Measurement 223 (2023) 113796

Contents lists available at ScienceDirect

Measurement

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

Measuring compound defect of bearing by wavelet gradient integrated spiking neural network

Jianping Xuan *, Zisheng Wang, Shaochen Li, Ang Gao, Chunlei Wang, Tielin Shi

School of Mechanical Science and Engineering, Huazhong University of Science and Technology, 430074, Wu'han, China

ARTICLE INFO

Keywords:
Spiking neural network
Wavelet gradient propagation mechanism
Compound defect detection
Multi-label learning

ABSTRACT

Compound defect is the common damage of equipment. However, measuring compound defect is difficult because of the complex correlation between multiple single-point defects. Fortunately, as the third generation of neural network that is closest to biological intelligence, the spiking neural network (SNN) has a huge potential on building classifiers. Moreover, time-frequency processing technologies for raw signals are often independent of the training of neural network, so that the end-to-end mode is not able to be built. Therefore, this paper tries to combine the wavelet packet transform (WPT) with SNN, and proposes a wavelet gradient integrated spiking neural network (WGI-SNN) framework. First, a wavelet gradient propagation mechanism (WGPM) is deduced, then recursive down-sampling wavelet networks (DSWNs) is designed to simulate the WPT. Secondly, a SNN model integrating DSWNs is constructed to measure compound defect samples. In addition, an experiment is implemented to verify the effectiveness and practicability of the proposed method.

1. Introduction

With the rapid development of digital technology, the demand for industrial upgrading is becoming increasingly urgent. In the process of industrial upgrading, the automatic defect detection function of production equipment is one of the evaluation criteria for intelligent degree [1–3]. Compared with observing the operating status of equipment by workers, unmanned intelligent defect identification can monitor and predict the defect status of key components in real time [4], to effectively reduce the probability of downtime and even safety accidents. Therefore, more and more scholars [5–7] devote themselves to the research of advanced defect detection methods.

Deep learning (DL) [8,9] is widely used as the basic model in defect detection due to its capability of extracting adaptively hidden information from input data. At present, DL models are designed mainly by using the artificial neural network (ANN), which is regarded as the second generation of neural network models. In the published deep models [10–14], the convolutional neural network (CNN) is one of the most popular framework of the artificial neural network (ANN) models, because it can directly learn the useful features by convolution operations and nonlinear activation functions. For example, Liang et al. [10] proposed a multi-scale dynamic adaptive residual network for dynamic feature selection, and Chen et al. [11] introduced a new residual deep subdomain adaptation network, extracting the transferable features from source and target domains.

Unfortunately, the ANN falls in the difficult position that the related models are not able to completely simulate the operating mechanism of biological brain neurons, because of the lack of biological interpretability [15]. As a result, with the rapid increase of difficult of defect detection, the detection performance of ANN has encountered great challenges. Consequently, to find the intelligent method based on good biological interpretability, spiking neural network (SNN) [16–18] becomes the important research object. SNN is named as the third generation of neural network models, due to that the model based on SNN is closest to biological intelligence [19,20]. Recently, there are a few applications [21–23] of SNN in the defect detection field. Zuo et al. [22] proposed an intelligent detection method for defect of rolling bearings based on SNN, then compared with the first and second generations of neural networks to prove the effectiveness of proposed method, and the final result demonstrated that the proposed method can obtain a satisfactory performance in bearing defect detection. Wang et al. [23] implemented the defect detection for bearing by using the improved SNN, designed a new encoding method to encode raw data into spike sequences, and validated the proposed method outperformed than the previous models. However, there are still few researches about applying SNN to detect compound defect.

In current manufacturing systems, key parts, such as bearings and tools, are easily damaged by multiple defects (namely compound defect) [24]. For instance, Shao et al. [25] used an improved convolutional deep belief network with compressed sensing to deal with

* Corresponding author.
E-mail address: jpxuan@hust.edu.cn (J. Xuan).

https://doi.org/10.1016/j.measurement.2023.113796
Received 10 July 2023; Received in revised form 12 October 2023; Accepted 31 October 2023
Available online 3 November 2023

compound defects and regarded every compound defect as a single defect class. However, the complex correlation between different defects is disregarded. Fortunately, multi-label learning [26,27] is appropriate to solve the aforementioned problem. For example, Yu et al. [28] designed a multi-label convolutional neural network (ML-CNN) for compound defect detection. Tan et al. [29] executed a comparative study of several traditional multi-label algorithms, including binary relevance (BR) [30] algorithm, classifiers chains (CC) [31] algorithm, and multi-label k-nearest neighbor (ML-KNN) [32] algorithm for simultaneous defect detection. Moreover, other multi-label algorithms, such as label powerset (LP) [33] and ranking support vector machine (Rank-SVM) [34], can learn compound defect features. In summary, though classic multi-label learning algorithms based on CNN or traditional machine learning can be used to achieve compound defect detection, their performances may not be best due to the limitation of ANN model. It is of great practical significance and theoretical value, to study how to complete the design of multi-label learning method using the SNN model.

Some signal processing methods, such as Fourier transform [35] and wavelet packet transform (WPT) [36], are currently regarded as the preprocessing approaches before training models. These preprocessing methods could enhance the detection capability of data-driven models. However, such an enhancement increases the dependence on signal processing knowledge while not conforming to the end-to-end training mode. Consequently, using network layers to implement the preprocessing function should be examined. Xiong et al. [37] integrated the WPT process into convolutional neural networks to implement the single-point defect detection. Nevertheless, this approach must define the shape of convolutional kernels in advance and cannot be directly adapted to any wavelet function. Li et al. [38] designed a down-sampling network implementing wavelet transform and proposed a wavelet integrated convolutional network for image classification and image segmentation. However, the WPT process was not implemented, and the down-sampling network lacked padding operations on the margin.

To construct the deep model with the good biological interpretability and meanwhile avoid the independence of time-frequency processing technologies and neural network algorithms, a wavelet gradient integrated spiking neural network (WGI-SNN) model is proposed in this paper to complete compound defect detection. The improvement of WGI-SNN model includes two aspects. For one thing, recursive down-sampling wavelet networks (DSWNs) are designed by analyzing the wavelet gradient propagation mechanism (WGPM) to implement the WPT process. For another, a SNN model with DSWNs is built by using integrate and fire neurons. Finally, an experiment is executed to discuss the difference between the WGI-SNN model and other state-of-the-art models, to prove the effectiveness and practicability of the proposed method. Overall, the main contributions of this study are listed as follows.

1. Recursive DSWNs are designed to implement the networks with WPT function of by deducing the forward and backward propagation processes of the WGPM. It can directly learn the wavelet feature in the gradient propagation of network, and leads to an end-to-end training mode.
2. A SNN model is constructed by using integrate and fire neurons. Specially, the proposed method uses the spike information as the training data, and meanwhile utilizes the conversion of ANN to SNN.
3. The effectiveness and practicability of the WGI-SNN framework are verified by discussing the difference between the proposed method and other advanced compound defect detection methods in an experiment.

The composition of this paper is organized as follows. Section 2 comprehensively describes related theories of WPT and SNN. Section 3 outlines the details of the proposed method. Section 4 executes the experiment to demonstrate the superiority of the proposed method. Section 5 concludes this paper.

2. Related theories

2.1. Theory of WPT

Wavelet transform (WT) technology extracting time and frequency information from raw signals is widely utilized in the signal processing field. However, only the low-frequency part is decomposed; thus, a considerable amount of detailed information is not well represented. As an extension of the WT, the WPT can simultaneously decompose the low- and high-frequency parts. Particularly, the low- and high-frequency parts of raw signals can form a wavelet packet coefficient matrix via the multiresolution analysis.

Fig. 1 illustrates an example of building a wavelet packet coefficient matrix by using the WPT with three decomposition levels. Particularly, the WPT can be implemented by executing a series of convolution operations with loss-pass and high-pass filters of an orthogonal wavelet [39]. In detail, the low-pass filter l and high-pass filter h can be defined as follows,

$$l(k) = \frac{1}{\sqrt{2}}\langle \phi(t), \phi(2t-k)\rangle \tag{1}$$

$$h(k) = \frac{1}{\sqrt{2}}\langle \psi(t), \psi(2t-k)\rangle \tag{2}$$

$$h(k) = (-1)^k l(1-k) \tag{3}$$

where ϕ is a scale function, ψ is a wavelet function, $\langle\cdot,\ \cdot\rangle$ is an inner product operation, and k, t are two variables.

Given one discrete 1D signal, approximation and detail coefficients can be calculated recursively as follows:

$$D_{j+1,2i}(k) = \sum_n l(n-2k) D_{j,i}(n) \tag{4}$$

$$D_{j+1,2i+1}(k) = \sum_n h(n-2k) D_{j,i}(n) \tag{5}$$

where $D_{0,0}$ represents the raw signal with N data points; $\{D_{j,i}(k), k = 1, 2, \ldots, N/2^j \ and \ i = 0, 1, \ldots, 2^j - 1\}$ represents the wavelet coefficients in the ith subband at jth decomposition level; $\{D_{j+1,2i}(k), k = 1, 2, \ldots, N/2^{j+1}\}$ represents the wavelet coefficients in the $(2i)$th subband at $(j+1)$th decomposition level; $\{D_{j+1,2i+1}(k), k = 1, 2, \ldots, N/2^{j+1}\}$ is the wavelet coefficients in the $(2i+1)$th subband at $(j+1)$th decomposition level.

2.2. Theory of SNN

2.2.1. Biological concept of neurons

There are over ten billion neurons in the brain of human, and is a very complex connection between them. In general, through reception, integration, transmission and output, neurons can transmit and process information to perceive environment and dominate actions [40]. From the point of view of biological behaviors, a neuron can be divided into three parts, namely dendrite, soma, and axon. Firstly, the dendrite transmits the information from other neurons to the soma. Secondly, the soma analyzes the cumulative input information, and generate an output signal after it reaches the threshold. Thirdly, the axon transmits the output signal from soma to other neurons.

During the transmission of information, there is a potential difference between the inside and outside of the neuron, which is called the membrane potential. On the one hand, when the input information does not exist, the membrane potential of neuron is called the resting potential. On the other hand, the membrane potential of the neuron will be changed after receiving the input signal. And when the potential reaches the threshold, a new spike signal is generated as the input of other neurons.

In summary, the input spike of a neuron can come from multiple neurons, and after the input spike meets the requirements, a new spike is generated and sent to the back neurons [40]. The transmission of information does not depend on the shape of the spike, but only on the time and number of input spike. Therefore, to have strong biological interpretability, SNN is built based on these descriptions [20].

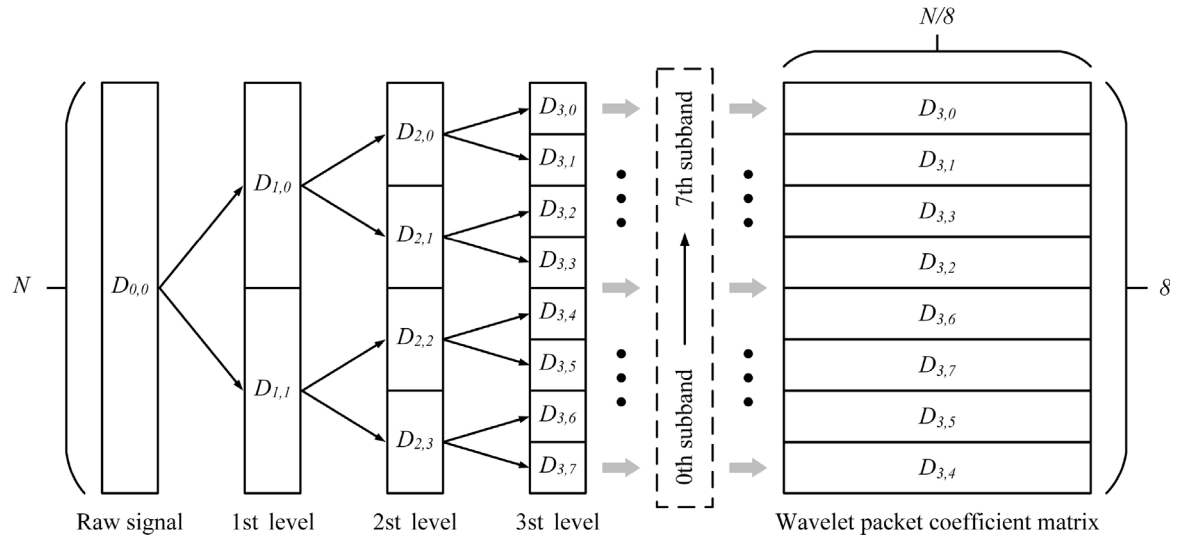

Fig. 1. Diagram of building a wavelet packet coefficient matrix using the WPT with three decomposition levels.

2.2.2. SNN

To design a neural network with good biological interpretability, SNN is proposed to imitate the above mentioned biological concept of neurons [41,42]. At present, SNN is closest to biological intelligence in all networks, and thus is regarded as the third generation of neural network models. SNN can reduce the distance between machine learning and neural science, by using the models being similar to the biological neuron mechanism. Fig. 2 shows the architecture of a multilayer spiking neural network. Generally, SNN has four characteristics as follows.

(1) The input of SNN is the spike signal, not the common continue value shown in ANN. As mentioned in the biological concept of neurons, spike signal is a discrete event.
(2) The effect of spike signal can be divided into two parts. On the one hand, the effect is positive when the spike signal produces a positive change for the membrane potential of neuron. On the other hand, the effect is negative when the spike signal produces a negative change for the membrane potential of neuron.
(3) The membrane potential of neurons is important, and it determines whether or not neurons is activated. After the neuron is activated, spike signal is output.
(4) The membrane potential of neuron will be reset, after spike signal is output by the neuron.

Integrate and fire [43,44] neuron is widely used in the structure of SNN. For the integrate and fire neuron, an action potential is defined as the status, when the membrane potential of neuron reaches the threshold. Overall, the law of potential change can be described as follows.

$$\tau_m \frac{dV}{dt} = E_L - V + R_m I \tag{6}$$

where τ_m is the time constant, E_L is the resting potential, R_m is the impedance, I is the input current, V is the real-time membrane potential of neuron, t is the time.

3. Proposed method

3.1. Design of recursive DSWNs

3.1.1. Deduction of the WGPM

One 1D signal $X = \{x_i\}_{i\in Z}$ can normally be decomposed into low-frequency part $D_{low} = \{d_k^{low}\}_{k\in Z}$ and high-frequency part $D_{high} = \{d_k^{high}\}_{k\in Z}$ using the WPT with one decomposition level [38]. Especially, two decompositions can be represented as follows:

$$d_k^{low} = \sum_i l_{i-2k} x_i \tag{7}$$

$$d_k^{high} = \sum_i h_{i-2k} x_i \tag{8}$$

where $L = \{l_k\}_{k\in Z}$ and $H = \{h_k\}_{k\in Z}$ are the low- and high-pass filters, respectively.

Concretely, assuming the raw signal contains N data points, the above equations can be reorganized as follows:

$$D_{low} = \mathcal{L}X \tag{9}$$

$$\frac{\partial D_{low}}{\partial X} = \mathcal{L}^T \tag{10}$$

$$D_{high} = \mathcal{H}X \tag{11}$$

$$\frac{\partial D_{high}}{\partial X} = \mathcal{H}^T \tag{12}$$

where $\mathcal{L} = \begin{pmatrix} l_1 & l_2 & l_3 & l_4 & l_5 & l_6 & \cdots & l_m \\ 0 & 0 & l_1 & l_2 & l_3 & l_4 & \cdots & l_{m-2} \\ 0 & 0 & 0 & 0 & l_1 & l_2 & \cdots & l_{m-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & l_{m-2u+2} \end{pmatrix}_{u\times m}$,

$\mathcal{H} = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 & \cdots & h_m \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & \cdots & h_{m-2} \\ 0 & 0 & 0 & 0 & h_1 & h_2 & \cdots & h_{m-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & h_{m-2u+2} \end{pmatrix}_{u\times m}$, $u = \frac{N}{2}$ and $m = N + n - 2$; the even number N is the length of the signal X; n is the number of wavelet filter coefficients. For example, using Daubechies 1 (DB1) wavelet, function $L = \frac{1}{\sqrt{2}}\{1,1\}$ and $H = \frac{1}{\sqrt{2}}\{1,-1\}$. If $n = 2$, $\mathcal{L} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{2}} \end{pmatrix}_{\frac{N}{2}\times N}$, $\mathcal{H} =$

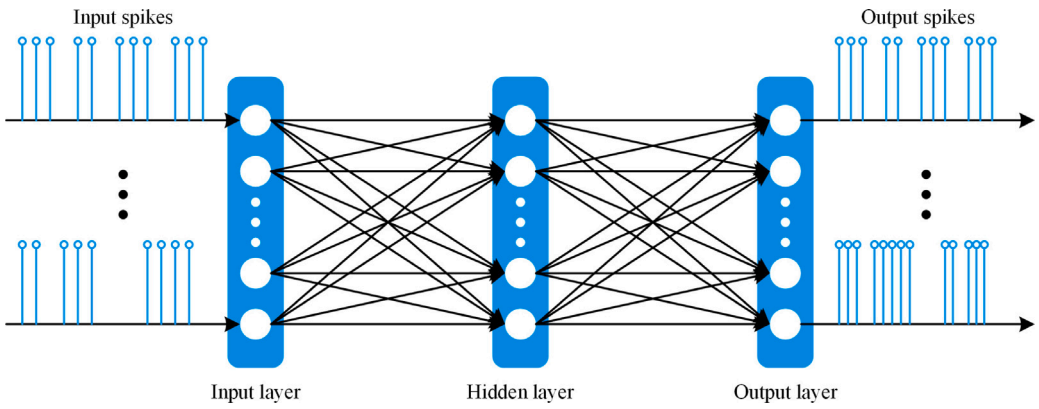

Fig. 2. Architecture of a multilayer spiking neural network.

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & -\frac{1}{\sqrt{2}} \end{pmatrix}_{\frac{N}{2}\times N}$$

. Moreover, if $n \neq 2$, then the left and right margins of signal X are padded with $\frac{n-2}{2}$ zero elements simultaneously.

The connected network layer can be designed on the basis of Eqs. (9) and (11). Similarly, the backward propagation of a fully connected network layer can be designed considering Eqs. (10) and (12). Therefore, a WGPM can be deduced by the forward and backward propagations.

3.1.2. Simulating the WPT process based on the WGPM

The WGPM can be implemented by a DSWN layer. Thus, a DSWN layer can represent the function of WPT with one decomposition level. This paper further discusses the WPT process with multiple decomposition levels to utilize the DSWN layer to represent completely the function of WPT.

Raw signal X can be decomposed as shown below by the WPT with one decomposition level.

$$X = \mathcal{L}^T D_{low} + \mathcal{H}^T D_{high} \tag{13}$$

Meanwhile, raw signal X can be computed as follows by the WPT with two decomposition levels:

$$\begin{aligned} X = {} & \mathcal{L}^T(\mathcal{L}^T D_{low} D_{low} + \mathcal{H}^T D_{low} D_{high}) \\ & + \mathcal{H}^T(\mathcal{L}^T D_{high} D_{low} + \mathcal{H}^T D_{high} D_{high}) \end{aligned} \tag{14}$$

where $D_{\text{low}} D_{\text{low}}$ and $D_{\text{low}} D_{\text{high}}$ are the low- and high-frequency components of D_{low}, respectively; $D_{\text{high}} D_{\text{low}}$ and $D_{\text{high}} D_{\text{high}}$ are the low- and high-frequency parts of D_{high}, respectively.

Functionally, the WPT with multiple decomposition levels can be implemented using recursive DSWNs, as shown in Fig. 3. The input data of each network layer are under-sampled channel by channel when developing the Python code implementing recursive DSWNs. In this paper, the number λ of under-sampling network layers is set to 6; therefore, recursive DSWNs can simulate the WPT with 6 decomposition levels.

3.2. 3.2 SNN model with recursive DSWNs

To construct an intelligent detection model with good biological interpretability, this paper designs a spiking neural network combined with the recursive DSWNs. Fig. 4 illustrates the brief architecture of the proposed method. First, raw signal is input into recursive DSWNs to obtain multiple frequency subbands. Second, these subbands are arranged to form a wavelet coefficient matrix. Third, the wavelet coefficient matrix is encoded as the spike, which is regarded as the input of SNN. Specially, SNN is composed of four convolutional layers, one fully connected layer, and one output layer. Finally, the output layer in SNN provides the multi-label prediction result. And the result is calculated by using the multi-label classifier of ML-CNN.

From the Fig. 4, it can be observed that the pooling layer is average, and the rectified linear unit (ReLU) layer in ANN is replaced with the spiking layer. The function of spiking layer is described in the Eq. (6), and the threshold is set as the value 1.

Table 1
Parameters of bearings used in the experiment.

Parameter	Value
Contact angle	0°
Roller number	8
Roller diameter	5 mm
Pitch diameter	20 mm
Manufacturer	NSK
Bearing specs	6200

4. Experiment

4.1. Dataset collection in laboratory experiment

An experiment is executed first to verify the effectiveness of the proposed framework in compound defect detection. Fig. 5 roughly shows the physical structure of the device in the laboratory. Moreover, the simplified diagram illustrates the details of the device in Fig. 6. Particularly, the device mainly includes a motor, an accelerometer placed near the drive end, and a bearing chock installing the defect bearing. A testing bearing is installed in the bearing chock near the driven end, and another normal bearing is placed in the bearing chock near the fan end. The raw dataset is collected under 13 operations, which contains 13 statuses of bearings. The sampling frequency and working speed are set to 49152 Hz and 2850 rpm, respectively. Table 1 lists the parameters of defect bearings used in the laboratory experiment. Table 2 comprehensively describes the details of 13 statuses. Fig. 7 illustrates one normal bearing and six single-point defect bearings, while compound defect can be obtained by combining several single-point defects.

Table 3 shows that each status of bearings is labeled with three sub-labels. The time-domain distribution of raw signals with 13 operations is drawn in Fig. 8. Furthermore, using data enhancement technology, the number of normal and defect samples in each status are 120 and 30, respectively, wherein every sample contains 4096 data points. Additionally, a K-fold cross validation is applied to implementing the data division following the above mentioned set.

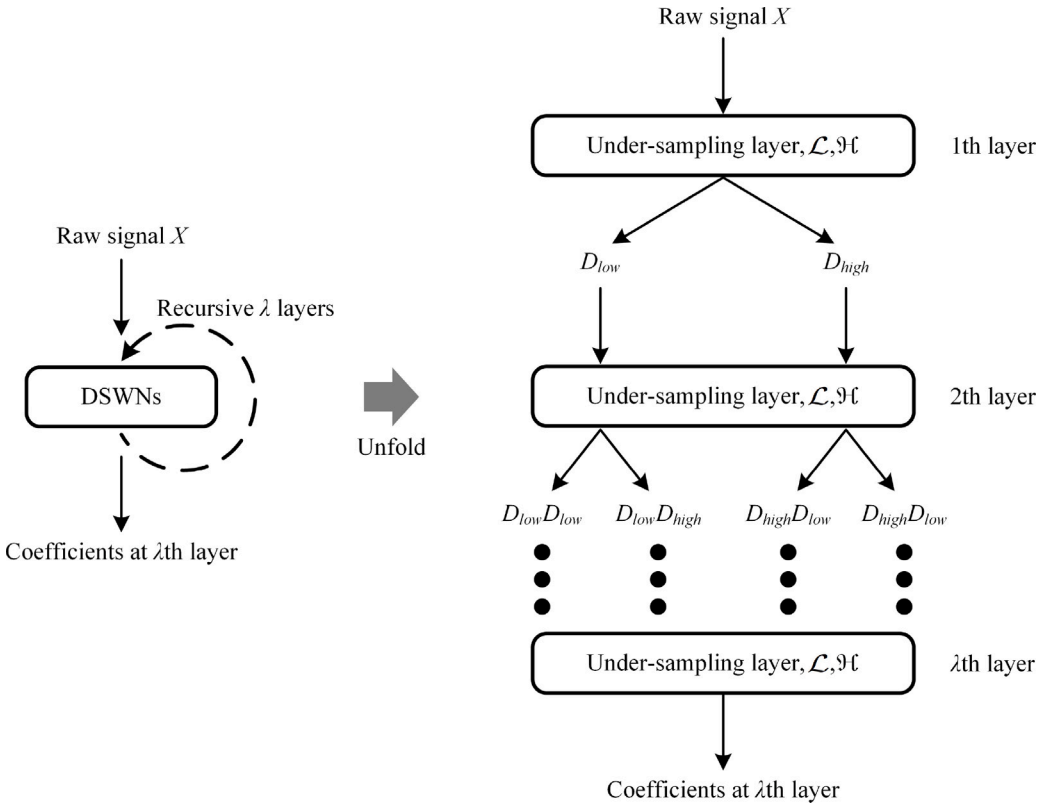

Fig. 3. Simulating the WPT with λ decomposition levels by recursive DSWNs.

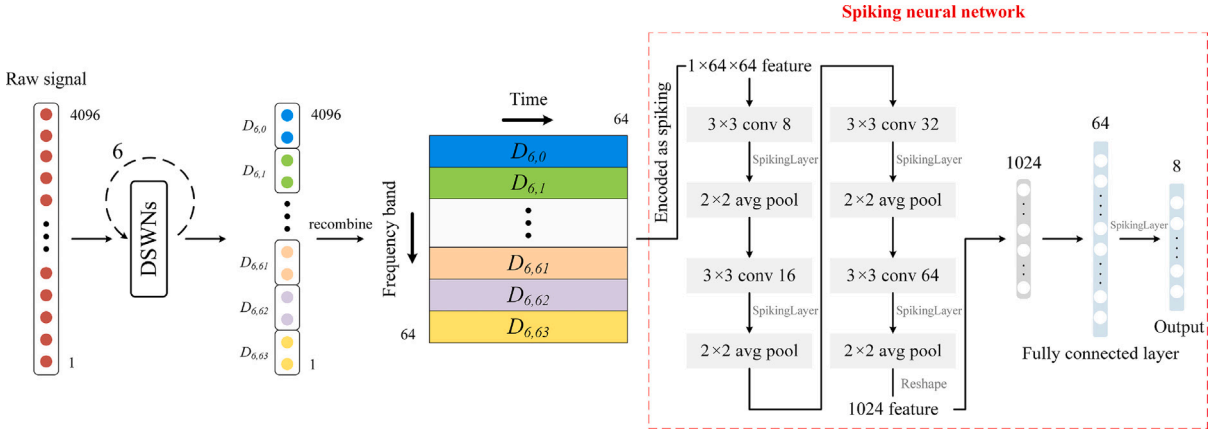

Fig. 4. Architecture of the proposed method.

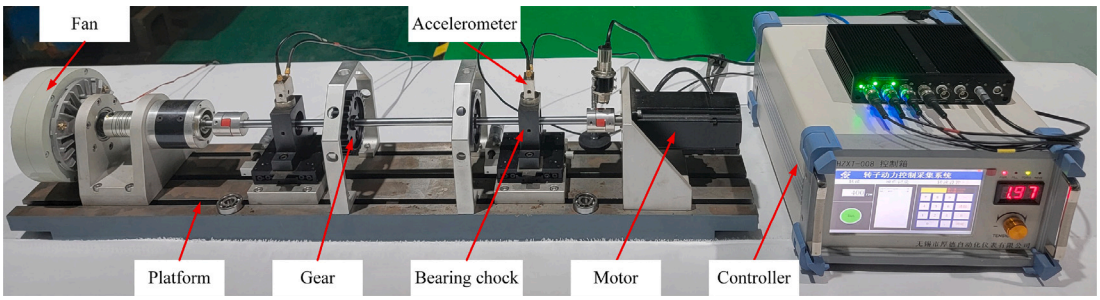

Fig. 5. Physical structure of the device in the experiment.

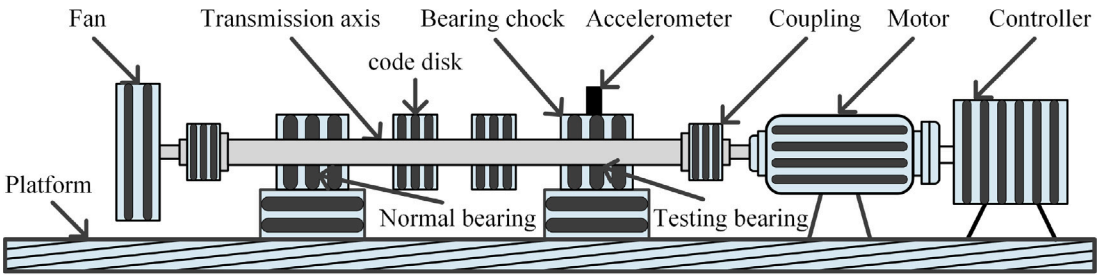

Fig. 6. Simplified diagram of the device in the experiment.

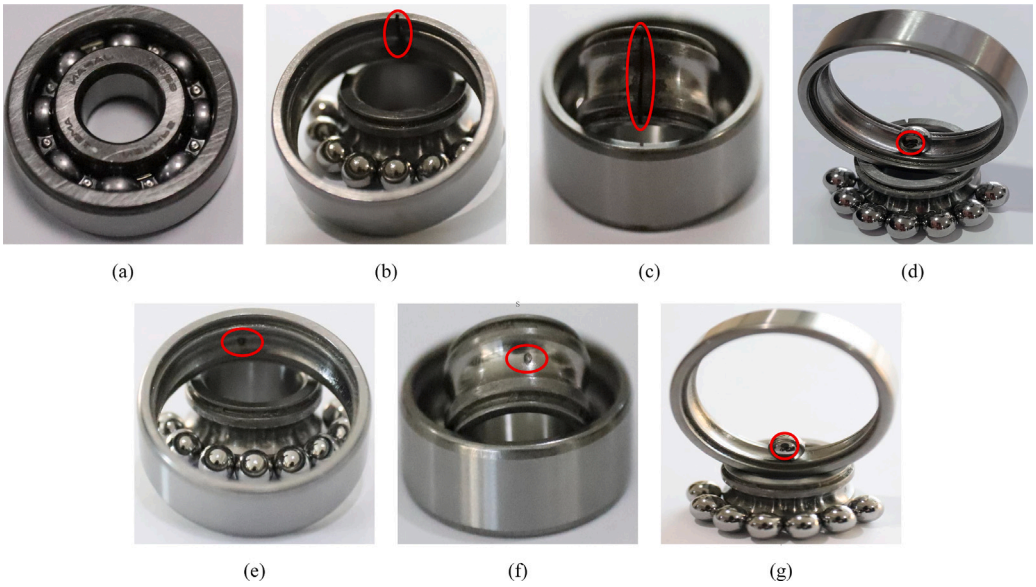

Fig. 7. Seven statuses of bearings in the experiment: (a) normal bearing, (b) outer race with crack, (c) inner race with crack, (d) ball with crack, (e) outer race with pitting corrosion, (f) inner race with pitting corrosion, and (g) ball with pitting corrosion.

Table 2
Detail of 13 statuses in the experiment.

Statuses	Labels	Fault types	Defect positions on bearing	Defect classes
1	O-C	Single point defect	Outer race (O)	Crack (C)
2	I-C	Single point defect	Inner race (I)	Crack (C)
3	B-C	Single point defect	Ball (B)	Crack (C)
4	O-P	Single point defect	Outer race (O)	Pitting corrosion (P)
5	I-P	Single point defect	Inner race (I)	Pitting corrosion (P)
6	B-P	Single point defect	Ball (B)	Pitting corrosion (P)
7	OI-C	Compound defect	Outer race and inner race (OI)	Crack (C)
8	IB-C	Compound defect	Inner race and ball (IB)	Crack (C)
9	BO-C	Compound defect	Ball and outer race (BO)	Crack (C)
10	OI-P	Compound defect	Outer race and inner race (OI)	Pitting corrosion (P)
11	IB-P	Compound defect	Inner race and ball (IB)	Pitting corrosion (P)
12	BO-P	Compound defect	Ball and outer race (BO)	Pitting corrosion (P)
13	N	–	Null	Normal (N)

4.2. Model configurations in the experiment

Eight methods in Table 4 are implemented in the laboratory experiment to recognize the dataset with compound defect. For one thing, five traditional multi-label models based on shallow structure, including BR, CC, LP, ML-KNN, and Rank-SVM, are executed, to validate the effectiveness of the proposed method relative to the traditional machine learning methods. For another, ML-CNN and WGI-ML-CNN based on ANN is also run as baseline models. Specially, the input data of proposed method is encoded in a range between 0 and 1. The computation of wavelet gradient is obtained by using the Python packet "PyWavelets" and DB1 wavelet function. All codes are developed with the PyTorch platform. Moreover, the code about SNN model is implemented by the Python packet "sinabs" [45].

4.3. Prediction results and discussions in the experiment

Eight methods are run for ten times on raw dataset, and related results are illustrated in Fig. 9. Moreover, the comparison of eight methods is shown in Table 5, which contains the minimum, average and maximum. All in all, related analysis can be summarized as follows.

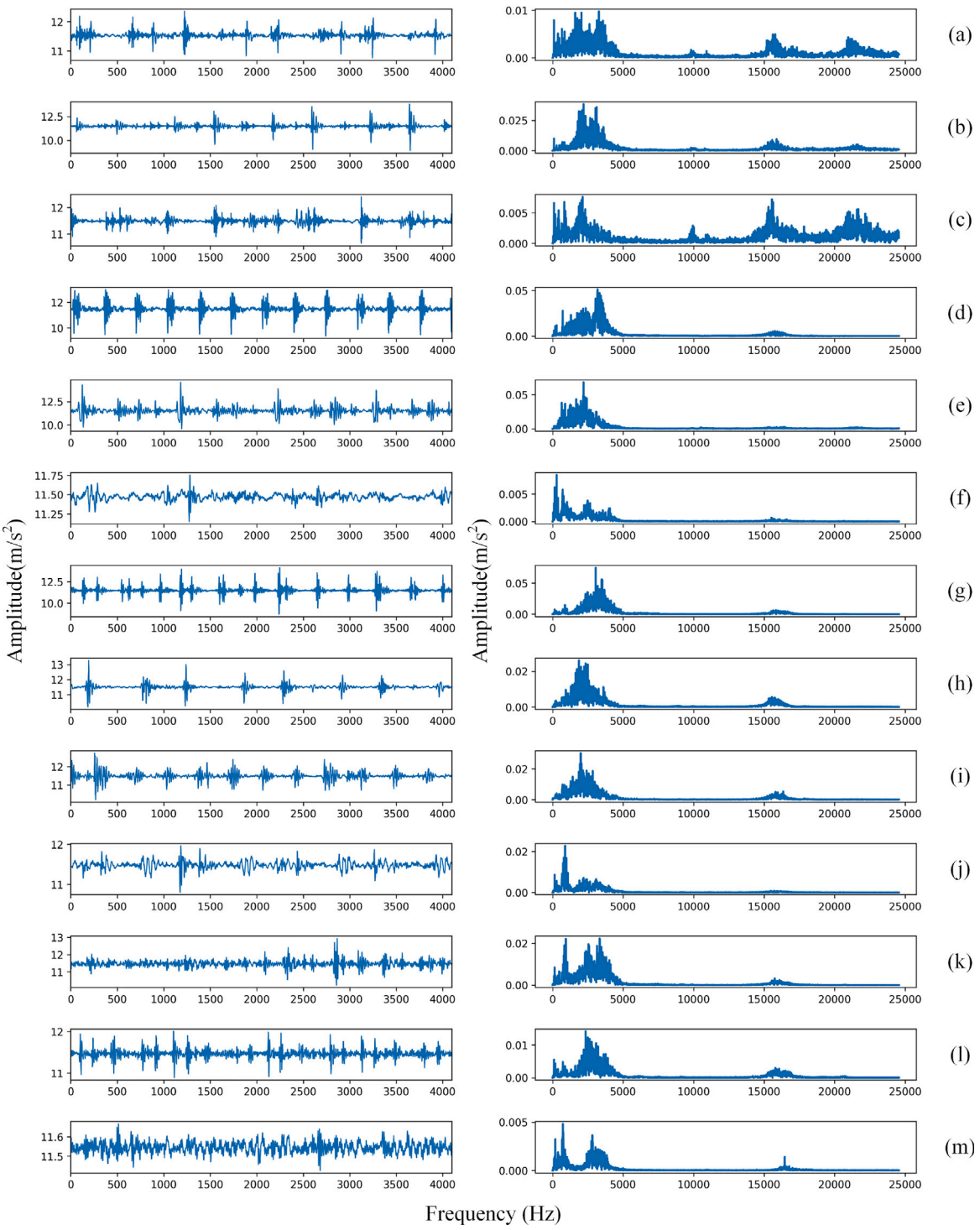

Fig. 8. Time-domain distribution of raw signals with 13 operations in experiment. (a) O-C, (b) I-C, (c) B-C, (d) O-P, (e) I-P, (f) B-P, (g) OI-C, (h) IB-C, (i) BO-C, (j) OI-P, (k) IB-P, (l) BO-P, (m) N.

(1) Accuracies of five traditional multi-label models, like BR, CC, LP, ML-KNN, and Rank-SVM, are substantially low. This phenomenon indicates that traditional multi-label models based on machine learning with shallow structure handle difficultly the task of compound defect detection.

(2) Two baseline models based on ANN, like ML-CNN and WGI-ML-CNN are also trained with the multi-label loss function. For the average of detection accuracies, WGI-ML-CNN is 2.13% higher than ML-CNN. Accordingly, recursive DSWNs integrating wavelet gradient can enhance the performance in recognizing compound defect of bearings.

(3) The proposed model called WGI-SNN is trained by raw data, where the output of recursive DSWNs is transformed into the spikes. The average accuracy of WGI-SNN is 5.64% and 3.51% higher than ML-CNN and WGI-ML-CNN, respectively. Consequently, it validates that the proposed method is effective, and can obtain a satisfactory detection performance in recognizing compound defect.

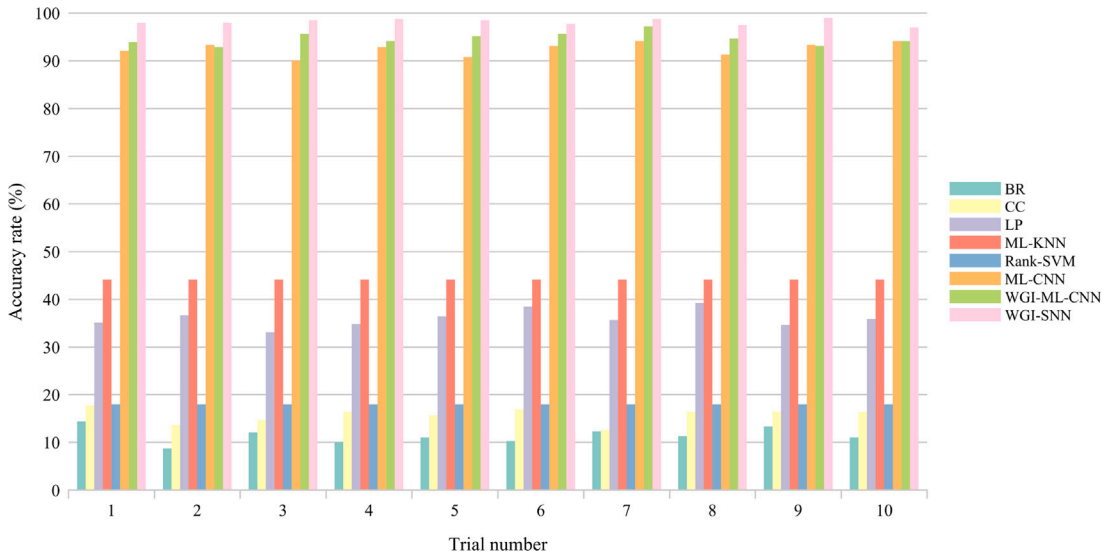

Fig. 9. Comparison of eight methods in the experiment.

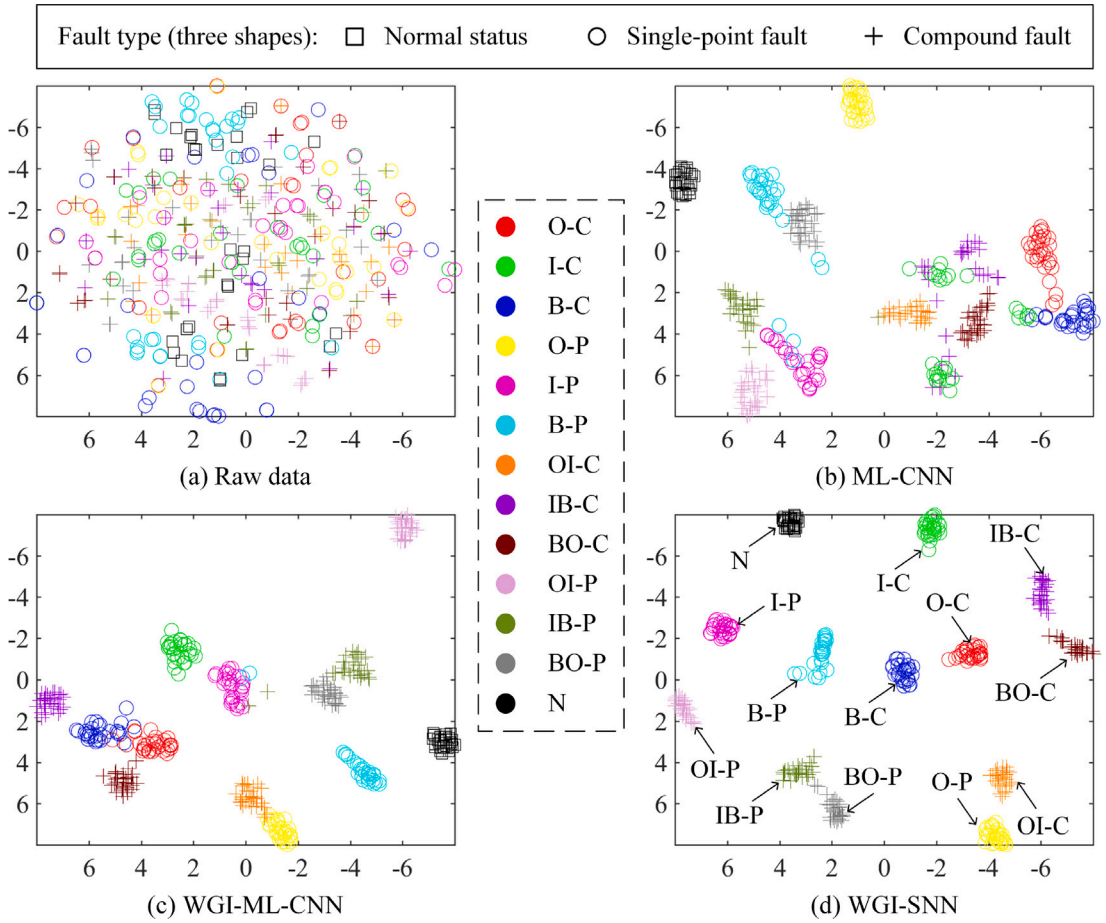

Fig. 10. 2D visualizations of defect feature extraction abilities of three methods in the experiment: (a) raw data, (b) ML-CNN, (c) WGI-ML-CNN, (D) WGI-SNN. Thirteen colors and three shapes of signs represent thirteen statuses and three defect types of the experimental bearing, respectively.

In addition, the experiment adopts the visualization technology called t-stochastic neighbor embedding (t-SNE) [46], to show the 2D visualizations of defect feature extraction abilities of three methods in Fig. 10. To explain clearly the difference of multiple classes of samples, thirteen colors are utilized to represent thirteen statuses of the experimental bearing. And three shapes are used to represent three defect types of the experimental bearing, which contains normal, single-point defect and compound defect. Fig. 10(a) shows the distribution of features of raw data. There is a serious overlap between symbols of different colors, and it indicates that sample are distinguished difficultly. Fig. 10(b) shows the distribution of features extracted by ML-CNN model from raw data. Fig. 10(c) shows the distribution of features

Table 3
Description of labeling each status of bearings with three sublabels in the experiment.

Statuses	Labels	Fault types	Multi-label *		
			1st sublabel	2st sublabel	3st sublabel
1	O-C	Single point defect	1 (Outer race)	5 (Placeholder)	6 (Crack)
2	I-C	Single point defect	2 (Inner race)	5 (Placeholder)	6 (Crack)
3	B-C	Single point defect	3 (Ball)	5 (Placeholder)	6 (Crack)
4	O-P	Single point defect	1 (Outer race)	5 (Placeholder)	7 (Pitting corrosion)
5	I-P	Single point defect	2 (Inner race)	5 (Placeholder)	7 (Pitting corrosion)
6	B-P	Single point defect	3 (Ball)	5 (Placeholder)	7 (Pitting corrosion)
7	OI-C	Compound defect	1 (Outer race)	2 (Inner race)	6 (Crack)
8	IB-C	Compound defect	2 (Inner race)	3 (Ball)	6 (Crack)
9	BO-C	Compound defect	3 (Ball)	1 (Outer race)	6 (Crack)
10	OI-P	Compound defect	1 (Outer race)	2 (Inner race)	7 (Pitting corrosion)
11	IB-P	Compound defect	2 (Inner race)	3 (Ball)	7 (Pitting corrosion)
12	BO-P	Compound defect	3 (Ball)	1 (Outer race)	7 (Pitting corrosion)
13	N	–	4 (Null)	5 (Placeholder)	8 (Normal)

Table 4
Model configurations of eight methods in the experiment.

	Methods	Parameter setting
Method 1	BR	BR model is trained directly by raw datasets with multiple sublabels. Specially, BR model is implemented using 120 decision trees.
Method 2	CC	CC model is trained directly by raw datasets with multiple sublabels. Specially, CC model is implemented using 120 decision trees.
Method 3	LP	LP model is trained directly by raw datasets with multiple sublabels. Specially, LP model is implemented using 120 decision trees.
Method 4	ML-KNN	ML-KNN model is trained directly by raw datasets with multiple sublabels. Specially, ML-KNN model is implemented using the KNN classifier with 9 neighbors of each input instance.
Method 5	Rank-SVM	Rank-SVM model is trained directly by raw datasets with multiple sublabels. Specially, Rank-SVM model is implemented using the SVM classifier with radial basis function.
Method 6	ML-CNN	ML-CNN model is trained as a multi-label classification directly by raw datasets with multiple sublabels. Specially, ML-CNN model is trained using the multi-label classification loss, while the size of output layer is set to the number of marked characters in all sublabels. Moreover, learning rate, batch size and training epoch are set to 0.00002, 12 and 100, respectively.
Method 7	WGI-ML-CNN	WGI-ML-CNN model is obtained by combining recursive DSWNs and ML-CNN. Learning rate, batch size and training epoch are set to 0.00002, 12 and 100, respectively.
Method 8	WGI-SNN	WGI-SNN framework is clearly described in Section 3. Learning rate, batch size and training epoch are set to 0.00006, 13 and 100, respectively.

Table 5
Results of all methods in compound defect experiment.

Methods	BR	CC	LP	ML-KNN	Rank-SVM	ML-CNN	WGI-ML-CNN	WGI-SNN
Minimum	8.72%	12.56%	33.08%	44.10%	17.95%	90.00%	92.82%	96.92%
Average	11.44%	15.67%	36.00%	44.10%	17.95%	92.49%	94.62%	98.13%
Maximum	14.36%	17.69%	39.23%	44.10%	17.95%	94.10%	97.18%	98.97%

extracted by WGI-ML-CNN model from raw data. Fig. 10(d) shows the distribution of features extracted by WGI-SNN model from raw data. The overlap in Fig. 10(d) is pretty slight, and the distribution of features extracted by WGI-SNN model is clearest. Therefore, it demonstrates that the extraction ability of WGI-SNN model is strongest compared with ML-CNN and WGI-ML-CNN models. Moreover, the comparison of Fig. 10(b) and (c) verifies that recursive DSWNs integrating wavelet gradient can improve powerfully the extraction ability for compound defect feature.

5. Conclusion

While handling compound defect detection in actual engineering application, the complex correlation between compound defect and single-point defect leads easily to a detection result with low accuracy. Current deep models based ANN are difficult to solve the problem. Due to the lack of biological interpretability, ANN has encountered great challenges. Besides, the training of the neural network is independent of the preprocessing process for raw signals, so that there is unable to complete an end-to-end mode. Therefore, this paper proposed a WGI-SNN model, combining SNN and wavelet gradient. In brief, the improvement of the proposed method recognizing the compound defect, can be summarized as two points. For one thing, by analyzing the computation of wavelet decomposition, WGPM is deduced. Furthermore, based on the deduced WGPM, recursive DSWNs are designed to simulate the WPT process. For another thing, a biological intelligent model with an end-to-end mode is implemented, by constructing a SNN model with recursive DSWNs. Finally, the effectiveness and practicability of the proposed method are demonstrated by the comparison of the WGI-SNN model and other advanced compound defect detection models minutely in an experiment.

CRediT authorship contribution statement

Jianping Xuan: Conception of the study, Methodology, Writing, Investigation, Funding acquisition. **Zisheng Wang:** Performed the experiment, Wrote the manuscript. **Shaochen Li:** Contributed significantly to analysis and manuscript preparation. **Ang Gao:** Performed the data analyses. **Chunlei Wang:** Performed the analysis with constructive discussions. **Tielin Shi:** Performed the analysis with constructive discussions.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant No. 52175094) and the National Key R&D Program of China (Grant No. 2020YFB2007700).

References

[1] Dongfang Zhao, Shulin Liu, Hongyi Du, Lu Wang, Zhonghua Miao, Deep branch attention network and extreme multi-scale entropy based single vibration signal-driven variable speed fault diagnosis scheme for rolling bearing, Adv. Eng. Inform. 55 (2023) 101844.
[2] Lin Zuo, Fengjie Xu, Changhua Zhang, Tangfan Xiahou, Yu Liu, A multi-layer spiking neural network-based approach to bearing fault diagnosis, Reliab. Eng. Syst. Saf. 225 (2022) 108561.
[3] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, Asoke K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Signal Process. 138 (2020) 106587.
[4] Kun Xu, Xianguang Kong, Qibin Wang, Shengkang Yang, Naining Huang, Junji Wang, A bearing fault diagnosis method without fault data in new working condition combined dynamic model with deep learning, Adv. Eng. Inform. 54 (2022) 101795.
[5] Hao Su, Ling Xiang, Aijun Hu, Yonggang Xu, Xin Yang, A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions, Mech. Syst. Signal Process. 169 (2022) 108765.
[6] Jipu Li, Ruyi Huang, Zhuyun Chen, Guolin He, Konstantinos C. Gryllias, Weihua Li, Deep continual transfer learning with dynamic weight aggregation for fault diagnosis of industrial streaming data under varying working conditions, Adv. Eng. Inform. 55 (2023) 101883.
[7] Kaige Su, Jianhua Liu, Hui Xiong, A multi-level adaptation scheme for hierarchical bearing fault diagnosis under variable working conditions, J. Manuf. Syst. 64 (2022) 251–260.
[8] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (2015) 436.
[9] Yaguo Lei, Bin Yang, Xinwei Jiang, Feng Jia, Naipeng Li, Asoke K. Nandi, Applications of machine learning to machine fault diagnosis: A review and roadmap, Mech. Syst. Signal Process. 138 (2020) 106587.
[10] Haopeng Liang, Jie Cao, Xiaoqiang Zhao, Multi-scale dynamic adaptive residual network for fault diagnosis, Measurement 188 (2022) 110397.
[11] Zuoyi Chen, Jun Wu, Chao Deng, Chao Wang, Yuanhang Wang, Residual deep subdomain adaptation network: A new method for intelligent fault diagnosis of bearings across multiple domains, Mech. Mach. Theory 169 (2022) 104635.
[12] Shengkang Yang, Xianguang Kong, Qibin Wang, Zhongquan Li, Han Cheng, Kun Xu, Deep multiple auto-encoder with attention mechanism network: A dynamic domain adaptation method for rotary machine fault diagnosis under different working conditions, Knowl.-Based Syst. 249 (2022) 108639.
[13] Hao Wu, Jimeng Li, Qingyu Zhang, Jinxin Tao, Zong Meng, Intelligent fault diagnosis of rolling bearings under varying operating conditions based on domain-adversarial neural network and attention mechanism, ISA Trans. 130 (2022) 477–489.
[14] Diwang Ruan, Jin Wang, Jianping Yan, Clemens Gühmann, CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis, Adv. Eng. Inform. 55 (2023) 101877.
[15] Amirhossein Tavanaei, Masoud Ghodrati, Saeed Reza Kheradpisheh, Timothée Masquelier, Anthony Maida, Deep learning in spiking neural networks, Neural Netw. 111 (2019) 47–63.
[16] Alberto Prieto, Beatriz Prieto, Eva Martinez Ortigosa, Eduardo Ros, Francisco Pelayo, Julio Ortega, Ignacio Rojas, Neural networks: An overview of early research, current frameworks and new challenges, Neurocomputing 214 (2016) 242–268.
[17] Abhronil Sengupta, Yuting Ye, Robert Wang, Chiao Liu, Kaushik Roy, Going deeper in spiking neural networks: VGG and residual architectures, Front. Neurosci. 13 (2019).
[18] Wei Fang, Zhaofei Yu, Yanqi Chen, Tiejun Huang, Timoth E.E. Masquelier, Yonghong Tian, Deep residual learning in spiking neural networks, in: 35th Conference on Neural Information Processing Systems, 2021, pp. 21056–21069.
[19] João D. Nunes, Marcelo Carvalho, Diogo Carneiro, Jaime S. Cardoso, Spiking neural networks: A survey, IEEE Access 10 (2022) 60738–60764.
[20] Kashu Yamazaki, Viet-Khoa Vo-Ho, Darshan Bulsara, Ngan Le, Spiking neural networks and their applications: A review, Brain Sci. 12 (7) (2022).
[21] Yasir Hassan Ali, Falah Y. H. Ahmed, Ahmed M. Abdelrhman, Salah M. Ali, Abdoulhdi A. Borhana, Raja Ishak Raja Hamzah, Novel spiking neural network model for gear fault diagnosis, in: 2022 2nd International Conference on Emerging Smart Technologies and Applications, 2022, pp. 1–6.
[22] Lin Zuo, Lei Zhang, Zhe-Han Zhang, Xiao-Ling Luo, Yu Liu, A spiking neural network-based approach to bearing fault diagnosis, J. Manuf. Syst. 61 (2021) 714–724.
[23] Jun Wang, Tianfu Li, Chuang Sun, Ruqiang Yan, Xuefeng Chen, Improved spiking neural network for intershaft bearing fault diagnosis, J. Manuf. Syst. 65 (2022) 208–219.
[24] Juan Xu, Long Zhou, Weihua Zhao, Yuqi Fan, Xu Ding, Xiaohui Yuan, Zero-shot learning for compound fault diagnosis of bearings, Expert Syst. Appl. 190 (2022) 116197.
[25] Haidong Shao, Hongkai Jiang, Haizhou Zhang, Wenjing Duan, Tianchen Liang, Shuaipeng Wu, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, Mech. Syst. Signal Process. 100 (2018) 743–765.
[26] Adane Nega Tarekegn, Mario Giacobini, Krzysztof Michalak, A review of methods for imbalanced multi-label classification, Pattern Recognit. 118 (2021) 107965.
[27] Shanling Han, Shoudong Zhang, Yong Li, Long Chen, The multilabel fault diagnosis model of bearing based on integrated convolutional neural network and gated recurrent unit, Int. J. Intell. Comput. Cybern. (2021).
[28] Chongchong Yu, Yaqian Ning, Yong Qin, Weijun Su, Xia Zhao, Multi-label fault diagnosis of rolling bearing based on meta-learning, Neural computing & applications 33 (10) (2021) 5393–5407.
[29] Yanghui Tan, Jundong Zhang, Hui Tian, Dingyu Jiang, Lei Guo, Gaoming Wang, Yejin Lin, Multi-label classification for simultaneous fault diagnosis of marine machinery: A comparative study, Ocean Eng. 239 (2021) 109723.
[30] Fengying Li, Xin Ma, Youqing Wang, A multi-label method of state partition and fault diagnosis based on binary relevance algorithm, 2020 IEEE 9th Data Driven Control and Learning Systems Conference, 2020, pp. 567–572.
[31] Jesse Read, Bernhard Pfahringer, Geoff Holmes, Eibe Frank, Classifier chains for multi-label classification, Mach. Learn. 85 (3) (2011) 333–359.
[32] Xin Ma, Yu Hu, Menghui Wang, Fengying Li, Youqing Wang, Degradation state partition and compound fault diagnosis of rolling bearing based on personalized multilabel learning, IEEE Trans. Instrum. Meas. 70 (2021) 1–11.
[33] Jesse Read, Antti Puurula, Albert Bifet, Multi-label classification with meta-labels, 2014 IEEE International Conference on Data Mining, pp. 941–946.
[34] Rene-Vinicio Sanchez, Pablo Lucero, Jean-Carlo Macancela, Mariela Cerrada, Rafael E. Vasquez, Fannia Pacheco, Multi-fault diagnosis of rotating machinery by using feature ranking methods and SVM-based classifiers, 2017 International Conference on Sensing, Diagnostics, Prognostics, and Control, IEEE, 2017, pp. 105–110.
[35] Ge Xin, Zhe Li, Limin Jia, Qitian Zhong, Honghui Dong, Nacer Hamzaoui, Jerome Antoni, Fault diagnosis of wheelset bearings in high-speed trains using logarithmic short-time Fourier transform and modified self-calibrated residual network, IEEE Trans. Ind. Inform. 18 (10) (2022) 7285–7295.
[36] Minghang Zhao, Myeongsu Kang, Baoping Tang, Michael Pecht, Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis, IEEE Trans. Ind. Electron. 66 (6) (2019) 4696–4706.
[37] Shoucong Xiong, Hongdi Zhou, Shuai He, Leilei Zhang, Qi Xia, Jianping Xuan, Tielin Shi, A novel end-to-end fault diagnosis approach for rolling bearings by integrating wavelet packet transform into convolutional neural network structures, Sensors 20 (2020).
[38] Qiufu Li, Linlin Shen, Sheng Guo, Zhihui Lai, Wavelet integrated CNNs for noise-robust image classification, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7245–7254.
[39] Zisheng Wang, Jianping Xuan, Tielin Shi, Alternative multi-label imitation learning framework monitoring tool wear and bearing fault under different working conditions, Adv. Eng. Inform. 54 (2022) 101749.
[40] Wulfram Gerstner, Werner M. Kistler, Richard Naud, Liam Paninski, Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition, Cambridge University Press, 2014.
[41] Shirin Dora, Nikola Kasabov, Spiking neural networks for computational intelligence: An overview, Big Data Cogn. Comput. 5 (4) (2021).
[42] Michael Hopkins, Garibaldi Pineda-García, Petruţ A. Bogdan, Steve B. Furber, Spiking neural networks for computer vision, Interface Focus 8 (4) (2018) 20180007.
[43] Aboozar Taherkhani, Ammar Belatreche, Yuhua Li, Georgina Cosma, Liam P. Maguire, T.M. McGinnity, A review of learning in biologically plausible spiking neural networks, Neural Netw. 122 (2020) 253–272.
[44] Jesus L. Lobo, Javier Del Ser, Albert Bifet, Nikola Kasabov, Spiking neural networks and online learning: An overview and perspectives, Neural Netw. 121 (2020) 88–100.
[45] Bodo Rueckauer, Iulia-Alexandra Lungu, Yuhuang Hu, Michael Pfeiffer, Shih-Chii Liu, Conversion of continuous-valued deep networks to efficient event-driven networks for image classification, Front. Neurosci. 11 (2017).
[46] Zisheng Wang, Jianping Xuan, Tielin Shi, Multi-source information fusion deep self-attention reinforcement learning framework for multi-label compound fault recognition, Mech. Mach. Theory 179 (2023) 105090.