RESEARCH ARTICLE

Xu WANG, Hongyang GU, Tianyang WANG, Wei ZHANG, Aihua LI, Fulei CHU

Deep convolutional tree-inspired network: a decision-tree-structured neural network for hierarchical fault diagnosis of bearings

Abstract The fault diagnosis of bearings is crucial in ensuring the reliability of rotating machinery. Deep neural networks have provided unprecedented opportunities to condition monitoring from a new perspective due to the powerful ability in learning fault-related knowledge. However, the inexplicability and low generalization ability of fault diagnosis models still bar them from the application. To address this issue, this paper explores a decision-tree-structured neural network, that is, the deep convolutional tree-inspired network (DCTN), for the hierarchical fault diagnosis of bearings. The proposed model effectively integrates the advantages of convolutional neural network (CNN) and decision tree methods by rebuilding the output decision layer of CNN according to the hierarchical structural characteristics of the decision tree, which is by no means a simple combination of the two models. The proposed DCTN model has unique advantages in 1) the hierarchical structure that can support more accuracy and comprehensive fault diagnosis, 2) the better interpretability of the model output with hierarchical decision making, and 3) more powerful generalization capabilities for the samples across fault severities. The multiclass fault diagnosis case and cross-severity fault diagnosis case are executed on a multicondition aeronautical bearing test rig. Experimental results can fully demonstrate the feasibility and superiority of the proposed method.

Xu WANG, Tianyang WANG (✉), Fulei CHU
State Key Laboratory of Tribology, Tsinghua University, Beijing 100084, China; Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China
E-mail: wty19850925@126.com

Xu WANG, Hongyang GU, Wei ZHANG, Aihua LI
High-Tech Research Institute of Xi'an, Xi'an 710025, China

1 Introduction

Bearings are widely used in rotating machinery, and their condition monitoring is crucial to the precision and reliability of mechanical systems [1]. In recent years, with the development of sensor technology and information science, the research of data-driven mechanical fault diagnosis has developed rapidly. In particular, the emergence of deep learning (DL) technology makes fault diagnosis based on deep neural network (DNN) redefine the most advanced performance [2,3]. Different from the top–down physics-based fault diagnosis methods, data-driven methods can resist the effect of environmental noise and equipment complexity, and update the model in a timely manner as the monitoring data increase to obtain more accurate fault recognition performance. Compared with traditional machine learning methods, DNN has more powerful data feature extraction capabilities and less reliance on prior knowledge or hand-made features. As bottom–up condition monitoring approaches, the DL-based fault diagnosis methods enjoy an evident advantage in saving resources and have attracted extensive attention due to their better effectiveness and robustness.

Researchers tackle data-driven fault diagnosis mainly for fault type discrimination and fault severity identification, where the former is to know the fault location of the components, and the latter tries to analyze the degradation level related to the physical size of defects [4,5]. The DL-based fault diagnosis approaches tend to learn the signal patterns associated with a particular fault type or severity by DNN methods, e.g., autoencoder [6], generative adversarial net [7], recurrent neural network [8], deep belief network [9], and

convolutional neural network (CNN) [10]. Generally, vibration signals, acoustic emission signals, electrical signals, temperatures, pressures, and sound signals can be used for condition monitoring and fault diagnosis of bearings. Among them, vibration signals are widely used in the fault diagnosis of bearings [11]. According to the different structures of the networks, the vibration data are usually transformed into different forms for analysis. For example, the time domain data or frequency domain data of the signal are generally used as the input of the recurrent neural network network, which is more suitable for the analysis of sequence data; the CNN model is suitable for processing high-dimensional data and performs well in analyzing the time-frequency distribution (TFD) of the signal, such as continuous wavelet transform (CWT) distribution [12], short-time Fourier transform distribution [13], and Chirplet transform [14,15]. More researchers regard the diagnosis problem as a single-level multiclassification problem and attempt to achieve higher classification accuracies by designing a more complex network. Most researchers ignore the logic of fault diagnosis and only focus on the judgment of fault type, but not the fault severity associated with the magnitude of the failure [16]. Few approaches consider fault types and severities together when transforming the diagnosis task into a common classification task for processing, where each fault mode and each fault severity are treated as a specific label. For example, Zhao et al. [17] converted the raw signals of bearings into grey images and directly adopted the CNN model for fault diagnosis. In the experiment, three fault types and three fault severities were considered at most, and the fault diagnosis task was transformed into a common 10-category classification task for processing. Analogously, Minhas et al. [18] recognized the different fault types and severities of bearings as a single-level multiclassification problem by the complementary ensemble empirical mode decomposition and support vector machine (SVM) classifier. Pan et al. [19] proposed a novel symplectic geometry matrix machine method for the classification of bearings with different fault types and severities. Wen et al. [20] adopted a hierarchical CNN model for the classification of bearings with different fault types and severities. These explorations are beneficial to obtaining precise fault recognition no matter which data form or network structure is used. However, considering the actual application scenarios faced by fault diagnosis, the following issues are often overlooked:

a) The logic of fault diagnosis is usually ignored. Although these models meet the demand of joint diagnosis of fault types and severities, they exponentially increase the complexity of the classification task and require more labeled data as well as complex models for the expected diagnosis performance. The substantial increase of classification complexity for these approaches brings greater challenges to the classification models.

Moreover, it is not in line with the logical cognition of experts to mix different fault attributes for identification.

b) Most of the works only consider the input and output but not the justifiable prediction of the intermediate process. Although DNN-based fault diagnosis networks have strong knowledge learning capabilities, explaining the discriminative details of intermediate decisions is still difficult, making the diagnosis results provided by DNN models often difficult to be trusted. The interpretability of the model has always been recognized as a topic worth exploring, which is of great importance for fault diagnosis tasks [21,22].

c) Data-driven fault diagnosis methods often have weak generalization ability for new categories but a strong dependence on the label information of the samples [23,24]. However, the fault severity of collected samples will not be exhaustive in the real case; consequently, the diagnostic model often fails in dealing with test samples belonging to unseen fault severity classes. The limitation of the cross-severity generalization becomes a large obstacle for existing models to be popularized and applied in engineering.

The existing research has several useful explorations on these problems. To deal with the first problem, the hierarchical diagnosis strategy that identifies the type and severity has been adopted in several works. However, the existing methods still stay at acquiring a hierarchical output by using a hierarchical Softmax classifier [25,26] rather than the hierarchical decision in the intermediate stage of the diagnosis models. In our view, a more effective approach would be to apply hierarchical decision rules to deal with this problem. The fault type should be judged first, and then the fault severity judgment can be made based on the prior knowledge provided by the fault type judgment. Regarding the second problem, the interpretability of DL models has always been continuously explored in various fields and still a difficult challenge worth numerous studies. An accepted way to improve the interpretability of the model is the estimation of the decision uncertainty [27,28]. For the third problem, the cross-severity identification of bearing faults is a new subject to the best of our knowledge. The hierarchical diagnosis framework and hierarchical decision rules can help the cross-severity generation of fault type diagnosis for these samples with unseen fault severities. The effective usage of fault type knowledge in the training data will greatly support the decision making of test samples across severities but still difficult to achieve with existing methods.

To address the mentioned issues, a novel deep convolutional tree-inspired network (DCTN) is explored for the hierarchical fault diagnosis of bearings. The proposed model effectively integrates the advantages of CNN and decision tree methods by designing an output decision layer similar to the decision tree structure to fine-tune the weights of the convolutional layers

reversely. The signals are converted into TFDs by the CWT method because the time-domain information is conducive to fault severity analysis, and the frequency domain information is more sensitive to different fault types [29,30]. The CNN-based architecture is used as a pre-training model. During pre-training, a Softmax classifier is connected to the backbone CNN. The powerful feature learning ability of CNN can ensure the effectiveness of fault-related feature extraction from the samples. After that, the Softmax classification layer and the fully-connected layer are replaced by the tree-structured decision layer to execute the hierarchical fault diagnosis decision in sequence. The hierarchical diagnosis helps reduce the task complexity of diagnosis and improve the accuracy of fault diagnosis. More importantly, the tree-inspired network designed in this paper enables the model to diagnose across fault severities of bearings. The multiclass fault diagnosis case and cross-severity fault diagnosis case are executed on a multicondition aero-engine bearing test rig to verify the feasibility and superiority of the proposed method. Given the state-of-the-art works in fault diagnosis, the proposed DCTN-based fault diagnosis approach has unique advantages in the following aspects:

a) The tree-structured hierarchy is helpful for a more accurate, comprehensive diagnosis decision-making. The judgment of the fault type provides a priori knowledge for fault severity diagnosis, which is beneficial to improving diagnosis accuracy. The multilevel decision information with the progressive determination of fault type and fault severity are more in line with maintenance cognition in engineering.

b) The interpretability of the model output is explored through the hierarchy structure. The decision tree model is one of the most interpretable machine learning methods, but its weak knowledge learning ability has always limited its application. The proposed DCTN model effectively integrates decision tree with the CNN model and can provide the hierarchy and uncertainties of decision-making to improve the interpretability of the model output.

c) The proposed model has more powerful generalization capabilities for samples with unseen fault severity categories. The final diagnosis decision of fault attributes is made from multiple views by the embedded hierarchical tree-structured decision layers. The trained model

can be generalized well for fault type diagnosis even if the sample has an unseen fault severity category. To our best knowledge, this paper carries out the first exploration in cross-severity fault diagnosis of bearings.

The rest of this paper is organized as follows. Section 2 presents the methodologies of the proposed DCTN model. Section 3 presents the DCTN-based fault diagnosis approach of bearings. Section 4 shows the two case studies for the fault diagnosis of aeronautical bearings and discusses in detail the superiority of the DCTN method over other related works. Finally, Section 5 presents the conclusions and conceivable future works.

2　Proposed deep convolutional tree-inspired network

Figure 1 shows the schematic view of the proposed DCTN model, which mainly consists of three convolutional layers, one pre-trained fully-connected layer, and one tree-structured decision layer. The proposed DCTN model takes CNN as the backbone network to learn the fault-related features in the TFDs of bearing signals. A tree-structured decision layer is then embedded into the pre-trained CNN to replace the fully-connected layer for fine-tuning. The weights of the fully-connected layer in the CNN are deduced to the node attribute representation of the tree-structured decision layer. Different types of nodes are given corresponding weights according to the logical relationship in the tree-structured decision layer. By defining a new supervision loss function and then fine-tuning the model weights, the leaf nodes and seed nodes can support the effectiveness of hierarchical fault diagnosis. The leaf nodes can also acquire an effective fault classification ability in the proposed tree-structured decision layer to deal with cross-severity fault diagnosis tasks, which lays the foundation for the better generalization of the model.

2.1　Learning the neural backbone of the seed nodes

Fault type discrimination and fault severity identification have an inherent logical relationship, which the decision making of the model should also correspond to. The decision tree model has good interpretability because the decision of each node has clear physical meanings.
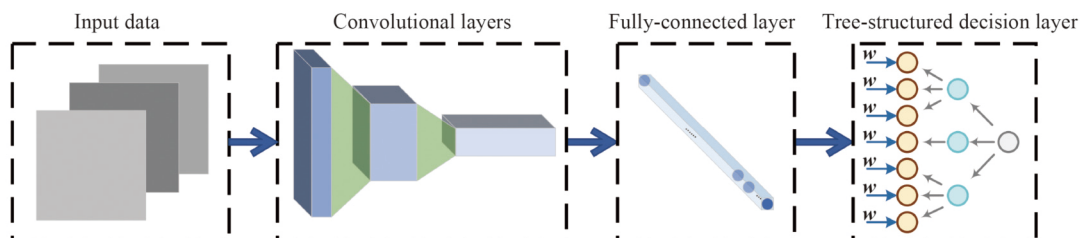

Fig. 1　Schematic view of the proposed deep convolutional tree-inspired network model.

However, weak knowledge learning ability has greatly limited its application for a long time [31–33]. Although decision trees are interpretable and simple to use, they are prone to overfitting, can be less robust to small changes in training data, and generally rely on heuristic algorithms. In recent years, many researchers have made several attempts to improve the performance of the decision tree model, such as the random forest model [34], the deep forest model [35], and several deep decision tree models [36,37]. These methods optimize the structure of the decision tree to improve the classification performance and retain interpretability. However, the performances of these methods are still not as good as the state-of-the-art DNN models even in small data sets.

As a feedforward neural network, CNN has shown strong feature extraction ability in the processing of sequence, image, and video data. Generally, the basic structure of a CNN includes two kinds of layers, one of which is the feature extraction layer. The input of each neuron is connected to the local receptive field of the previous layer, and the local features are extracted by the feature extraction layers. Once the local feature is extracted, the positional relationship between it and other features is also determined. The second is the feature mapping layer, in which each layer is connected with multiple feature maps, and each feature map corresponds to a classification plane. With the deepening of exploration, several studies have found that embedding the decision tree model into DNN models can better guarantee the recognition accuracy of the model [38–40]. Hence, the tree-structured decision layer is embedded into the CNN model for ensuring the performance of hierarchical fault diagnosis and reinforcing the interpretability of the model. The constructed DCTN model takes the convolutional layers as the backbone. Inspired by the decision tree, a tree-structured decision

layer is embedded in the backbone model to provide hierarchical diagnosis logic and is endowed with the ability to diagnose across fault severities. To understand the distribution of embedded features extracted by the convolutional layers better, the weights of the final fully connected layer are used to induce the hierarchy and embedded decision rules.

The constructed CNN neural backbone owns three convolutional layers and one fully-connected layer. The convolutional layers can learn the fault-related features from the TFDs of the signals by model training. The fully-connected layer can reduce the dimension of the features learned from the convolutional layers to adapt to the dimension of the seed nodes in the tree-structured decision layer. The layer details of the backbone CNN are shown in Table 1, where N represents the number of samples, $R \times R$ represents the dimension of the TFD, and K is usually equal to the number of sample categories. The weight of the model is initialized by pre-training to guarantee the accuracy of the entire model. During pre-training, a Softmax classifier is connected to the backbone CNN. After that, the Softmax classification layer and the fully-connected layer are replaced by the tree-structured decision layer for fine-tuning. The Adam optimization algorithm is used to optimize the model and speed up the convergence. The cross-entropy function is used as the loss function in the model pre-training.

The cross-entropy of the prediction loss $H(p, q)$ is

$$H(p, q) = -\sum_x \{p(x) \log q(x) + [1 - p(x)] \log [1 - q(x)]\}, \qquad (1)$$

where x is the input features of the Softmax classifier, $p(\cdot)$ is the probability distribution of the predicted output, and $q(\cdot)$ is the probability distribution of the actual output.

Table 1 Layer details of backbone CNN

Layer	Set	Output shape
Input	–	$N \times R \times R \times 1$
2D convolution layer	Kernel size: 1×1, channel: 16, stride: 1	$N \times R \times R \times 16$
Batch normalization	Feature number: 16, eps: 10^{-5}	$N \times R \times R \times 16$
ReLU activation	–	$N \times R \times R \times 16$
2D max pooling layer	Kernel size: 2×2	$N \times R/2 \times R/2 \times 16$
2D convolution layer	Kernel size: 3×3, channel: 32, stride: 1	$N \times R/2 \times R/2 \times 32$
Batch normalization	Feature number: 32, eps: 10^{-5}	$N \times R/2 \times R/2 \times 32$
ReLU activation	–	$N \times R/2 \times R/2 \times 32$
2D max pooling layer	Kernel size: 2×2	$N \times R/4 \times R/4 \times 32$
2D convolution layer	Kernel size: 3×3, channel: 64, stride: 1	$N \times R/4 \times R/4 \times 64$
Batch normalization	Feature number: 64, eps: 10^{-5}	$N \times R/4 \times R/4 \times 64$
ReLU activation	–	$N \times R/4 \times R/4 \times 64$
Adaptive average pooling layer	Kernel size: 1×1	$N \times 1 \times 1 \times 64$
Fully-connected layer	Batch size: 64×1×1, out features: K, no bias	$N \times K$

2.2 Interpreting the tree-structured decision layer

Inspired by the decision tree model, the sequential decision of the hierarchical fault diagnosis is carried out by the tree-structured decision layer. The structure of the decision sequence is designed according to the underlying logic of the fault diagnosis task. Two decision hierarchies are generated, yielding the two main fault diagnosis levels, namely, fault type and fault severity. Figure 2 shows that the first hierarchy is to determine the fault type of the input sample to acquire the corresponding superclass attribute, whereas the second hierarchy is to determine the fault severity of the input sample to acquire the corresponding subclass attribute. The parentheses indicate the calculation method of prediction probability at this node. The input–output relationship between leaf nodes and root nodes is not as simple as that in the neural network.

According to the structure of the decision tree, the decision nodes in the first decision level are defined as the leaf nodes, and the decision nodes in the second decision level are defined as the seed nodes. The number of leaf nodes is consistent with the number of sample fault types, whereas the number of seed nodes is consistent with the number of sample categories K. The seed nodes correspond to the weights of the fully-connected layer of the pre-trained backbone CNN. The weight matrix $W \in \mathbb{R}^{K \times L}$ of the fully-connected layer is obtained by the back-propagation training with Softmax classifier. According to the network structure setting in Table 1, the dimension value L of the sample feature after convolutional layers should be 64.

In the pre-training stage, distance d_j between the feature and each classification hyperplane is

$$d_j = \frac{w_j \cdot x}{\|w_j\|}, \ j = 1, 2, \ldots, K, \tag{2}$$

where $w_j \in \mathbb{R}^{1 \times L}$ is the weight vector of the jth vector in weight matrix W of the fully-connected layer, and $x \in \mathbb{R}^{L \times 1}$ refers to the input feature vector of the tree-structured decision layer, which is also the output of the final convolutional layer. The multiclassification model sets a hyperplane for each category and divides the feature space through multiple hyperplanes. One region corresponds to one category. Distance d_j refers to the similarity between the test sample and the labeled sample in the jth hyperplane. The prediction scope $z_j = w_j \cdot x$ corresponding to K categories can be acquired by the fully-connected layer. Then, it is mapped to prediction probabilities by the Softmax classifier as

$$\hat{y}_j = \text{softmax}(z_j) = \frac{e^{z_j}}{\sum\limits_{j=1}^{K} e^{z_j}}, \tag{3}$$

where $\hat{y}_j$ refers to the predicted probability for the jth category and satisfies

$$\sum_{j=1}^{K} \hat{y}_j = 1. \tag{4}$$

The weight of seed nodes directly adopts from the pre-trained fully-connected layer. In this way, the identification ability of the subclass is equivalent to that of the pre-trained CNN model, which guarantees the performance of the model. The corresponding weight sw_i ($i = 1, 2,$ or 3) of the ith leaf node is obtained by adding the weight of its seed nodes. Taking the structure shown in Fig. 2 as an example, the following relationship can be obtained:

$$\begin{cases} sw_1 = \sum\limits_{j=1}^{3} w_j, \\ sw_2 = w_4, \\ sw_3 = \sum\limits_{j=5}^{7} w_j, \\ \sum\limits_{i=1}^{3} (sw_i \cdot x) = 1, \\ \sum\limits_{j=1}^{7} (w_j \cdot x) = 1. \end{cases} \tag{5}$$
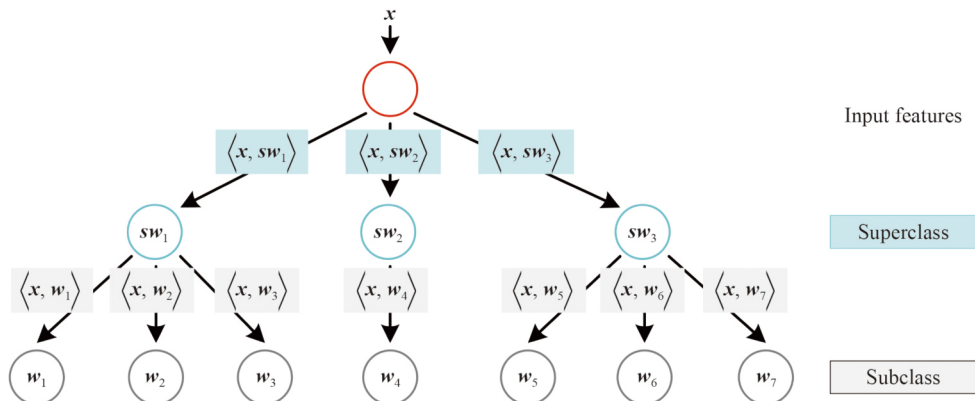

Fig. 2 Weight propagation of tree-structured decision layer.

2.3   Fine-tuning with decision loss

Fine-tuning is of great importance to improve the overall performance of the model. If only the weights obtained by the pre-training model are used and the weights are determined according to Eq. (5), the accuracy of the overall model will be the same as that of the pre-training, and the advantages of the hierarchical structure cannot be exerted. The Softmax function at each decision node is used to generate the corresponding decision probabilities because probabilities are naturally better interpretable. Taking the structure in Fig. 2 as an example, the following relation is met after the fine-tuning:

$$\begin{cases} \sum_{j=1}^{3}\left(\boldsymbol{w}_j^* \cdot \boldsymbol{x}\right) = 1, \\ \boldsymbol{w}_4^* = s\boldsymbol{w}_2, \\ \sum_{j=5}^{7}\left(\boldsymbol{w}_j^* \cdot \boldsymbol{x}\right) = 1, \end{cases} \tag{6}$$

where $\boldsymbol{w}_j^* \in \mathbb{R}^{1 \times L}$ refers to the weight vector of the jth tree-structured decision layer after fine-tuning. The classification of the superclass can provide prior knowledge for the identification of the subclass through the fine-tuning. The DCTN model designed in this paper fine-tunes the weights of the backbone model and the tree-structured decision layer, and performs Softmax classification on all nodes to make the final fault diagnosis decision according to the path probabilities. In detail, the probability of correct prediction for seed nodes is defined as P (subclass). The probability of correct prediction of leaf nodes is defined as P (superclass). Hence, the probability of overall correct prediction of the model is calculated as

$$P(\ell) = P(\text{subclass}|\text{superclass}) \cdot P(\text{superclass}), \tag{7}$$

where $P(\ell)$ refers to the path probabilities of the tree-structured decision layer, ℓ refers to the overall prediction. The final class prediction is defined as

$$q(\hat{\ell}) = \arg\max P(\ell), \tag{8}$$

where $q(\hat{\ell})$ is the predicted probabilities of the tree-structured decision layer. The loss function $H^*(p(\cdot), q(\cdot), p(\cdot), q(\cdot))$ of the tree-structured decision layer is calculated based on the cross-entropy function as

$$H^*\left(p(k), q(\hat{k}), p(\ell), q(\hat{\ell})\right) \\ = H\left(\{p(k)\}_{k=1}^{K}, \{q(\hat{k})\}_{k=1}^{K}\right) + \omega H\left(\{p(\ell)\}_{\ell=1}^{K}, \{q(\hat{\ell})\}_{\ell=1}^{K}\right), \tag{9}$$

where $p(k)$ refers to the true labels of the pre-trained network, $q(\hat{k})$ refers to the predicted probabilities of the pre-trained network, and $p(\ell)$ refers to the true labels of the tree-structured decision layer. The first term on the right side of the equation represents the same cross-entropy function as the pre-trained network to maintain the effectiveness of the original training. The second term is the newly added loss term, which corresponds to all predictions related to the tree decision path probabilities. Super parameter ω is the weight adjusting the pre-trained decision and tree-structured decision.

3   Proposed DCTN-based fault diagnosis approach of bearings

To analyze the ability of multi-fault identification and the capacity of generalization for the superclasses of the proposed DCTN-based fault diagnosis approach, two fault diagnosis tasks, namely, multiclass fault diagnosis and cross-severity fault diagnosis of aeronautical bearings, are carried out. Different fault diagnosis networks are designed corresponding to different tasks, which are described in detail in Subsections 3.2 and 3.3.

3.1   Aeronautical bearing test rig

The bearing dataset is collected by the Politecnico di Torino rolling bearing test rig [41], which is shown in Fig. 3. The aeronautical bearing at the B1 position can be easily removed from its support, allowing checking the response of the system when installing bearings with different fault types and severities. The bearings of the spindle are grease lubricated, whose temperature is limited by a liquid refrigeration circuit.

Table 2 shows the serial number of the damaged bearing, the fault locations and severities, the subclass labels, and the superclass labels. Among them, the superclass is determined according to the location of the fault, marked as N in the table for no fault, I for the inner ring fault, and R for the outer ring fault. Rockwell tools are used to produce localized defects on the elements, resulting in conical indentations on the inner ring or individual rollers. The set fault size is shown in Table 2. Given such a small fault severity, observing the specific fault size is difficult with the existing signal processing methods. The *XYZ* triaxial sensor of the B1 bearing is installed at the A1 position. According to experience, the signals collected in the *Y* direction can better reflect the health status of the bearing.

The operating condition details of the aeronautical bearing are shown in Table 3. Data are collected from aeronautical bearings operating under 17 loads and speeds with a sampling frequency of 51.2 kHz and a sampling time of 10 s. The raw signals of the seven bearings are shown in Fig. 4. The raw signal is unstable and contains some noise in the real case. In our view, this instability of the raw signals makes directly using the signals in the time domain for fault diagnosis difficult. To obtain a better expression of fault features, TFD is used as the basic data form for analysis.
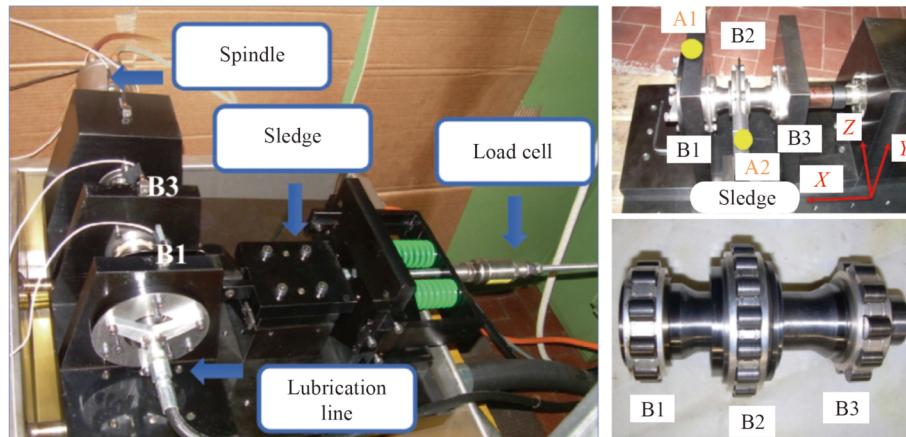
Fig. 3 Overview of Politecnico di Torino rolling bearing test rig.

Table 2 Fault set details of aeronautical bearings

Serial number	Fault location	Fault size/μm	Superclass	Subclass
N-1	No defect	–	N	1
I-2	On the inner ring	450	I	2
I-3	On the inner ring	250	I	3
I-4	On the inner ring	150	I	4
R-5	On a roller	450	R	5
R-6	On a roller	250	R	6
R-7	On a roller	150	R	7

Table 3 Operating condition details of aeronautical bearings

Number	Load/N	Speed/(r·min⁻¹)
C1	0	6×10^3
C2	1000	6×10^3
C3	1400	6×10^3
C4	1800	6×10^3
C5	0	12×10^3
C6	1000	12×10^3
C7	1400	12×10^3
C8	1800	12×10^3
C9	0	18×10^3
C10	1000	18×10^3
C11	1400	18×10^3
C12	1800	18×10^3
C13	0	24×10^3
C14	1000	24×10^3
C15	1400	24×10^3
C16	0	3×10^4
C17	1000	3×10^4

3.2 Time–frequency analysis based on CWT

The CWT method can effectively represent the local characteristics of signals in the time–frequency domain and has proven to be quite suitable for fault analysis of mechanical equipment [13,42]. For signal $s(t)$ in time t and the specified mother wavelet ψ, the CWT function is defined as follows:

$$\text{CWT}(s(t)) = |a|^{-1/2} \int_{-\infty}^{+\infty} \left[s(t)\psi^*\left(\frac{t-b}{a}\right) \right] \mathrm{d}t, \qquad (10)$$

where $a > 0$ is the stretch factor, $b \geqslant 0$ is the shift factor, and $*$ refers to the conjugate operation. The complex Morlet wavelet with bandwidth frequency and center frequency of 3 is selected as the mother wavelet, the scale sequence length is set as 256, and a and b are set as 2 and 5 empirically, respectively.

The TFD of 0.1 s length signals collected in the Y direction from the bearing under condition C17 is shown in Fig. 5. The occurrence of faults is usually accompanied by the increase of the impact components in the time–frequency domain. Furthermore, the impact component distribution varies when the fault location of the bearing is different. The impacts of the I-2 and R-5 bearings locate in different frequency bands. For bearings with the same fault position, the vibration component in the signal increases gradually with the increase of the fault size, that is, from 150 to 450 μm. However, distinguishing bearings N-1, I-4, and R-7 only by observation from the TFDs is difficult. Therefore, more effective models are needed to distinguish different fault bearings. In general, CWT-TFD can effectively characterize the difference of bearing signals in different health states, which lays a good foundation for fault diagnosis.

3.3 DCTN-based hierarchical multiclass fault diagnosis network

The designed DCTN-based hierarchical multiclass fault diagnosis network is shown in Fig. 6. The DCTN-based hierarchical multiclass fault diagnosis approach differs from the existing methods in the strategy for hierarchical decision making. In decision making, the different fault

types corresponding to the superclass labels of the samples are first distinguished. Then, the different fault severities corresponding to each type with the same subclass label of the samples are distinguished. The fully-connected layer in the backbone CNN is replaced with the tree-structured decision layer in the fine-tuning. The training and test samples of the model correspond to seven categories listed in Table 2.
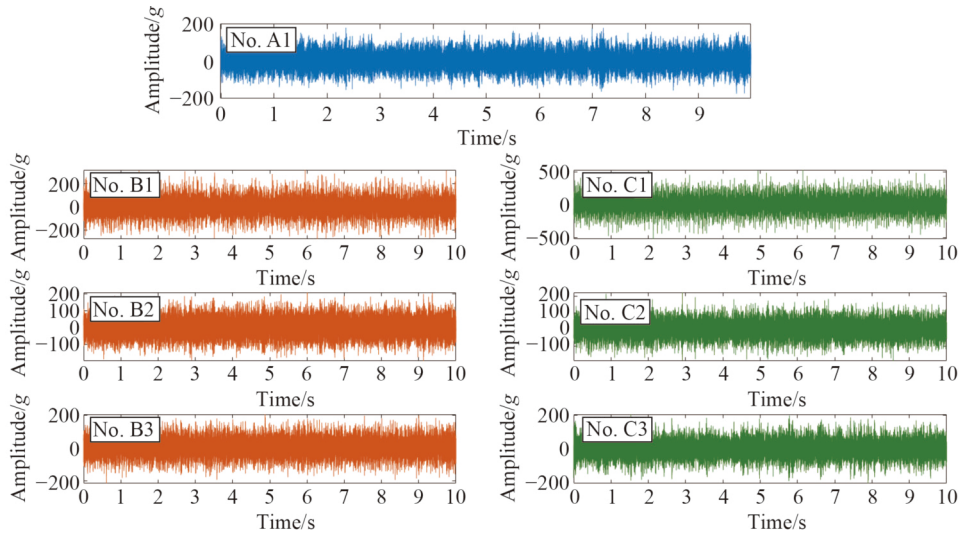

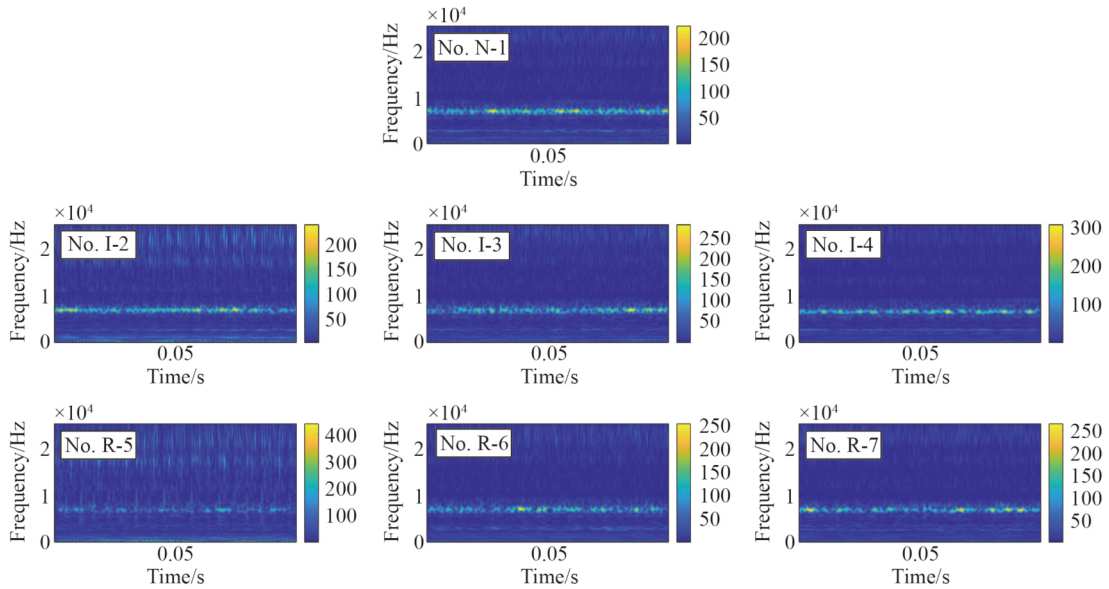**Fig. 4** Raw signals of bearings under different health states.

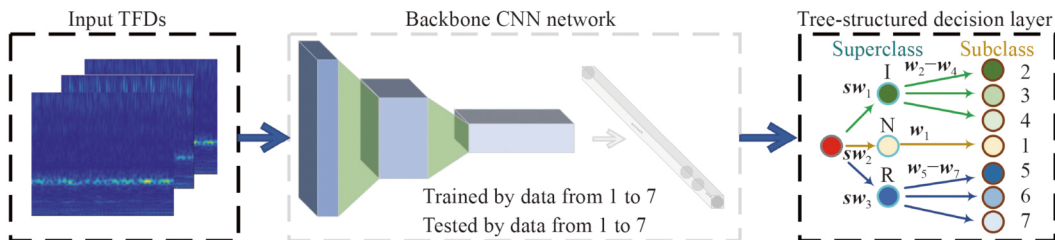**Fig. 5** CWT-based TFD of bearings under different health states.

Fig. 6 Designed DCTN-based hierarchical multiclass fault diagnosis network.

3.4   DCTN-based cross-severity fault diagnosis network

The cross-severity fault diagnosis approach is a new attempt for fault diagnosis models. The designed DCTN-based cross-severity fault diagnosis network is shown in Fig. 7. The cross-severity samples corresponding to the same superclass labels are identified for decision making. For example, bearing I-3 has the same super class as I-2 and I-4, which correspond to the fault located on the inner ring, and bearing R-6 has the same superclass as R-5 and R-7 corresponding to the fault located on the roller. The model shown in Fig. 6 is trained by the samples from bearings N-1, I-2, I-4, R-5, and R-7, and tested by the samples from bearings I-3 and R-6. The purpose of prediction is to identify successfully the superclass labels of the test bearings, whose node weights are initialized by seed nodes that do not contain the corresponding subclass labels of the test bearings.

4   Case studies

This section discusses two fault diagnosis cases, namely, multiclass bearing fault diagnosis task and cross-severity bearing fault diagnosis task. The first case is to verify the diagnosis performance of the proposed DCTN model. The adopted strategy of hierarchical decision-making is expected to be conducive to improve the precision of fault identification. The second case is to verify the generalization ability across different fault severities. The proposed model is built by the Pytorch framework and implemented on a computer with 64-bit Windows 10 system, RAM of 16 GB i5 CPU, and NVIDIA RTX 2080 GPU.

4.1   Case one: multiclass fault diagnosis of bearings

This subsection discusses the diagnosis task of seven fault categories that belong to different fault types and severities. The input data of the model is the TFD matrix generated by the CWT method. For convenience of processing, all the TFD matrices are normalized to the dimension of 100×100 as the standard input of the network model by bilinear interpolation. Each fault category corresponds to 100 signal samples, several of

which are randomly selected as the training set and the rest as the test set. In the training, the batch size of the model is set as 16, and the learning rate is set as 0.01. Moreover, 10% of the training data are randomly extracted as the validation data set. All the training processes can achieve convergence within 200 epochs. To deal with measurement error, the results of fault identification accuracy given in the experiment are averaged after 10 measurements.

Figure 8 illustrates the fault diagnosis performance of bearings under operating condition C17. The ratio of the training data in the whole dataset ranges from 0.1 to 0.9. Theoretically, using more samples in model training is more conducive to the model achieving higher recognition accuracies. The analysis under different training sample sizes is beneficial to comparing the fault diagnosis performance of the proposed model more comprehensively. Parameter sensitivity analysis is also executed as the proportion weight ω in Eq. (9) and is set as 0.1, 0.5, 1, 5, 10, 50, 100, and 500. Theoretically, parameter ω cannot be very large or very small. A large ω will lead to the reduction or even loss of the feature learning ability achieved by pre-training in the fine-tuning of the model, whereas a small ω will prevent the advantages of the designed hierarchical decision from being reflected.

Figure 8 shows that the results are consistent with the above analysis. When the training data ratio is higher than 0.5, the model can achieve an accuracy of 100% under different ω settings. When the ratio is less than 0.5, the recognition accuracy shows a downward trend along with the decrease of the training data. In comparison, the fault diagnosis performance is unsatisfactory when ω is 0.1, 0.5, or 1. When ω is 100 or 500, the performance of the model is relatively better but not as good as the performance of the model when $\omega = 10$. Therefore, we can conclude that the hierarchical diagnosis strategy of the model proposed in this paper is beneficial to improving the accuracy of fault diagnosis, but the knowledge learning ability of the CNN model should also be retained. A better fault diagnosis performance can be reached by the reasonable allocation of the two parts in the final decision.

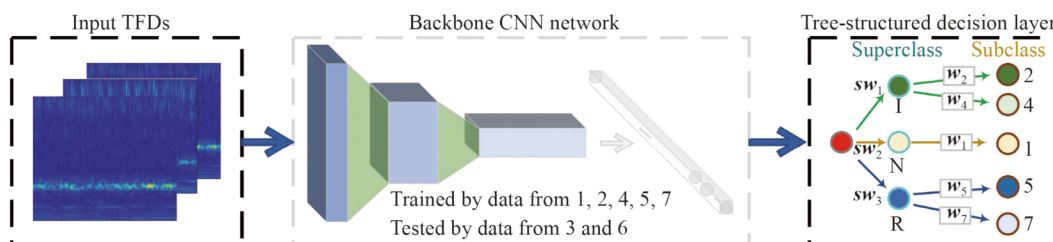Table 4 shows the fault diagnosis performance of bearings with different training data ratios under 17

Fig. 7   Designed DCTN-based cross-severity fault diagnosis network.

operating conditions. The mean accuracies on the right of the table correspond to the recognition accuracy rate under specific conditions. The mean accuracy corresponding to the data collected under condition C4 is the lowest, which is 97.08%. The mean accuracies of the data collected under conditions C5 and C13 are the highest, which are both 99.89%. The performance of fault diagnosis varies under different working conditions, but the overall identification can be considered very effective. The mean accuracies on the bottom of the table correspond to the recognition accuracy rate under a specific ratio. In a comprehensive consideration of all working condition data, the mean recognition accuracy also increases with the increase of the ratio of training samples, and it can be maintained at 100% when the ratio exceeds 0.7. When the ratio is 0.1, the mean recognition rate is the lowest, which is 96.04%. Overall, the proposed DCTN-based multiclass bearing fault diagnosis approach is effective under different operating conditions.

To analyze the performance of the proposed DCTN-based fault diagnosis approach more objectively, it is compared with seven other typical fault diagnosis
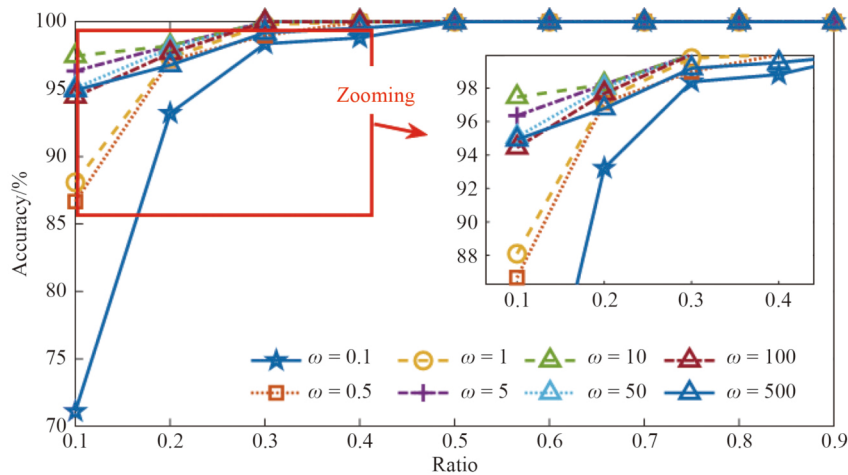

Fig. 8 Fault diagnosis performance of bearings under different ratios and ω settings.

Table 4 Fault diagnosis accuracy of bearings with different training data ratios

Condition	Fault diagnosis accuracy/%									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
C1	96.19	93.21	99.39	100.0	100.0	100.0	100.0	100.0	100.0	98.75
C2	96.67	93.93	96.53	100.0	99.43	98.57	100.0	100.0	100.0	98.35
C3	96.83	88.57	96.94	96.19	99.43	99.29	100.0	100.0	100.0	97.47
C4	97.46	89.81	95.71	92.38	99.43	98.93	100.0	100.0	100.0	97.08
C5	99.03	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.89
C6	97.94	96.43	94.08	100.0	99.71	100.0	100.0	100.0	100.0	98.68
C7	93.65	95.00	96.12	100.0	100.0	100.0	100.0	100.0	100.0	98.31
C8	93.81	95.18	95.31	99.05	98.29	100.0	100.0	100.0	100.0	97.96
C9	97.78	99.82	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.73
C10	97.14	97.68	99.39	100.0	100.0	100.0	100.0	100.0	100.0	99.36
C11	93.81	98.93	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.19
C12	93.81	98.93	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.19
C13	99.05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.89
C14	93.81	95.54	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.82
C15	94.44	98.21	100.0	100.0	99.43	100.0	100.0	100.0	100.0	99.12
C16	93.81	99.82	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.29
C17	97.46	98.21	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.52
Mean	96.04	96.43	98.44	99.27	99.75	99.81	100.0	100.0	100.0	

approaches:

1) The TFD-CNN approach that has the same input and structure as the pre-trained backbone network.

2) The TFD-local binary convolutional neural network (LBCNN) approach that has the same input as the proposed approach and uses the LBCNN for fault identification. The used network structure is the same as the model in Ref. [12].

3) The TFD-PCA-SVM approach that uses the principal component analysis (PCA) method to acquire the sample features from TFDs and adopts the SVM method for fault identification. The penalty parameter and kernel parameters in SVM are selected automatically by grid searching.

4) The TFD-PCA-KNN approach that uses the PCA features and the k-nearest neighbor (KNN) method for fault identification.

5) The TFD-PCA-extreme learning machine (ELM) approach that uses the PCA features and the ELM method for fault identification. The weight matrix and bias of hidden layers in the ELM model can be adjusted automatically.

6) The time-features-SVM approach that uses 14 typical time-domain features of bearing signals, namely, maximum value, minimum value, mean value, peak-to-peak value, rectified mean value, variance, standard deviation, kurtosis, skewness, root-mean-square, corrugation factor, crest factor, impulse factor, and margin factor, and the SVM method for fault identification.

7) The time-features-KNN approach that uses 14 typical time-domain features and the KNN method for fault identification.

8) The time-features-ELM approach that uses 14 typical time-domain features and the ELM method for fault identification.

9) The raw-data-wide deep convolution neural network

(WDCNN) approach that uses the raw signal as input and WDCNN [43,44] for fault identification. The used network structure is the same as the model in Ref. [44].

The fault diagnosis performance of various approaches under different ratios of training data is shown in Fig. 9. The time–frequency analysis exhibits evident advantages over the typical time-domain analysis method with the same classifiers, showing that TFD is an effective data analysis form for the joint diagnosis of bearing fault type and bearing fault severity. The DL models, namely, CNN, WDCNN, LBCNN, and DCTN, perform better than other methods in accuracy. Although the SVM, KNN, and ELM models are all typical small-sample-analysis methods, the diagnosis performance is not satisfactory when the sample size is small. The proposed DCTN model shows overall higher fault diagnosis accuracies than the CNN model with the same convolutional layers, which fully illustrates that the proposed hierarchical decision strategy is beneficial to improving the decision-making ability of the model.

4.2   Case two: cross-severity fault diagnosis of bearings

The cross-severity fault diagnosis task attempts to identify the fault type of the test samples with fault categories that are unseen for the training samples. The set of the cross-severity fault diagnosis tasks is shown in Table 5. Specifically, for the aeronautical bearings with failures on the inner ring or a roller, the monitored signals corresponding to different fault severities are used for training and testing. The test data in each task correspond to two fault types, namely, defined data superclass I and R, and the same fault severity, namely, 150, 250, or 450 μm. The training data contain all the fault types but lack the fault severity of the test data.
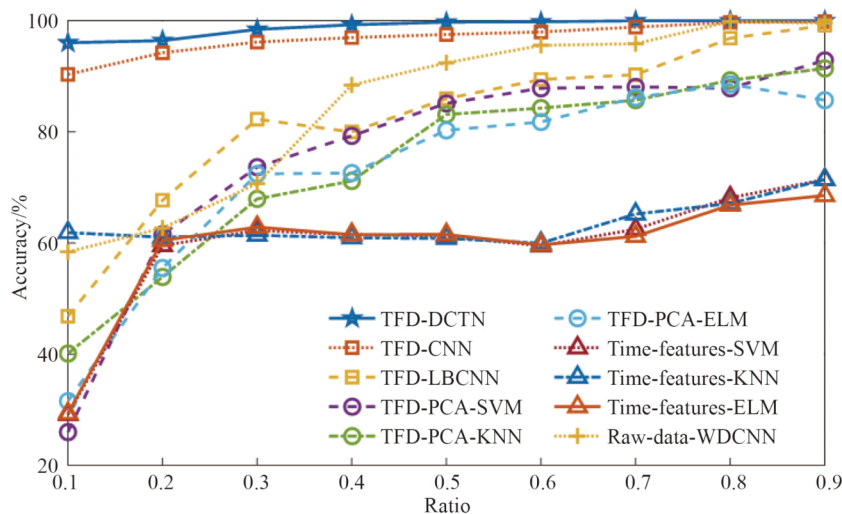
Figure 10 shows the predicted superclass labels of the

Fig. 9   Fault diagnosis performance of different approaches under different ratios of training data.

test samples under condition C17 with six fault categories. The model parameters used in the experiment are consistent with those set in Section 4.1. Cross-severity fault diagnosis is effective because most of the predictions of the corresponding labels are correct, especially for the bearing samples corresponding to R-5, where the predictions of the superclass labels are all correct. In addition, the incorrect predictions of the samples corresponding to labels I-4 and R-7 are identified as label N, which is reasonable for these two sets of data corresponding to small faults of bearings. The predicted labels show the effectiveness of the DCTN model in superclass identification.

Figure 11 shows the prediction probabilities of the test samples with six fault categories in three tasks. The bar chart shows the mean of the prediction probabilities of all the samples corresponding to each category. The error bars show the range of the prediction probabilities for each superclass label. The prediction probability of the

superclass category corresponding to the test sample is the largest, which is the fundamental basis for the realization of cross-severity fault diagnosis because all decisions are inferred according to the probability value. The correct prediction probability of samples corresponding to R-5 category is close to 1, which is the best prediction performance among the six categories. The predicted labels and probabilities fully demonstrate the effectiveness of the proposed DCTN model for cross-severity fault diagnosis tasks, which can support better generalization of the model.

The proposed DCTN-based cross-severity fault diagnosis approach can reduce the requirements for labeled data in practical application and is more consistent with engineering needs. Moreover, the three cross-severity fault diagnosis tasks listed in Table 5 are performed using the comparison methods selected in Section 4.1. The fault diagnosis accuracies of all the methods are shown in Table 6. The fault diagnosis accuracies of each category, mean accuracies of each task, and mean accuracies of each approach are listed in Table 6. The highest mean prediction accuracies for the whole work and each task are shown in bold form. The following conclusions can be drawn from the results in Table 6:
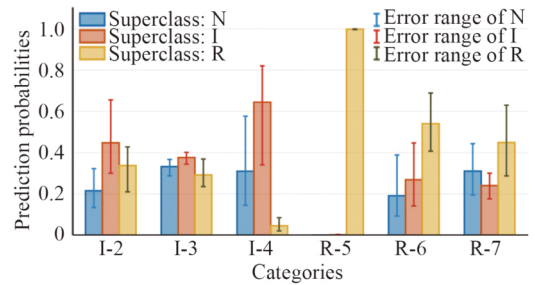
a) Most methods are completely ineffective in the

Table 5　Set of cross-severity fault diagnosis tasks

Task	Categories of training bearings	Categories of test bearings
1	N-1, I-3, I-4, R-6, R-7	I-2, R-5
2	N-1, I-2, I-4, R-5, R-7	I-3, R-6
3	N-1, I-2, I-3, R-5, R-6	I-4, R-7
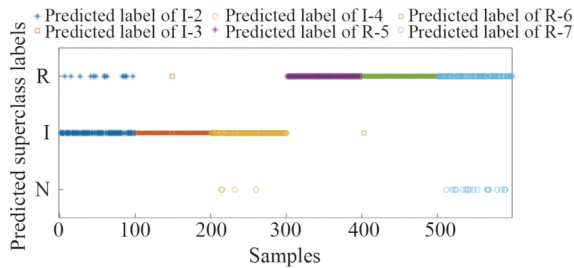
Fig. 10　Predicted superclass labels of cross-severity fault diagnosis tasks.

Fig. 11　Predicted probabilities of cross-severity fault diagnosis tasks.

Table 6　Fault diagnosis accuracies of different approaches in cross-severity fault diagnosis tasks

Approach	Fault diagnosis accuracy/%									
	I-2	R-5	Task 1	I-3	R-6	Task 2	I-4	R-7	Task 3	Mean
TFD-DCTN	86.00	100.0	93.00	99.00	99.00	99.00	96.00	83.00	89.50	93.83
TFD-CNN	2.00	98.00	50.00	2.00	11.00	6.50	9.00	1.00	5.00	20.50
TFD-LBCNN	8.00	98.00	53.00	5.00	97.00	51.00	0.00	100.0	50.00	51.33
TFD-PCA-SVM	97.00	0.00	48.50	0.00	100.0	50.00	36.00	0.00	18.00	38.83
TFD-PCA-KNN	19.00	93.00	56.00	0.00	96.00	48.00	0.00	0.00	0.00	34.67
TFD-PCA-ELM	97.00	0.00	48.50	20.00	52.00	36.00	0.00	1.00	0.50	28.33
Time-features-SVM	100.0	37.00	68.50	77.00	0.00	38.50	58.00	38.00	48.00	51.67
Time-features-KNN	92.00	22.00	57.00	100.00	0.00	50.00	92.00	22.00	57.00	54.67
Time-features-ELM	98.00	18.00	58.00	40.00	0.00	20.00	69.00	31.00	50.00	42.67
Raw-data-WDCNN	12.00	100.0	56.00	24.00	96.00	60.00	0.00	100.0	50.00	55.33

identification of most fault categories when the corresponding results of samples in each category are analyzed separately. The recognition rate of several methods reaching 100% with ineffective fault classification is realized. For example, the diagnosis accuracy of the TFD-PCA-SVM approach in the R-6 fault category is 100%, but the accuracy of I-3 is 0 in the same task. Analysis of the predicted labels reveals that the model identifies the labels of all the samples as R, that is, the classifier loses its discriminability. Hence, the mean recognition accuracy of this method in Task 2 is 50%, but this approach is not valid for this task.

b) The mean fault diagnosis accuracy of the proposed DCTN-based approach is remarkably higher than that of other methods. The mean fault diagnosis accuracy of the proposed DCTN-based approach is up to 93.83%. The results show that the proposed DCTN method is more suitable for cross-severity fault diagnosis tasks. In the three tasks, the proposed approach achieves the highest recognition accuracy of 99%. In comparison, most of the accuracies of the other methods are less than 50%, and several methods are completely ineffective with an accuracy of 0. Overall, the effectivity of the fault diagnosis approach proposed in this paper is verified in each task. Compared with the other approaches, it shows an evident advantage in the cross-severity fault diagnosis task.

5 Conclusions

Aiming at the problems of poor interpretability and weak generalization ability that commonly exist in the deep-learning-based fault diagnosis methods, this paper proposes a DCTN-based hierarchical fault diagnosis method that effectively merges the advantage of decision tree and the CNN model. The proposed DCTN model uses the convolutional layers in the CNN model for sample characterization and replaces the fully-connected layer in the CNN model with a novel tree-structured decision layer, in which the leaf nodes and seed nodes are set for fault type and fault severity identification, respectively. The ability of hierarchical decision making is given to these nodes in the model through pre-training and fine-tuning with exclusive loss functions. The final fault diagnosis decision is made according to the overall path probabilities in the tree structure.

Hierarchical multiclass fault diagnosis experiments and cross-severity fault diagnosis experiments are executed to analyze the generalization of the proposed model. The proposed DCTN-based fault diagnosis approach achieves a relatively higher multiclass recognition performance. In particular, the diagnosis accuracy of this model is even higher than that of the backbone CNN, indicating that the hierarchical decision-making strategy adopted in the model is beneficial to fault diagnosis. Moreover, the proposed method shows a more powerful generalization ability in the cross-severity fault diagnosis experiments, which is meaningful in practice because the collection of the training samples has difficulty covering all fault severities. The experiment highlights the effectiveness and superiority of the proposed method in fault diagnosis. This paper makes a useful exploration of the decision interpretability of the fault diagnosis model, and more importantly, provides a feasible way to realize cross-severity fault diagnosis of bearing. All these are beneficial to improving the confidence level of the fault diagnosis model and facilitating the solution of practical engineering problems. As a complete data-driven method, the proposed model has few limitations on the applied objects; thus, it also has better generalizability for other devices.

The purpose of this work is not to provide a complete solution but rather to suggest an alternative approach to deliver improved interpretability and generalization performance of bearing fault diagnosis. Several issues are still worthy of further exploration: 1) In terms of model interpretability, the proposed method still has difficulty explaining whether the convolutional layers have learned useful fault-related knowledge or in which way the model can effectively learn the knowledge. Therefore, the interpretability of the CNN model and other DNN models in fault diagnosis needs to be explored further. Certainly, this is a very challenging task that many researchers attempt to break through. 2) In terms of cross-severity fault diagnosis task, the diagnosis results of the proposed method remain in the accurate judgment of super class labels, that is, fault types. It would be more meaningful if the approximate range of the fault severity to which the test sample belongs can be accurately identified, which can be the direction of our next efforts.

Nomenclature

Abbreviations

CNN	Convolutional neural network
CWT	Continuous wavelet transform
DCTN	Deep convolutional tree-inspired network
DL	Deep learning
DNN	Deep neural network
ELM	Extreme learning machine
KNN	k-nearest neighbor
LBCNN	Local binary convolutional neural network
PCA	Principal component analysis
SVM	Support vector machine
TFD	Time–frequency distribution
WDCNN	Wide deep convolutional neural network

Variables

a	Stretch factor
b	Shift factor
$\mathrm{CWT}(s(t))$	CWT time−frequency function of signal $s(t)$
$d_j\ (j = 1, 2, \ldots, K)$	Distance between the feature and each classification hyperplane
$H(p, q)$	Cross-entropy loss function
$H^*(p(\cdot), q(\cdot), p(\cdot), q(\cdot))$	Loss function of the tree-structured decision layer
K	Number of sample categories
ℓ	Overall prediction
L	Feature dimension of the fully-connected layer
N	Number of samples
$p(\cdot)$	Probability distribution of the predicted output
$p(k)$	True labels of the pre-trained network
$p(\ell)$	True labels of the tree-structured decision layer
$P(\ell)$	Path probabilities of the tree-structured decision layer
P (subclass)	Probability of correct prediction for seed nodes
P (superclass)	Probability of correct prediction of leaf nodes
$q(\cdot)$	Probability distribution of the actual output
$q(\hat{k})$	Predicted probabilities of the pre-trained network
$q(\hat{\ell})$	Predicted probabilities of the tree-structured decision layer
R	Dimension of the TFD matrix
$s(t)$	Signal in time t
sw_j	Weight vector of the jth leaf note
w_j	Weight vector of the jth vector in weight matrix W of the fully-connected layer
w_j^*	Weight vector of the jth tree-structured decision layer after fine-tuning
W	Weight matrix
x	Input features of the Softmax classifier in the cross-entropy loss
$\boldsymbol{x}$	Input feature vector of the tree-structured decision layer
$\hat{\boldsymbol{y}}$	Prediction probabilities by the Softmax classifier
$\hat{\boldsymbol{y}}_j$	Predicted probability for the jth category
z_j	Prediction scope corresponding to K categories
ω	Weight adjusting the pre-trained decision and tree-structured decision
ψ	Mother wavelet

References

1. Chen X F, Wang S B, Qiao B J, et al. Basic research on machinery fault diagnostics: past, present, and future trends. Frontiers of Mechanical Engineering, 2018, 13(2): 264–291

2. Zheng P, Wang H, Sang Z, et al. Smart manufacturing systems for Industry 4.0: conceptual framework, scenarios, and future perspectives. Frontiers of Mechanical Engineering, 2018, 13(2): 137–150

3. Hoang D T, Kang H J. A survey on deep learning based bearing fault diagnosis. Neurocomputing, 2019, 335: 327–335

4. Lei Y G, Yang B, Jiang X, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap. Mechanical Systems and Signal Processing, 2020, 138: 106587

5. Zhou D H, Zhao Y H, Wang Z D, et al. Review on diagnosis techniques for intermittent faults in dynamic systems. IEEE Transactions on Industrial Electronics, 2020, 67(3): 2337–2347

6. Wu X Y, Zhang Y, Cheng C M, et al. A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. Mechanical Systems and Signal Processing, 2021, 149: 107327

7. Liang P F, Deng C, Wu J, et al. Single and simultaneous fault diagnosis of gearbox via a semi-supervised and high-accuracy adversarial learning framework. Knowledge-Based Systems, 2020, 198: 105895

8. An Z H, Li S M, Wang J R, et al. A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network. ISA Transactions, 2020, 100: 155–170

9. Zhong T, Qu J F, Fang X Y, et al. The intermittent fault diagnosis of analog circuits based on EEMD-DBN. Neurocomputing, 2021, 436: 74–91

10. Zhao D Z, Wang T Y, Chu F L. Deep convolutional neural network based planet bearing fault classification. Computers in Industry, 2019, 107: 59–66

11. Lu S L, Yan R Q, Liu Y B, et al. Tacholess speed estimation in order tracking: a review with application to rotating machine fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 2019, 68(7): 2315–2332

12. Cheng Y W, Lin M X, Wu J, et al. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. Knowledge-Based Systems, 2021, 216: 106796

13. Li M F, Wang T Y, Kong Y, et al. Synchro-reassigning transform

for instantaneous frequency estimation and signal reconstruction. IEEE Transactions on Industrial Electronics, 2021 (in press)

14. Li M F, Wang T Y, Chu F L, et al. Scaling-basis Chirplet transform. IEEE Transactions on Industrial Electronics, 2020, 68(9): 8777–8788

15. Li M F, Wang T Y, Chu F L, et al. Component matching chirplet transform via frequency-dependent chirp rate for wind turbine planetary gearbox fault diagnostics under variable speed condition. Mechanical Systems and Signal Processing, 2021, 161: 107997

16. Cerrada M, Sánchez R V, Li C, et al. A review on data-driven fault severity assessment in rolling bearings. Mechanical Systems and Signal Processing, 2018, 99: 169–196

17. Zhao J, Yang S P, Li Q, et al. A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network. Measurement, 2021, 176: 109088

18. Minhas A S, Kankar P K, Kumar N, et al. Bearing fault detection and recognition methodology based on weighted multiscale entropy approach. Mechanical Systems and Signal Processing, 2021, 147: 107073

19. Pan H Y, Yang Y, Zheng J D, et al. A fault diagnosis approach for roller bearing based on symplectic geometry matrix machine. Mechanism and Machine Theory, 2019, 140: 31–43

20. Wen L, Li X, Gao L. A new two-level hierarchical diagnosis network based on convolutional neural network. IEEE Transactions on Instrumentation and Measurement, 2020, 69(2): 330–338

21. Amorim J P, Abreu P H, Reyes M, et al. Interpretability vs. complexity: the friction in deep neural networks. In: Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow: IEEE, 2020, 20006226

22. Yang Z B, Zhang J P, Zhao Z B, et al. Interpreting network knowledge with attention mechanism for bearing fault diagnosis. Applied Soft Computing, 2020, 97: 106829

23. Rauber T W, da Silva Loca A L, Boldt F de A, et al. An experimental methodology to evaluate machine learning methods for fault diagnosis based on vibration signals. Expert Systems with Applications, 2021, 167: 114022

24. Wu Y, Jin W D, Li Y, et al. A novel method for simultaneous-fault diagnosis based on between-class learning. Measurement, 2021, 172: 108839

25. Stock M, Nguyen B, Courtens W, et al. Otolith identification using a deep hierarchical classification model. Computers and Electronics in Agriculture, 2021, 180: 105883

26. Lu C, Wang Z Y, Zhou B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Advanced Engineering Informatics, 2017, 32: 139–151

27. Liu P, Zhang Y, Zhang X Y, et al. Evaluation of measurement uncertainty of oxygen in titanium alloys based on Monte Carlo method. Journal of Physics: Conference Series, 2020, 1605: 012135

28. Kraus M, Feuerriegel S. Forecasting remaining useful life: interpretable deep learning approach via variational Bayesian inferences. Decision Support Systems, 2019, 125: 113100

29. Gangsar P, Tiwari R. Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: a state-of-the-art review. Mechanical Systems and Signal Processing, 2020, 144: 106908

30. Wang X, Wang T Y, Ming A B, et al. Semi-supervised hierarchical attribute representation learning via multi-layer matrix factorization for machinery fault diagnosis. Mechanism and Machine Theory, 2022, 167: 104445

31. Blanco-Justicia A, Domingo-Ferrer J, Martínez S, et al. Machine learning explainability via microaggregation and shallow decision trees. Knowledge-Based Systems, 2020, 194: 105532

32. Sagi O, Rokach L. Explainable decision forest: transforming a decision forest into an interpretable tree. Information Fusion, 2020, 61: 124–138

33. Vamsi I, Sabareesh G R, Penumakala P K. Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-stationary loading. Mechanical Systems and Signal Processing, 2019, 124: 1–20

34. Cabrera D, Sancho F, Sánchez R V, et al. Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition. Frontiers of Mechanical Engineering, 2015, 10(3): 277–286

35. Zhou Z H, Feng J. Deep forest: towards an alternative to deep neural networks. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017). 2017, 3553–3559

36. Humbird K D, Peterson J L, McClarren R G. Deep neural network initialization with decision trees. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(5): 1286–1295

37. Jiang S H, Mao H Y, Ding Z M, et al. Deep decision tree transfer boosting. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(2): 383–395

38. Kontschieder P, Fiterau M, Criminisi A, et al. Deep neural decision forests. In: Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE, 2015, 1467–1475

39. Zhang Q S, Yang Y, Ma H T, et al. Interpreting CNNs via decision trees. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019, 6254–6263

40. Roy D, Panda P, Roy K. Tree-CNN: a hierarchical deep convolutional neural network for incremental learning. Neural Networks, 2020, 121: 148–160

41. Daga A P, Fasana A, Marchesiello S, et al. The Politecnico di Torino rolling bearing test rig: description and analysis of open access data. Mechanical Systems and Signal Processing, 2019, 120: 252–273

42. Zhou P, Peng Z K, Chen S Q, et al. Non-stationary signal analysis based on general parameterized time–frequency transform and its application in the feature extraction of a rotary machine. Frontiers of Mechanical Engineering, 2018, 13(2): 292–300

43. Zhang W, Peng G, Li C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors (Basel), 2017, 17(2): 425

44. Jiang Y, Feng C, He B, et al. Actuator fault diagnosis in autonomous underwater vehicle based on neural network. Sensors and Actuators. A, Physical, 2021, 324: 112668