Measurement 214 (2023) 112806

Contents lists available at ScienceDirect

Measurement

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

An ensemble method with DenseNet and evidential reasoning rule for machinery fault diagnosis under imbalanced condition

Gang Wang [a,b,c], Yanan Zhang [a], Feng Zhang [a], Zhangjun Wu [a,b,c,*]

[a] School of Management, Hefei University of Technology, Hefei, Anhui, People's Republic of China
[b] Key Laboratory of Process Optimization and Intelligent Decision-Making (Hefei University of Technology), Ministry of Education, Hefei, Anhui, People's Republic of China
[c] Ministry of Education Engineering Research Center for Intelligent Decision-Making & Information System Technologies, Hefei 230009, People's Republic of China

ARTICLE INFO

Keywords:
Fault Diagnosis
Deep Learning
Imbalance Learning
Evidential Reasoning Rule
DenseNet

ABSTRACT

Fault diagnosis is of significant importance for intelligent manufacturing as it can increase production efficiency and decrease the uncertain breakdown risk of machines. Previous studies have extensively utilized different types of shallow statistical features as well as deep representation features to comprehensively describe the fault information for achieving better performances of fault diagnosis. However, making full use of the combination of such shallow statistical features and deep representation features under the class-imbalance situation in real-world applications becomes a challenge in current research of fault diagnosis. To remedy the above issue, an Imbalanced Ensemble Method with DenseNet and Evidential Reasoning Rule, namely IEMD-ER, is proposed to incorporate both human experience and machine wisdom for machinery fault diagnosis under the class-imbalance situation. To this end, the shallow statistical features with human experience are extracted by several signal processing techniques, while the modified DenseNet is adopted to extract the deep representation features with machine wisdom. Based on these features, multiple diverse base classifiers are produced by leveraging the importance of different features. The outputs of each base classifier are adaptively fused using the Evidential Reasoning Rule as the final fault diagnosis results. Extensive experiments are conducted on the real-world bearing datasets collected by Paderborn University to validate the proposed method. The experimental results proved the superiority of the proposed IEMD-ER.

1. Introduction

Since uncertain machine breakdown may result in serious system failure and additional financial losses, timely and intelligent fault diagnosis techniques have always been highly demanded during the maintenance of machines [1,2]. In the past decades, with the increasing accessibility of massive industrial big data, machine learning-based fault diagnosis has become a significant competitive differentiator and has been proven effective in many applications of fault diagnosis, such as bearing, gearbox, and electric motor [3–5].

Generally, machine learning-based fault diagnosis techniques consist of two essential processes, i.e., feature extraction and fault pattern recognition. In the feature extraction stage, the features extracted from different domains, including the time domain, frequency domain, and time–frequency domain, are considered as shallow statistical features that have been reported to be effective in fault diagnosis [6–8]. For example, Ciabattoni et al. utilized the statistical spectral analysis to extract time-domain features for fault diagnosis and empirically proved that the time-domain features are reliable indicators to reflect the fault types [9]. Wang et al. extracted the fault-induced features by a hierarchical frequency-domain feature extraction method and proved their effectiveness for representing fault information [10]. Li et al. used a time–frequency approach based on a generalized synchrosqueezing transform to map the signals to time–frequency picture features for the diagnosis of the gearbox, and their experimental study demonstrated the efficiency of such time–frequency domain features in fault diagnosis [11]. However, these shallow statistical features only contain limited shallow information about the fault pattern, which is difficult to describe the internal information about the fault patterns especially when the complexity of the machinery systems is increasing [12]. Therefore, in-depth mining should be carried out to reveal the complex internal information in the signal [13]. Recently, deep learning methods

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China.
E-mail address: wuzhangjun@hfut.edu.cn (Z. Wu).

https://doi.org/10.1016/j.measurement.2023.112806
Received 23 June 2022; Received in revised form 22 March 2023; Accepted 24 March 2023
Available online 3 April 2023

have proved their superiority in feature representation learning, as they could extract deep representation features from the collected data automatically with the layered network. In previous studies, various deep learning methods, for instance, Stacked Autoencoder (SAE), Convolutional Neural Network (CNN), and Deep Belief Network (DBN) have been adopted for deep representation feature extraction in fault diagnosis [14–16]. For instance, Hyunseok et al. employed the Denoising SAE method to diagnose the fault of rotating machinery online, while its usefulness was verified by three case studies [17]. Yi et al. modified the original logistic units of DBN and applied it to the diagnosis for planetary gearboxes, the results show it has superior performance in fault diagnosis [18]. Wang et al. adopted the CNN-based bearing fault diagnosis method with SE blocks, and the experimental results proved the superior performance of their method [19]. Among these mentioned deep learning methods, CNNs are more suitable for signal data processing due to the convolutional and pooling function are computationally efficient during exhibiting the periodic characteristics contained in the signals [20]. With the architectures of CNN becoming increasingly deep for capturing more complex information, the vanishing-gradient phenomena often occur in traditional network topologies of CNN [21]. As an efficient solution to such phenomena, DenseNet alternatively uses the dense connection between different layers to enhance the information flow and extract high-quality features even if the depth of the network is excessively deep [22]. Considering that the structure of DenseNet is suitable to balance the quality of the features and the depth of the network, a modified DenseNet with 1-D kernels is utilized in this paper to extract deep internal information from the raw 1-D signal data. Although the shallow statistical and deep representation features can indicate the fault patterns at various levels, how to comprehensively use them to enhance the fault diagnosis performance is still a challenging issue especially when they contain redundant components [23].

In the stage of fault pattern recognition, different intelligent fault diagnosis approaches based on machine learning, such as Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Artificial Neural Networks (ANN), have been successfully applied and achieved superiority performance in previous studies [24–26]. For instance, Qin et al. developed a weighted ELM that can quickly and accurately detect the fault status by considering feature reliability [27]. Rauber et al. constructed a SVM-based fault diagnosis technique with the features vector including, statistical parameters, and envelope features, to distinguish the fault types of rotating machinery [28]. Mojtaba et al. integrated the ANN and discrete wavelet transform to establish a fast diagnosing system for multifunctional spoilers with higher reliability [29]. However, in the real-world application of fault diagnosis, the samples in the normal state are often far more than the samples in fault states due to that the mechanical systems work in the normal state at most times [30]. Consequently, the conventional machine learning methods will inevitably suffer the class imbalance problem. In previous studies, sampling methods, cost-sensitive learning methods, and ensemble methods are three frequently used methods to address the class imbalance problem [31–33]. However, sampling methods may lose some sample information during the sampling process [34]. And it is also difficult to determine the cost weight when employing cost-sensitive learning methods [35]. Thus, ensemble methods become a popular choice to solve the class-imbalance problem due to their generalization ability and stability. For example, Wang et al. employed an ensemble method with a heterogeneous structure for identifying the fault types of planetary gearbox and the results showed the fault diagnosis accuracy has been significantly improved [36]. Li et al. identified the faults by assembling different classification results based on different feature extractors and their effectiveness and applicability were validated by the experimental results [37]. Zhang et al. adopted an ensemble learning-based fault diagnosis approach which contains several incremental support vector machines, engineering tests, and experimental results used to validate their method [38]. In ensemble learning, generating base learners and aggregating their outputs are two main fundamental steps. The former step can be established by the instance partitioning approaches or feature partitioning approaches. The instance partitioning approaches, such as Bagging and Boosting, may perform poorly when the features are high-dimensional [39]. Instead, the feature partitioning approaches perform better under the situation of the high-dimensional features [40]. Random Subspace (RS), one of the representative feature partitioning-based approaches, trains each base learner on the feature subset randomly selected from the raw features. To reduce the negative influence of redundant features being randomly selected into the same feature subset, it is necessary to take some prior knowledge of the data, such as the importance of features, into account during feature partitioning. As an efficient feature weighting method, Random Forest (RF) is suitable to assign the importance to different features [41]. Besides, to aggregate the output of different base learners, Majority Voting and Average are two frequently used strategies in previous studies, nevertheless the relative weight between different base learners and their inherent properties have not been fully exploited [42]. Different from the traditional aggregating strategies, the Evidential Reasoning (ER) Rule can incorporate the weight and reliability simultaneously when aggerating different prediction results, and thus improves the performance of fault diagnosis [43,44]. Therefore, in this paper, the RF is introduced to the feature sampling process to assign different weights to features and the ER rule is introduced for aggregating the final diagnosis result.

Inspired by the above-mentioned analysis, an Imbalanced Ensemble Method with DenseNet and Evidential Reasoning Rule, i.e., IEMD-ER, is proposed in this paper. Firstly, different features including time-domain features, frequency-domain features, time–frequency domain features extracted by statistical methods, as well as deep representation features obtained using DenseNet are the input of the proposed method. Secondly, an improved ensemble method composed of random subspace technique and RF is proposed which can choose high-quality feature subspaces and produce superiority base classifiers to defeat the class imbalanced problem. Finally, to get robust diagnosis results, the fault diagnosis results are generated by combining the outputs of different base classifiers with the ER rule, meanwhile, their reliabilities and weights are considered simultaneously. Experiments were conducted on the bearing datasets provided by Paderborn University, and the experimental results show the superior performance of the proposed method to other intelligent fault diagnosis techniques.

The main contributions of this study are summarized as follows:

(1) An enhanced and robust ensemble learning-based framework that can enhance the machinery fault diagnosis performance is presented. In the framework, the feature fusion of different domain features and class-imbalance problems are well-considered simultaneously in this framework.
(2) An imbalanced ensemble method with DenseNet and ER rule is proposed in this study. In this method, the shallow statistical features, such as time domain features, frequency domain features, and time–frequency domain features, as well as deep representation features are obtained and assigned weights first. Multiple diverse base classifiers are generated based on the feature subspaces generated from weighted features. Finally, the output of each base classifier is adaptively aggregated by the ER Rule to obtain the final diagnosis results.
(3) The bearing vibration dataset obtained from Paderborn University is adopted for the evaluation of the proposed IEMD-ER, and the empirical study based on the dataset proves that the proposed IEMD-ER is an efficient fault diagnosis method compared with other commonly used methods.

The remainder of this study is organized as follows. Section 2 gives details information of the framework. Section 3 will give the experiment procedure in detail. The experimental results and the discussion are reported in Section 4. Finally, Section 5 gives the conclusion and

indicates future works.

2. Proposed intelligent fault diagnosis method

2.1. Framework of IEMD-ER

Timely and intelligent fault diagnosis is of crucial importance to keep rotating machinery operating reliability and security. Previous studies have extensively explored machine learning-based fault diagnosis methods. However, most of the studies only adopted shallow statistics features or deep representation features, which may lose the internal feature learning ability or the domain knowledge. Furthermore, in real-world applications, machines usually operate in the normal status, which is to say, the normal samples are generally more than faulty samples in general, which will cause the class-imbalance problem. Therefore, a novel IEMD-ER method for machinery fault diagnosis is proposed, in which the shallow statistical features and deep representation features are obtained for multiple diverse base classifiers training meanwhile the different importance of the features are being considered as well. To get the final diagnosis results, the outputs produced by these base classifiers are then adaptively fused using the Evidential Reasoning Rule. Fig. 1 displays the framework of the proposed IEMD-ER, and the components of the proposed method are presented as follows:

(1) Data Acquisition: Vibration signal data with various fault types of machinery, including three inner fault types and two outer fault types are collected and divided into samples respectively.
(2) Feature extraction: signal processing methods are used to extract shallow statistical features which can describe the different fault patterns in different domains, including time-domain, frequency-domain, and time–frequency domain. Meanwhile, a modified DenseNet with 1-D kernels is adopted to obtain deep representation features from signal data.
(3) Model Construction: An improved RS approach is used to generate different feature subsets for training base classifiers, which incorporates the SMOTE sampling strategy and considers the importance of different features simultaneously [45]. Then, the outputs of different base classifiers are combined by the ER rule where the weight and reliability of each base classifier are considered simultaneously during the fusing process.

2.2. Feature extraction

The collected signal data can be expressed in different aspects of the fault diagnosis field. To fully reveal the fault patterns, the time-domain, frequency-domain, and time–frequency domain features considered as shallow statistical features are extracted firstly. More specifically, the features in the time domain and frequency domain are acquired using numerical formulations, meanwhile, the features in the time–frequency domain are obtained using WPT and EEMD. To better represent the deep fault information, the deep representation features are extracted by utilizing the modified DenseNet with 1-D kernels to represent the deep internal information of the signal data.

2.2.1. Feature extraction of shallow statistical

The shallow statistical features can be categorized into time-domain features, frequency-domain features, and time–frequency domain features. The details of the shallow statistical feature extraction are given below.

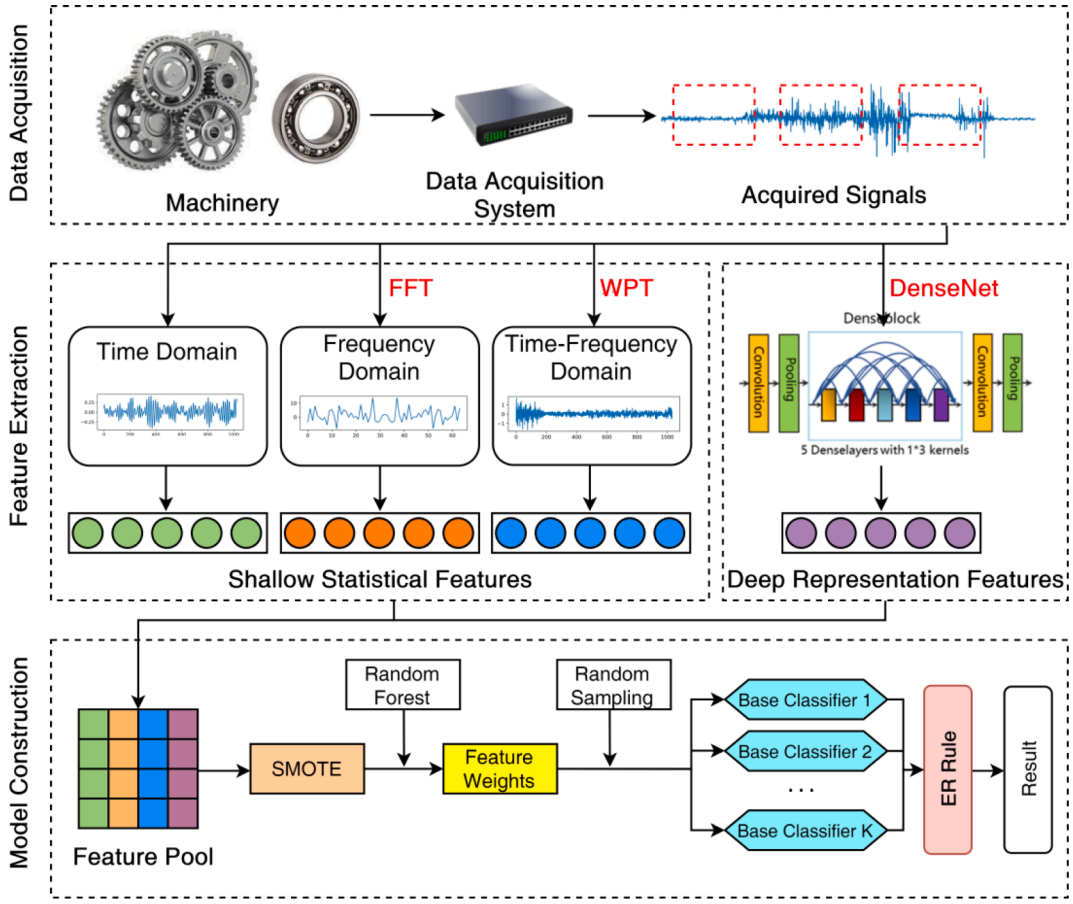

Fig. 1. Framework of the proposed method.

(1) **Time-domain features**

As commonly adopted features, Time-domain features obtained by numerical formulations are sensitive to early faults and have been proven to be useful to represent machinery faults. Therefore, some classical time-domain features, including mean, maximum (MAX), minimum (MIN), root mean square (RMS), skewness value (SV), kurtosis value (KV), and mean of absolute amplitude (MAA) are obtained from the raw vibration signals.

(2) **Frequency-domain features**

Different from time-domain features, frequency-domain features show the distribution characteristics of the signal and can discover some useful information from different aspects. In previous studies, the vibration signal can transform from the time domain into the frequency domain with the help of the fast Fourier transform (FFT). On this basis, some commonly used frequency-domain features are extracted, including the maximum of frequency (MAXF), skewness value of frequency (SVF), kurtosis value of frequency (KVF), root variance of frequency (RVF), and mean of frequency (MEANF), etc.

(3) **Time-frequency domain features**

Time-frequency domain features provide better-localized defect information and can analyze the nonstationary signal data both in the time and frequency views, which is more comprehensive than time-domain and frequency-domain features. In previous studies, ensemble empirical mode decomposition (EEMD) and wavelet package transforms (WPT) are commonly used signal processing techniques. Thus, EEMD and WPT are employed to extract the time–frequency features.

2.2.2. Feature extraction of deep representation

The shallow statistical features extracted from signal data by the mathematical formulates and signal processing techniques only capture the shallow information. The performance of these extracted features is affected by expert knowledge, which will perform uncertainly in fault diagnosis. Thus, it is important to take the deep information into account to fully describe the fault patterns. As mentioned before, CNNs are suitable to learn the deep representations from the vibration signal data because they can perform well on the periodic signal data. Thus, the CNNs are employed to extract the deep representation features in this paper.

Usually, the CNNs generate the output x_l of the l^{th} layer by using a set of nonlinear functions H to the outputs of the previous layer, which is given as:

$$x_l = H(x_{l-1}) \tag{1}$$

In detail, $H(\cdot)$ can be the Convolution operation, Pooling operation, rectified linear units (ReLU), or Batch Normalization (BN). For example, the convolution operation can be formulated as:

$$y_i = w^T x_{i:i+k-1} + b \tag{2}$$

where w denotes the convolutional kernel vector, $x_{i:i+k-1}$ means the subsignal data with length k of the input signal data $\times$ from i to $i+k$-1, and b is the bias term.

The deep representation features, i.e., the output of the last hidden layer, can be extracted through consecutive nonlinear operations. However, some details of the original data tend to disappear in the last layer of the network when the depth of the CNN network goes deeper and deeper. To address the issue mentioned before, DenseNet can strengthen the information flow by connecting each layer to all the previous layers.

As a variation of CNN, DenseNet can strengthen the feature propagation and increase the training efficiency with dense connection and fewer parameters. Thus, deep and high-quality features can be obtained with the special architecture. The input of l^{th} layer is the output from all preceding layers,$\mathrm{x}_0, \ldots, x_l$:

$$x_l = H_l([\mathrm{x}_0, \ldots, x_l]) \tag{3}$$

The input of the DenseNet is a segment of the raw signals, which is illustrated in Fig. 2.

The first convolutional layer is used to learn the initial representation of the original signal data without any transformation. In this paper, the standard DenseNet is modified for getting better performance in the features extraction, 1) To process the raw vibration signal data directly, 1-D convolutional kernels are used directly instead of 2-D convolutional kernels. 2) The convolutional kernels of previous layers of the network are wide meanwhile the followings in different blocks are small. Using wide convolutional kernels and multilayer small convolutional kernels can deepen the network and help to acquire better representations of the input signals, thereby improving the network's performance. To enhance the generalization ability, batch normalizations and dropout are used, which are applied after the convolutional operations. The deep representation features can be obtained using several consecutive Denseblocks. Each Denseblcok has 5 Denselayers which consist of several convolutional and batch norm layers. Finally, the deep representation features are obtained by gathering the outputs from the hidden layer of the modified DenseNet.

2.3. Generation of base classifiers

In real-world applications, the normal class samples are usually far more than fault class samples. Directly using the combination of shallow statistical and deep representation features will cause high-dimension problems to a certain extent, and another drawback is typical classifiers are always biased toward the major class based on these samples. To tackle this problem, the improved RS method is adopted due to it can select different low-dimensional feature sets instead of using all the features to construct base classifiers, which can enhance the diversity and robustness of the method. However, the base classifiers constructed based on the randomly selected feature subsets may have poor performance in the fault diagnosis. Thus, taking the importance of different features and minor class samples into account during the process of feature selection is essential, which can make the important features are easily selected than other redundant features. Therefore, an improved semi-random base classifier generation method is proposed in this study, making important features easily selected and guaranteeing the diversity and accuracy of the generated base classifiers. For dealing with the imbalanced dataset, we incorporate the SMOTE into the base classifier generation process to balance the dataset before determining the importance of different features.

The SMOTE algorithm is an improved instance sampling method, which can produce new samples based on the characteristic of samples belonging to the minor classes. The main idea of SMOTE is to select the neighbors of minor samples, and then generate new samples from the selected minor class samples [46]. For the sample x belonging to the minor class, the SMOTE algorithm first selects a defined number of nearby neighbors of x, and generates new samples based on these selected neighbors then, which is given in equation (4):

$$x_{new} = x + rand(0,1) \times (\tilde{x} - x) \tag{4}$$

where $\tilde{x}$ means the neighbors of the sample x.

After balancing the dataset, a predefined number of feature subspaces, which are also named feature subsets, can be selected to construct the base classifiers. If the features are selected arbitrarily, it is risky to assemble some poor base classifiers, which will have a negative influence on the final diagnosis result. Thus, some prior knowledge from the features should be considered which can be used to guide the feature subsets generation during the construction of the ensemble model. In

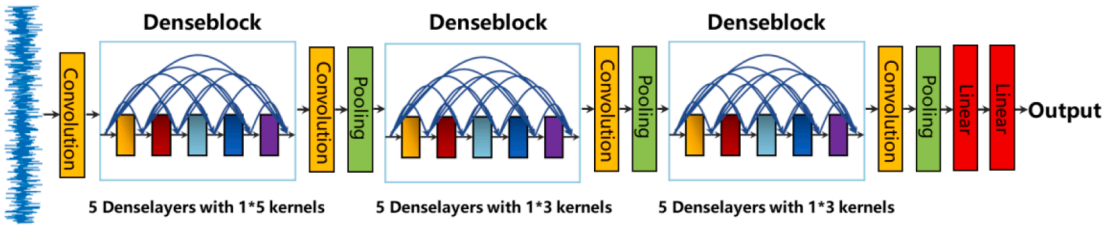

Fig. 2. Structure of the modified DenseNet.

other words, it is important to find a way to assign meaningful weights to different features. Therefore, in the improved semi-random RS method, the importance of features is determined by calculating the contribution of each base classifier when adding noise to a feature in the training process of the RF [47]. Then, a predefined number of feature subspaces is produced by considering the weight of features. Finally, based on each feature subset, the decision tree is employed as the base classifier to give the prediction results. Meanwhile, the improved semi-random RS method can ensure that the features with large weights have a high probability of being selected. Table 1 gives the pseudo-code of the feature importance calculation and the base classifier generation procedure.

2.4. Combination of results using Evidential Reasoning rule

After the training of base classifiers with different feature subspaces, the proposed method tries to combine the prediction results of each base classifier to get the diagnosis results. Instead of using a majority voting strategy or average strategy to fuse the perdition results, which cannot take the relative importance between different classifiers and their intrinsic characteristics into account, the proposed IEMD-ER employed an adaptive ER rule to fuse different prediction results of the base classifiers.

Table 1
Feature importance calculation and base classifiers generation.

Feature importance calculation and base classifiers generation
Input: Training set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$; Random subspace rate:$r$ OOB_t: out-of-bag data using bootstrap of tree t; Features $F = \{f_1, f_2, \ldots, f_m\}$; a predefined number of base classifiers as K; feature subset fs_k; **Process**: 1. create T trees based on D; 2. **for** feature f_m in F: 3. **for** a tree in T: 4. calculate the error $errOOB1_{t,f_m}$ of this tree using OOB_t; 5. add some noise to the feature; 6. calculate the error $errOOB2_{t,f_m}$ of this tree using OOB_t; 7. calculate the importance of f_m in this tree by ($errOOB2_{t,f_m} - errOOB1_{t,f_m}$); 8. **end for** 9. normalizing the importance of feature $w_m = \frac{\sum_{t=1}^{\lvert T\rvert} \lvert errOOB2_{t,f_m} - errOOB1_{t,f_m}\rvert}{\lvert T\rvert}$ 10. **end for** 11. get the importance of all features $w = \{w_1, w_2, \ldots, w_m\}$ 12. **for** k in K: 13. features number $= round(m \times r)$ 14. while $\lvert fs_k\rvert <$ *featurenumber* do 15. generate a random number *random* between 0 and 1; 16. set $sum = 0$; 17. **for** p in M: 18. $sum = sum + w_p$ 19. **if** $sum \geq random$ and $p-th$ feature $\cap f_k = \varnothing$: 20. add the p-th feature to f_k; 21. $w_p = min(w)$ 22. **end if** 23. **end for** 25. training base classifier C_k on the training set with generated feature subset fs_k; 26. **end for** **Output**: Base classifiers $C = \{C_1, C_2, \ldots, C_K\}$.

The ER rule is used to fuse the evidence or information from the perspective of probability, which is developed based on the Dempster-Shafer evidence theory and has been widely employed in several scenarios, such as decision making and fault diagnosis [39,48]. In machinery fault diagnosis, the prediction result of a base classifier is considered as a piece of evidence. And each prediction result is generated by one feature subset sampled independently with replacement, which can meet the requirement of the ER rule. Thus, ER rule is adopted as the results fusion strategy in this paper. To fuse the different prediction results, the weight and reliability of each base classifier should be defined in advance. Based on the concept of ER rule, the weight determines the relative importance among different evidence, and reliability shows the intrinsic characteristics of the information source. The accuracy is used to judge the base classifier by comparing it with other base classifiers. Thereby, the average train accuracy of base classifiers is used to represent the relative importance of each base classifier in this paper. Furthermore, in human cognition, a man will be reliable if the man is similar to others. Thus, we denote the average similarity of a base classifier with other base classifiers as the reliability of the classifier. The calculation of weight and reliability are given in Equations (5) and (6).

$$weight_j = \frac{acc\_train_j}{\sum_{i=1}^{k} acc\_train_i}, j \neq i \tag{5}$$

$$reliability_j = \frac{\sum_{i=1, i\neq j}^{k-1} (y_j \cdot y_i)}{k-1}, j \neq i \tag{6}$$

where y_j is the predicted result of all the train samples of a base classifier j, $y_j \cdot y_i$ means y_j multiply y_i, k means the number of base classifiers. According to the definition above, the weight and reliability of each base classifier can be determined automatically without the need for any prior knowledge about the datasets. A brief description of the procedure of the ER rule is given as follows, and the detailed process of the ER rule can be found in [44].

As the base assumption of the ER rule, a set of mutually exclusive assumptions $\Omega = \{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ should be considered first. In this paper, $\Omega = \{Healthy; Inner\ Fault\ Type\ 1, 2, 3; Outer\ Fault\ Type\ 1, 2\}$ denotes different categories of faults. $P(\Omega)$ means the power set composed of all subsets of Ω. In ER rule, the ignorance is defined as the uncertainty of a situation, which means no probability can be assigned to a situation σ or any subsets of σ. The prediction result of the base classifier in probabilistic format can be regarded as a piece of evidence, which can be modeled by a distribution defined in the power set by:

$$e_j = \left\{ (\sigma, p_{\sigma,j}), \forall \sigma \subseteq \Omega, \sum_{\sigma \in \Omega} p_{\sigma,j} = 1 \right\} \tag{7}$$

where $(\sigma, p_{\theta,j})$ indicates that the σ is supported by evidence e_j with the probability of $p_{\sigma,j}$. In this paper, σ is directly related to a certain type of machinery's health condition. The weight of each base classifier is denoted as $weight_j$, and the reliability of each base classifier is considered as $reliabiliy_j$. The weight and reliability of each base classifier are considered as evidence simultaneously, the support degree of σ for e_j is identified as Equations (8) and (9):

$$\widetilde{m}_{\sigma,j} = \begin{cases} 0 & \sigma = \phi \\ \widetilde{w}_j p_{\sigma,j} & \sigma \subseteq \Omega, \sigma \neq \phi \\ 1 - \widetilde{w}_j & \sigma = P(\Omega) \end{cases} \tag{8}$$

$$\widetilde{w}_j = \frac{weight_j}{1 + weight_j - reliability_j} \tag{9}$$

where $\widetilde{w}_j$ means the new hybrid weight. The combined evidence $e(2)$, which can support σ with the probability of $p_{\sigma,e(2)}$, can be calculated using two pieces of evidence e_1 and e_2. The calculations are given as Equations (10) and (11):

$$p_{\sigma,e(2)} = \begin{cases} 0 & \sigma = \phi \\ \dfrac{\widehat{m}_{\sigma,e(2)}}{\sum_{D \subseteq \Omega} \widehat{m}_{D,e(2)}} & \sigma \subseteq \Omega, \sigma \neq \phi \end{cases} \tag{10}$$

$$\widehat{m}_{\sigma,e(2)} = \left[(1 - r_2)m_{\sigma,1} + (1 - r_1)m_{\sigma,2}\right] + \sum_{B \cap C = \sigma; B, C \subseteq \Omega} m_{B,1} m_{C,2} \forall \sigma \subseteq \Omega \tag{11}$$

Based on $B \cap C = \sigma$, the conjunctive belief $m_{B,1}$ and $m_{C,2}$ of σ is reason by using the latter term of equation (11). After obtaining the evidence $e(2)$ by using evidence e_1 and e_2, the fusion of the remaining pieces of evidence can be calculated by applying the same procedure, which is given in Equations (12) and (13):

$$\widehat{m}_{\sigma,e(j)} = \left[\left(1 - r_j\right)m_{\sigma,e(j-1)} + m_{P(\sigma),e(j-1)} m_{\sigma,j}\right] + \sum_{B \cap C = \sigma; B, C \subseteq \Omega} m_{B,e(j-1)} m_{C,j}, \forall \sigma \subseteq \Omega \tag{12}$$

$$\widehat{m}_{P(\Omega),e(j)} = \left(1 - r_j\right)m_{P(\Omega),e(j)} \tag{13}$$

where $\widehat{m}_{P(\Omega),e(j)}$ denotes the residual support of evidence $e(j)$ that can be assigned to a certain fault type σ.

3. Experiment setup

3.1. Experimental data

The real-world bearing datasets generated in a specific test rig at the Chair of Design and Drive Technology, Paderborn University is adopted to verify the proposed method. The acceleration of the bearing housing is measured on the adapter at the top end of the rolling bearing module using a piezoelectric accelerometer. The signal is digitalized and saved synchronously to the MCS with a sampling rate of 64 kHz [49]. After the experiments, 12 bearings are run with artificial damages and 14 bearings run with damages from accelerated lifetime tests, and 6 bearings run without any damages. To verify the proposed method, 6 different bearings were used in this study: 3 bearings with inner faults, 2 bearings with outer faults, and 1 healthy bearing. And they are labeled with 0 to 5. To evaluate the proposed method under different class-imbalanced levels, four datasets IR_{10}, IR_{20}, IR_{50}, and IR_{100} with different imbalanced ratios are created from the original Paderborn University Bearing Dataset. The detail of these datasets is given in Table 2. The samples of the normal state are regarded as the major class, and others of different faults are regarded as the minor class. The details about the number of major class samples, imbalanced ratios, and fault types are specified in Table 1, where the imbalanced ratio of 10, 20, 50, and 100 denotes that the samples of the major class are 10, 20, 50, and 100 times larger compared with the samples of the minor classes. In addition, each condition is intercepted into one sample by every 1024 points to generate enough samples.

Table 2
Details of the Datasets.

Class	Number of Samples			
	IR_{10}	IR_{20}	IR_{50}	IR_{100}
Healthy	1500	1500	1500	1500
Inner Fault Type 1	150	75	30	15
Inner Fault Type 2	150	75	30	15
Inner Fault Type 3	150	75	30	15
Outer Fault Type 1	150	75	30	15
Outer Fault Type 2	150	75	30	15
Total	2250	1875	1650	1575
Ratio	10	20	50	100

3.2. Feature extraction

3.2.1. Shallow statistical features extraction

To describe the fault pattern comprehensively, 16 time-domain features and 12 frequency-domain features are extracted from the signal data first. To better describe the features, the corresponding definitions of the features are listed in Table 3 and Table 4. In Table 3, x_i in each equation means the collected vibration signal value at i, and N means the length of the collected vibration signal. Meanwhile, the y_l and L in Table 4 have the same meaning with x_i and N for clearer expression.

Nevertheless, time–frequency domain features are also obtained to describe the fault types. EEMD and WPT are adopted to obtain the time–frequency domain features in this paper. For the EEMD, the ensemble number is set to 100, and the stand deviation of the noise data is set to 0.2. After the decomposition of the vibration signal, the energy entropy of the intrinsic mode function (IMF) and residue signal in the EEMD is obtained as features. For the WPT, the original vibration data at the 5th level decomposed by the mother wavelet 'DB4', and the node energy can be regarded as the features. Thus, 32 wavelet packet features and 10 features extracted by EEMD are consequently obtained as the time–frequency domain features. Fig. 3. shows the reconstructed signals by different nodes of EEMD.

3.2.2. Deep representation features extraction

DenseNet with 1-D kernels is utilized to obtain the deep representation features automatedly. The output of the last hidden layer is selected as the deep representation feature to describe the fault patterns. In the modified DenseNet, different convolution kernel sizes and block structures are used to construct the network for getting better feature extraction ability. The hyperparameters of the DenseNet are determined by cross-validation which is given in Table 5.

3.3. Evaluation metrics

For a given classification problem, there are four types of results, which are true positive *(TP)*, true negative *(TN)*, false positive *(FP)*, and false negative *(FN)*. To evaluate the methods in this paper, two kinds of commonly used metrics are adopted based on these possible results, namely MG-Mean and Macro-F1 [50,51]. MG-Mean and Macro-F1 are effective metrics for evaluating the overall performance of a multi-class classification problem. If the classifier performs equally well on all classes, it can achieve high MG-Mean and Macro-F1 scores. The calculation of MG-Mean and Macro-F1 are given as:

$$MG - Mean = \sqrt[n]{\prod_{i=1}^{n} recall_i} \tag{14}$$

$$Macro - F1 = \frac{\sum_{i=1}^{n} \frac{2 \times precision_i \times recall_i}{precision_i + recall_i}}{n} \tag{15}$$

where $recall_i$ means the Recall of class i, $precision_i$ means the Precision of class i, and n means the number of classes, i.e. fault types.

3.4. Experimental procedure

To show the effectiveness of IEMD-ER, some well-known methods,

Table 3
Features extracted in time-domain.

Mean	$X_{\text{mean}} = \frac{1}{N}\sum_{i=1}^{N} x_i$	Mean of absolute amplitude	$X_{\text{mav}} = \frac{1}{N}\sum_{i=1}^{N} \lvert x_i \rvert$
Root variance	$X_{rv} = \left(\frac{1}{N}\sum_{i=1}^{N}(x_i - X_{\text{mean}})^2\right)^{1/2}$	Max	$X_{max} = max(x_i)$
Min	$X_{min} = min(x_i)$	Root Mean square	$X_{rms} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{1/2}$
Square root of amplitude	$X_{sra} = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i \rvert}\right)^2$	Kurtosis value	$X_{kv} = \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - X_{\text{mean}}}{X_{rv}}\right)\right)^4$
Skewness value	$X_{sv} = \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - X_{\text{mean}}}{X_{rv}}\right)\right)^3$	Peak-peak value	$X_{ppv} = X_{max} - X_{min}$
Crest factor	$X_{cf} = max(\lvert x_i \rvert)/X_{rms}$	Impulsive factor	$X_{if} = max(\lvert x_i \rvert)/X_{abs}$
Margin factor	$X_{mf} = max(\lvert x_i \rvert)/X_{\text{sra}}$	Kurtosis factor	$X_{kf} = X_{kv}/X_{rms}^4$
Shape factor	$X_{shf} = X_{kv} / X_{abs}$	Skewness factor	$X_{skf} = X_{kv}/X_{rv}^3$

Table 4
Features extracted in frequency-domain.

Mean of frequency	$X_{\text{meanf}} = \frac{1}{L}\sum_{l=1}^{L} y_l$	Root variance of frequency	$X_{rvf} = \left(\frac{1}{L}\sum_{l=1}^{L}(y_l - X_{\text{meanf}})^2\right)^{1/2}$
Maximum of frequency	$X_{maxf} = max(y_l)$	Minimum of frequency	$X_{minf} = min(y_l)$
Root mean square of frequency	$X_{rmsf} = \left(\frac{1}{L}\sum_{l=1}^{L} y_l^2\right)^{1/2}$	Skewness value of frequency	$X_{svf} = \left(\frac{1}{L}\sum_{l=1}^{L}\left(\frac{y_l - X_{mcanf}}{X_{rvf}}\right)\right)^3$
Kurtosis value of frequency	$X_{kvf} = \left(\frac{1}{L}\sum_{l=1}^{L}\left(\frac{y_l - X_{\text{meanf}}}{X_{\text{ruf}}}\right)\right)^4$	Skewness factor of frequency	$X_{skff} = X_{kvf}/X_{rvf}^3$
Kurtosis factor of frequency	$X_{kff} = X_{kvf}/X_{rmsf}^4$	Frequency center	$x_{fc} = \sum_{l=1}^{L}(f_l \cdot y_l)/X_{\text{meanf}}$
Root mean square with frequency	$x_{rmsf} = \left(\frac{1}{L}\sum_{l=1}^{L}\left(f_l^2 \cdot y_l\right)/X_{\text{meanf}}\right)^{1/2}$	Root variance with frequency	$x_{rvf} = \left(\frac{1}{L}\sum_{l=1}^{L}(f_l - X_{fc})^2 \cdot y_l/X_{\text{meanf}}\right)^{1/2}$

including DT, Adaboost, Bagging, and RF are selected to analyze the same dataset. And the standard RS with ER is also adopted as the compared method to show the improvement of the proposed IEMD-ER. In addition, some sampling methods are also selected to demonstrate the advantage of ensemble methods to defeat the imbalanced data. Thus, some sampling methods, such as SMOTE, Random under sampling with Boost (RusBoost), and Random over sampling with Boost (RosBoost) are also selected as the compared methods.

To validate the methods, 80 % of the dataset is chosen as the training set, and the remaining as the testing set. Furthermore, in order to improve the robustness of the fault diagnosis result, all the methods conducted on the dataset were using cross-validation. In detail, the training set will be divided into five folds, of which four folds are taken for training and the remains for testing. The parameters will be tuned in the procedure of training. The parameter settings with the best MG-Mean and Macro-F1 are chosen as the final parameters perform on the testing set. The detailed parameter settings of the proposed IEMD-ER and comparison methods are given in Table 6.

4. Results and discussions

4.1. Results

After the experiments, the results of IEMD-ER and other compared methods are shown in Tables 7 and 8. The mean and standard deviation of MG-Mean and Macro-F1 are calculated by using the output of ten times fivefold cross-validation. As shown in Table 7 and Table 8, it is clear that the proposed method achieves the best performance. i.e., 98.29 % (IR_{10}), 97.71 % (IR_{20}), 97.20 % (IR_{50}), and 95.60 % (IR_{100}) for the MG-Mean and 98.33 % (IR_{10}), 97.84 % (IR_{20}), 97.43 % (IR_{50}), and 96.33 % (IR_{100}) for the Macro-F1. Meanwhile, the IEMD-ER has better performance, since its standard deviations are small while the mean is high. Besides, as shown in the tables, the ensemble methods have better performance than the single method. Hence, using the combined features can improve the performance of MG-Mean and Macro-F1 under class-imbalance situations. The proposed IEMD-ER method performed well for the fault diagnosis task. In short, the results shown in the tables have proved the usefulness of the proposed IEMD-ER in fault diagnosis.

4.2. Discussions

4.2.1. Evaluation of different domain features and their combination

The results of using shallow statistical features, deep representation features, as well as their combination are reported to verify the rationality of different domain features. It is clear in Fig. 4, the internal correlation of shallow statistical features is strong than that of deep representation features, which indicates the shallow statistical features may contain some redundant information thus influencing the performance of the fault diagnosis model.

Figs. 5 and 6 illustrate the MG-Mean and Macro-F1 of shallow statistical features (F1), deep representation features (F2), and their combination (F3). As for single-domain features, the performance of the methods based on deep representation features is better compared with those of the shallow statistical features under different imbalance ratios. Besides, the performance will decrease while the imbalanced ratio is increasing. For instance, the MG-Mean and Macro-F1 of shallow statistical features under different imbalanced ratios are 2.69 %, 6.2 %, and 7.53 % as well as 2.31 %, 5.00 %, and 4.77 % lower than the average MG-Mean and Macro-F1 of IR_{10}. However, for the combined features it is only 0.57 %, 1.00 %, and 2.60 % with G-Mean as well as 0.49 %, 0.90 %, and 2.00 % with Macro-F1 lower than the average of IR_{10} under the same situation. It can be seen that the stability of the performance of shallow statistical features tends to decrease when the imbalance ratio increases. Hence, the combined features are performed stable due they

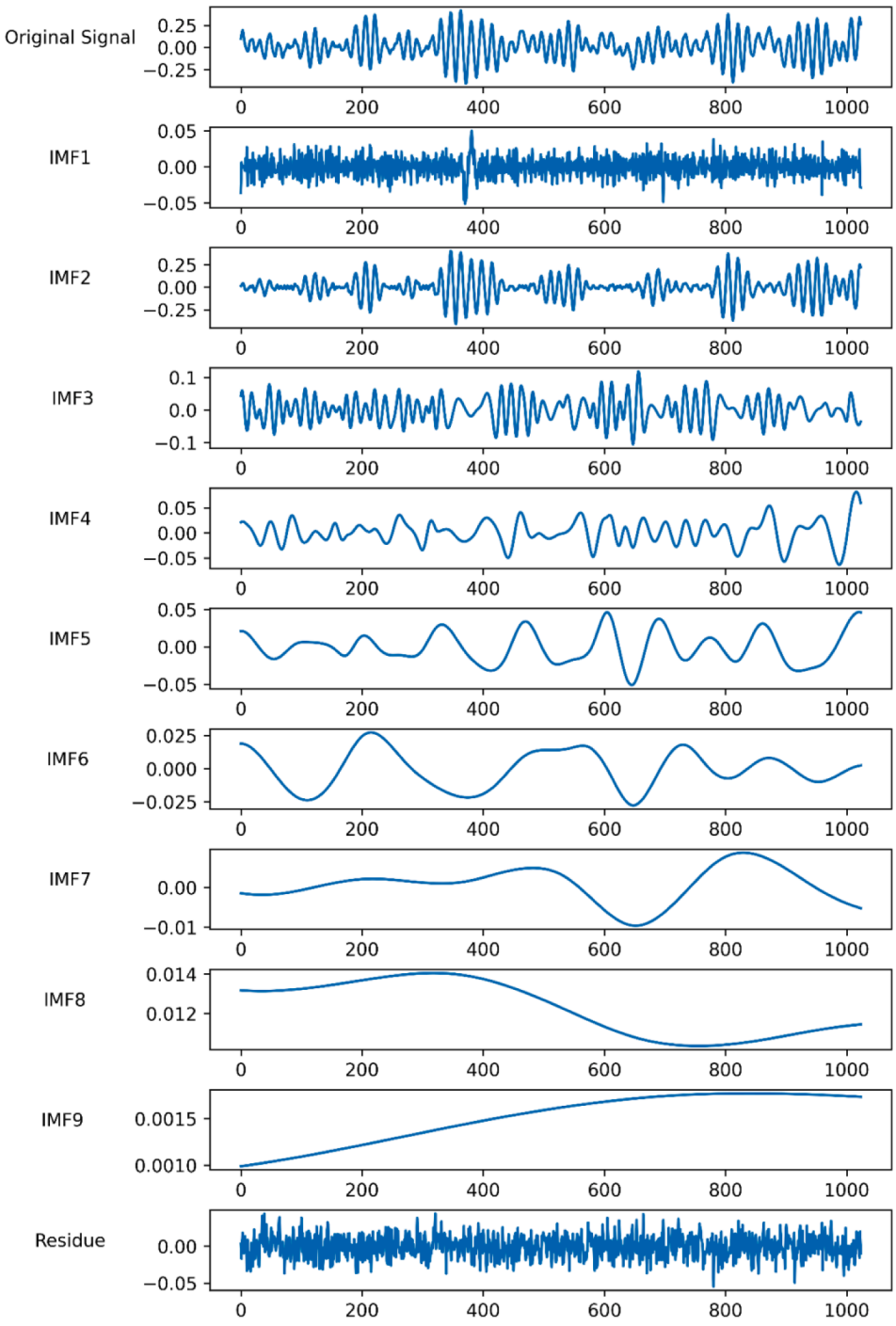

Fig. 3. Reconstructed Signals.

Table 5
Hyperparameter Settings of the Modified DenseNet.

Name	Valve
Layers	34
Dense blocks	3
Dense layer of each Block	5
Growth Rate	3
Kernel Size	5,3
Drop Rate	0.6

can represent the deep structure information in the signal data. Furthermore, most of the methods perform better with the combined features, which shows the effectiveness of the combined features. Specifically, in IR_{10}, the MG-Mean, and Macro-F1 of the proposed method is improved by 5.12 %, 4.89 % compared with shallow statistical features, 0.75 %, and 0.77 % compared with deep representation features. The improvements of MG-Mean and Macro-F1 show the usefulness of combined features in fault diagnosis. In the meantime, the performance of the proposed method with different imbalance ratios has improved by 0.48 % (IR_{10}), 0.60 % (IR_{20}), 1.29 % (IR_{50}), and 1.62 % (IR_{100}) with MG-Mean as well as 0.46 % (IR_{10}), 0.63 % (IR_{20}), 2.08 % (IR_{50}), and 1.31 % (IR_{100}) with Macro-F1 compared with the standard RS with ER. However, not all the methods can properly deal with the fused features, the

Table 6
Parameters setting.

Methods	Parameters
DT	Max Depth: 5, Min Samples Split: 25;
Adaboost	Number of the Base Classifiers: 10; Base Classifier: DT;
Bagging	Number of the Base Classifiers: 10; Base Classifier: DT;
RF	Number of Base Classifiers: 10; Bootstrap: True; Base classifier: DT, Criterion: Gini;
SMOTE	K:10;
RusBoost	Sampling Strategy: {1:10,1:20,1:50,1:100};
RosBoost	Sampling Strategy: {1:10,1:20,1:50,1:100};
RS-ER	Subspace Rate: {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}; Number of Ensembles: 10; Base Classifier: DT;
IEMD-ER	Subspace Rate: {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0}; Number of ensembles: 10; Base classifier: DT.

results of Adaboost on combined features are worse than on deep representation features, which indicates the noise in the shallow statistical features will decrease the diagnosis performance. In summary, the performance of fault diagnosis can be improved with the fused features, and the proposed method can address the high-dimensional and class-imbalance problem to a certain extent.

4.2.2. Evaluation of different aggregation strategies

In this paper, the ER rule is employed to fuse the outputs of different base classifiers. Other fusion strategies such as major voting (Voting), weighted voting (W-Voting), and evidence reasoning approach (ERA) are selected for comparison to prove the usefulness of the ER rule [52]. All the hyperparameters of the proposed method are set to the same except for the fusion strategy for the fair evaluation. As presented in Fig. 7, the values of G-Mean and Macro-F1 with different fusion strategies demonstrate that the proposed IEMD-ER has superior performances than other fusion strategies. And the proposed method is more stable than other methods based on the standard variance. The performance of W-Voting and ERA is lower than the Voting strategy when the imbalance ratio becomes high, which indicates that the high imbalance ratio will increase the conflict between different evidence and influence the fusing procedure. In addition, compared with the W-Voting and ERA, the ER rule performs better, which means ER rule can improve its performance by taking the weight and reliability into account under an imbalanced situation. Therefore, the proposed IEMD-ER can enhance the performance of fault diagnosis meanwhile considering the feature fusion and imbalance problem simultaneously.

5. Conclusion and future work

The fault diagnosis of machinery is of importance in the industry field for reducing uncertain machine breakdown risk. In this study, an Imbalanced Ensemble method with DenseNet and Evidential Reasoning rule (IEMD-ER) is proposed to enhance the performance of machinery fault diagnosis. The proposed IEMD-ER takes all the features into a unified model, which is different from previous studies that only incorporate the shallow statistical features or deep representation features independently. The proposed IEMD-ER can fuse the shallow statistical features and deep representation features by utilizing an improved semi-random subspace method. Meanwhile, the SMOTE is introduced to the proposed method to defeat the imbalanced problem and the ER rule is adopted for the prediction results fusion of base classifiers. Experiments are conducted on the bearing datasets provided by Paderborn University to validate the effectiveness of the proposed IEMD-ER based.

Although the effectiveness of the proposed IEMD-ER has been proved in this work, there are still some aspects that need to be further enhanced. Firstly, the RF is adopted to effectively select the feature subsets, meanwhile, other methods can also be introduced to generate the feature subsets. Secondly, it is worthwhile to validate the proposed method under different class-imbalance ratios, as different imbalance datasets can be adopted to enhance the effectiveness of IEMD-ER in real applications. Thirdly, in the real-world applications of fault diagnosis, parallel computing methods should be employed to improve computational efficiency.

Table 7
MG-Mean of the methods.

Methods	MG-Mean							
	IR_{10}		IR_{20}		IR_{50}		IR_{100}	
	Mean (%)	Std (%)	Mean (%)	Std (%)	Mean (%)	Std (%)	Mean (%)	Std (%)
DT	94.89	1.40	93.38	2.31	87.06	4.55	86.90	8.52
Adaboost	95.80	1.43	94.67	2.16	89.43	5.20	81.47	14.63
Bagging	97.36	1.31	96.09	1.96	95.40	3.64	91.13	14.31
RF	97.59	1.27	96.33	1.85	95.37	3.55	93.70	5.68
SMOTE	96.63	1.62	95.04	2.41	93.43	4.75	90.26	7.53
RusBoost	96.87	1.24	95.08	2.78	92.81	5.56	89.23	14.96
RosBoost	95.81	2.01	94.55	2.50	91.76	4.61	88.90	14.45
RS-ER	97.81	1.17	97.11	2.05	95.91	3.82	94.07	6.03
IEMD-ER	**98.29**	0.96	**97.71**	1.82	**97.20**	2.66	**95.69**	4.42

Table 8
Macro-F1 of the methods.

Methods	Macro-F1							
	IR_{10}		IR_{20}		IR_{50}		IR_{100}	
	Mean (%)	Std (%)	Mean (%)	Std (%)	Mean (%)	Std (%)	Mean (%)	Std (%)
DT	95.17	1.25	94.00	2.10	88.59	3.95	89.51	6.52
Adaboost	95.97	1.30	94.96	2.00	90.37	4.40	85.56	6.98
Bagging	97.45	1.22	96.42	1.81	95.93	3.20	93.83	5.30
RF	97.71	1.18	96.57	1.67	95.76	3.21	94.75	4.45
SMOTE	96.70	1.32	95.20	2.22	94.05	4.03	93.26	5.79
RusBoost	96.60	1.29	94.39	2.84	89.98	7.41	87.51	9.07
RosBoost	95.96	1.87	94.68	2.25	92.57	3.96	91.11	5.70
RS-ER	97.87	1.12	97.21	1.94	95.35	3.69	95.02	4.93
IEMD-ER	**98.33**	0.91	**97.84**	1.70	**97.43**	2.43	**96.33**	3.63

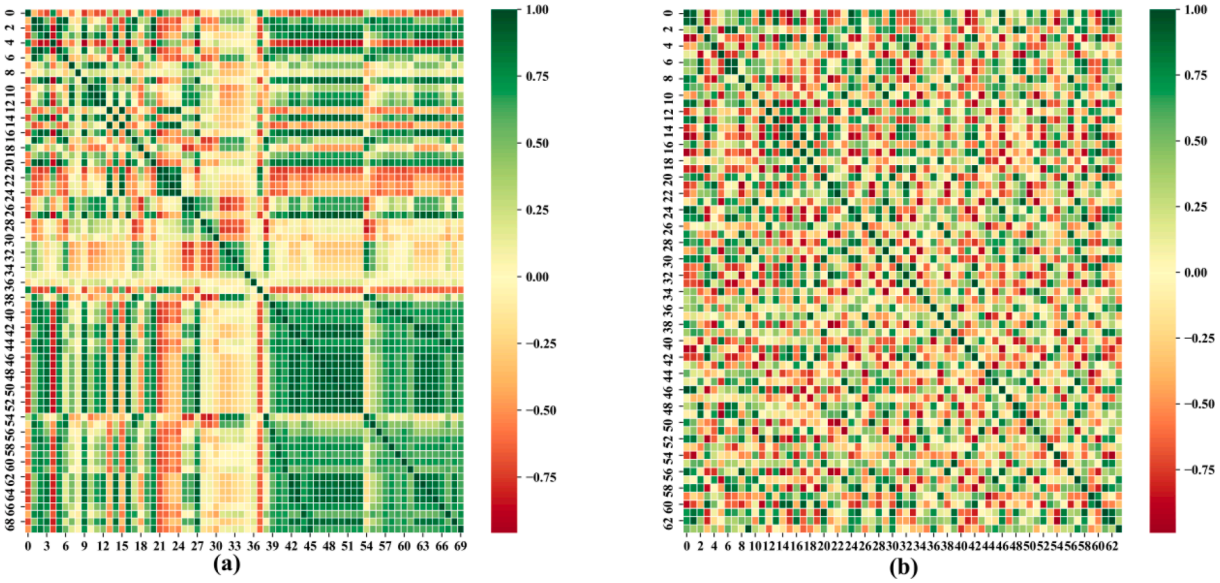

Fig. 4. Correlation of different features in IR_{10}: (a) Shallow statistical features;(b) Deep representation features.

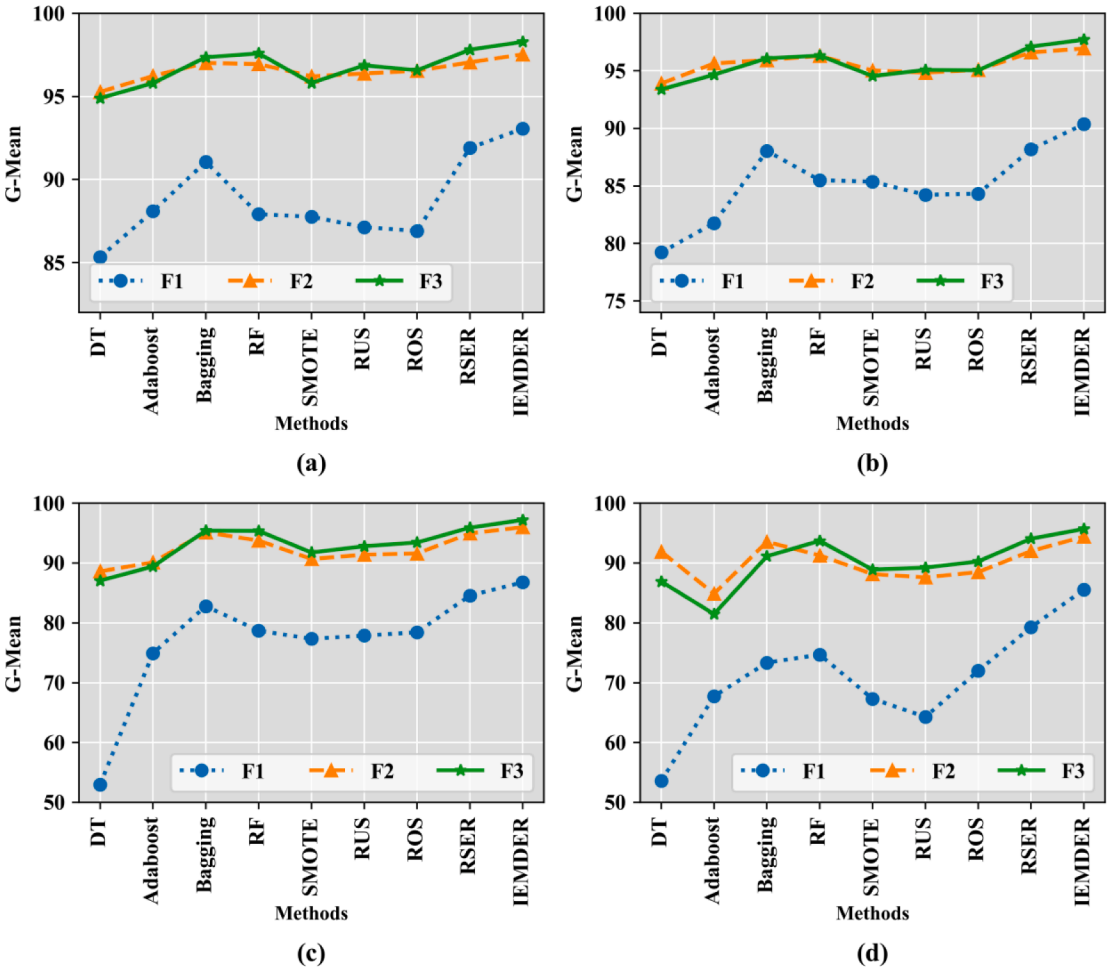

Fig. 5. MG-Mean comparisons of different features under four imbalance ratios: (a) IR_{10}; *(b)*IR_{20}; (c) IR_{50}; (d) IR_{100}.

CRediT authorship contribution statement

Gang Wang: Conceptualization, Formal analysis. **Yanan Zhang:** Conceptualization, Methodology. **Feng Zhang:** Software, Visualization. **Zhangjun Wu:** Conceptualization, Methodology.

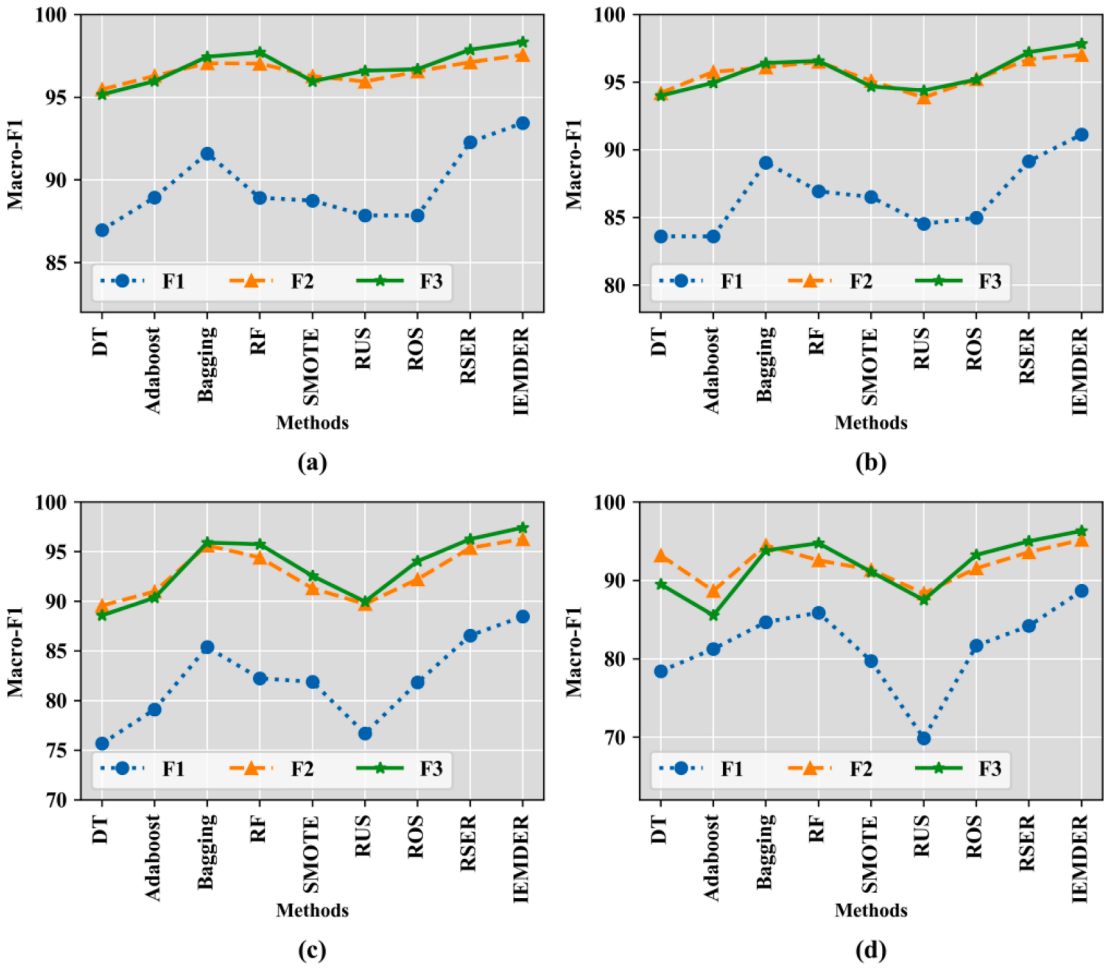

Fig. 6. Macro-F1 comparisons of different features under four datasets: (a) IR_{10}; (b)IR_{20}; (c) IR_{50}; (d) IR_{100}.

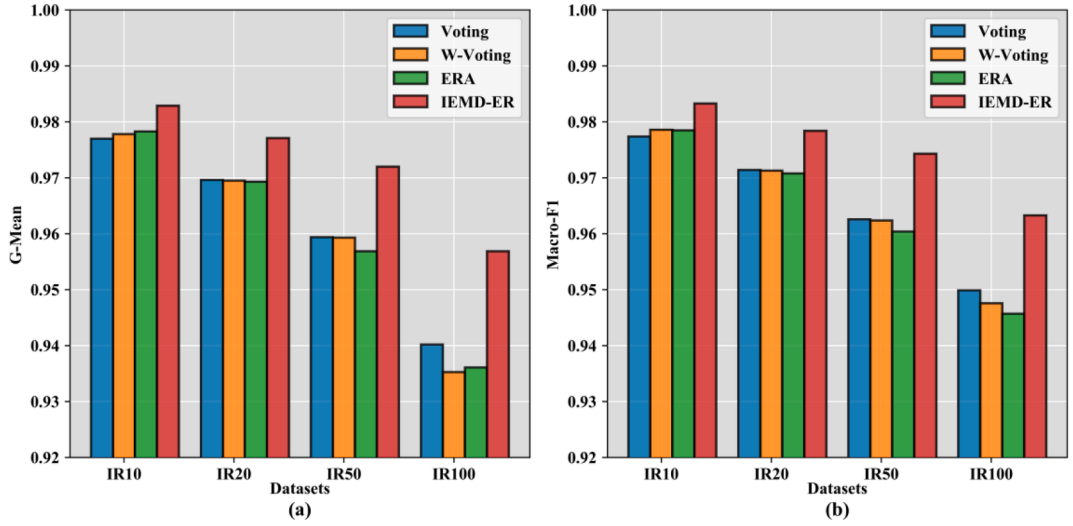

Fig. 7. MG-Mean and Macro-F1 comparisons for different aggregation strategies under four datasets: (a) MG-Mean; (b) Macro-F1.

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.

Data availability

Data will be made available on request.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (72071062), Science Fund for Distinguished Young Scholars of AnHui (2208085J12), Anhui Provincial Key Research and Development Program (202104a05020038), and Fundamental Research Funds for the Central Universities (PA2021KCPY0032).

References

[1] H. Cao, F. Fan, K. Zhou, Z. He, Wheel-bearing fault diagnosis of trains using empirical wavelet transform, Measurement. 82 (2016) 439–449, https://doi.org/10.1016/j.measurement.2016.01.023.
[2] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: A review, Measurement. 48 (2014) 292–305, https://doi.org/10.1016/j.measurement.2013.11.012.
[3] X. Ji, Y. Ren, H. Tang, C. Shi, J. Xiang, An intelligent fault diagnosis approach based on Dempster-Shafer theory for hydraulic valves, Measurement. 165 (2020), 108129, https://doi.org/10.1016/j.measurement.2020.108129.
[4] J. Li, X. Yao, X. Wang, Q. Yu, Y. Zhang, Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis, Measurement. 153 (2020), 107419, https://doi.org/10.1016/j.measurement.2019.107419.
[5] R.F.R. Junior, I.A. dos Santos Areias, M.M. Campos, C.E. Teixeira, L.E.B. da Silva, G.F. Gomes, Fault detection and diagnosis in electric motors using 1d convolutional neural networks with multi-channel vibration signals, Measurement. 190 (2022) 110759. 10.1016/j.measurement.2022.110759.
[6] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R.E. Vásquez, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing. 168 (2015) 119–127, https://doi.org/10.1016/j.neucom.2015.06.008.
[7] L. Ai, J. Wang, X. Wang, Multi-features fusion diagnosis of tremor based on artificial neural network and D-S evidence theory, Signal Process. 88 (2008) 2927–2935, https://doi.org/10.1016/j.sigpro.2008.06.018.
[8] C. Wang, M. Gan, C. Zhu, Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory, J. Intell. Manuf. 29 (2018) 937–951, https://doi.org/10.1007/s10845-015-1153-2.
[9] L. Ciabattoni, F. Ferracuti, A. Freddi, A. Monteriu, Statistical Spectral Analysis for Fault Diagnosis of Rotating Machines, IEEE Trans. Ind. Electron. 65 (2018) 4301–4310, https://doi.org/10.1109/TIE.2017.2762623.
[10] B. Wang, C. Ding, Hierarchical Frequency-Domain Sparsity-Based Algorithm for Fault Feature Extraction of Rolling Bearings, IEEE Trans. Instrum. Meas. 69 (2020) 6228–6240, https://doi.org/10.1109/TIM.2020.2972083.
[11] C. Li, M. Liang, Time–frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform, Mech. Syst. Signal Process. 26 (2012) 205–217, https://doi.org/10.1016/j.ymssp.2011.07.001.
[12] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement. 111 (2017) 1–10, https://doi.org/10.1016/j.measurement.2017.07.017.
[13] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 95 (2017) 187–204, https://doi.org/10.1016/j.ymssp.2017.03.034.
[14] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (2017) 1693–1702, https://doi.org/10.1109/TIM.2017.2669947.
[15] C. Lu, Z.Y. Wang, W.L. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377–388, https://doi.org/10.1016/j.sigpro.2016.07.028.
[16] L. Wen, X. Li, L. Gao, Y. Zhang, A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method, IEEE Trans. Ind. Electron. 65 (2018) 5990–5998, https://doi.org/10.1109/TIE.2017.2774777.
[17] H. Oh, J.H. Jung, B.C. Jeon, B.D. Youn, Scalable and Unsupervised Feature Engineering Using Vibration-Imaging and Deep Learning for Rotor System Diagnosis, IEEE Trans. Ind. Electron. 65 (2018) 3539–3549, https://doi.org/10.1109/TIE.2017.2752151.
[18] Y. Qin, X. Wang, J. Zou, The Optimized Deep Belief Networks with Improved Logistic Sigmoid Units and Their Application in Fault Diagnosis for Planetary Gearboxes of Wind Turbines, IEEE Trans. Ind. Electron. 66 (2019) 3814–3824, https://doi.org/10.1109/TIE.2018.2856205.
[19] H. Wang, J. Xu, R. Yan, R.X. Gao, A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN, IEEE Trans. Instrum. Meas. 69 (2019) 2377–2389, https://doi.org/10.1109/tim.2019.2956332.
[20] W. Huang, J. Cheng, Y. Yang, G. Guo, An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis, Neurocomputing. 359 (2019) 77–92, https://doi.org/10.1016/j.neucom.2019.05.052.
[21] R. Liu, F. Wang, B. Yang, S.J. Qin, Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions, IEEE Trans. Ind. Inform. 16 (2019) 3797–3806, https://doi.org/10.1109/TII.2019.2941868.
[22] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, K. Weinberger, Convolutional Networks with Dense Connectivity, IEEE Trans. Pattern Anal. Mach. Intell. 44 (2019) 8704–8716, https://doi.org/10.1109/tpami.2019.2918284.
[23] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks, IEEE Trans. Ind. Electron. 63 (2016) 7067–7075, https://doi.org/10.1109/TIE.2016.2582729.
[24] J. Wang, Y. Zhang, F. Zhang, W. Li, S. Lv, M. Jiang, L. Jia, Accuracy-improved bearing fault diagnosis method based on AVMD theory and AWPSO-ELM model, Measurement. 181 (2021), 109666, https://doi.org/10.1016/j.measurement.2021.109666.
[25] K. Shao, W. Fu, J. Tan, K. Wang, Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing, Measurement. 173 (2021), 108580, https://doi.org/10.1016/j.measurement.2020.108580.
[26] B. Qiu, Y. Lu, L. Sun, X. Qu, Y. Xue, F. Tong, Research on the damage prediction method of offshore wind turbine tower structure based on improved neural network, Measurement. 151 (2020), 107141, https://doi.org/10.1016/j.measurement.2019.107141.
[27] Q. Hu, A. Qin, Q. Zhang, J. He, G. Sun, Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA, IEEE Sens. J. 18 (2018) 8472–8483, https://doi.org/10.1109/JSEN.2018.2866708.
[28] T.W. Rauber, F. De Assis Boldt, F.M. Varejão, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, IEEE Trans. Ind. Electron. 62 (2015) 637–646, https://doi.org/10.1109/TIE.2014.2327589.
[29] M. Kordestani, M.F. Samadi, M. Saif, K. Khorasani, A new fault diagnosis of multifunctional spoiler system using integrated artificial neural network and discrete wavelet transform methods, IEEE Sens. J. 18 (2018) 4990–5001, https://doi.org/10.1109/JSEN.2018.2829345.
[30] C. Wang, H. Li, K. Zhang, S. Hu, B. Sun, Intelligent fault diagnosis of planetary gearbox based on adaptive normalized CNN under complex variable working conditions and data imbalance, Measurement. 180 (2021), 109565, https://doi.org/10.1016/j.measurement.2021.109565.
[31] R. Chen, J. Zhu, X. Hu, H. Wu, X. Xu, X. Han, Fault diagnosis method of rolling bearing based on multiple classifier ensemble of the weighted and balanced distribution adaptation under limited sample imbalance, ISA Trans. 114 (2021) 434–443, https://doi.org/10.1016/j.isatra.2020.12.034.
[32] Y. Zhang, X. Li, L. Gao, L. Wang, L. Wen, Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning, J. Manuf. Syst. 48 (2018) 34–50, https://doi.org/10.1016/j.jmsy.2018.04.005.
[33] A. Iranmehr, H. Masnadi-Shirazi, N. Vasconcelos, Cost-sensitive support vector machines, Neurocomputing. 343 (2019) 50–64, https://doi.org/10.1016/j.neucom.2018.11.099.
[34] W. Qian, S. Li, A novel class imbalance-robust network for bearing fault diagnosis utilizing raw vibration signals, Measurement. 156 (2020), 107567, https://doi.org/10.1016/j.measurement.2020.107567.
[35] S. Wang, L.L. Minku, X. Yao, Resampling-based ensemble methods for online class imbalance learning, IEEE Trans. Knowl. Data Eng. 27 (2014) 1356–1368, https://doi.org/10.1109/TKDE.2014.2345380.
[36] Z. Wang, H. Huang, Y. Wang, Fault diagnosis of planetary gearbox using multi-criteria feature selection and heterogeneous ensemble learning classification, Meas. J. Int. Meas. Confed. 173 (2021), 108654, https://doi.org/10.1016/j.measurement.2020.108654.
[37] Y. Li, Y. Song, L. Jia, S. Gao, Q. Li, M. Qiu, Intelligent Fault Diagnosis by Fusing Domain Adversarial Training and Maximum Mean Discrepancy via Ensemble Learning, IEEE Trans. Ind. Inform. 17 (2021) 2833–2841, https://doi.org/10.1109/TII.2020.3008010.
[38] X. Zhang, B. Wang, X. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, Knowl.-Based Syst. 89 (2015) 56–85, https://doi.org/10.1016/j.knosys.2015.06.017.
[39] G. Wang, F. Zhang, B. Cheng, F. Fang, DAMER: a novel diagnosis aggregation method with evidential reasoning rule for bearing fault diagnosis, J. Intell. Manuf. 32 (2021) 1–20, https://doi.org/10.1007/s10845-020-01554-5.
[40] X. Mao, F. Zhang, G. Wang, Y. Chu, K. Yuan, Semi-random subspace with Bi-GRU: Fusing statistical and deep representation features for bearing fault diagnosis, Measurement. 173 (2021), 108603, https://doi.org/10.1016/j.measurement.2020.108603.
[41] Z. Chai, C. Zhao, Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification, IEEE Trans. Ind. Inform. 16 (2020) 54–66, https://doi.org/10.1109/TII.2019.2915559.
[42] X.-B. Wang, X. Zhang, Z. Li, J. Wu, Ensemble extreme learning machines for compound-fault diagnosis of rotating machinery, Knowl.-Based Syst. 188 (2020), 105012, https://doi.org/10.1016/j.knosys.2019.105012.
[43] J.B. Yang, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, Eur. J. Oper. Res. 131 (2001) 31–61, https://doi.org/10.1016/S0377-2217(99)00441-5.
[44] J.-B. Yang, D.-L. Xu, Evidential reasoning rule for evidence combination, Artif. Intell. 205 (2013) 1–29, https://doi.org/10.1016/j.artint.2013.09.003.
[45] W. Zhang, X. Li, X.-D. Jia, H. Ma, Z. Luo, X. Li, Machinery fault diagnosis with imbalanced data using deep generative adversarial networks, Measurement. 152 (2020), 107377, https://doi.org/10.1016/j.measurement.2019.107377.
[46] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[47] Y. Wang, D. Wang, X. Ye, Y. Wang, Y. Yin, Y. Jin, A tree ensemble-based two-stage model for advanced-stage colorectal cancer survival prediction, Inf. Sci. 474 (2019) 106–124, https://doi.org/10.1016/j.ins.2018.09.046.
[48] X. Xu, Z. Zhao, X. Xu, J. Yang, L. Chang, X. Yan, G. Wang, Machine learning-based wear fault diagnosis for marine diesel engine by fusing multiple data-driven

models, Knowl.-Based Syst. 190 (2020) 105324, https://doi.org/10.1016/j.knosys.2019.105324.
[49] C. Lessmeier, J. Kimotho, D. Zimmer, W. Sextro, Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification, in: 2016 European Conference of the Prognostics and Health Management vol. 3, 2016, pp. 5–8.
[50] Y. Sun, M.S. Kamel, Y. Wang, Boosting for learning multiple classes with imbalances class distribution, in: Sixth International Conference on Data Mining, 2006, pp. 592–602, https://doi.org/10.1109/ICDM.2006.29.
[51] Y. Yu, L. Guo, H. Gao, Y. Liu, T. Feng, Pareto-optimal Adaptive Loss Residual Shrinkage Network for Imbalanced Classification of Machinery Fault Diagnostics, IEEE Trans. Ind. Inform. 18 (2021) 2233–2243, https://doi.org/10.1109/TII.2021.3094186.
[52] J.B. Yang, M.G. Singh, An Evidential Reasoning Approach for Multiple-Attribute Decision Making with Uncertainty, IEEE Trans. Syst. Man Cybern. 24 (1994) 1–18, https://doi.org/10.1109/21.259681.