A Decision Fusion SWT-RF Method for Rolling Bearing Enhanced Diagnosis Under Low-Quality Data

Jiayu Chen, *Member, IEEE*, Cuiying Lin, Qinhua Lu, Chaoqi Yang, Peng Li, *Member, IEEE*, Pingchao Yu, and Hongjuan Ge

***Abstract*— Low-quality data, including insufficient samples and low signal-to-noise ratio (SNR), restrict the effective application of intelligent diagnostic methods based on deep learning. In this study, a new rolling bearing enhanced diagnostic method is proposed based on swin-transformer (SWT) and ReliefF (RF) combined with Dempster–Shafer (DS) evidence theory. First, SWT is improved for adaptive deep feature mining of the vibration signal. Meanwhile, to enhance the quality of features, RF is introduced to optimize the deep fault features output by the global average pooling layer, which helps improve the classifier performance. Then, the DS evidence theory-based decision fusion strategy is designed to realize the fusion of different axial signals at the decision level, which enhances the fault knowledge threshold and further improves the diagnostic ability. Finally, the bearing cases with data collected from the accelerated life degradation and different distributions are studied. The results reveal that the proposed method can adaptively mine fault features with low-quality data and realize efficient enhanced diagnosis.**

***Index Terms*— Adaptive feature learning, low signal-to-noise ratio (SNR), multisensors, rotating machinery.**

I. Introduction

As THE crucial equipment of mechanical equipment, such as aeroengines, gas turbines, and wind turbines, rolling bearings play an important role in the rotor system [1], [2]. However, rolling bearings usually work in extremely harsh environments, such as high temperature, high pressure, and high speed. It is prone to various failures and even causes huge economic losses and disastrous accidents. Therefore, it is crucial to effectively diagnose faults in rolling bearings [3].

Due to the development of deep learning, deep neural network (DNN), such as convolutional neural networks (CNNs) [4], [5], residual networks [6], [7], and efficient networks [8], [9], has been widely used in the field of fault diagnosis. Moreover, their deep architecture can realize adaptive feature extraction of original data, which is suitable for end-to-end direct diagnosis. However, in harsh environments, it faces the problem of low-quality data, including insufficient samples [10], [11] and low signal-to-noise ratio (SNR) [12]. This degrades the performance of the direct diagnostic method based on DNNs. For fault diagnosis of low-quality data, Chen et al. [11] proposed an improved deep-fused CNN approach that integrates fusion feature importance and diversity. Li et al. [13] improved the support matrix machine and successfully achieved fault diagnosis by fusing vibration signals with infrared thermography. Hou et al. [14] developed a Siamese multiscale residual feature fusion network specifically designed for fault diagnosis in scenarios with limited sample sizes. Therefore, the key to realize effective fault diagnosis under low-quality data is to mine diverse deep fault features, of which the typical solutions, basically, can be divided into two categories, respectively, model optimization and data fusion.

For the model aspect, more and more models based on DNNs have been developed to increase diagnostic performance, such as transfer learning-based models, ensemble learning-based models, and generative adversarial network (GAN)-based models [15]. Recently, inspired by the transformer in the field of natural language processing [16], Google proposed a vision transformer (ViT), which enhanced the model performance by introducing a self-attention mechanism. It achieved great success in the field of image classification [17], speech recognition, and even chatbot, like Generative Pre-trained Transformer-4 (GPT-4) [18]. It also provided a new perspective for the research of fault diagnosis under low-quality data. Ding et al. [19] put forward a method based on a self-attention mechanism and time–frequency transformer. Fan et al. [20] put forward a method based on improved vibration gray-scale texture image and ViT, and realized bearing fault diagnosis under low-quality data. Lv et al. [21] used improved

Manuscript received 24 July 2023; revised 28 November 2023; accepted 18 December 2023. Date of publication 5 January 2024; date of current version 30 January 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52102474, Grant U2233205 and Grant 72371215, in part by the China Postdoctoral Science Foundation under Grant 2023M731663, in part by the Fundamental Research Funds for the Central Universities under Grant XCXJH-20230744, in part by the Reliability Assurance Center of Chinese Academy of Sciences under Grant CRAC-ZZKT-KY-2022-03, and in part by the Sichuan Science and Technology Program under Grant 2023YFSY0003. The Associate Editor coordinating the review process was Dr. Gabriele Patrizi. *(Corresponding author: Jiayu Chen.)*

Jiayu Chen, Qinhua Lu, Chaoqi Yang, Pingchao Yu, and Hongjuan Ge are with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: jiayu_chen@nuaa.edu.cn; lu_qinhua@nuaa.edu.cn; yang_chaoqi@nuaa.edu.cn; yupingchao@nuaa.edu.cn; nuaa1102b@163.com).

Cuiying Lin is with the Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Quanzhou 362200, China (e-mail: cuiying.lin@nuaa.edu.cn).

Peng Li is with the Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China (e-mail: lp@csu.ac.cn).

Digital Object Identifier 10.1109/TIM.2024.3350130

ViT to realize bearing fault diagnosis under low SNR. However, when the sample is input into the encoder, ViT is trained by splitting the sample into multiple patches, which not only leads to a sharp increase in calculation, but also results in the drowning of significant features.

To solve the problem of computation redundancy in ViT, Liu et al. [22] proposed a swin-transformer (SWT) with a hierarchical design. Inspired by the multiscale idea of CNNs, SWT can realize hierarchical feature extraction. Different from dividing the sample into multiple patches to reduce the dimension in ViT, SWT uses a sliding window to reduce the sample dimension. In this way, the attention calculation can be completed within the window instead of global calculation. It not only improves the classification accuracy, but also saves the calculation cost [22]. In addition, compared to other deep learning models, SWT sets itself apart by incorporating the self-attention mechanism. This allows the model to selectively emphasize the features that are more crucial for the given task. As a result, SWT exhibits remarkable diagnostic performance, even when confronted with low-quality data.

Nevertheless, in SWT, there are dozens of time differences between the fault features output at its full connection layer and the final diagnostic results, which decreases the classifier performance. For this problem, one effective way is to optimize the fault features extracted from the model. ReliefF (RF) is put forward by Robnik-Sikonja and Kononenko [23], which calculates the weight of each feature to determine its importance and then realizes feature optimization. Furthermore, RF is insensitive to noise and can first learn the features with strong correlation in the feature set, which improves the diagnostic performance of classifier.

For the data aspect, the solution is to enrich the knowledge threshold of training samples when the model performance has reached the optimal level [13]. The enhanced diagnostic method based on decision fusion can realize fault feature extraction of multiple knowledge thresholds by acquiring the original information of multiple sensors and finally adopts multiple diagnostic results for decision fusion. The decision fusion exhibits remarkable robustness, enabling it to make sound decisions even in the event of an abrupt sensor failure. Its notable advantage lies in its high fault tolerance, ensuring the reliability and continuity of decision-making processes. Currently, the enhanced diagnostic approach that capitalizes on decision fusion frequently relies on a diverse range of time–frequency analysis methods to refine and enhance the information embedded within the signal [13], [24]. Subsequently, DNNs are used in fault diagnosis, and finally, decision fusion is carried out by Dempster–Shafer (DS) evidence theory [25], [26]. However, there are two challenges associated with this approach. On the one hand, preprocessing the signal such as time–frequency analysis increases the complexity of diagnosis. This requires the operator to have sufficient prior knowledge, which makes the fault tolerance of the model relatively low [27]. On the other hand, multiple or even multiclass basic models increase the complexity of the whole method, which will lead to an increase in the model calculation and reduce its practicability. Therefore, the key to enhanced diagnosis based on decision fusion is to reduce the method complexity, while containing data information diversity, thus realizing end-to-end direct diagnosis based on original data.

To take advantage of model optimization and data fusion in feature extraction as well as results evaluation, this study proposes a decision fusion SWT and RF method to realize effective intelligent diagnosis for bearing faults. The main contributions of this study are as follows.

1) To avoid the dependence of the basic model on the prior knowledge of signal processing under low-quality data, an SWT is improved for adaptive deep feature mining of the original vibration signal.
2) RF is introduced into SWT to optimize the deep fault features of the global average pool layer, which helps SWT to bridge the gap between fault features and final diagnostic cases.
3) To overcome the problem of limited fault knowledge threshold, a decision fusion strategy is proposed and integrated into SWT-RF to realize the fusion of different axial signals at the decision level, which further improves the diagnostic performance.
4) The proposed method can adaptively extract deep fault features, improve the expression ability of significant features, and enhance the fault knowledge threshold, so that it can realize efficient diagnosis under low-quality data.

The remainder of this article is organized as follows. A brief introduction to related work is provided in Section II. The proposed method is described in detail in Section III. Section IV presents the experimental results. Section V concludes this article.

II. Related Studies

A. Swin-Transformer

Fig. 1 shows the architecture of SWT. As shown in Fig. 1(a), the input images are first divided into several 4×4 patches, and then, they are mapped into word symbols in the linear embedding (LE) layer to prepare for the input of two successive SWTs in Fig. 1(b). SWT introduces the window-based multihead self-attention (W-MSA) instead of the multihead self-attention (MSA). The time complexity of MSA and W-MSA is as follows:

$$\Omega(\text{MSA}) = 4h \cdot wC^2 + 2(h \cdot w)^2 C \tag{1}$$

$$\Omega(\text{W-MSA}) = 4h \cdot wC^2 + 2M^2 h \cdot wC \tag{2}$$

where $h \cdot w$ is the number of input patches, C is any dimension after the projection of the LE layer, and m is the window size. Different from MSA, the corresponding computational complexity of W-MSA shows a corresponding linear relationship with the number of patches, which is far lower than the square relationship of MSA. Therefore, it is crucial to improve the computational efficiency of the model.

Shifted window-based multihead self-attention (SW-MSA) aims to realize information exchange between windows. In successive SWT blocks, W-MSA and SW-MSA are used alternately. As shown in Fig. 1(b), if W-MSA is used in the nth layer, then SW-MSA is needed in the $n + 1$ layer. The computations for successive SWT blocks are

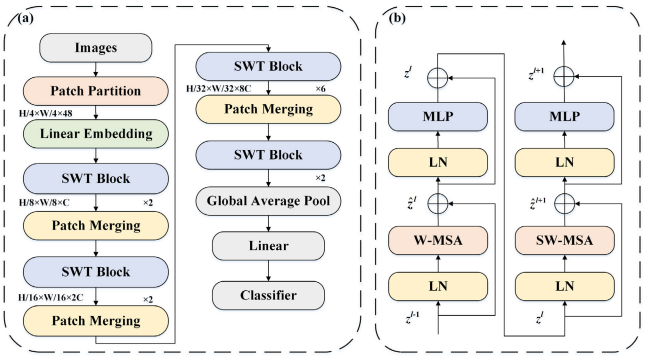

Fig. 1. Model architecture of the SWT. (a) SWT architecture. (b) Two successive SWT blocks.

TABLE I
PSEUDOCODE OF RF

Algorithm1: RF

Input: $M_D = [X_1, X_2, X_3 \ldots, X_j]$, $X_j \in R$, g and k.
Output: $w[e_i]$
Initialize: $w[e_i] = 0$
For $j = 1$ *to* g *do*
  Sample m_j is randomly selected in M_D.
  Search for the k nearest neighbor sample H of the same class as m_j, and search the k nearest neighbor sample S of m_j from different classes.
  For $e = 1$ *to* i *do*
  Update the $w[e_i]$ by Eq.(7).
  End for
  Sort the $w[e_i]$.
End for

defined by the following equations, in which (3) and (4) denote, respectively, the output features of W-MSA and multilayer perception (MLP) within the first block, while (5) and (6) correspond to the output features of SW-MSA and MLP within the second block:

$$\hat{z}^l = \text{W-MSA}\big(\text{LN}\big(z^{l-1}\big)\big) + z^{l-1} \tag{3}$$

$$z^l = \text{MLP}\big(\text{LN}\big(\hat{z}^l\big)\big) + \hat{z}^l \tag{4}$$

$$\hat{z}^{l+1} = \text{SW-MSA}\big(\text{LN}\big(z^l\big)\big) + z^l \tag{5}$$

$$z^{l+1} = \text{MLP}\big(\text{LN}\big(\hat{z}^{l+1}\big)\big) + \hat{z}^{l+1} \tag{6}$$

where $\hat{z}^l$ and z^l denote the output features of the W-MSA and the MLP for block l, respectively. Also, $\hat{z}^{l+1}$ and z^{l+1} denote the output features of the SW-MSA and the MLP for block $l+1$, respectively.

B. ReliefF

Assuming that the label set of a given diagnostic task is Y, there is a label space with $Y = [y_1, y_2, y_3, \ldots, y_i]$ and $y_i \in R$ is the class label space. The dataset for fault diagnosis is M_{D}, where $M_{\text{D}} = [X_1, X_2, X_3, \ldots, X_j]$, and $X_j \in R$ represents the jth sample. After SWT training, the corresponding fault features are extracted. The fault features of a sample are $E = [e_1, e_2, e_3, \ldots, e_k]$, and $e_k \in R$ represents the kth feature. The goal of RF is to randomly select a sample X_j from M_{D}, then find the K nearest neighbor sample H from the sample similar to X_j, and find the K nearest neighbor sample S from the sample different from X_j. Finally, it will update the feature weight $w[e_i]$ [28]. Its pseudocode is shown in Table I.

In addition, its mathematical process is as follows:

$$\begin{aligned} w[e_i] &= -\sum_{j=1}^{k} \text{diff}\big(e, X_i, H_j\big)/(g \cdot k) + \sum_{C \neq \text{class}(R_i)} \\ &\quad \times \left[\frac{P(Y)}{1 - P(\text{class}(X_i))} \sum_{j=1}^{k} \text{diff}\big(e_i, X_i, M_j(Y)\big)\right]/(g \cdot k) \end{aligned} \tag{7}$$

$$\text{diff}\big(e, X_i, H_j\big) = \begin{cases} 0, & \text{if } e \text{ is discrete and } X_i[e] = H_j[e] \\ \dfrac{\big|X_i[e] - H_j[e]\big|}{\max(e) - \min(e)}, & \text{if } e \text{ is continuous} \\ 1, & \text{if } e \text{ is discrete and } X_i[e] \neq H_j[e] \end{cases} \tag{8}$$

where class(X_i) represents the class label, and diff(e, X_i, H_j) represents the distance between the X_i and H_j on the feature e. For g and k represent the number of sample points and the nearest neighbor number, respectively, they are generally determined according to the actual situation of the dataset. $P(Y)$ represents the probability of class Y. The contributions of different classes of samples are weighted by the prior probability of their $P(Y)$, and $M_j(Y)$ represents the jth sample in class Y.

C. DS Evidence Theory

Since it is difficult for a single sensor to fully reflect the working condition of the bearing and to provide redundancy to ensure the proper functioning of the system, the DS evidence theory [29] is introduced in this study, which contains the following basic definitions.

1) Basic Probability Distribution Function: The basic probability distribution function is also known as mass function. Under the recognition framework Θ (the elements in Θ are mutually exclusive), if the mapping mass: $2^{\Theta} \to [0, 1]$ satisfies the following process:

$$\text{mass}(\emptyset) = 0, \quad \sum_{A \subset \Theta} \text{mass}(A) = 1. \tag{9}$$

2) Confidence Function: If the mapping Bel:$2^{\Theta} \to [0, 1]$ satisfies

$$\text{Bel}(A) = \sum_{A \subset B} \text{mass}(A). \tag{10}$$

It is a confidence function defined on Θ. The set function Bel(A) represents the confidence given to subset A and all its subsets.

Assuming that two sensors are given in a fault diagnostic task, their confidence functions are Bel_1 and Bel_2, respectively. The corresponding basic probabilities are assigned to mass_1 and mass_2, respectively. Then, the basic probability distribution function of the system is as follows:

$$\text{mass}(A) = \frac{1}{K} \sum_{A_1 \cap A_2 = A} \text{mass}_1(A_1)\text{mass}_2(A_2) \tag{11}$$

where k is a normalization factor and has the following definitions:

$$K = \sum_{A_1 \cap A_2 = \emptyset} \text{mass}_1(A_1)\text{mass}_2(A_2). \tag{12}$$

III. Method

To realize effective fault diagnosis under low-quality data, a decision fusion enhanced diagnostic method based on SWT-RF and DS is proposed in this article. It can not only adaptively extract deep fault features, but also capture the diverse information of different knowledge thresholds for complementary fusion. In this section, three important parts of this method are introduced, in which the SWT and RF are used for deep fault features extraction and optimization, respectively, and a decision fusion strategy based on the DS evidence theory is used to obtain diagnostic results.

A. SWT-Based Deep Fault Feature Extraction

Consider a vibration signal consisting of N points, $\mathbf{X} = [x_1, x_2, x_3, \ldots, x_N]$, and $x_i \in \mathbf{R}$ represents the ith data point of the sample. It is converted into a vibration signal image of $H \times W \times 3$, and then input into the SWT. If the whole image is divided into several $n \times n$ patches, the transformation of an X_d vibration signal image is formulated as follows:

$$\begin{cases} N = H \times W, \quad \left(H \geq W \in \mathbf{N}^*\right) \\ (H, W) = \arg\min(|H - W|) \\ X = H \times W \times 3 \end{cases} \tag{13}$$

$$X_d = \left[\frac{H}{n}, \frac{W}{n}, 3n^2\right] \tag{14}$$

where H and W represent the length and width of the vibration signal image, respectively; X represents the initial dimension of the vibration signal image; and X_d represents the dimension that the vibration signal image is divided into several patches.

Then, the X_d dimensional signal is transformed linearly by LE. This step is to convert the vector dimension of the signal into a value that can be processed by SWT. Here, the hyperparameter C is introduced, which is set as 96 according to the experience of the hyperparameter setting [19]. Therefore, after LE processing, the X_d dimensional signal is formulated as follows:

$$X_{d_1} = \text{LE}\left(\text{flatten}\left[\frac{H}{n}, \frac{W}{n}, C\right]\right) \tag{15}$$

where flatten represents the feature flattening operation in LE.

The output data of the LE layer are input into the first SWT block, and the output dimension does not change during this process. Then, three pairs of patch merging (PM) and SWT blocks are downsampled to reduce the resolution. Considering the downsampling rate θ_i ($i = 2, 3, 4$) in three stages, which are usually defined as $1/2n$, $1/4n$, and $1/8n$, the dimension of the output feature is formulated as follows:

$$\begin{cases} X_{d_2} = \text{Swim}\left(\text{PM}\left[\frac{H}{2n}, \frac{W}{2n}, 2C\right]\right) \\ X_{d_3} = \text{Swim}\left(\text{PM}\left[\frac{H}{4n}, \frac{W}{4n}, 4C\right]\right) \\ X_{d_4} = \text{Swim}\left(\text{PM}\left[\frac{H}{8n}, \frac{W}{8n}, 8C\right]\right) \end{cases} \tag{16}$$

where X_{d_2}, X_{d_3}, and X_{d_4} represent the deep feature dimensions output by the SWT block, respectively, in the θ_i ($i = 2$, 3, and 4) stages.

Finally, deep features are input to the global average pooling (GAP) layer and the linear layer (LL). Then, the final deep fault features are output, which are recorded in E as follows:

$$E = \text{LL}\left(\text{GAP}\left[X_{d_4}\right]\right). \tag{17}$$

B. RF-Based Feature Optimization

According to (17), the model generates $8C$ fault features for a single sample, which are then input into the classifier to get the class labels. For the classifier, the feature length of a single sample is too long to highlight the significant features and finally leads to the decline of model diagnostic performance. Therefore, it is necessary to optimize its features.

To optimize the deep fault features extracted by SWT, RF is applied, and its formulation is as follows:

$$w[e_i] = \text{ReliefF}(D \cdot E, Y), \quad i \leq D \tag{18}$$

where $D \cdot E$ represents the fault feature set of the sample, Y represents the class label set of the sample, and $w[e_i]$ represents the weight of the feature.

Then, the fault features are sorted based on the feature weights, of which the features close to 0 are filtered out. The filtered algorithm can be described as follows:

$$F = \text{Sorted}(w[e_i]), \quad w[e_i] \leq 0.1 \tag{19}$$

where e_i represents a single feature of the sample, and F represents the optimized features of the sample.

C. DS Evidence Theory-Based Decision Fusion Strategy

Assuming that the testing set M_{D} contains n fault classes, the confusion matrix can be described as follows:

$$\text{CM}_a = \begin{bmatrix} \text{cm}_{11} & \text{cm}_{12} & \ldots & \text{cm}_{1n} \\ \text{cm}_{21} & \text{cm}_{22} & \ldots & \text{cm}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cm}_{n1} & \text{cm}_{n2} & \ldots & \text{cm}_{nn} \end{bmatrix} \tag{20}$$

where $a(1, 2, \ldots, p)$ is the serial number of the sensor, and the diagonal elements $\text{cm}_{ii} (i = 1, 2, \ldots, n)$ represent the probability of each fault class that can be correctly classified, while the other elements cm_{ij} ($j = 1, 2, \ldots, n$) represent the percentage of j-class samples classified as i-class versus samples classified as i-class.

To measure the uncertainty of classification ability and determine the contribution rate to the diagnostic results, the credibility of the classifier can be described as follows:

$$\omega_{ij} = \text{cm}_{ij} \tag{21}$$

$$\omega_{aj} = \mathrm{cm}_{jj} / \sum_{i=1}^{n} \mathrm{cm}_{ij} \tag{22}$$

$$\gamma = \sum_{i=1}^{n} \omega_{ii} \tag{23}$$

where ω_{aj} is the local credibility of the classifier identifying the j-class, and γ is the global credibility of the classifier.

Therefore, for the output result P_{aj} of a certain classifier, the basic probability assignment is as follows:

$$p'_{aj} = \omega_{aj} p_{aj} / \sum_{j=1}^{4} \omega_{aj} p_{aj} \tag{24}$$

$$\mathrm{mass}_a(S_1, S_2, \ldots, S_n, \Theta) = \left(\gamma_a p'_{a1}, \gamma_a p'_{a2}, \ldots, \gamma_a p'_{an}, 1-\gamma_a\right) \tag{25}$$

where $\Theta = (S_1, S_2, \ldots, S_n, \Theta)$ is the recognition framework, and S_n is the nth fault class of rolling bearings.

D. Framework of the Proposed Method

The enhanced diagnostic method is based on the proposed SWT-RF and decision fusion strategy to realize efficient diagnosis under the low-quality data condition, typically like insufficient samples with low SNR. It can fully use the original data collected by different axial signal sensors, extract effectively deep fault features, enhance the expression ability of significant features, and realize the information complementarity under different information thresholds. The framework of the proposed method is shown in Fig. 2, and its detailed procedure can be divided into five steps.

Step 0: Vibration signal data are collected by a plurality of axial sensors and then divided into training and testing sets.

Step 1: Vibration signal samples are converted into the images of $H \times W \times 3$ and are input to the SWT model, which can be divided into several $n \times n$ dimensional patches. Then, the deep fault features are adaptively extracted in four stages of SWT. Finally, $8C$ features for each sample are output in the GAP layer.

Step 2: To enhance the information expression ability of significant features and improve the diagnostic performance of the classifier, the features' output by SWT is optimized by RF. Finally, the optimized features are input to the classifier for fault diagnosis, and the basic probability assignment of each working condition is output.

Step 3: The DS evidence theory is used to fuse the two-axial data, and the diagnostic performance of the whole model is improved based on the complementary information features of different knowledge thresholds.

Step 4: The diagnostic performance of the proposed method is tested by using testing samples from different time sources, and the final enhanced diagnostic results are obtained.

TABLE II
XJTU-SY FAULTY BEARING INFORMATION

Class	Operating condition	Element	Class	Operating condition	Element
1	35 Hz/12 kN	OR	**6**	40 Hz/10 kN	OR
2	35 Hz/12 kN	OR	**7**	40 Hz/10 kN	IR+Ball+ Cage+ OR
3	35 Hz/12 kN	OR	**8**	40 Hz/10 kN	IR
4	35 Hz/12 kN	Cage	**9**	40 Hz/10 kN	IR
5	35 Hz/12 kN	IR+OR	**10**	40 Hz/10 kN	OR

IV. CASE STUDY

A. Description of the Experimental Data

To better evaluate the diagnostic performance of the proposed method, XJTU-SY bearing dataset was used for comparative experiments [30]. The fault data are collected from the accelerated life test rather than fault injection, which is closer to the situation of low-quality data and can verify the practicality of the proposed method. As shown in the test rig in Fig. 3, the testbed is composed of an alternating current (ac) motor, a motor speed controller, a support shaft, two support bearings, a hydraulic loading system, and so on. In addition, the sampling frequency is set to 25.6 kHz and the operating conditions include different radial force and rotating speed. In addition, the experimental rig contains two sensors with different axes, which can carry out the enhanced fault diagnosis experiment of multisensor fusion.

In this case, ten classes data of bearing healthy conditions are selected from two axial (X and Y) acceleration sensors to form the dataset, containing a training set and a testing set. Specific information on the dataset used in this case is presented in Table II, in which OR represents the outer ring fault and IR represents the inner ring fault. The sampling length of a single sample is set to 1024, and 200 samples of each class condition are collected to form the training set, including 2000 samples in total. Moreover, to simulate the real situation with insufficient samples, training and testing samples are collected from different sources. Therefore, 2000 samples are selected from another experiment to form the testing set. Additionally, according to Step 1 of the procedure, samples are converted into $32 \times 32 \times 3$ dimensional images, so that they can be input into the SWT model.

B. Ablation Experiment

To verify the effectiveness of the proposed decision fusion SWT-RF model, an ablation experiment has been carried out. Both X- and Y-axes data are used, and the SWT-RF and SWT are used for verification. In this experiment, the hyperparameter settings of SWT are referenced to [22], and the detailed hyperparameters are shown in Table III. In addition, according to the characteristics of the bearing fault diagnosis with the input image size of 32×32, the downsampling rates are set as 2, 4, 16, and 32 for the corresponding four stages, and the size of patches is set as 2×2.

The experiment has been repeated ten times to avoid particularity, and the average diagnostic accuracy and standard deviation are shown in Fig. 4.

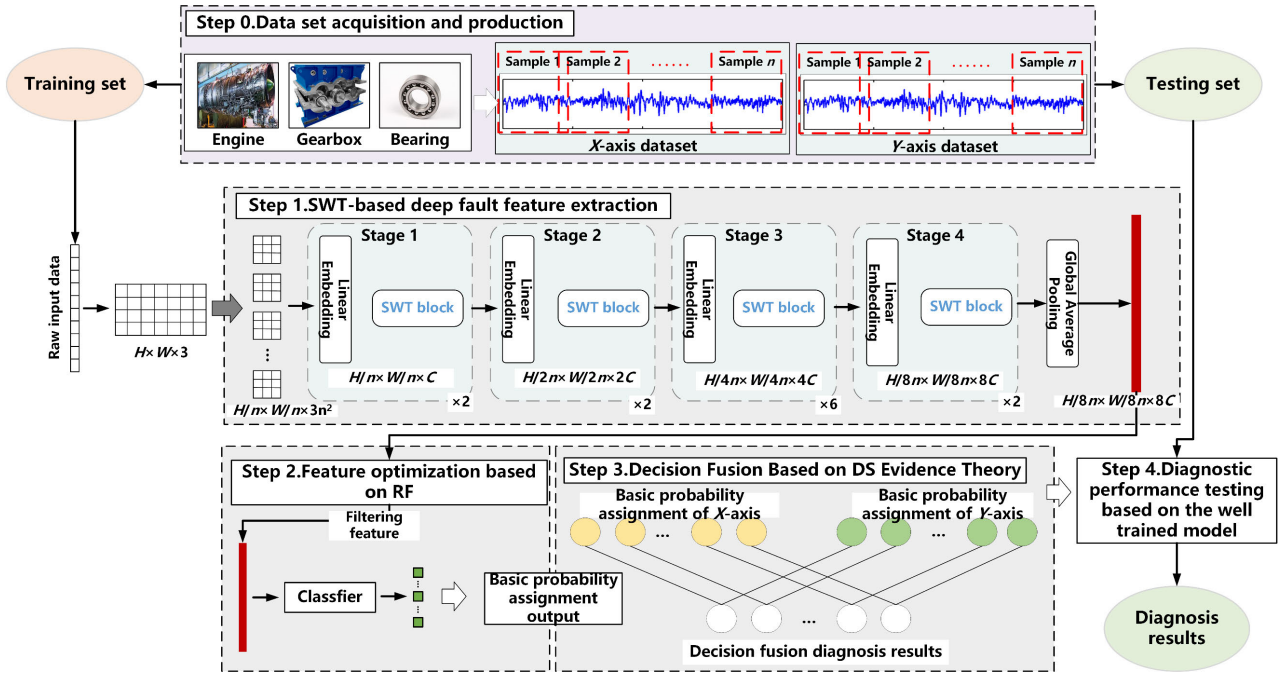

Fig. 2. Frame of the proposed enhanced diagnostic method.

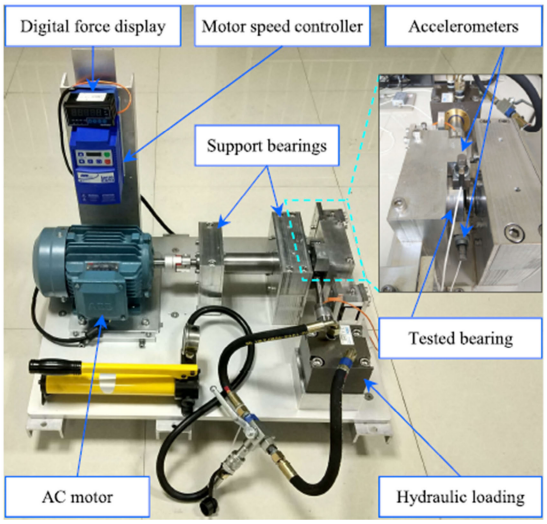

Fig. 3. Bearing accelerated life test rig of the XJTU-SY.

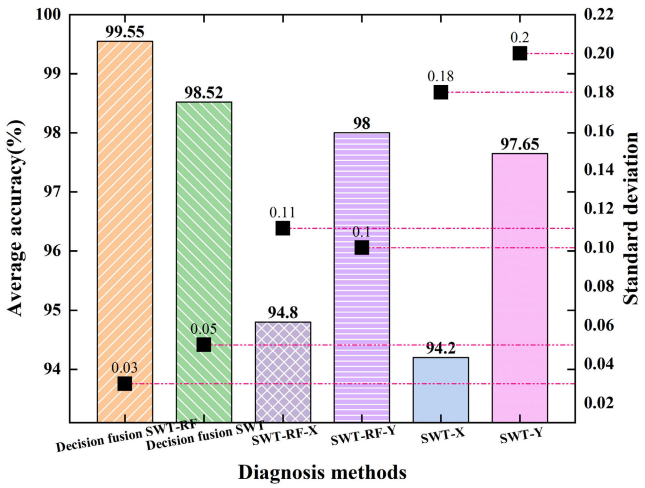

Fig. 4. Comparisons of the diagnostic results of different methods.

TABLE III
SWT HYPERPARAMETER

NO.	Hyperparameter	Stage 1	Stage 2	Stage 3	Stage 4
1	Down-sampling rate	2	4	8	16
2	Output size	16×16	8×8	4×4	2×2
3	Layer numbers	2	2	6	2
4	Window size	7×7			
5	Output dimension	96	192	384	768
6	Head numbers	3	6	12	24

The results in Fig. 4 reveal that the average diagnostic accuracies of the six mentioned methods are 99.55%, 98.52%, 94.8%, 98%, 94.2%, and 97.65%, respectively, in which the proposed decision fusion SWT-RF method reaches the highest diagnostic accuracy. Meanwhile, it is worth noting that the standard deviation of the decision fusion SWT-RF method is only 0.03, which is the smallest among all the methods. These results show that the proposed method has better diagnostic accuracy and stability than the other compared methods.

In addition, the results obtained by the proposed method are better than those of SWT-RF-X, while they are also better than those of SWT-RF-Y. This validates the effectiveness of the integrated DS evidence theory-based decision fusion strategy. Similarly, by comparing the results obtained by the proposed decision fusion SWT-RF and decision fusion SWT, SWT-RF-X, and SWT-X, as well as the results obtained by the SWT-RF-Y and SWT-Y, the effectiveness of the proposed RF-based feature optimization has been validated.

In detail, after introducing the RF-based feature optimization, the performance of the SWT model is improved to a

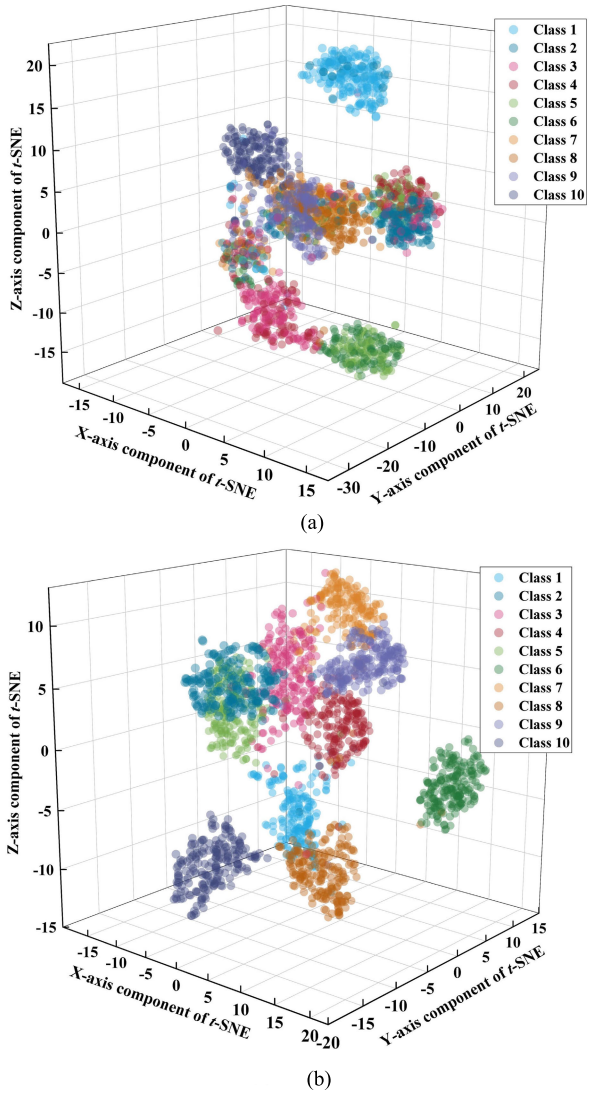

Fig. 5. Visualization comparison of the extracted feature before and after the RF based on t-SNE. (a) Before RF. (b) After RF.

certain extent. Not only the average diagnostic accuracies for the X- and Y-axes are improved by 0.6% and 0.35%, respectively, but also the stability of the model is enhanced. The standard deviations for SWT-RF-X and SWT-RF-Y are both improved and are only 0.11 and 0.1, which are 0.07 and 0.1 lower than those before feature optimization, respectively.

To further reflect the superiority of the proposed method, this article introduces t-distributed Stochastic Neighbor Embedding (t-SNE) to visually show the deep feature extraction ability by using the X axial data, as shown in Fig. 5. Fig. 5(a) and (b) shows the visualization of extracted feature based on t-SNE before and after RF feature optimization, respectively. Although the basic clustering can be realized before feature optimization shown in Fig. 5(a), its separability is poor due to the interference of weak correlation features, which is also the fundamental reason for the low diagnostic performance based on the SWT. Moreover, the RF-based feature optimization can achieve the clear feature clustering and separate effectively the individual features. This significantly enhances the distribution of feature information and improves the diagnostic performance.

For the sensors in different axes can contain different fault information, this study adopts the X- and Y-axes data for decision fusion, aiming at further improving the diagnostic performance of the model. To illustrate the effectiveness of the DS-based decision fusion strategy, the confusion matrices of the proposed decision fusion SWT-RF, SWT-RF-X, and SWT-RF-T are output, and their results are compared in Fig. 6. In Fig. 6, the oblique typeface symbolizes the diagnostic accuracy. To denote notable enhancements in accuracy resulting from decision fusion, circular markers are employed, accompanied by directional arrows indicating the fusion process. Complementary data along the Y-axis is utilized to augment the diagnostic outcome classes, demarcated by brown circles. Similarly, the X-axis data are employed as supplementary information to enrich the diagnostic class, distinguished by the green circle.

Comparing the results obtained by SWT-RF-X and SWT-RF-Y methods shown in Fig. 6(a) and (c), it can be found that the SWT-RF-X can achieve better performance for class 9 fault condition, of which the diagnostic accuracy is 98.5% versus 92%. In contrast, the SWT-RF-Y performs better for classes 1, 3, 4, and 7 with 100%, 100%, 99%, and 96% diagnostic accuracy, respectively. For SWT-RF-X method, they are only 90.5%, 83.5%, 91%, and 92.5%, respectively.

In general, the diagnostic accuracy obtained by one axis performs better than that of the other axis can be summarized for three main reasons.

1) *Bearing Installation:* When the bearing is installed more accurately or tightly along one axis, its vibration signal tends to be more sensitive to faults.
2) *Mechanical Structural Design:* The design of the mechanical structure may facilitate the propagation of vibration to one axis.
3) *Uneven Wear of Bearings:* Nonuniform wear of bearings along different axes may lead to a more distinct vibration signal in one axis.

In this case, the X-axis is potentially more effective in detecting faults associated with axial loads. Meanwhile, as the load undergoes variations, the Y-axis exhibits higher diagnostic accuracy, particularly evident in classes 2 and 6. Additionally, the Y-axis is more sensitive to faults related to rolling elements, which is noticeable in classes 1 and 3.

It can be summarized that different axial data carry different fault information, making it more sensitive to certain categories of fault modes. Therefore, using different axial data for feature extraction and optimization can help the model learn more fault information and distinguish the difference between different fault modes. This is witnessed by the results obtained by the decision fusion SWT-RF method shown in Fig. 6(b). It takes advantage of the complementary features of information in X- and Y-axes data and can achieve almost 100% for all the ten classes. This further proves the effectiveness of the decision fusion strategy under the principle of information complementarity.

C. Comparison With Methods on Different Fusion Levels

Typically, there are three fusion methods, data fusion, feature fusion, and decision fusion, respectively. To verify

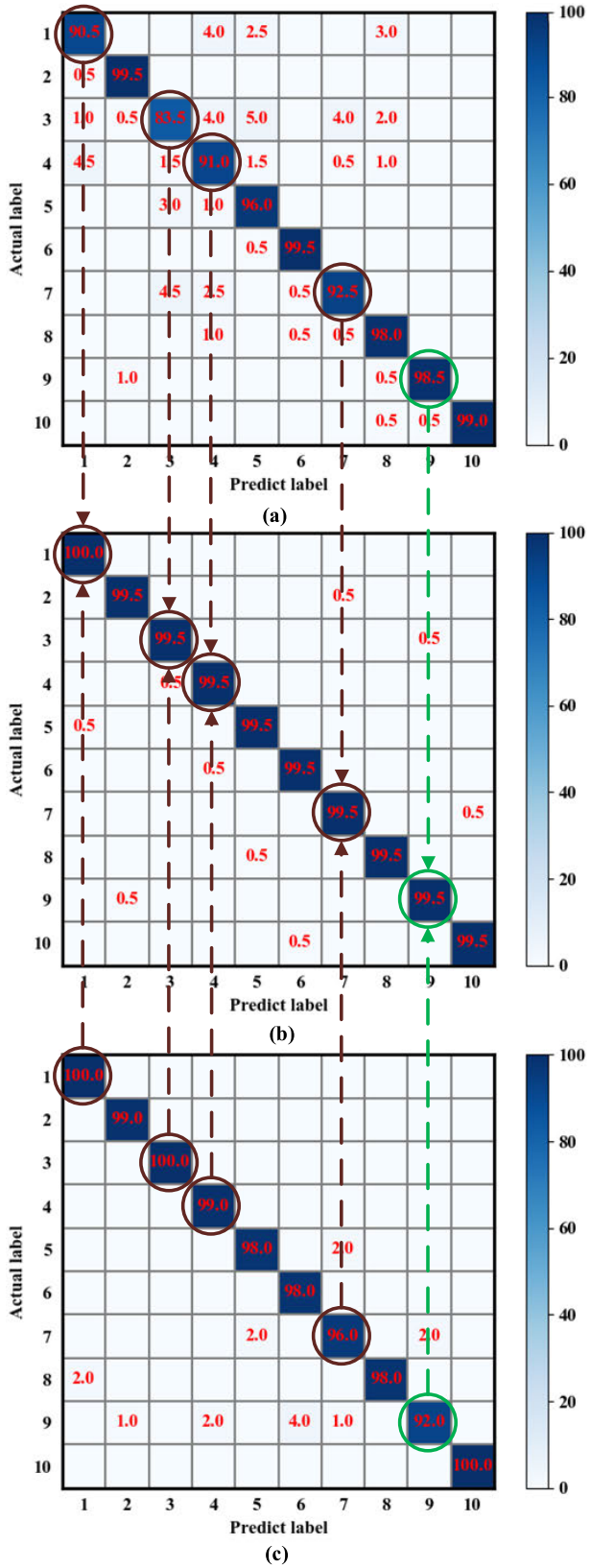

Fig. 6. Comparison of the diagnostic confusion matrices for the three different methods. (a) SWT-RF-X method. (b) Proposed decision fusion SWT-RF method. (c) SWT-RF-Y method.

the superiority of decision fusion in different fusion levels, the data and feature fusion methods are used for comparative experiments. The data fusion model adopts 2-D-gcforest [31]

TABLE IV
COMPARISON OF THREE DIFFERENT FUSION METHODS

Fusion level	Method	Sample length	Data form	Accuracy
Data	2D-gcforest [31]	2048	2D image	71.4%
Feature	FCNN [32]	1024	2D image and vibration signal	95%
Decision	Decision fusion SWT-RF	1024	2D image	99.55%

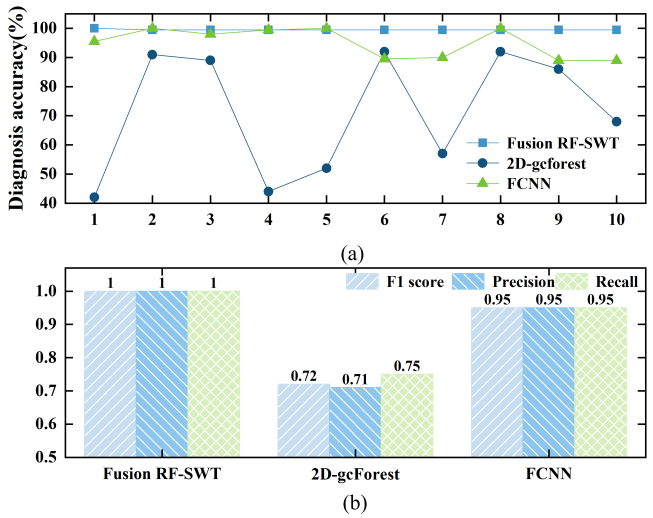

Fig. 7. Comparisons of the diagnostic results of the three different fusion-level methods. (a) Labels of classes. (b) Diagnosis method.

and the feature fusion model adopts the fusion convolutional neural network (FCNN) [32]. Because this article studies the direct diagnostic method based on the vibration signal without any preprocessing, the data form used in this experiment is the original vibration signal and the 2-D image. To quantitatively evaluate the performance of different methods, three evaluation indexes are introduced, respectively, F1 score, precision, and recall rate. The diagnostic results of the three methods are shown in Table IV and Fig. 7.

As shown in Table IV and Fig. 7, among the three fusion levels, the 2-D-gcforest based on data fusion has the lowest diagnostic accuracy, and its accuracy is only 71.4%. Moreover, its F1 score and recall rate are only 0.72 and 0.75, respectively. On the one hand, this indicates that the fusion of a single data form at the data level leads to unsatisfactory diagnostic results. On the other hand, it also indicates its poor diagnostic robustness and stability.

In terms of the feature fusion method, the FCNN can achieve 95% diagnostic accuracy, which reaches an acceptable level. However, there is still a difference of 4.55% in diagnostic accuracy compared to the proposed method. Their differences in diagnostic accuracy are mainly reflected in classes 1, 6, 7, 9, and 10, as shown in Fig. 7(a). The biggest difference occurs in classes 9 and 10, which are both almost 10.5%. In addition, although the 2-D-gcforest has poor diagnostic performance, it can reach 92% diagnostic accuracy for class 6, which is 2.5% higher than that of FCNN.

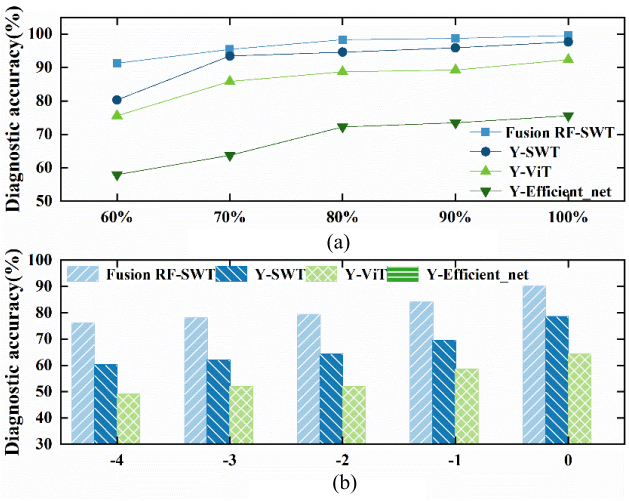

Fig. 8. Comparisons of the diagnostic results for the proposed method and the three mainstream methods. (a) Proportion of training samples. (b) Different SNRs.

For the proposed decision fusion method, its diagnostic accuracy is the highest shown in Table IV, which is 99.55%. Meanwhile, its F1 score, accuracy, and recall all reach the maximum value of 1 shown in Fig. 7(b). It validates its superiority over the data and feature fusion methods.

D. Comparisons Under Low-Quality Data Condition

To verify the robustness of the proposed decision fusion SWT-RF method under low-quality data, the comparative experiments were carried out under different training sample ratios and SNRs. In this experiment, the mainstream intelligent fault diagnostic methods based on deep learning architecture were used for comparison, including SWT, ViT, and Efficient net. Finally, the experimental results are shown in Fig. 8.

1) Condition of Insufficient Samples: To evaluate the ability of the proposed method under insufficient samples, comparative experiments under the proportions of training sample on testing sample varying from 60% to 100% were carried out. As shown in Fig. 8(a), the average diagnostic accuracies of the proposed decision fusion SWT-RF are 91.3%, 95.4%, 98.3%, and 99.55% when the proportions are 60%, 70%, 80%, 90%, and 100%, respectively. Especially, although the proportion is only 60%, the diagnostic accuracy can still reach 91.3%. These results validate the effectiveness of the proposed method under insufficient samples and unbalanced healthy-faulty conditions.

In addition, when the proportion is 60%, the diagnostic accuracies of the SWT, ViT, and Efficient net are 80.3%, 75.55%, and 58.2%, respectively, which are obviously lower than that of the proposed method. This shows that the efficiency of intelligent diagnostic method with a single knowledge threshold will be greatly reduced under unbalanced datasets. Moreover, the performance of the proposed method is superior to the other mainstream methods under insufficient sample condition.

2) Condition of Low SNR: To evaluate the robustness of the proposed method under low SNR, comparative experiments under different degrees of Gaussian white noise injected into the vibration signal were carried out. The signal pitch ratio is introduced to represent the noise level, which is defined as

$$\mathrm{SNR} = 10\log_{10}(P_r/P_n) \tag{26}$$

where P_r and P_n represent the power of the original vibration signal and the power of the additional noise, respectively. As shown in Fig. 8(b), the diagnostic results of the proposed method are higher than those of SWT, ViT, and Efficiency net under different SNRs. Its average diagnostic accuracy is 76.1%, 78%, 79.4%, 84.1%, and 90.2% under SNRs of −4, −3, −2, −1, and −0 dB, respectively. These results are the best of all methods. It shows that the proposed method has good robustness to low SNR signals.

TABLE V
COMPARISON OF DIAGNOSTIC RESULTS FOR THE XJTU-SY BEARING DATASET

Method	Training-testing proportion	Number of classes	Accuracy (%)
LSTM-Cascade CatBoost [33]	3:1	4	99.33
PDC-LR-HCNN [34]	7:3	10	99.33
DFCNN without data preprocessing [11]	1:1	10	83.04
Fusion SWT-RF	1:1	10	99.55

E. Comparison With Existing Intelligent Diagnostic Methods

To verify the superiority of the proposed method in bearing fault diagnosis, as well as its generalization over different bearing dataset, an experiment compares the proposed method with two existing diagnostic methods based on the XJTU-SY bearing dataset, including long short-term neural network cascade forest (LSTM-Cascade CatBoost) [33], adaptive resource allocation DNN (PDC-LR-HCNN) [34], and DFCNN [11], of which the detailed diagnostic results are shown in Table V.

As shown in Table V, LSTM-Cascade CatBoost and PDC-LR-HCNN train the model based on a large training-testing proportion, 300% and 233%, respectively. However, in this situation, they only achieve 99.33% diagnostic accuracy, which is still 0.22% lower than that of the proposed method based on 99.55% proportion. For a fair comparison, the data preprocessing process has been eliminated for the DFCNN method, which is also not involved in the other methods. Then, its diagnostic accuracy is only 83.04%, which is the lowest result. When incorporated with the data preprocessing process, the DFCNN can improve greatly the diagnostic accuracy [11]. However, it marginally surpasses the proposed method by only 0.45%, while it needs more time spent on data preprocessing. In addition, it requires special knowledge of the signal processing, such as Complementary Ensemble Empirical Mode Decomposition (CEEMD) and short time fourier transform (STFT). More importantly, when the proposed method is introduced with the data preprocessing, it can also realize 100% diagnostic accuracy. Therefore, these results show that the proposed method can complete an efficient fault diagnosis on the XJTU-SY bearing dataset, which validates its generalization and superiority.

V. Conclusion

To realize effective diagnosis under low-quality data, this study proposes a decision fusion method based on SWT and RF to enhance the diagnosis of rolling bearing faults. First, the vibration signals of multiple sensors are directly input to the improved SWT model to adaptively extract their deep feature. This avoids the dependence of the basic model on the prior knowledge of signal processing under low-quality data. Then, the RF is introduced to optimize the extracted deep fault features to improve the expression ability of significant features. In addition, DS decision fusion is integrated to fuse the basic probability assignment of multisensors. Thus, the performance of multiknowledge threshold diagnosis is improved at the decision level. Finally, five comparative experiments are conducted, and the results verify the effectiveness and generalization of the proposed decision fusion SWT-RF method for fault diagnosis of rolling bearing, as well as its superiority over the existing intelligent diagnostic methods.

It is an important challenge to effectively use the transformer as the basic model for the data distribution imbalance problem. To bridge this gap, it is a good research direction to combine the SWT with transferring learning.

References

[1] W. Deng, Z. Li, X. Li, H. Chen, and H. Zhao, "Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3508509, doi: 10.1109/TIM.2022.3159005.

[2] S. Lu, R. Yan, Y. Liu, and Q. Wang, "Tacholess speed estimation in order tracking: A review with application to rotating machine fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 7, pp. 2315–2332, Jul. 2019, doi: 10.1109/TIM.2019.2902806.

[3] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, M. E. Orchard, and M. Saif, "Critical wind turbine components prognostics: A comprehensive review," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9306–9328, Dec. 2020, doi: 10.1109/TIM.2020.3030165.

[4] S. Gao, S. Shi, and Y. Zhang, "Rolling bearing compound fault diagnosis based on parameter optimization MCKD and convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3508108, doi: 10.1109/TIM.2022.3158379.

[5] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019, doi: 10.1007/s11265-018-1378-3.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[7] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020, doi: 10.1109/TII.2019.2943898.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, in Proceedings of Machine Learning Research, Long Beach, CA, USA, vol. 97, Jun. 2019, pp. 6105–6114.

[9] B. Hu, J. Tang, J. Wu, and J. Qing, "An attention EfficientNet-based strategy for bearing fault diagnosis under strong noise," *Sensors*, vol. 22, no. 17, p. 6570, Aug. 2022, doi: 10.3390/s22176570.

[10] L. Zhang, H. Zhang, and G. Cai, "The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3514212, doi: 10.1109/tim.2022.3178483.

[11] J. Chen, C. Lin, B. Yao, L. Yang, and H. Ge, "Intelligent fault diagnosis of rolling bearings with low-quality data: A feature significance and diversity learning method," *Rel. Eng. Syst. Saf.*, vol. 237, Sep. 2023, Art. no. 109343, doi: 10.1016/j.ress.2023.109343.

[12] H. Fang et al., "CLFormer: A lightweight transformer based on convolutional embedding and linear self-attention with strong robustness for bearing fault diagnosis under limited sample conditions," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 3504608, doi: 10.1109/tim.2021.3132327.

[13] X. Li, J. Cheng, H. Shao, K. Liu, and B. Cai, "A fusion CWSMM-based framework for rotating machinery fault diagnosis under strong interference and imbalanced case," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5180–5189, Aug. 2022, doi: 10.1109/TII.2021.3125385.

[14] Z.-G. Hou, H.-W. Wang, S.-L. Lv, M.-L. Xiong, and K. Peng, "Siamese multiscale residual feature fusion network for aero-engine bearing fault diagnosis under small-sample condition," *Meas. Sci. Technol.*, vol. 34, no. 3, Mar. 2023, Art. no. 035109, doi: 10.1088/1361-6501/aca044.

[15] W. Li, X. Zhong, H. Shao, B. Cai, and X. Yang, "Multi-mode data augmentation and fault diagnosis of rotating machinery using modified ACGAN designed with new framework," *Adv. Eng. Informat.*, vol. 52, Apr. 2022, Art. no. 101552, doi: 10.1016/j.aei.2022.101552.

[16] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Long Beach, CA, USA: Curran, 2017, pp. 6000–6010.

[17] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," Oct. 2020, pp. 21–22, *arXiv:2010.11929*.

[18] J. A. Baktash and M. Dawodi, "Gpt-4: A review on advancements and opportunities in natural language processing," 2023, *arXiv:2305.03195*.

[19] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time–frequency transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings," *Mech. Syst. Signal Process.*, vol. 168, Apr. 2022, Art. no. 108616, doi: 10.1016/j.ymssp.2021.108616.

[20] H. W. Fan, N. G. Ma, X. H. Zhang, C. Y. Xue, J. T. Ma, and Y. Yan, "New intelligent fault diagnosis approach of rolling bearing based on improved vibration gray texture image and vision transformer," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, early access, p. 14, May 2022. [Online]. Available: http://sage.cnpereading.com/paragraph/article/?doi=10.1177/09544062221085871, doi: 10.1177/09544062221085871.

[21] D. Lv, H. Wang, and C. Che, "Multiscale convolutional neural network and decision fusion for rolling bearing fault diagnosis," *Ind. Lubrication Tribol.*, vol. 73, no. 3, pp. 516–522, May 2021, doi: 10.1108/ilt-09-2020-0335.

[22] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. 18th IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002, doi: 10.1109/iccv48922.2021.00986.

[23] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct./Nov. 2003, doi: 10.1023/a:1025667309714.

[24] J. Fei, X. Lv, Y. Cao, and S. Li, "A hierarchical decision fusion diagnosis method for rolling bearings," *Appl. Sci.*, vol. 11, no. 2, p. 739, Jan. 2021, doi: 10.3390/app11020739.

[25] Y. Fu, Y. Liu, and Y. Yang, "Multi-sensor GA-BP algorithm based gearbox fault diagnosis," *Appl. Sci.*, vol. 12, no. 6, p. 3106, Mar. 2022, doi: 10.3390/app12063106.

[26] X. Tang, Z. Xu, and Z. Wang, "A novel fault diagnosis method of rolling bearing based on integrated vision transformer model," *Sensors*, vol. 22, no. 10, p. 3878, May 2022, doi: 10.3390/s22103878.

[27] J. Y. Chen, J. J. Cui, C. Y. Lin, and H. J. Ge, "An intelligent fault diagnostic method based on 2D-gcForest and $L_{2,p}$-PCA under different data distributions," *IEEE Trans. Ind. Inform.*, vol. 18, no. 10, pp. 6652–6662, Oct. 2022, doi: 10.1109/TII.2022.3168325.

[28] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 483–519, Mar. 2013, doi: 10.1007/s10115-012-0487-8.

[29] R. R. Murphy, "Dempster–Shafer theory for sensor fusion in autonomous mobile robots," *IEEE Trans. Robot. Autom.*, vol. 14, no. 2, pp. 197–206, Apr. 1998, doi: 10.1109/70.681240.

[30] B. Wang, Y. Lei, N. Li, and N. Li, "A hybrid prognostics approach for estimating remaining useful life of rolling element bearings," *IEEE Trans. Rel.*, vol. 69, no. 1, pp. 401–412, Mar. 2020, doi: 10.1109/TR.2018.2882682.

[31] X. Liu et al., "Deep forest based intelligent fault diagnosis of hydraulic turbine," *J. Mech. Sci. Technol.*, vol. 33, no. 5, pp. 2049–2058, May 2019, doi: 10.1007/s12206-019-0408-9.

[32] F. Xue, W. Zhang, F. Xue, D. Li, S. Xie, and J. Fleischer, "A novel intelligent fault diagnosis method of rolling bearing based on two-stream feature fusion convolutional neural network," *Measurement*, vol. 176, May 2021, Art. no. 109226, doi: 10.1016/j.measurement.2021.109226.

[33] M. M. Yang, W. Z. Liu, W. X. Zhang, M. Wang, and X. Fang, "Bearing vibration signal fault diagnosis based on LSTM-cascade CatBoost," *J. Internet Technol.*, vol. 23, no. 5, pp. 1155–1161, 2022, doi: 10.53106/160792642022092305024.

[34] S. Ning and K. Du, "Research on intelligent fault diagnosis of rolling bearing based on adaptive resource allocation deep neural network," *IEEE Access*, vol. 10, pp. 62920–62931, 2022, doi: 10.1109/ACCESS.2022.3182467.

Jiayu Chen (Member, IEEE) received the B.S. degree in transportation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013, and the Ph.D. degree in system engineering from Beihang University, Beijing, China, in 2019.

Since 2019, he has been an Assistant Professor with the College of Civil Aviation and the Civil Aviation Key Laboratory of Aircraft Health Monitoring and Intelligent Maintenance, Nanjing University of Aeronautics and Astronautics. His research interests include prognostics and health management, intelligent fault diagnosis, and maintainability design and analysis for complex systems, such as aviation equipment and armored vehicle.

Cuiying Lin received the B.S. degree in traffic management from the Civil Aviation Fight University of China, Chengdu, China, in 2020, and the M.S. degree in transportation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2023.

Her research interests include the intelligent fault diagnosis of rotating machinery, such as turbine wind gearbox, bearing, and pump.

Qinhua Lu received the B.S. degree in electrical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2020, where he is currently pursuing the M.S. degree in transportation engineering with the College of Civil Aviation.

His research interests include the intelligent fault diagnosis of rotating machinery, health management, and maintenance strategy for complex system.

Chaoqi Yang is currently pursuing the B.S. degree in transportation engineering with the College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

His research interests include the intelligent fault diagnosis of rotating machinery under unbalanced data.

Peng Li (Member, IEEE) received the B.S. degree in control engineering from Central South University, Changsha, China, in 2013, the M.S. degree in reliability engineering from Beihang University, Beijing, China, in 2016, the Ph.D. degree in computer application technology from the University of Chinese Academy of Sciences, Beijing, in 2022, and the Ph.D. degree in systems engineering from the City University of Hong Kong, Hong Kong, China, in 2023.

Since 2016, he has been an Assistant Professor with the Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing. His research interests include prognostic and health management, and reliability engineering.

Pingchao Yu received the B.S. degree in transportation engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013, and the Ph.D. degree from Beihang University, Beijing, China, in 2019.

Since 2019, he has been an Associate Professor with the Nanjing University of Aeronautics and Astronautics. His research interests include vibration analysis, nonlinear rotor dynamics, and joint structure dynamics.

Hongjuan Ge received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively, and the Ph.D. degree in electrical machinery and electrical appliances from the Nanjing University of Aeronautics and Astronautics, Nanjing, in 2007.

She is currently a Professor with the College of Civil Aviation and also a Senior Researcher with the Civil Aviation Key Laboratory of Aircraft Health Monitoring and Intelligent Maintenance, Nanjing University of Aeronautics and Astronautics. Her research interests include the fault diagnosis of aviation equipment, aircraft airworthiness technology, and management.