RESEARCH ARTICLE

One Dimensional Convolutional Neural Networks Using Sparse Wavelet Decomposition for Bearing Fault Diagnosis

**XIAOFAN LIU [ID], (Student Member, IEEE), JASON CENTENO,
JUAN ALVARADO, AND LIZHE TAN [ID], (Senior Member, IEEE)**
Department of Electrical and Computer Engineering, Purdue University Northwest, Hammond, IN 46323, USA

Corresponding author: Lizhe Tan (lizhetan@pnw.edu)

ABSTRACT This paper proposes a novel algorithm for bearing fault diagnosis using sparse wavelet decomposition for feature extraction combined with a multi-scale one dimensional convolutional neural network (1-D CNN). The proposed algorithm consists of three stages. The first stage determines bearing fault frequency bands according to bearing physical parameters and constructs a sparse wavelet decomposition structure. The second stage decomposes a raw bearing signal into multi-resolution signals based on a decomposition structure achieved at the second stage. Finally, the decomposed multi-resolution signal features are fed into the sub-neural networks according to the multi-scale 1-D CNN (MSCNN) network, and then the outputs of the final convolutional/polling layers are concatenated into a single channel which is further used as the input to a fully connected layer. In comparison with the other bearing fault diagnosis methods, our proposed algorithm can achieve a higher classification accuracy of 99.85% using the Case Western Reserve University (CWRU) bearing dataset. The proposed algorithm is successfully validated via our designed experiments.

INDEX TERMS 1-D CNNs, sparse wavelet decomposition, multi-scale neural network, bearing fault diagnosis.

I. INTRODUCTION

Rolling element bearings in rotating machines are widely employed in industrial machinery and found anywhere from household appliances to heavy duty machines. The rolling element bearings are critical components. A failure of rolling element bearings in a rotating machine is one of the most frequent reasons for machine breakdown [1], [2], [3], [4], [5]. Using diagnostic methods to detect bearing defects at an early stage with proper maintenance can greatly extend a working life of rotating parts in machinery and reduce the economic loss due to downtime [2], [6], [7], [8], [9].

Generally, any rolling bearing in operation generates vibration signals due to the unevenness of the forces acting on internal components. If there is damage occurred in bearing

---

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan [ID].

components, the bearing will produce a periodic pulse signal that is different from the one produced by a normal operating condition. Therefore, diagnostics of bearing defects can be achieved by analyzing and monitoring the acquired vibration signals [10], [11] on a regular base. Common methods for bearing fault diagnosis by employing vibration signals are classified into traditional signal processing-based methods, machine learning-based methods, and deep learning-based methods. In traditional signal processing-based methods, the vibration signals are analyzed in time, frequency, or time-frequency domains to derive vibration signatures and then using the achieved characteristics of the vibration signatures to determine the type of defects [12]. Common signal processing techniques include fast Fourier transform (FFT), wavelet transform (WT), empirical mode decomposition (EMD), and wavelet packet transform (WPT) [14], [15], [16], [17], [18]. Time-domain analysis often determines

the bearing condition by directly investigating some scalar indicators in vibration signals [19]. These scalar indicators usually include, but are not limited to, signal peak value, root-mean-square (RMS), crest factor, kurtosis, spectral kurtosis, and clearance factor [20], [21], [22], [23]. However, the time domain-based signal analysis is difficult to handle non-smooth vibration signals. This fact leads to attention and development of frequency domain-based signal analysis methods. M. Cocconcelli *et al.* [24] used the short-time Fourier transform (STFT) to process the original vibration signal and analyzed the results of the conversion in frequency domain to conduct the detection of bearing defects. Similarly, J.-H. Lee *et al.* [25] used the Wigner-Ville distribution in their study to perform the frequency domain analysis of vibration signals for mechanical fault detection. Although various signal processing methods can effectively extract the features, that is, vibration signatures, present in the original signal, the determination of defect status for these features is still conducted manually. This process relies heavily on judgment rules developed by experts using the characteristics exhibited from typical defects. However, when bearing defects produce atypical or obscure characteristics, these manually developed judgment rules may produce erroneous judgments due to the inability to describe these characteristics.

On the other hand, machine learning can discover unique patterns hidden in data that are difficult to be detected manually. A machine learning model is capable of making accurate judgments based on unique data patterns learned from old data when processing completely new data. M.-Y. Chow *et al.* [26] used an artificial neural network (ANN) to detect early mechanical faults occurring in the rotors of small and medium-sized motors. A. Malhi and R. X. Gao [27] used the principal component analysis (PCA) to automatically extract the vibration signal features most relevant to the defect state and classify the features using a feedforward neural network (FFNNs). Y. S. Wang *et al.* [28] proposed an intelligent engine fault diagnosis method by employing the Hilbert-Huang transform (HHT) with the support vector machine (SVM) algorithm. Similarly, D. H. Pandya *et al.* [29] applied the HHT to process the acoustic signals of bearings with the difference in which the determination of a type of bearing defects was done using the K-nearest neighbor algorithm (KNN). F. Shen *et al.* [30] used a transfer learning-based singular value decomposition (SVD) model to accomplish the diagnosis of bearing defects. Although the machine learning-based bearing defect diagnosis algorithms can achieve good results in most cases, they have difficulty to extract more abstract features due to their shallow model, which limits for a further accuracy improvement of diagnosis performance. On the other hand, the methods with deep learning-based bearing fault diagnosis achieve the improved performance and have gradually become an active research area in recent years.

A deep learning model usually relies on a deeper neural network structure and it requires a larger amount of data to conduct network training. Although the deep learning requires greater computational load as well as a larger amount of training data requirements, such problems can be solved with an increase in computer performance and the number of available open-source datasets in recent years.

Most applications of deep learning models for bearing defect diagnosis tend to use the raw signal directly for end-to-end learning. M. Zhao *et al.* [31] proposed a novel deep neural network structure called deep residual shrinkage networks to improve the accuracy of fault diagnosis. However, using the signal features extracted from processing the raw signals may significantly improve the bearing diagnostic accuracy by using the deep learning neural networks. Our previous research [32] using the input features generated from STFT and employing the structure of a convolutional neural network (CNN) optimized by the particle swarm optimization (PSO) algorithm has achieved improved classification results based on public datasets. Again, a combination of STFT and CNN with tunable input size for bearing fault diagnosis are explored in [33]. Furthermore, evaluation of deep learning neural networks such as CNN and long-short term memory (LSTM) with the processed input features using the various signal transforms, including STFT, Cepstrum, WPT, and EMD has been reported in [34]. Although all above mentioned methods have obtained good results on multiple public datasets, these deep learning models only performed at a single scale level of input signals. There is still a room for further improvement in terms of classification accuracy and computational load. To proceed for development, multiscale neural network-based algorithms are recently proposed. W. Chen and K. Shi [35] used an end-to-end multiscale convolutional neural network (MSCNN) to classify the types of bearing defects. In their approach, the original signal, the smoothed signal, and the down-sampled signal constitute three different scales of inputs. G. Jiang *et al.* [36] also used the MSCNN to investigate the health status detection of wind turbine gearboxes. As matter of fact in [36], the MSCNN simply used the down sampled signals at different scales as the inputs, it achieved excellent detection results. Furthermore, Y. Yao *et al.* [37] used a one-dimensional convolutional layer to decompose the original signal into multi-scale features to operate in an end-to-end form for improving the practicality. In addition, their method introduces an attention mechanism to improve the final detection. Our initial results using 1-D CNN, auto-encoder, deep residual network (ResNet), and gate recurrent unit (GRU) along with input multi-scale features using down sampled signals, EMD, and WPT are reported in [38], where 1-D CNN with the WPT signal decomposition demonstrates the best classification accuracy. However, the methods proposed in our studies [38] did not explore how to construct multi-resolution signal features in different frequency bands, since the signal characteristics in bearings often happen to be located in several specific frequency zones [32], [39], [40], [41]. These frequency zones are designated by the low frequency, bearing defect frequency, bearing natural frequency, and high frequency zones, respectively, which can be determined from the shaft speed

and bearing physical parameters. Furthermore, the frequency bands from these zones are very narrow and sparse relative to the signal sampling rate in the data acquisition system. Therefore, it is of practical significance to extract signal features from these frequency zones and construct them as signal multi-resolution features to improve the neural networks' diagnosis performance.

To extract the signals in the bearing characteristic frequency bands, we propose to enhance the sparse WPT decomposition algorithm from our previous work [42], [43], [44] by searching an optimal sparse decomposition structure, considering the fact that the path for WPT decomposition follows Gray codes. Once the sparse WPT decomposition is completed, the multi-resolution features can be applied to the sub-neural networks in parallel and then followed by a fully connected layer in the 1-D CNN network. The contributions of this paper are listed as follows:

a. A framework using multi-resolution signal features in the bearing characteristic frequency bands and multi-scale 1-D CNN is proposed.

b. An improved sparse WPT decomposition method for extracting signals in the sparse multi-frequency bands is developed. This method first uses Gray codes to generate all the paths required for completed signal decomposition for each frequency band. Then a sparse decomposition path is found based on the lowest cost of the decomposition operations.

c. A multi-scale 1-D CNN network is designed using the multi-resolution signal features from each bearing characteristic frequency band.

d. The proposed method is validated using public dataset as well as dataset from our designed experiments.

II. BEARING FAULT CHARACTERISTICS

Bearing faults come from roughness fault and single-point fault [32], [39], [40], [41]. The roughness fault is due to the condition in which the surface of a bearing has considerably degraded over a large area. In fact, it is a type of distributed and noncyclic fault caused by improper lubrication, erosion, or pollution. The roughness fault does not contain identifiable characteristic frequency components. On the other hand, most importantly, a single-point fault is a localized fault caused by a small hole, a pit, crack, or a missing a piece of material. When a localized defect in the bearing repeatedly strikes other parts of the bearing during rotation, the successive strikes can produce a series of impulse responses, whose amplitudes are modulated [1]. Thus, the characteristic fault frequency components created by a periodic impact vibration are produced, which can be analyzed to detect the types of bearing faults. Our research focus on the detection and diagnosis of single-point fault from ball bearings.

A. BEARING DEFECTIVE FREQUENCY BANDS

Fig. 1 shows the structure of a typical ball bearing, which consists of inner race, outer race, cage, and balls. The characteristic bearing defect frequencies can be determined from
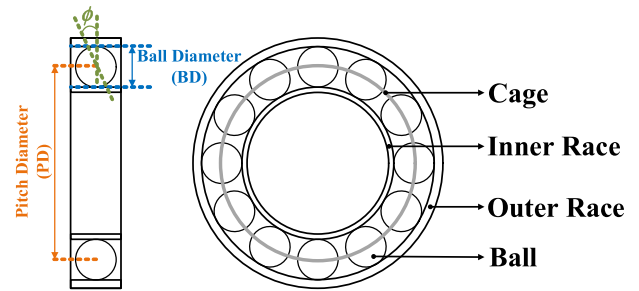

FIGURE 1. The structure of a typical bearing.

Equations (1)-(4), respectively.

$$f_{OD} = \frac{n}{2} f_{rm} \left(1 - \frac{BD}{PD} \cos \phi \right) \qquad (1)$$

$$f_{ID} = \frac{n}{2} f_{rm} \left(1 + \frac{BD}{PD} \cos \phi \right) \qquad (2)$$

$$f_{BD} = \frac{PD}{2BD} f_{rm} \left(1 - \left(\frac{BD}{PD} \right)^2 \cos^2 \varphi \right) \qquad (3)$$

$$f_{Cage} = \frac{1}{2} f_{rm} \left(1 - \frac{BD}{PD} \cos \phi \right) \qquad (4)$$

where f_{OD}, f_{ID}, f_{BD}, and f_{Cage} are the outer race, inner race, ball, and cage frequencies, respectively. n and f_{rm} are the number of balls and shaft speed, respectively. BD and PD are the ball diameter and rolling pitch diameter, respectively. Note that ϕ is the contact angle between the ball and bearing as shown in Fig. 1, which is approximately to be 0~10° [13].

The bearing fault frequencies can be grouped into the following four zones: low frequency, bearing defect frequency, bearing nature resonance frequency, and high frequency zones as shown in Fig. 2.
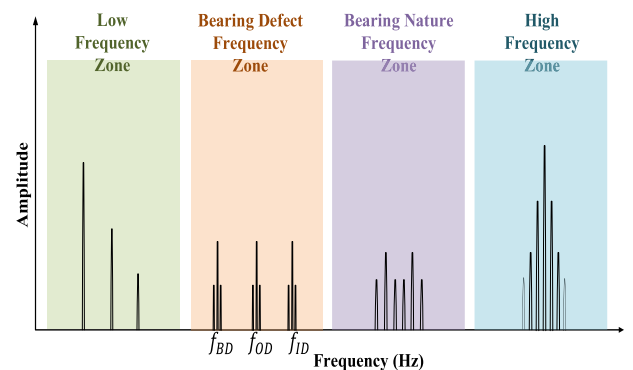

FIGURE 2. The bearing characteristic frequency zones.

The bearing defect frequency zone has the low-frequency components coming from the entry of the ball/roller in the vicinity of defect and high-frequency component caused by the impact of rolling element on the trailing edge of the defect. Again, note that a full-band vibration signal contains not only frequencies from the shaft speed and bearing defects but also the natural frequencies and high frequency noise caused by the bearing rotation. If the full-band vibration signals were

TABLE 1. Bearing defect frequencies of bearing fault.

Name	Fault frequencies (Multiple of shaft speed)	Fault frequencies (Hz)
f_{OD}	3.585	107.55
f_{ID}	5.415	162.45
f_{BD}	2.357	70.71
f_{Cage}	0.3983	11.949

TABLE 2. Characteristic frequency bands for training process.

Band Name	Frequency Range (Hz)
Band 1	$7.50 - 37.50$
Band 2	$63.21 - 169.95$
Band 3	$169.95 - 500.00$

used directly as training data, the detection effectiveness of the model may eventually be diminished by learning features that are irrelevant to the labels. Therefore, extracting features from the characteristic frequency bands and using the extracted features as training data samples for the model will greatly improve the final learning and detection accuracies.

As an illustration using the physical parameters of test bearing provided by the CWRU bearing dataset, the shaft speed of f_{rm} is set to 30 Hz (1800 rpm) since a slight variation between 29.93 and 29.95 Hz in the rotational speed of the bearing rolls during data acquisition and the contact angle is used as $\phi \approx 10^0$ [13], the corresponding fault frequency values for f_{OD}, f_{ID}, f_{BD}, and f_{Cage} by Equations (1)-(4) are shown in Table 1.

From Table 1, the band between f_{BD} and f_{ID} is a target band for a single-point fault, which is between 70.71 Hz and 162.45 Hz.

B. OTHER RELEVANT FREQUENCY BANDS
In addition to the fault frequency band where the inner race, outer race, and ball frequency components are located, the proximity band of f_{rm} also contains the corresponding characteristic information, including the cage frequency. The fault characteristic information in this frequency band comes from the effect of bearing defects on the rolling. In our research, a frequency band from $0.5f_{rm}$ to f_{rm}, which is 15 Hz to 30 Hz, is considered as the low frequency zone. An additional frequency band from 162.46 Hz to 500 Hz is considered to contain information about the defect characteristics in the high-frequency band. The reason for choosing 500 Hz as the maximum frequency is to prevent the invalid information from the high frequency noise which may affect the validity of the training data. After obtaining the required three frequency bands, each band is expanded by $0.25f_{rm}$ from its band frequency edges in order to ensure the sufficient frequency margins to cover the required frequency range. Three frequency bands and their respective frequency edges for extracting training data samples are listed in Table 2.

III. SPARSE WAVELET DECOMPOSITION
A. GRAY CODES IN WAVELET DECOMPOSITION
To extract the corresponding information of the characteristic and relevant frequency bands from the full-band information of the original signal, a complete wavelet packet decomposition (WPT) is first applied. It decomposes the original

signal into two sub-bands at low and high frequencies and continues the decomposition for each subband at the next level. Fig. 3 displays the complete decomposition structure from decomposing a signal with a frequency band of 0~200 Hz up to three levels.
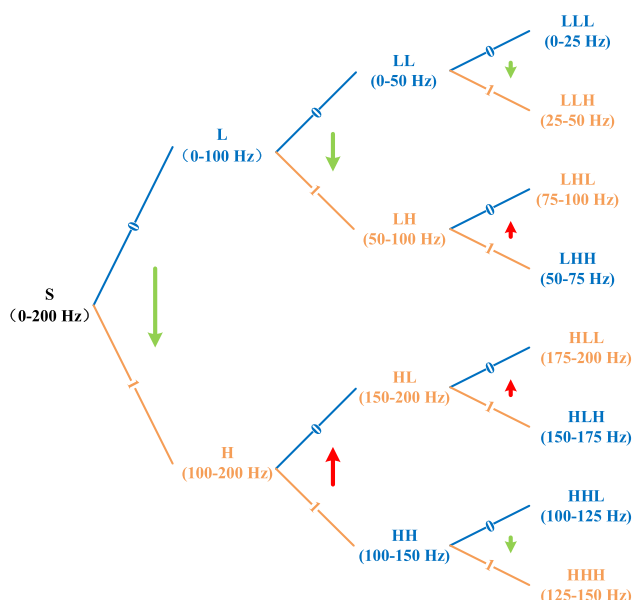

FIGURE 3. The subband tree structuring using WPT.

As shown in Fig. 3, symbol "0" represents the low-frequency decomposition while "L" represents the resulting low-frequency subband. Symbol "1" denotes the high-frequency decomposition while "H" represents the high-frequency subband. The arrows represent the direction of the frequency distribution at the current decomposition level, with the arrows pointing from low to high frequencies. A green arrow indicates no change in the direction of frequency distribution, while a red arrow indicates a direction change of the frequency distribution. According to Fig. 3, the complete decomposition paths follow Gray codes [41], [42], [43] and can be easily seen in Table 3 for decomposition with three levels. Note that the results in Table 3 can be extended to n-level decomposition in which 2^n sub-bands can be arranged in an order from low to high frequencies according to n-bit Gray codes in the generated order.

Using this approach, the decomposition paths for the target frequency band can be easily obtained. As an illustration, if a target band has the frequency range from 75 Hz to 125 Hz, the

TABLE 3. Complete wavelet decomposition.

Frequency Band (Hz)	Gray Code	Subband Name
0 – 25	000	LLL
25 – 50	001	LLH
50 – 75	011	LHH
75 – 100	010	LHL
100 – 125	110	HHL
125 – 150	111	HHH
150 – 175	101	HLH
175 - 200	100	HLL

required decomposition paths found from Table 3 or Fig. 3 are 010 and 110, as well as shown in Fig. 4.
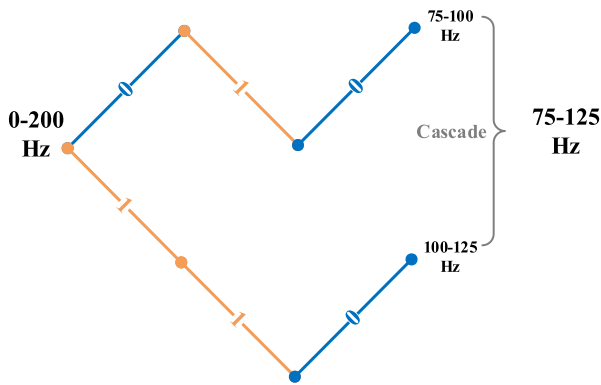

FIGURE 4. WPT decomposition path using gray codes.

For bearing fault detection, we can use this approach to extract the multi-resolution signal features for a given bearing frequency characteristic band. As illustrative example, we can extract features from the band between 7.5 Hz and 37.5 Hz with the given full frequency band from 0 to 6000 Hz at a sampling rate of 12,000 Hz. We will describe our optimized method in the next section.

B. OPTIMAL SPARSE WTP DECOMPOSITION

To obtain the multi-resolution signal features using a sparse WTP decomposition for characteristic frequency bands as listed in Table 2, our proposed method is divided into three steps: determining the optimal decomposition level; finding a sparse decomposition path structure; and performing decomposition execution.

First, for given the sampling frequency of f_s, and target frequency band ideally ranging from f_L to f_H, the minimum number of decomposition levels $N_{\min}$ is determined by

$$N_{\min} = \log_2 \frac{0.5 f_s}{f_H - f_L} \qquad (5)$$

Let $N_{\max}$ be the maximum number of decomposition levels:

$$N_{\max} = N_{\min} + L_{\max} \qquad (6)$$

where L_{max} is the number of maximum allowable steps. We can search for the optimal N_{best} so that the following error

is minimized:

$$err = \min \{\max(f_L - f_{Lt}, f_{Ht} - f_H)\} \qquad (7)$$

where f_{Lt} and f_{Ht} are the actual lower frequency and the upper frequency edges, respectively. They can be determined by

$$f_{Lt} = \left\lfloor \frac{f_L}{\Delta f} \right\rfloor \Delta f \qquad (8)$$

$$f_{Ht} = \left\lceil \frac{f_H}{\Delta f} \right\rceil \Delta f \qquad (9)$$

Note that Δf is the band resolution at decomposition of level N, given below:

$$\Delta f = \frac{f_s}{2^N}, \quad N \in [N_{\min}, N_{\max}] \qquad (10)$$

Let us define the desired minimum error as

$$err_{\min} = \lambda(f_H - f_L) \qquad (11)$$

where the empirical constant λ determines the minimum error as a percentage of the target frequency band. The algorithm for searching the optimal decomposition level N_{best} can be illustrated in the following pseudo-code.

Algorithm About Finding the Best Decomposition Level

Input: N_{min}; N_{max}; err_{min}
Output: N_{best}
  1: **for** $N = N_{min}: N_{max}$ **do**
  2:     **if** $err < err_{min}$ **then**
  3:         $N_{best} = N$
  4:     **else**
  5:         $N_{best} = N_{max}$
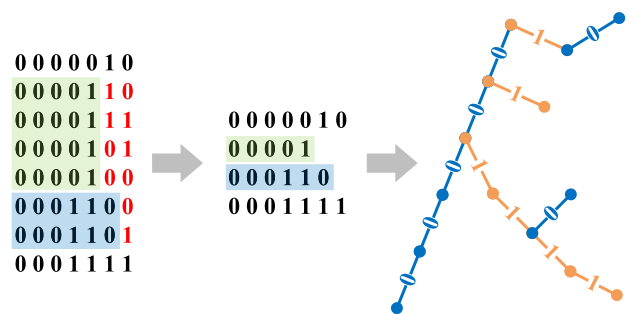 6:   **end if**
  7: **end for**

FIGURE 5. Simplification process in the coding table and subband tree.

With N_{best}, the second step is to find a sparse decomposition structure. This can be illustrated using an example shown in Fig. 5 for the case of extracting band 3. The following parameters are used: $f_s = 6,000$ Hz, $f_L = 169.95$ Hz, $f_H = 500.00$ Hz, $L_{\max} = 6$, and $\lambda = 0.15$, the algorithm finds $L_{best} = 7$, $f_{Lt} = 140.63$ Hz and $f_{Ht} = 515.63$ Hz, respectively. Eight 7-bit Gray codes covering this frequency band, beginning with 0000010 and ending with 0001111 are listed in Fig. 5. Again, note that any consecutive two Gray codes

with least significant bit ending with 1 and 0 can be merged and the corresponding branch can be pruned by one level. Repeating such a process leads to four Gray codes as shown in Fig. 5. Eventually, a sparse decomposition structure using the four Gray codes is achieved and displayed in Fig 5. As can be seen, without the decomposition structure reduction, there are $7 \times 8 = 56$ decomposition operations. However, with the obtained sparse structure, only 13 operations are required.

Note that the extracted signal features are in a nature of sparse multi-resolution. The algorithm can be repeated for the other two bands similarly. The combined and simplified final sparse structure for three bands with a further reduced number of decomposition operations is displayed in Fig. 6. As shown in Fig. 6, the highest level is 10 and there are 31 decomposition operations in total.
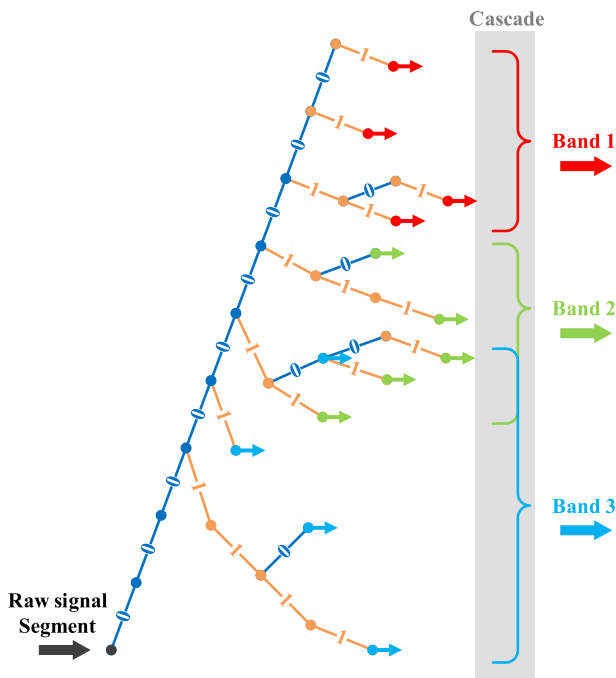

FIGURE 6. Processing raw signal segments using optimized route coding.

The third step is to perform signal extraction by applying the final sparse decomposition structure. The original raw signal with duration of 10 seconds at 12,000 Hz sampling rate is first pre-processed to have zero mean and then normalized. 512 segments with the same length are retrieved at the random starting locations. The length of each segment is determined to have no less than $2^{N_{best}+3}$ data samples in the segment. The requirement of the minimum length ensures that the original signal retains a data sufficient length to prevent loss of signal characteristics after N_{best} level decomposition. Each segment is decomposed into three frequency bands based on the final sparse structure as shown in Fig. 6. After decomposition, the multi-resolution signal features from each band forms a feature vector. The feature vectors from all the three frequency bands are used as the input for the 1-D MSCNN network in training and detection processes.
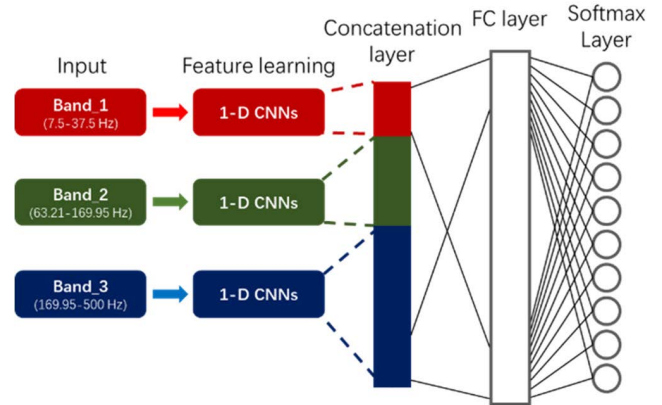

FIGURE 7. Structure of the multi-scale 1-D CNN.

TABLE 4. Information for each layer in the multi-scale 1-D CNN.

Layer Name	Parameters （Kernel Size, No. of Feature Maps, Activation Function, Strides, Padding）
Conv0_1	Conv1D (3×1, 32, ReLU, 1, Same)
Conv1_1	Conv1D (3×1, 32, ReLU, 1, Same)
Conv2_1	Conv1D (3×1, 32, ReLU, 1, Same)
MXP0_1	MaxPooling1D (2, None, None, 1, None)
MXP1_1	MaxPooling1D (2, None, None, 1, None)
MXP2_1	MaxPooling1D (2, None, None, 1, None)
Conv0_2	Conv1D (3×1, 64, ReLU, 1, Same)
Conv1_2	Conv1D (3×1, 64, ReLU, 1, Same)
Conv2_2	Conv1D (3×1, 64, ReLU, 1, Same)
MXP0_2	MaxPooling1D (2, None, None, 1, None)
MXP1_2	MaxPooling1D (2, None, None, 1, None)
MXP2_2	MaxPooling1D (2, None, None, 1, None)
CONCATE	None
BN	None
FLATTEN	None
FC	Dense (128, ReLU, None, None)
OUTPUT	Dense (10, Softmax)

IV. MULTI-SCALE 1-D CNN

In order to use three extracted feature vectors, our proposed method uses three parallel 1-D convolutional/polling layers separately. Fig. 7 displays the proposed structure. As shown in Fig. 7, each 1-D CNN extracts and learn feature data in its frequency range. The outputs from three 1-D CNNs in parallel are combined to form a single channel, that is, concatenation layer, which is fed into a fully connected layer and the softmax layer as the output layer. All the convolutional layers and first fully connected layer are activated with the ReLU function. The softmax function is used as the classifier at the output layer with 10 nodes, each represents a bearing class. Table 4 lists the detailed parameters of the designed multi-scale 1-D CNN structure. Note that maximum polling operation is used for all the polling layers.

V. EXPERIMENTAL RESULTS

In this section, the proposed method was first tested using the Case Western Reserve University (CWRU) bearing dataset and the obtained classification results were compared with

TABLE 5. The data selected form the CWRU bearing dataset.

Fault Depth	Fault Name	Class Name
0.021	**Ball**	**No. 1 (21-Ball)**
	Inner Race	**No. 2 (21-IR)**
	Outer Race 12	**No. 3 (21-OR)**
	Outer Race 6	
	Outer Race 3	
0.014	**Ball**	**No. 4 (14-Ball)**
	Inner Race	**No. 5 (14-IR)**
	Outer Race 6	**No. 6 (14-OR)**
0.07	**Ball**	**No. 7 (07-Ball)**
	Inner Race	**No. 8 (07-IR)**
	Outer Race 3	**No. 9 (07-OR)**
	Outer Race 6	
	Outer Race 12	
0	**Normal**	**No. 0 (Normal)**

TABLE 6. Comparison with other similar methods proposed in [38].

Method Name	Accuracy	F1-score	Recall
Raw data (1 channel)	0.9487	0.9632	0.9614
Downsample (4 channels)	0.9812	0.9864	0.9861
EMD (4 channels)	0.9864	0.9860	0.9847
WPT (4 channels)	0.9870	0.9916	0.9917
Proposed method (3 channels)	**0.9988**	**0.9988**	**0.9989**

TABLE 7. Comparison with other published methods.

Method	Accuracy (%)
ANN [26]	95.00
APF-KNN [29]	96.67
SVD [30]	85.00
DRSN-CW [31]	89.57
MSCNN [36]	98.53
Multi-scale CNN [37]	93.30
Proposed Method	**99.88**

the other mainstream methods. Second, the proposed method is further validated using our own testbed.

A. EXPERIMENTS USING CWRU BEARING DATASET

The hardware platform uses an X86 workstation with the following configuration: CPU Intel i7-6900k, GPU NVIDIA GTX 1080, RAM 128GB DDR3. In the MATLAB environment, data pre-processing, a sparse decomposition structure, decomposition execution, and feature vector generation are conducted. Tensorflow 2.6 is adopted as the main machine learning tool for training and evaluating multi-scale 1-D CNN neural network model. We choose nine bearing data files consisting of three sets of the inner race, outer race, and ball defects at the fault depth levels of 0.021, 0.014, and 0.07 inches, respectively, and one normal data file from the CWRU bearing dataset. For each fault depth level, we selected data files containing the ball, inner race, and outer race defects. Table 5 shows more detailed information and 10 class labels.

The extracted signal features as proposed in Section III from the CWRU bearing dataset are randomly divided into three parts for training, testing, and validation in the ratio of 6:2:2. These three parts of data samples are then used to train the neural network model 10 times independently, and the obtained average results are compared with our previous 1-D CNN methods [38] as shown in Table 6.

As shown in Table 4, the proposed method obtains the accuracy as 0.9988, averaged macro F-1score and Recall as 0.9988 and 0.9989, respectively, which are higher than what we achieved using our previous methods. Table 7 includes comparisons of our proposed algorithm with the other published methods. Note that only the accuracy values

are compared. As the result, our proposed method achieves the best performance in terms of classification accuracy.
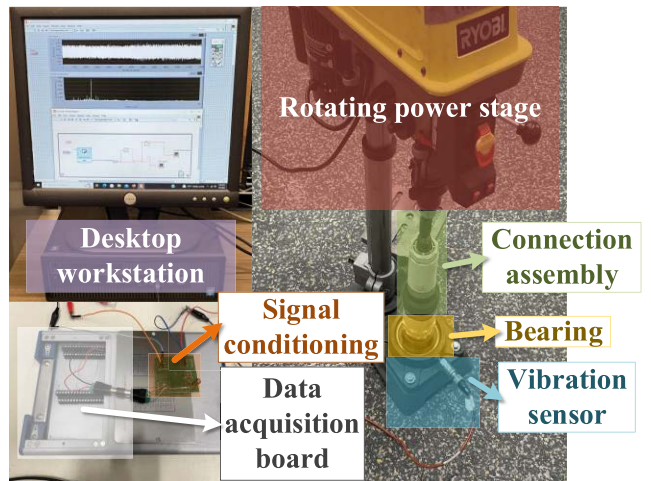

FIGURE 8. Test platform.

B. DESIGNED EXPERIMENTS

Similar to [45], [46], Fig. 8 shows a data acquisition platform (testbed) established in our Engineering Design Studio at Purdue University Northwest to validate our proposed algorithm. It consists of a rotating power stage which is modified from a press drilling machine (Ryobi), joint assembly, bearing unit, a vibration sensor (CTC AC 102-1A accelerometer with 100 mV/G, ±10%), signal condition circuit (deigned PCB circuit), data acquisition board (NI USB-6356), and desktop workstation. Note that a joint assembly is used between the drill and bearing unit and a vibration sensor is installed at the

six o'clock position of the bearing mounting. The vibration signals are captured via the designed signal conditioning circuit and sent to a data acquisition board using LabVIEW. Our complete platform is shown in Fig. 8.

The bearing defects are artificially created on the inside of outer race, the outside of inner race, and the ball surface, respectively. The locations and shapes of defects are shown in Fig. 9.

FIGURE 9. Locations and shapes of bearing defects.

In our experiments, a normal bearing and three fault bearings are driven at a rotational speed of 630 rpm, respectively. The vibration signals are collected at a sampling rate of 12,000 Hz with a duration of 10 seconds (120,000 samples) for each acquisition. The data collected from the four bearings are aggregated to form a self-collected bearing dataset. Details of the bearing physical parameters and the self-collected bearing dataset are shown in Table 8. Typical plots of the acquired vibration signals for the normal bearing, inner race defect, outer race defect, and ball defect are displayed in Fig. 10.
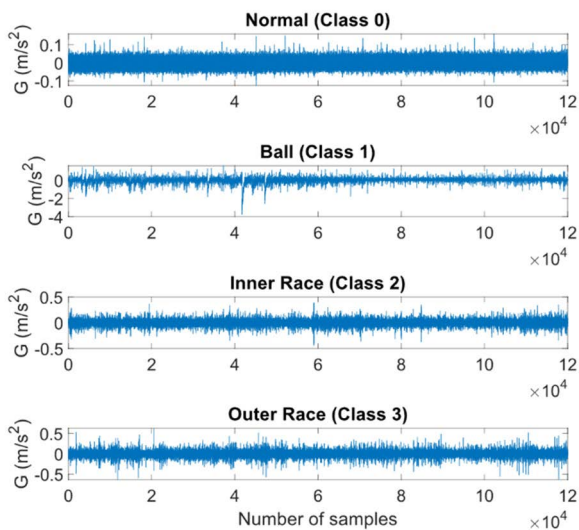

FIGURE 10. Typical acquired vibration signals for various bearing faults.

Based on Section II and using the bearing contact angle as $\phi = 0^0$, the frequency ranges for three sparse decomposition bands (band 1, band 2, band 3) are determined as 2.62∼13.13 Hz, 25.02∼64.76 Hz, and 64.76∼500 Hz, respectively. 10 independent training sessions are conducted on our proposed multi-scale 1-D CNN. The training and

TABLE 8. Bearing parameters in self-collected bearing dataset.

Bearing parameters		Fault location	Class name
Ball diameter	13.838 mm	Ball	No.1
Pitch diameter	75.4 mm	Inner Race	No.2
Rotor speed	630 rpm	Outer Race	No.3
Sampling rate	12000 Hz	Normal	No.0

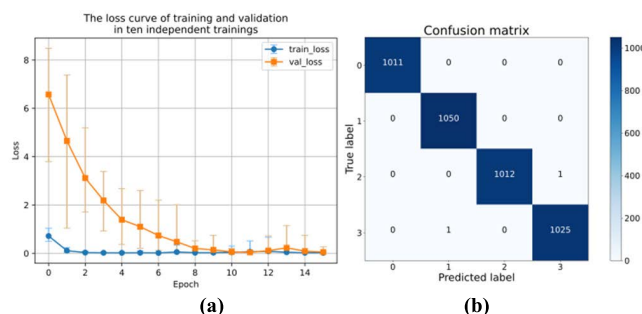validation loss curves with a confusion matrix are shown in Fig. 11a and Fig. 11b, respectively.

FIGURE 11. (a) Loss function curves for 10 independent trainings; (b) Confusion matrix.

Our proposed method demonstrates an excellent performance on our self-collected bearing dataset and achieves an average accuracy of 99.95% after 10 independent runs of training. In addition, the average values for macro F-1 score, precision, and recall are approximately the same, that is, 99.95%. Since our proposed method has a significant reduction of data dimension, it is suitable for real-time processing applications.

The proposed algorithm offers a superior performance because of the sparse wavelet decomposition stage in which the obtained multi-resolution signal features are targeted on the bearing characteristic frequency bands. Note that this method achieves the 1-D MSCNN network model based on the shaft speed, which is the limitation similar to the other standard CNN methods used for bearing fault diagnosis. Our future work will investigate the performance of the proposed method using the multiple metrics for different public datasets and different scenarios in machine fault detection.

VI. CONCLUSION

We have proposed a novel algorithm for bearing fault diagnosis using sparse wavelet decomposition for feature extraction combined with multi-scale neural networks. The algorithm first determines a sparse wavelet decomposition structure based on bearing characteristic frequency zones. The second stage decomposes a raw bearing signal into multi-resolution signal features. Finally, the algorithm applies decomposed signal features to sub-neural networks based on the one-dimensional convolutional neural network (1-D CNN). In comparison with other bearing fault diag-

nosis methods, our proposed method can achieve a higher classification accuracy of 99.85% using the Case Western Reserve University (CWRU) bearing dataset. The effectiveness of our proposed method is also validated via our designed experiments with 99.95% classification accuracy on our self-collected bearing dataset. Our proposed algorithm has a potential for applications in real-time training and classification.

ACKNOWLEDGMENT

REFERENCES

[1] B. R. Robert, *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. 2nd ed. Hoboken, NJ, USA: Wiley, 2021.

[2] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Oxford, U.K.: Butterworth-Heinemann, 2017.

[3] R. N. Bell, D. W. McWilliams, P. O'Donnell, C. Singh, and S. J. Wells, "Report of large motor reliability survey of industrial and commercial installations, Part I," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 853–864, Jul. 1985.

[4] R. N. Bell, C. R. Heising, P. O'Donnell, S. J. Wells, and C. Singh, "Report of large motor reliability survey of industrial and commercial installations, Part II," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 865–872, Jul. 1985.

[5] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "DDeep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.

[6] W. Huang, J. Cheng, and Y. Yang, "Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection," *Mech. Syst. Signal Process.*, vol. 114, pp. 165–188, Jan. 2019.

[7] M. E. H. Benbouzid, M. Vieira, and C. Theys, "Induction motors' faults detection and localization using stator current advanced signal processing techniques," *IEEE Trans. Power Electron.*, vol. 14, no. 1, pp. 14–22, Jan. 1999.

[8] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984–993, Oct. 2000.

[9] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Trans. Ind. Appl.*, vol. 47, no. 1, pp. 34–46, Jan./Feb. 2011.

[10] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, pp. 1–8, Jul. 2014.

[11] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case Western Reserve University data," *J. Failure Anal. Prevention*, vol. 16, no. 2, pp. 271–284, Feb. 2016.

[12] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *IEEE Access*, vol. 8, pp. 93155–93178, 2020.

[13] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[14] L. Tan and J. Jiang, *Digital Signal Processing: Fundamentals and Applications*. 3rd ed. New York, NY, USA: Academic, 2018.

[15] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.

[16] Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *J. Sound Vib.*, vol. 253, no. 5, pp. 1087–1100, Jun. 2002.

[17] C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, "Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data," *Neurocomputing*, vol. 409, pp. 35–45, Oct. 2020.

[18] C. Castejón, M. Jesás, J. J. Gómez, J. Carlos García-Prada, A. J. O. Nez, and H. Rubio, "Automatic selection of the WPT decomposition level for condition monitoring of rotor elements based on the sensitivity analysis of the wavelet energy," *Int. J. Acoust. Vib.*, vol. 20, no. 2, pp. 95–100, Jun. 2015.

[19] P. Gupta and M. K. Pradhan, "Fault detection analysis in rolling element bearing: A review," *Mater. Today, Proc.*, vol. 4, no. 2, pp. 2085–2094, 2017.

[20] N. Tandon and B. C. Nakra, "Detection of defects in rolling element bearings by vibration monitoring," *J. Mech. Eng. Division*, vol. 73, pp. 271–282, Jan. 1993.

[21] D. Dyer and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *J. Mech. Des.*, vol. 100, no. 2, pp. 229–235, Apr. 1978.

[22] R. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, Dec. 1983, pp. 607–610.

[23] A. Shrivastava and S. Wadhwani, "Development of fault detection system for ball bearing of induction motor using vibration signal," *Int. J. Sci. Res.*, vol. 2, no. 5, pp. 256–259, Jun. 2012.

[24] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "STFT based approach for ball bearing fault detection in a varying speed motor," in *Condition Monitoring of Machinery in Non-Stationary Operations*. Springer, Oct. 2012, pp. 41–50.

[25] J.-H. Lee, J. Kim, and H.-J. Kim, "Development of enhanced Wigner–Ville distribution function," *Mech. Syst. Signal Process.*, vol. 15, no. 2, pp. 367–398, Mar. 2001.

[26] M.-M. Chow, P. M. Mangum, and S. O. Yee, "A neural network approach to real-time condition monitoring of induction motors," *IEEE Trans. Ind. Electron.*, vol. 38, no. 6, pp. 448–453, Dec. 1991.

[27] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[28] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, pp. 1–9, Jan. 2014.

[29] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4137–4145, 2013.

[30] F. Shen, C. Chen, R. Yan, and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prognostics Syst. Health Manag. Conf. (PHM)*, Oct. 2015, pp. 1–6.

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

[32] J. Chen, J. Jiang, X. Guo, and L. Tan, "A self-adaptive CNN with PSO for bearing fault diagnosis," *Syst. Sci. Control Eng.*, vol. 9, no. 1, pp. 11–22, Dec. 2020.

[33] J. Chen, J. Jiang, X. Guo, and L. Tan, "An efficient CNN with tunable input-size for bearing fault diagnosis," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 625–634, Jan. 2021.

[34] Y. Cai, L. Tan, and J. Chen, "Evaluation of deep learning neural networks with input processing for bearing fault diagonosis," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2021, pp. 140–145.

[35] W. Chen and K. Shi, "Multi-scale attention convolutional neural network for time series classification," *Neural Netw.*, vol. 136, pp. 126–140, Apr. 2021.

[36] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.

[37] Y. Yao, S. Zhang, S. Yang, and G. Gui, "Learning attention representation with a multi-scale CNN for gear fault diagnosis under different working conditions," *Sensors*, vol. 20, no. 4, pp. 1233–1253, Feb. 2020.

[38] X. Liu, Y. Cai, Y. Song, and L. Tan, "Bearing fault diagnosis based on multi-scale neural networks," presented at the IEEE Int. Conf. Electro Inf. Technol. (EIT), May 2022.

[39] I. El-Thalji and E. Jantunen, "Fault analysis of the wear fault development in rolling bearings," *Eng. Failure Anal.*, vol. 57, pp. 470–482, Nov. 2015.

[40] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Math. Problems Eng.*, vol. 2017, Jul. 2017, Art. no. 8617315.

[41] J. R. Stack, T. G. Habetler, and R. G. Harley, "Fault classification and fault signature production for rolling element bearings in electric machines," *IEEE Trans. Ind. Appl.*, vol. 40, no. 3, pp. 735–739, May 2004.

[42] N. Chimitt, W. Misch, L. Tan, A. Togbe, and J. Jiang, "Comparative study of simple feature extraction for single-channel EEG based classification," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2017, pp. 166–170.

[43] J. Dai, V. Vijayarajan, X. Peng, L. Tan, and J. Jiang, "Speech recognition using sparse discrete wavelet decomposition feature extraction," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, May 2018, pp. 0812–0816.

[44] J. Dai, Y. Zhang, J. Hou, X. Wang, L. Tan, and J. Jiang, "Sparse wavelet decomposition and filter banks with CNN deep learning for speech recognition," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2019, pp. 98–103.

[45] Y. Zhou, Y. Dong, H. Zhou, and G. Tang, "Deep dynamic adaptive transfer network for rolling bearing fault diagnosis with considering cross-machine instance," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[46] B. Yang, Y. Lei, F. Jia, N. Li, and Z. Du, "A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines," *IEEE Trans. Ind. Electron.*, vol. 67, no. 11, pp. 9747–9757, Nov. 2020.

JUAN ALVARADO received the B.S. degree in electrical engineering from Purdue University Northwest, Hammond, IN, USA, in 2022.

He is currently an Electric Drive Test System Engineer at General Motors. He has been elected and will serve as the President for the SHPE Detroit Professional Chapter, in 2022. His research interests include PWM motor drive inverters and PCB design. He is a member of the Society of Hispanic Engineers and has served as the President for the Purdue Northwest Student Chapter.

XIAOFAN LIU (Student Member, IEEE) received the B.S. degree in electrical engineering from the Shandong University of Agriculture and Engineering, Jinan, China, in 2019. He is currently pursuing the M.S. degree in electrical and computer engineering with Purdue University Northwest, Hammond, IN, USA.

His research interests include the areas of intelligent fault diagnosis of the rotating machinery, stereo depth estimation, drivers' emotion recognition, and pose estimation.

JASON CENTENO received the B.S. degree in electrical engineering from Purdue University Northwest, Hammond, IN, USA, in 2022. He is a member of the Society of Hispanic Engineers. He served as a Secretary for the Purdue Northwest Student Chapter.

LIZHE TAN (Senior Member, IEEE) received the B.S. degree from Southeast University, Nanjing, China, in 1984, and the M.S. degree in engineering mechanics and the M.S. and Ph.D. degrees in electrical engineering from the University of New Mexico, Albuquerque, NM, USA, in 1987, 1989, and 1992, respectively.

He is currently a Professor and the Department Chair with the Department of Electrical and Computer Engineering, Purdue University Northwest, Hammond, IN, USA. He authored or coauthored two textbooks *Digital Signal Processing: Fundamentals and Applications* (Elsevier, Third Edition, 2018) and *Analog Signal Processing and Filter Design* (Linus Publications, Second Edition, 2016). He holds a granted U.S. patent. His research interests include the areas of signal processing for machine learning applications, control systems, computer vision, and robotics.

• • •