109/TMECH.2019.2928967.

[6] Cheng Y, Zhu H, Wu J, Shao X. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks. 2019;15(2):987–97. http://dx.doi.org/10.1109/TII.2018.2866549.

[7] Ghalamchi B, Jia Z, Mueller MW. Real-time vibration-based propeller fault diagnosis for multicopters. IEEE/ASME Trans Mechatronics 2020;25(1):395–405. http://dx.doi.org/10.1109/TMECH.2019.2947250.

[8] Liang P, Deng C, Wu J, Yang Z, Zhu J, Zhang Z. Compound Fault Diagnosis of Gearboxes via Multi-label Convolutional Neural Network and Wavelet Transform. Comput Ind 2019. http://dx.doi.org/10.1016/j.compind.2019.103132.

[9] Wang Z, Yao L, Chen G, Ding J. Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals. ISA Trans 2021;114:470–84. http://dx.doi.org/10.1016/j.isatra.2020.12.054, URL https://www.sciencedirect.com/science/article/pii/S0019057820305735.

[10] Kim M, Ko JU, Lee J, Youn BD, Jung JH, Sun KH. A domain adaptation with semantic clustering (DASC) method for fault diagnosis of rotating machinery. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.03.002, URL https://www.sciencedirect.com/science/article/pii/S0019057821001312.

[11] Liu Z, Peng D, Zuo MJ, Xia J, Qin Y. Improved Hilbert–Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.07.011, URL https://www.sciencedirect.com/science/article/pii/S0019057821003785.

[12] Wang B, Zhang X, Xing S, Sun C, Chen X. Sparse representation theory for support vector machine kernel function selection and its application in high-speed bearing fault diagnosis. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.01.060, URL https://www.sciencedirect.com/science/article/pii/S0019057821000707.

[13] Zheng J, Pan H, Tong J, Liu Q. Generalized refined composite multiscale fuzzy entropy and multi-cluster feature selection based intelligent fault diagnosis of rolling bearing. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.05.042, URL https://www.sciencedirect.com/science/article/pii/S0019057821003050.

[14] Martin-Diaz I, Morinigo-Sotelo D, Duque-Perez O, Osornio-Rios RA, Romero-Troncoso RJ. Hybrid algorithmic approach oriented to incipient rotor fault diagnosis on induction motors. ISA Trans 2018;80:427–38. http://dx.doi.org/10.1016/j.isatra.2018.07.033, URL https://www.sciencedirect.com/science/article/pii/S0019057818302891.

[15] Liang S, Zhang S, Huang Y, Zheng X, Cheng J, Wu S. Data-driven fault diagnosis of FW-UAVs with consideration of multiple operation conditions. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.07.043, URL https://www.sciencedirect.com/science/article/pii/S0019057821004109.

[16] Wei Y, Yang Y, Xu M, Huang W. Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest. ISA Trans 2021;109:340–51. http://dx.doi.org/10.1016/j.isatra.2020.10.028, URL https://www.sciencedirect.com/science/article/pii/S0019057820304237.

[17] Zhang PB, Yang ZX. A new learning paradigm for random vector functional-link network: RVFL+. Neural Netw 2020. http://dx.doi.org/10.1016/j.neunet.2019.09.039, arXiv:1708.08282.

[18] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012. http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0001284.

[19] Zhang W, Li X, Ding Q. Deep residual learning-based fault diagnosis method for rotating machinery. ISA Trans 2019;95:295–305. http://dx.doi.org/10.1016/j.isatra.2018.12.025, URL https://www.sciencedirect.com/science/article/pii/S0019057818305202.

[20] Shao H, Jiang H, Wang F, Wang Y. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. ISA Trans 2017;69:187–201. http://dx.doi.org/10.1016/j.isatra.2017.03.017, URL https://www.sciencedirect.com/science/article/pii/S0019057817303592.

[21] Xu Z, Li C, Yang Y. Fault diagnosis of rolling bearings using an improved multi-scale convolutional neural network with feature attention mechanism. ISA Trans 2021;110:379–93. http://dx.doi.org/10.1016/j.isatra.2020.10.054, URL https://www.sciencedirect.com/science/article/pii/S0019057820304572.

[22] Fenton N, Wang W. Risk and confidence analysis for fuzzy multicriteria decision making. Knowl-Based Syst 2006;19(6):430–7. http://dx.doi.org/10.1016/j.knosys.2006.03.002, URL https://www.sciencedirect.com/science/article/pii/S0950705106000694.

[23] Milne R, Nicol C, Travé-Massuyès L. Tiger with model based diagnosis: initial deployment. Knowl-Based Syst 2001;14(3):213–22. http://dx.doi.org/10.1016/S0950-7051(01)00099-5, URL https://www.sciencedirect.com/science/article/pii/S0950705101000995.

[24] Liu Z-h, Lu B-l, Wei H-l, Chen L, Li X-h. A Stacked Auto-Encoder Based Partial Adversarial Domain Adaptation Model for Intelligent Fault Diagnosis of Rotating Machines. IEEE Trans Ind Inf 2020;3203(MMD). http://dx.doi.org/10.1109/TII.2020.3045002.

[25] Zhang C, Xu L, Li X, Wang H. A Method of Fault Diagnosis for Rotary Equipment Based on Deep Learning. In: Proceedings - 2018 Prognostics and system health management conference, PHM-Chongqing 2018. IEEE; 2019, p. 958–62. http://dx.doi.org/10.1109/PHM-Chongqing.2018.00171.

[26] Dongzhu Z, Hua Z, Shiqiang D, Yafei S. Aero-engine Bearing Fault Diagnosis Based on Deep Neural Networks. In: ICMAE 2020 - 2020 11th International conference on mechanical and aerospace engineering, no. 978. 2020, p. 145–9. http://dx.doi.org/10.1109/ICMAE50897.2020.9178886.

[27] Zhang W, Li X, Ma H, Luo Z, Li X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. Knowl-Based Syst 2021;213:106679. http://dx.doi.org/10.1016/j.knosys.2020.106679, URL https://www.sciencedirect.com/science/article/pii/S095070512030808X.

[28] Karabadji NEI, Seridi H, Bousetouane F, Dhifli W, Aridhi S. An evolutionary scheme for decision tree construction. Knowl-Based Syst 2017;119:166–77. http://dx.doi.org/10.1016/j.knosys.2016.12.011, URL https://www.sciencedirect.com/science/article/pii/S0950705116305056.

[29] Ji X, Yang B, Tang Q. Seabed sediment classification using multibeam backscatter data based on the selecting optimal random forest model. Appl Acoust 2020. http://dx.doi.org/10.1016/j.apacoust.2020.107387.

[30] Zhou ZH, Feng J. Deep forest. Natl Sci Rev 2019. http://dx.doi.org/10.1093/nsr/nwy108, arXiv:1702.08835.

[31] Li X, Jiang H, Wang R, Niu M. Rolling bearing fault diagnosis using optimal ensemble deep transfer network. Knowl-Based Syst 2021;213:106695. http://dx.doi.org/10.1016/j.knosys.2020.106695, URL https://www.sciencedirect.com/science/article/pii/S0950705120308248.

[32] Chen H, Li C, Yang W, Liu J, An X, Zhao Y. Deep balanced cascade forest: An novel fault diagnosis method for data imbalance. ISA Trans 2021. http://dx.doi.org/10.1016/j.isatra.2021.07.031, URL https://www.sciencedirect.com/science/article/pii/S0019057821003980.

[33] Liu X, Huang H, Xiang J. A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine. Knowl-Based Syst 2020;195:105653. http://dx.doi.org/10.1016/j.knosys.2020.105653, URL https://www.sciencedirect.com/science/article/pii/S0950705120301039.

[34] Hou Q, Wang J, Shen Y. Multiple sensors fault diagnosis for rolling bearing based on variational mode decomposition and convolutional neural networks. In: 2020 11th International conference on prognostics and system health management (PHM-2020 Jinan). 2020, p. 450–5. http://dx.doi.org/10.1109/PHM-Jinan48558.2020.00087.

[35] Quinlan JR. Bagging, boosting, and C4.5. In: Proceedings of the national conference on artificial intelligence. 1996.

[36] Bartlett PL, Maass W. Vapnik-Chervonenkis Dimension of Neural Nets. In: The handbook of brain theory and neural networks. 2002.

[37] Bartlett PL, Mendelson S. Rademacher and Gaussian complexities: Risk bounds and structural results. J Mach Learn Res 2003. http://dx.doi.org/10.1162/153244303321897690.

[38] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014. http://dx.doi.org/10.1109/TSP.2013.2288675.

[39] Luo M, Li C, Zhang X, Li R, An X. Compound feature selection and parameter optimization of ELM for fault diagnosis of rolling element bearings. ISA Trans 2016. http://dx.doi.org/10.1016/j.isatra.2016.08.022.

[40] Liu C, Zhu L, Ni C. Chatter detection in milling process based on VMD and energy entropy. Mech Syst Signal Process 2018;105:169–82. http://dx.doi.org/10.1016/j.ymssp.2017.11.046, URL https://www.sciencedirect.com/science/article/pii/S0888327017306349.

[41] Wen X, Lu G, Liu J, Yan P. Graph modeling of singular values for early fault detection and diagnosis of rolling element bearings. Mech Syst Signal Process 2020;145:106956. http://dx.doi.org/10.1016/j.ymssp.2020.106956, URL https://www.sciencedirect.com/science/article/pii/S0888327020303423.