



Rolling bearing fault diagnosis based on feature fusion with parallel convolutional neural network

Mingxuan Liang¹ · Pei Cao² · J. Tang² 

Received: 12 September 2020 / Accepted: 18 November 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

Deep learning has seen increased application in the data-driven fault diagnosis of manufacturing system components such as rolling bearing. However, deep learning methods often require a large amount of training data. This is a major barrier in particular for bearing datasets whose sizes are generally limited due to the high costs of data acquisition especially for fault scenarios. When small datasets are employed, over-fitting may occur for a deep learning network with many parameters. To tackle this challenge, in this research, we propose a new methodology of parallel convolutional neural network (P-CNN) for bearing fault identification that is capable of feature fusion. Raw vibration signals in the time domain are divided into non-overlapping training data slices, and two different convolutional neural network (CNN) branches are built in parallel to extract features in the time domain and in the time-frequency domain, respectively. Subsequently, in the merged layer, the time-frequency features extracted by continuous wavelet transform (CWT) are fused together with the time-domain features as inputs to the final classifier, thereby enriching feature information and improving network performance. By incorporating empirical feature extraction such as CWT, this proposed method can effectively enable deep learning even with dataset size limitation in practical bearing diagnosis. The algorithm is validated through case studies on publicly accessible experimental rolling bearing datasets. A wide range of dataset sizes is tested with cross-validation, and influencing factors on network performance are discussed. Compared with existing methods, the proposed approach not only possesses higher accuracy but also exhibits better stability and robustness as training dataset sizes and load conditions vary. The concept of feature fusion through P-CNN can be extended to other fault diagnosis applications in manufacturing systems.

Keywords Fault identification · Parallel convolutional network · Wavelet · Reduced dataset · Rolling bearing

1 Introduction

Intelligent fault diagnosis of rotating machinery plays an important role in smart manufacturing, and rolling bearings are one type of key components where a small defect may cause catastrophic consequences. The timely and accurate classification of fault conditions of rolling bearings has attracted a lot of research interests in manufacturing community [1–3]. Vibration signals measured by off-the-shelf sensors are commonly used in condition monitoring. However, the anomalies

in rolling bearing responses induced by faults are generally insignificant. Owing to complicated nonlinear behaviors, e.g., radial clearance, contact mechanics, and rolling friction, raw time-domain signals exhibit nonlinear dynamic effects as well as transient phenomena. This is further compounded by noises and various uncertainties [4]. For a long time, the condition monitoring and fault identification of rolling bearings were primarily based on frequency-domain methods, such as fast Fourier transform (FFT) [5]. While FFT is undoubtedly a powerful tool in analyzing steady-state responses and elucidating random noise effects, its capability on transient responses is limited since time-frequency information is almost entirely lost in nonstationary data after performing FFT. To address this issue, a number of time-frequency domain approaches have been developed, such as short-time Fourier transform (STFT) [6] and continuous wavelet transform (CWT) [7]. Although these time-frequency methods and empirical methods have shown more promise than earlier

✉ J. Tang
jjiong.tang@uconn.edu

¹ College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China

² Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA

frequency-domain methods, their effectiveness depends highly on the subjective selection of specific features during the analysis [8, 9]. Moreover, human interpretation is frequently required. For example, while wavelet methods have been popular, there is no consensus in existing literature on which type of wavelet is most suitable for fault feature extraction [10, 11].

Because of the rapid advancements of machine learning techniques as well as computing power, automated fault diagnosis methods, including stacked auto-encoder (SAE) network [12, 13], principal component analysis (PCA) [14], support vector machine (SVM) [15], and deep belief network (DBN) [16], have seen increasing implementations for rolling bearings. With the help of these techniques, it is possible to develop diagnosis and classification approaches by training with raw time-domain vibration signals where fault features are implicitly embedded, thereby avoiding the subjective feature selection and human interpretation [17, 18]. As such, data-driven methods based on large-scale fault datasets, such as deep learning, have become attractive [19]. Weimer et al. [20] employed a deep convolutional neural network (CNN) to overcome the difficulties of re-defining manual fault features for each new case in the production process, which improves the automation and accuracy of monitoring. In the investigations by Ince et al. [21] and Abdeljaber et al. [22], 1-dimensional convolutional neural networks were used to detect motor faults by truncating the time-domain vibration raw signals of rolling bearing into the same length as training data, and the results indicated higher accuracy than model-based methods. In most intelligent fault diagnosis networks, feature extraction methods such as dimension reduction, downsampling, and convolutional operation are all designed to more effectively extract implicit fault features [23, 24]. Indeed, since the first introduction of the real sense of the CNN model by LeCun et al., the CNN has shown excellent performance in feature extraction and classification when large amounts of training data are available [25]. Many deep neural networks have been developed based on LeCun's model, such as AlexNet [26], GoogLeNet [27], and VGG [28]. There have been recent efforts of employing CNN in the intelligent fault identification of rolling bearings. For example, Wen et al. [29] converted raw time-domain vibration signals into uniformly distributed images with a sequence that are subsequently trained by 2-dimensional CNN for rolling bearing diagnosis.

On one hand, the accuracy of CNNs continues to improve owing to their algorithmic advancement, especially for applications with large-scale datasets. On the other hand, for rolling bearing diagnosis investigations, fault datasets are generally limited, and the acquisition of large amounts of training data comes at a high cost. A number of workarounds have been attempted to produce sufficient training information with limited time-domain vibration signal. Some data-enhanced methods have been developed for the case of rolling bearing,

such as overlap truncation and interpolation operation of raw signal [30]. Although it yields many training data slices, the overlap truncation on time-domain vibration signal may introduce inter-dependency of data used in training. Ideally, the training data should be divided into non-overlapping data slices, which should be as short as possible to make sure more training data slices can be obtained from the raw time-domain vibration signal. Meanwhile, the minimum length of data slice should be no less than one revolution in order to maintain the integrity of training data. The data slices produced via the above method are often still small datasets of limited size. Some studies have indicated that this may again lead to over-fitting, as too many parameters for a deep learning network need to be trained based on a small dataset. Aiming at overcoming the over-fitting problem, Zhang et al. [31] and Cao et al. [32] designed a two-stage transfer learning network to identify faults in gears and bearing systems. First, fault features are extracted based on a pre-trained deep CNN. Then, well-trained parameters are transferred to predict fault types from the time-domain vibration signal. The results showed that this method can improve the robustness to some extent even when the training data size decreases.

The above literature review indicates that while deep learning such as CNN is appealing, the limited amount of training dataset poses significant challenges. In particular for rolling bearing, although the number of training data slices can be multiplied by employing truncation of time-domain vibration signal, the size of fault dataset is still limited based on non-overlapping truncation and thus over-fitting may occur [31–35]. It is worth noting that, in contrast, the traditional time-frequency analysis is not strictly constrained by the amount of fault data, as time-frequency features can be extracted empirically in reduced dataset, especially for data containing transient signals [7, 36, 37]. In other words, even though there exists some subjectivity in the manual selection/analysis of time-frequency features through such as wavelet transform, it may indeed lead to valid results of fault diagnosis [38, 39]. Intuitively, one may hypothesize that fusing a certain amount of such manually extracted feature information into the deep-learning framework may relieve the issue of dataset size. The subjectivity in manual feature extraction may in turn be mitigated by using the classifier operation in the deep-learning. Such an information fusion strategy, however, has not been attempted in the fault diagnosis of rolling bearing. In fact, to improve the accuracy of rolling bearing fault diagnosis with a reduced dataset, especially in the case of load condition changing, multi-domain information fusion in deep learning could be a worthwhile approach to explore [40].

In this research, we tackle the problem of rolling bearing fault diagnosis with limited dataset through deep-learning. Most recently, there has been the suggestion of formulating multi-channel CNN, i.e., parallel

convolutional neural network (P-CNN) architecture, to facilitate information fusing in deep learning, which in certain cases could be more effective than solely deepening the network [41, 42]. Here in this research, we adopt this concept to develop a new approach of rolling bearing fault diagnosis in which the time-frequency features extracted by continuous wavelet transform (CWT) are fused with the fault features obtained by a separate general CNN branch as the input. An intelligent classifier is designed for the classification of fused features. We intend to solve the over-fitting issue when training is practiced on dataset with limited size and at the same time reduce the subjectivity in manual extraction of features. The proposed approach is compared with the general CNN trained with time-domain data only and other intelligent diagnosis methods under different load conditions and dataset sizes, and the influencing factors on the identification performance are discussed in depth.

The rest of the paper is organized as follows. In Section 2, the components of the proposed approach are outlined. In Section 3, we first introduce the pre-processing of the time-domain vibration raw signals by non-overlapping truncation, and then present the architecture and specific parameters of the proposed network. Section 4 gives validation of the fault diagnosis performance where publicly accessible bearing fault dataset is used. The size of training dataset is adjusted, and comparison with respect to some conventional methods is carried out. Concluding remarks are summarized in Section 5.

2 Methodology components

2.1 Feature extraction based on CWT

As a powerful variant of STFT, CWT can overcome the limitations faced by STFT to analyze transient, nonstationary vibration signal by changing the translation value and scale value of mother wavelets continuously. The mother wavelet $\psi(t)$ can generally be represented as.

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (1)$$

The CWT of a time-domain vibration signal $x(t)$ under the scale factor $a \in \mathbb{R}^{+*}$ and the translation factor $b \in \mathbb{R}$ can be given by the following integral.

$$X_{\omega}(a, b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (2)$$

where the over-bar on mother wavelet function $\psi(t)$ indicates the complex conjugate operation. Parameter a is the wavelet scale that characterizes the frequency, and parameter b represents the time or space position. In order to visually

characterize the relationship between time, frequency, and wavelet magnitude, the CWT results of the time-domain vibration signal can be expressed by wavelet scalogram in the following manner.

$$Sc_x(a, b) = \frac{1}{|b|} \left| \int_{-\infty}^{\infty} x(t) \psi\left(\frac{t-b}{a}\right) dt \right|^2 \quad (3)$$

Here, the time-frequency feature in the wavelet scalogram will be further extracted using CNN, and the newly obtained feature maps will be fused with feature maps extracted from the time-domain vibration raw signal by another different CNN branch to improve the accuracy and robustness of the network for various dataset sizes. In this research, without loss of generality, the Gaussian derivative wavelets with an eighth order derivative function are employed as the mother wavelet.

$$\psi(t) = C \exp^{-t^2} \quad (4)$$

where C is an order-dependent normalization constant.

2.2 Convolutional neural network (CNN)

2.2.1 Convolutional layer

As one of the very effective machine learning networks, CNN can achieve better learning by sharing convolutional kernel and employing [downsampling](#) to improve the robustness. The convolutional layer mainly plays the role of extracting features from a local region on the feature map, in which various convolutional kernels correspond to different feature extractors. In order to calculate the output Y^p , the convolutional operations on the inputs $X^1, X^2, \dots, X^D$ are implemented using convolutional kernels $W^{p,1}, W^{p,2}, \dots, W^{p,D}$. Then, the convolutional operation results are summed together, and a scalar offset value b^p will be added after that. The output of the convolutional layer Z^p can be obtained as.

$$Z^p = \mathbf{W}^p \otimes \mathbf{X} + b^p = \sum_{d=1}^D W^{p,d} \otimes X^d + b^p \quad (5)$$

where $\otimes$ represents the convolutional operation, and $\mathbf{W}^p \in \mathbb{R}^{m \times n \times D}$ is the convolution kernel. Based on a nonlinear activation function, the output feature map Y^p can be represented as.

$$Y^p = f(Z^p) \quad (6)$$

where $f(\cdot)$ denotes the nonlinear activation function. In this paper, the rectified linear unit (ReLU) function is employed as the activation function, which can make the network training faster than the previous used Sigmoid activation function [43].

2.2.2 Pooling layer

The pooling layers are usually designed after the convolutional layer, in order to control over-fitting effectively by collecting valid information from the feature map extracted from the convolutional layer. As an effective downsampling method, the max-pooling layers are the most common pooling operation to extract the maximum value from each neuron in the feature map, which can be described as the following.

$$Y_{m,n}^d = \max_{i \in R_{m,n}^d} x_i \quad (7)$$

where x_i denotes the activation value of each neuron in the regions $R_{m,n}^d$ on the feature map.

3 Architecture of the proposed network

The architecture of CNN consists of convolutional layer, pooling layer, full connection, and classifier, and is usually built by serial stacking. Figure 1 outlines the typical architecture of CNN for fault diagnosis of rolling bearing. It is a single-branch serial system from the input of image feature maps to the output of prediction vector. Essentially, the data slices of time-domain raw signal are imaged as 2-D feature maps to be fed into the network. The convolution layers and the max-pooling layers can be regarded as nonlinear feature extraction part in the network. The convolution layers are followed by the nonlinear activation function, and the full connection layer mainly plays the role of classifier. Classical deep learning networks are built upon this architecture to facilitate intelligent learning of large-scale datasets through deepening the networks [26–28]. In reality, because of the over-fitting problem, the accuracy and robustness of the general CNN may decrease rapidly when the dataset size decreases [44, 45].

Recent studies indicated that multi-channel CNN, i.e., parallel convolutional neural network (P-CNN), can be effective in certain applications [41, 42]. Based on the literature review provided in Section 1, we hypothesize that this P-CNN concept could be employed to tackle the issue of limited dataset size in fault diagnosis of rolling bearing. As shown in Fig. 2, this proposed network consists of two parallel branches, each

being a convolutional extraction process with specific parameters. The time-domain raw signal is analyzed twice in parallel by these two different convolutional operations. The time-domain raw signal is divided into training data slices in non-overlapping manner and fed to the P-CNN. In one branch, the training data slices are processed directly one by one by the time-domain CNN branch. In the other branch, the training data slices are first processed by CWT operation to extract time-frequency features, and the wavelet scalogram results then be analyzed by another CNN branch subsequently. Finally, the time-domain features and the time-frequency features are fused together in the merge layer as the input to the classifier. In this network, the output dimension of the two branches can be adjusted by the pooling layer and the bilinear interpolation of wavelet scalogram for improved feature fusion from the two branches. Table 1 presents the parameters of the proposed network.

As shown in Fig. 2 and Table 1, three consecutive convolutional operations with a kernel size of 3×3 are performed in the two different CNN branches, and the number of convolution kernels is doubled compared to the upper convolution kernels as the network deepens. The ReLU function is used as activation function following each convolutional operation. The max-pooling is employed to reduce the parameters of feature maps to prevent over-fitting after the consecutive convolutional operations. In the branch for wavelet scalogram, the pixel of wavelet scalogram is resized based on bilinear interpolation. Then, the new feature maps of wavelet scalogram are further extracted with a dimension of 10×10 by the consecutive convolutional operations and the max-pooling operations. In the merge layer, the time-frequency features from wavelet scalogram are fused with time-domain features to enhance the fault feature, and the consecutive convolutional operations with 64 kernels are employed to continue to extract new features after the merge layer. The dropout technology with a probability 50% is used according to [46] to prevent over-fitting.

The enhanced features are classified in two consecutive full connection layers. The weights in each layer are randomly initialized and trained for optimization. Since the learning efficiency is very low for Sigmoid loss function when the neuron outputs are close to 0 or 1, the cross entropy loss function is introduced to improve training efficiency.

Fig. 1 Representative architecture of general CNN

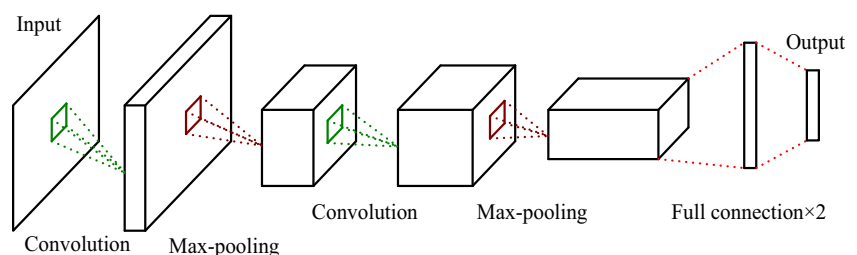


Table 1 Specific parameters of the proposed network

Operation	Time-domain raw signal slices			
CNN branch	Time domain		Wavelet scalogram	
Pre-processing	Imaging		CWT operations	
Parallel CNN	Conv_ReLU repeat 3 times	$8 \times 3 \times 3$	Conv_ReLU repeat 3 times	$8 \times 3 \times 3$
			Max pooling	2×2
			Conv_ReLU repeat 3 times	$16 \times 3 \times 3$
			Max pooling	2×2
			Conv_ReLU repeat 3 times	$32 \times 3 \times 3$
	Max- pooling	2×2	Max-pooling	2×2
Fusion	New features maps			
CNN	Conv_ReLU repeat 3 times		$64 \times 3 \times 3$	
Feature classification	Dropout		50%	
	Fully connection1		32	
	Fully connection2		10	
	Softmax		N/A	
	Classifier		Cross entropy	

$$L(p_n, \hat{p}_n) = -\frac{1}{n} \sum_{n=1}^n [p_n \log \hat{p}_n + (1-p_n) \log (1-\hat{p}_n)] \quad (8)$$

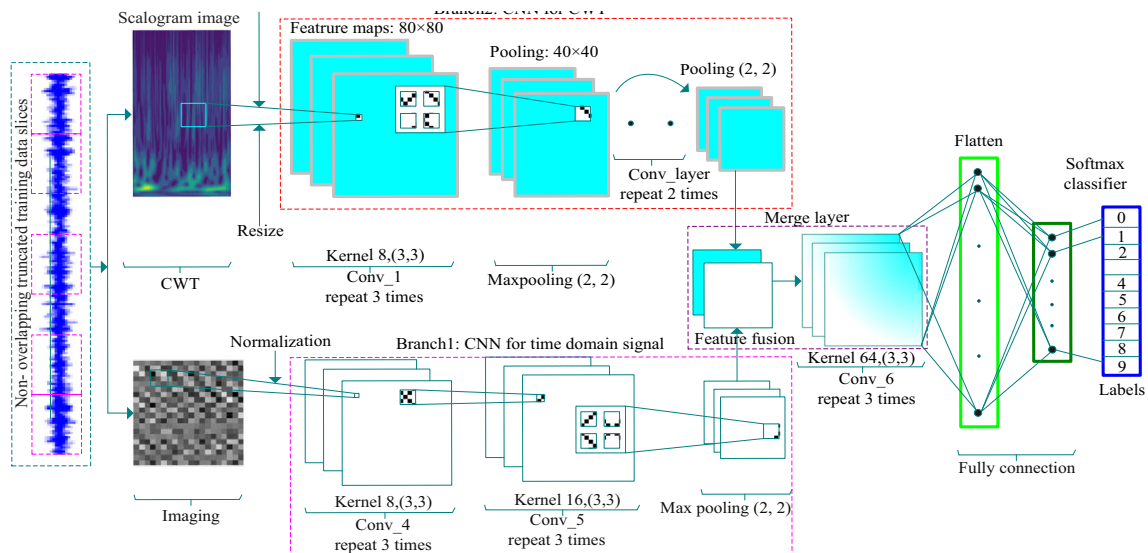
where p_n and $\hat{p}_n$ are the inputs of the loss function. $p_n \in [0, 1]$ represents the label value of the training data slices. $\hat{p}_n = \sigma(x_n) \in [0, 1]$ represents the predicted results of the training data slices, and its specific expression can be given as following,

$$\sigma(x_n) = \frac{1}{1 + e^{x_n}} \quad (9)$$

4 Fault identification result analysis

4.1 Fault dataset of rolling bearing

In this research, the fault dataset of rolling bearing comes from Case Western Reserve University (CWRU) bearing data center (<http://csegroups.case.edu/bearingdatacenter>). The dataset has been analyzed comprehensively by various researchers [30, 47]. The experimental setup for the data acquisition of rolling bearing from CWRU is outlined in Fig. 3. It mainly consists of an induction motor, torque transducer, dynamometer, accelerometer, and control electronics. The tested bearing is deep groove

**Fig. 2** Proposed network architecture

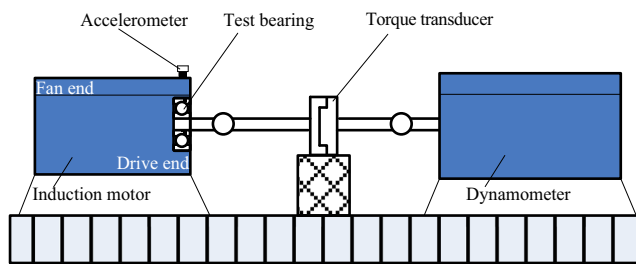


Fig. 3 Experimental setup for vibration signal of rolling bearing

ball bearing (6205-2RS JEM SKF), which is installed at the drive end of an induction motor.

In CWRU dataset, single point defects with three different sizes of 0.007 in., 0.014 in., and 0.021 in. are manufactured by electric discharge machining on different parts of the bearing, as shown in Fig. 4a and Fig. 4b. The numbers 1, 2, and 3 indicate that the faults are on the ball element (BA), inner ring (IR), and outer ring (OR), respectively. Specifically, the fault of outer ring is located at 6 o'clock position when the bearing is installed on the bearing housing. There is no restriction for the fault location on the inner ring and rolling element. Various time-domain vibration signals were recorded under three different motor loads, 1 hp, 2 hp, and 3 hp, respectively. The signals were recorded by an accelerometer that was attached at the drive end of the induction motor with magnetic base. Time-domain vibration raw signals with a length of no less than ten seconds were collected by a 16-channel data recorder for each fault condition under a sampling frequency of 12 kHz.

The dataset containing the time-domain vibration signals from the healthy bearing and from bearings with various faults is analyzed, which are summarized in Table 2. The healthy bearing is marked with label 0, and the other 9 fault cases are assigned with 9 different labels from 1 to 9, respectively. There are 10 different conditions in this dataset. In each condition, the first 10-s vibration raw signal is divided in the non-overlapping manner into 300 data slices, and each data slice has a length of 400 data points. There are totally 3000 data slices, corresponding to 10 conditions, in this dataset.

The training data slices of vibration signal with a length of 400 under various conditions are shown in Fig. 5a. The 2-D

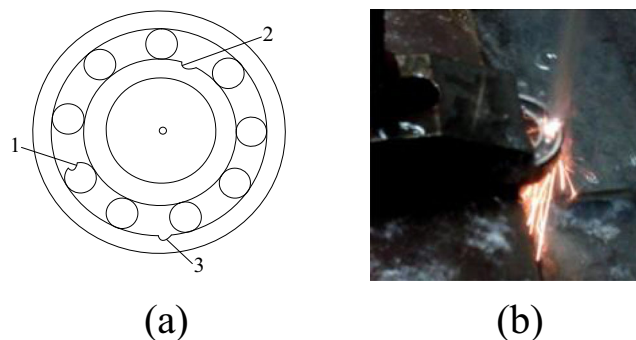


Fig. 4 Three fault scenarios of rolling bearing

Table 2 Indices of rolling bearing fault conditions

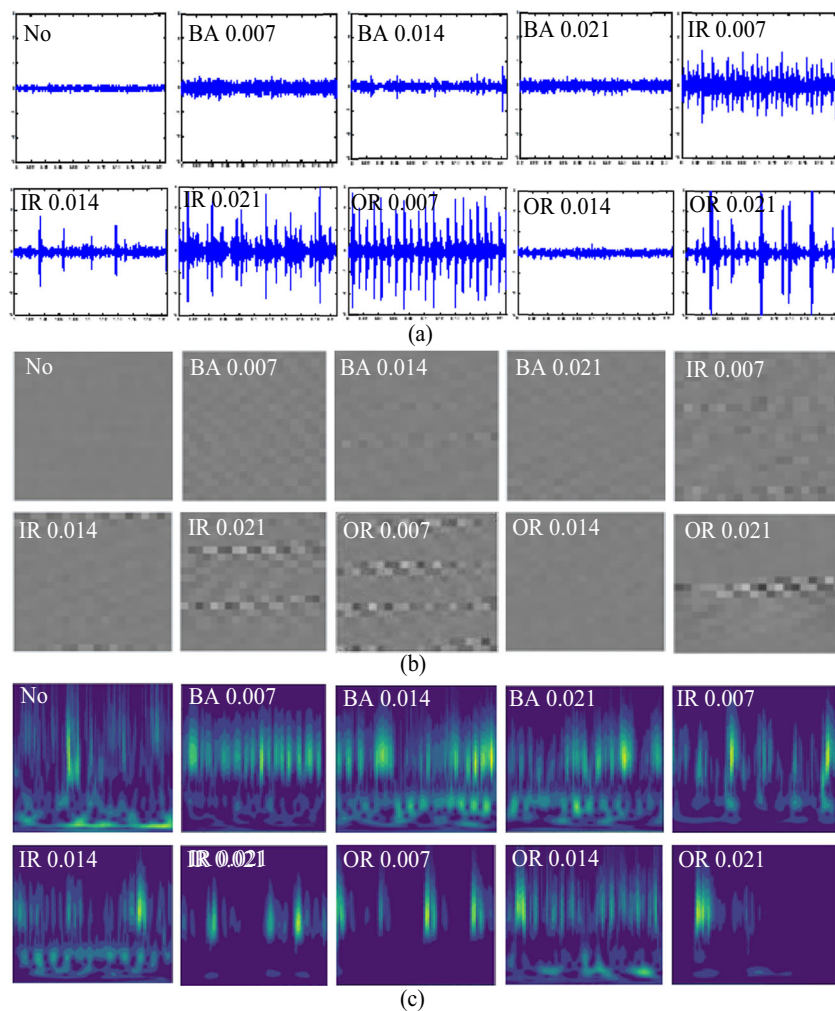
Fault type		Fault size/inch	Dataset size	Label
Normal bearing		No	300	0
Fault bearing	OR	0.007	300	1
		0.014	300	2
		0.021	300	3
	IR	0.007	300	4
		0.014	300	5
		0.021	300	6
	BA	0.007	300	7
		0.014	300	8
		0.021	300	9

image feature maps of the training data slices are shown in Fig. 5b. The corresponding wavelet scalograms of training data slices are shown in Fig. 5c. In the proposed network, each training data slice from raw vibration signal is introduced into the network in the beginning, and they are employed as two different training sample forms to ensure that they can be fed to the two different branches of the network. As indicated, the data slices will be sent to one CNN branch for time-domain signal processing. At the same time, they will also be processed by wavelet extraction and then send to another CNN branch designed for wavelet scalogram.

It can be clearly observed that it is not possible to distinguish different bearing faults directly from the raw vibration signal. Especially when the faults are on the rolling elements, the differences on the vibration amplitudes are not obvious for these three fault sizes 0.007 in., 0.014 in., and 0.021 in. The time-domain raw signals are displayed as the images in Fig. 5b with a pixel of $n \times n$ after normalization. Here, n is the edge length of each image in Fig. 5b, and n^2 represents the length of the data slice in Fig. 5a. Although it becomes even more difficult to identify the fault feature manually from the images, we can utilize 2-D convolutional operations to extract useful fault features. Furthermore, the time-frequency features displayed on the wavelet scalogram after wavelet analysis may improve classification in the following layer after the feature fusion of time-domain feature and time-frequency domain feature.

4.2 Basic validation of the proposed network

As described above, the time-domain vibration raw signals of the rolling bearings are collected under three different motor load conditions. In order to validate the performance of the proposed network under these different conditions, we define the vibration raw data under 1 hp, 2 hp, and 3 hp motor loads as dataset A, dataset B, and dataset C, respectively. Meanwhile, the proposed network is first trained on 80% data

Fig. 5 Different signal forms under various fault sizes

slices of each dataset, and the well-trained network is tested using the other 20% testing data slices from the three different datasets, as shown in Table 3. We employ 9 different dataset scenarios to indicate these combinations. For instance, the scenario $B \rightarrow C$ means the network is trained on dataset B, and tested by 20% random data slices from dataset C.

Table 3 Fault dataset scenario

Training dataset	Testing dataset	Scenario
A (1 hp)	A	$A \rightarrow A$
	B	$A \rightarrow B$
	C	$A \rightarrow C$
B (2 hp)	A	$B \rightarrow A$
	B	$B \rightarrow B$
	C	$B \rightarrow C$
C (3 hp)	A	$C \rightarrow A$
	B	$C \rightarrow B$
	C	$C \rightarrow C$

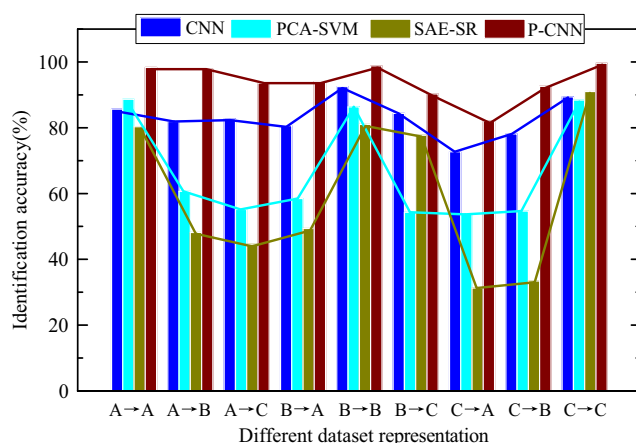
Example: $B \rightarrow C$ means the network is trained on dataset B, and tested on dataset C

In this research, several conventional methods for bearing fault identification, including general CNN, support-vector machine (SVM), and stacked auto-encoder (SAE) network, are employed to compare with the proposed network. Here, the general CNN is applied through the direct training and testing on the datasets of time-domain raw signal. When SVM is used, the time-domain signals are usually dimension-reduced by PCA, and we refer to this method as PCA-SVM for short. In some investigations [48], the labeled bearing fault datasets can also be predicted by SAE network followed by Softmax regression, which is referred to as SAE-SR network hereafter. In this subsection, these methods are all tested three times on various dataset scenarios. The identification accuracy is defined as the ratio of the number of correctly identified data to the total number of testing data used. The average identification results of three tries are reported in Table 4, and the trends of average identification accuracy in various data scenarios are shown in Fig. 6.

Table 4 Fault identification results in basic validation

Dataset scenarios	CNN	PCA-SVM	SAE-SR	P-CNN
Average identification accuracy (%)				
A → A	85.67	88.67	80.33	98.33
A → B	81.88	60.67	48.17	97.79
A → C	82.71	55.17	44.58	93.54
B → A	80.63	58.50	49.38	94.04
B → B	92.50	86.33	81.00	98.83
B → C	84.38	54.33	77.71	90.28
C → A	72.71	53.67	31.25	82.00
C → B	78.13	54.67	33.42	92.58
C → C	89.40	88.40	91.00	99.60
Mean	83.11	66.71	59.65	93.58

Clearly, the P-CNN method proposed in this paper shows the best comprehensive performance. Its mean value of average identification accuracy and the average identification accuracy on each dataset scenarios are higher than those of other methods. Employing the proposed P-CNN that fuses wavelet features with time-domain features can enhance the fault identification accuracy for CNN. It can be observed that these four networks all show better performance when being trained and tested on the same datasets from the same load conditions (e.g., A → A, B → B, and C → C), the identification accuracy of which is significantly higher than that of being trained and tested on different datasets (e.g., A → B, A → C, B → A, C → A, B → C, and C → B). All four networks show the worst performance in the dataset scenarios A → C and C → A. From the comparison between the four networks in Fig. 6, we can observe that while P-CNN has the best performance, the trends of average identification accuracy for these four methods are actually similar. The average identification accuracy of general CNN is much higher than that of PCA-SVM and SEA-SR. Especially when the training dataset and the

**Fig. 6** Trends of identification accuracy

testing dataset are different, the average identification accuracy drastically decreases for PCA-SVM and SEA-SR.

4.3 Further analysis with various dataset sizes

The overarching goal of this research is to explore fault identification of rolling bearing with limited size of training dataset. Here in order to highlight the performance of the proposed network, we examine the results under various dataset sizes. We train and test the network by adjusting the dataset size as 70%, 40%, 10%, and 5% of the total dataset in dataset C. Eight percent data slices of each reduced dataset are employed as training data randomly, and the rest 20% data slices are taken as testing data. In this part of the research, the four networks mentioned above are all tested five times to perform validation on the fault datasets with reduced sizes, and the identification results are shown in Table 5.

We can observe that (1) the testing accuracy of the proposed P-CNN on each reduced dataset is higher than that of the other three methods; and (2) the average accuracy of P-CNN is also better. The average identification accuracy of P-CNN is 99.60% based on the training and testing on the total dataset (3000 data slices), which is much higher than those of general CNN (89.40%), PCA-SVM (88.40%), and SAE-SR (91.00%). As the size of the dataset decreases, the identification accuracy of the other three conventional methods decreases significantly. For 5% dataset trained by SAE-SR, the average identification accuracy decreases to 29.24% which is the worst performance amongst all cases. This is a 61.76% decrease from the identification accuracy based on the total dataset. In comparison, the proposed P-CNN can maintain a high identification accuracy when the dataset size decreases. Even for the 5% dataset size, the average identification accuracy can still reach 94.17% which is only 5.43% less than that based on the total dataset. These results demonstrate that the proposed P-CNN built upon the fusion of time-frequency domain features and time-domain features yield a much higher accuracy compared with the general CNN, PCA-SVM, and SAE-SR methods when the size of training dataset decreases.

The identification results for different dataset sizes are also illustrated through box-plots for comparison of consistency/stability amongst various networks, as shown in Fig. 7. It can be seen that the identification results of the proposed P-CNN are generally more consistent for five tests, and the variance is smaller. Except for the 40% dataset case where the PCA-SVM results exhibit smaller variance, the proposed P-CNN leads to both higher accuracy and better consistency in all cases. For that 40% dataset case, P-CNN yields much better accuracy than PCA-SVM. Compared with general CNN training based on solely time-domain raw signal, the proposed P-CNN can significantly improve both the accuracy and stability.

Table 5 Comparison of identification results under various sizes of training dataset

Method	CNN	PCA-SVM	SAE-SR	P-CNN
Dataset size	Identification accuracy (%)			
100% (3000)	90.33	89.67	91.50	99.50
	91.67	89.67	91.67	99.50
	92.17	85.83	90.33	99.50
	87.67	88.83	91.50	99.67
	85.17	88.00	90.00	99.83
	Average: 89.40	Average: 88.40	Average: 91.00	Average: 99.60
70% (2100)	87.38	86.19	89.76	99.52
	82.62	85.95	90.00	98.81
	92.38	85.95	85.71	98.33
	76.67	87.38	90.24	99.76
	83.33	88.10	78.81	100.00
	Average: 84.48	Average: 86.71	Average: 86.90	Average: 99.28
40% (1200)	75.42	86.67	68.75	97.08
	85.00	86.67	68.33	97.92
	80.83	86.26	80.83	99.17
	70.83	86.25	78.75	97.08
	70.42	86.25	75.00	99.58
	Average: 76.50	Average: 86.42	Average: 74.33	Average: 98.17
10% (300)	78.33	61.67	46.67	95.00
	70.00	61.67	50.00	96.67
	70.00	61.67	55.00	96.67
	70.00	63.33	55.00	98.33
	75.00	63.33	45.00	98.33
	Average: 72.67	Average: 62.33	Average: 50.33	Average: 97.00
5% (150)	70.00	56.67	26.32	93.33
	66.67	56.67	23.23	96.67
	60.00	56.67	26.67	96.67
	66.67	60.00	30.00	96.67
	43.33	60.00	40.00	90.00
	Average: 61.33	Average: 58.00	Average: 29.24	Average: 94.17

4.4 Analysis of network performance with respect to manual feature extraction methods

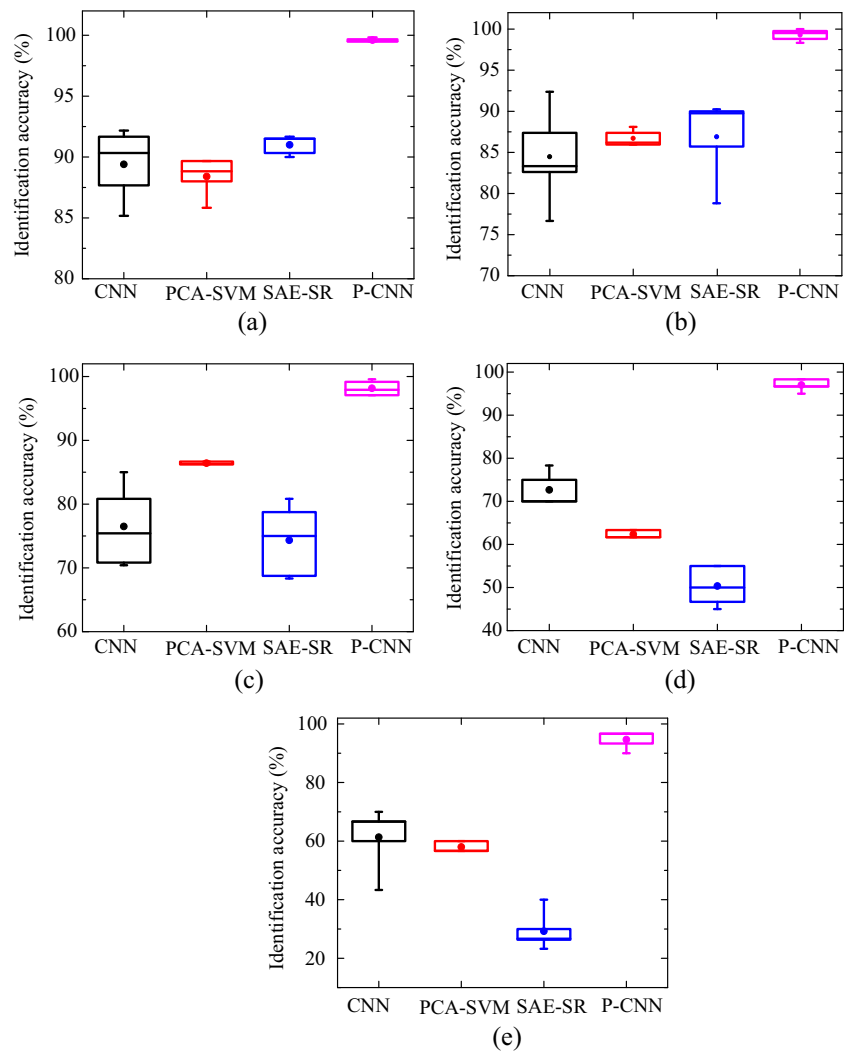
One key idea behind the proposed P-CNN is the fusion of manually selected features, such as those identified through CWT followed by one CNN, into those features identified directly from time-domain raw signals by another CNN. Indeed, in this current investigation, the selection of CWT for manual feature extraction is based on the phenomenological investigation on several representative feature extraction methods in which CWT exhibits the best performance. In this

sub-section, we report the details of such phenomenological investigation which can shed some lights on the underlying physical reason why CWT is selected and how it benefits the P-CNN performance.

Alternatively, empirical mode decomposition (EMD) and fast Fourier transform (FFT) can be used for manual feature extraction in this proposed P-CNN. Figure 8 plots comparison of the average identification accuracies of the proposed P-CNN when fused with the feature maps from CWT (i.e., the method proposed in this research), from EMD and from FFT. Also plotted in Fig. 8 are the average identification accuracy of general CNN by direct training on the time-frequency features from wavelet scalogram (general CNN with CWT) and that of general CNN by direct training on time-domain raw signal (general CNN with raw signal). For each method, three tests are carried out, and the average accuracy is plotted. It is worth noting that the architecture of general CNN is based on serial stacking as shown in Fig. 1. While the identification accuracies of P-CNN fused by the feature maps from FFT, EMD, and CWT are all above 90% when training is conducted on the total dataset, those corresponding for fusing from FFT and EMD decrease sharply as the dataset sizes decrease. Clearly, the time-frequency features from CWT extraction lead to better performance in terms of information fusion in P-CNN. Meanwhile, the general CNN for time-domain raw signal consistently exhibits the worst identification accuracy under various dataset sizes. Although the general CNN with CWT can improve the identification accuracy to some extent compared with the general CNN for time-domain raw signal, the performance is still much lower than that of the proposed P-CNN which fuses the time-frequency features with the time-frequency domain features extracted from CWT. This comparison highlights the significance of information fusion. That is, the fused results outperform those by general CNN using time-domain information along or using time-frequency domain information along.

Compared with the Fourier transform, wavelet functions are not unique, and there are many different mother wavelets that can be employed for feature extraction. An important factor in the proposed P-CNN methodology is the selection of mother wavelet that can yield optimal performance. For instance, each order derivative of the Gaussian function can be regarded as a wavelet function for fault feature extraction of rolling bearing, and the negative-normalized second-order derivative of the Gaussian function is a special case in the Gaussian wavelet function family known as Hermitian wavelet. Morlet wavelet is composed of a complex exponential (carrier) multiplied by a Gaussian window (envelope). Some previous investigations indicate that the Morlet wavelet could be effective in analyzing the vibration responses of rolling bearing, since it is capable of extracting short pulses that occur when the bearing exhibits faults [49]. Hereafter, we explore the effects of different wavelet functions on the identification accuracy of the proposed network to gain an understanding of the feature fusion in P-

Fig. 7 Identification results under various sizes of training dataset illustrated by box plots. **a** 100%, **b** 70%, **c** 40%, **d** 10%, and **e** 5%



CNN. Figure 9 presents the identification accuracies under different wavelet functions used in the proposed P-CNN. We employ dataset C with different dataset sizes. The results of the first order to the eight-order derivative wavelets of the Gaussian function are compared with those of the Morlet wavelet and the Hermitian wavelet. As can be seen from Fig. 9, overall, the

eight-order derivative wavelet leads to the best performance, which is the underlying reason that it is selected to facilitate CWT based time-frequency feature extraction for information fusing in the proposed P-CNN. The accuracies of the parallel convolutional network with all these wavelets generally decrease as the dataset size decreases. This is especially true for

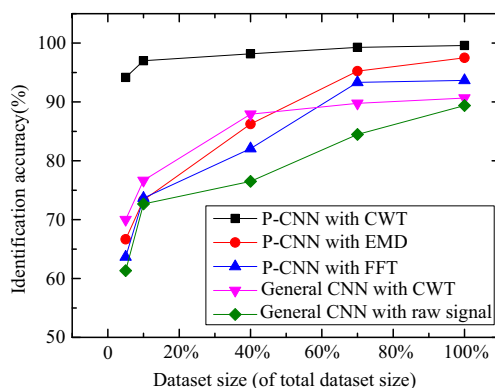


Fig. 8 Identification accuracy comparison using different features

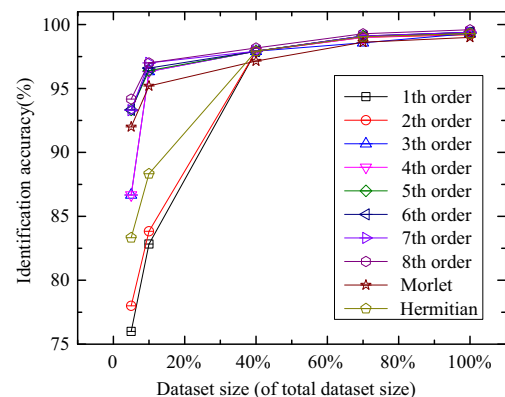


Fig. 9 Effect of different wavelet functions to the performance of P-CNN

the first four order derivative wavelets of the Gaussian function. The accuracies are significantly lower than those of the higher order derivative wavelets when training is conducted on less than 40% of the total dataset. Above the fourth-order derivative, however, increasing the derivative order does not significantly increase the identification performance. As the negative-normalized second-order derivative wavelet of the Gaussian function, the Hermitian wavelet shows a better performance than the second-order derivative wavelet for the same dataset, but its accuracy is slightly inferior to that of Morlet wavelet function when dataset size decreases.

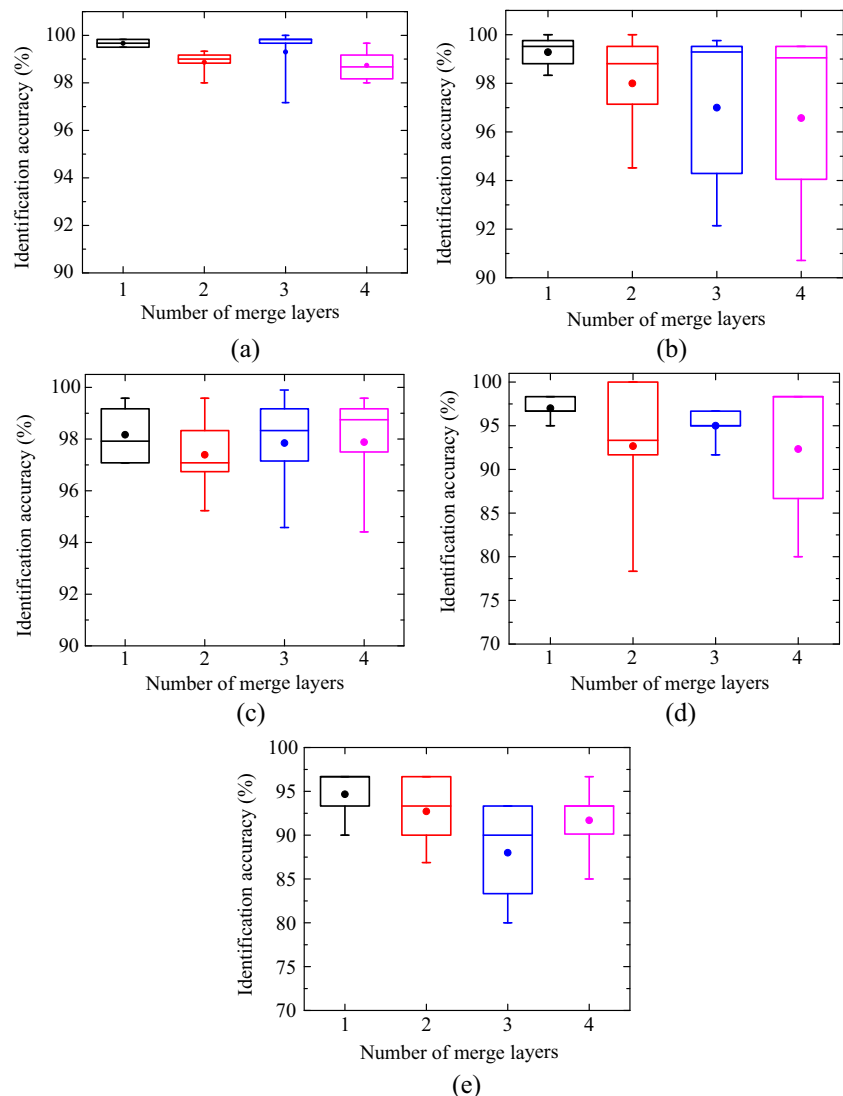
The preceding discussions demonstrate that fusion of time-domain features and time-frequency features can effectively improve the identification. The specific role of feature fusion in the proposed warrants some further analysis. To evaluate the effects of fusion of time-frequency features extracted from CWT on identification accuracy, we adopt respectively 1 to 4 merge layers in the proposed P-CNN. Once again, we employ

dataset C with different dataset sizes, and perform 5 tests for each dataset condition. The results are plotted in Fig. 10. Interestingly, it can be observed that the identification results of the network fluctuate significantly as the number of the merge layer increases, regardless the dataset sizes. Indeed, the parallel network shows higher average accuracy with one merge layer of feature fusion, and the stability generally decreases as the number of merge layer increases. Therefore, although feature fusion can improve the accuracy of the proposed parallel network, too many merge layers is not recommended. This analysis provides the rationale for the merge layer setup in the proposed network.

5 Concluding remarks

This research aims at tackling the issue of limited dataset size in deep learning based fault diagnosis of rolling bearing

Fig. 10 Effect of number of merge layers under different dataset sizes. **a** 100%, **b** 70%, **c** 40%, **d** 10%, and **e** 5%



employed in manufacturing systems. A parallel convolutional neural network (P-CNN) is formulated, which is capable of fusing time-domain features identified by one CNN and the time-frequency domain features extracted by CWT and subsequently identified by another CNN. The network performance is examined for a wide range of dataset sizes using publicly accessible rolling bearing data with various faults and multiple load conditions. The case analyses indicate that the proposed P-CNN outperforms a series of conventional methods. Both the identification accuracy and result consistency are improved with the information fusion. The key influencing factors, including the selection of manual feature extraction approach and the network structure, are discussed. The proposed P-CNN exhibits the promising aspect of achieving satisfying fault identification performance with reduced dataset size.

Authors' contributions Mingxuan Liang, Pei Cao, and Jiong Tang worked together to generate the conception of the work. Mingxuan Liang carried out algorithm development and data analysis and interpretation, and played the lead role in drafting the paper with Pei Cao's support. Jiong Tang provided advisement to Mingxuan Liang and Pei Cao, and also provided critical revision of the paper.

Funding Mingxuan Liang acknowledges the supports by the National Natural Science Foundation of China under Grant 51705494 and the Natural Science Foundation of Zhejiang Province, China, under Grant LQ17E050005. Pei Cao and J. Tang acknowledge the support by the National Science Foundation, USA, under Grant 1741171.

Compliance with ethical standards

Conflict of interest The authors declare they have no conflict of interest.

Ethical approval Not applicable.

Consent to participate Not applicable.

Consent to publication Not applicable.

References

- Attoui I, Oudjani B, Boutasseta N, Fergani N, Bouraiou A (2020) Novel predictive features using a wrapper model for rolling bearing fault diagnosis based on vibration signal analysis. *Int J Adv Manuf Technol* 106:9–12
- Wang H, Chen J, Zhou Y, Ni G (2019) Early fault diagnosis of rolling bearing based on noise-assisted signal feature enhancement and stochastic resonance for intelligent manufacturing. *Int J Adv Manuf Technol* 1–4, 1–7
- Xu Q, Lu S, Jia W, Jiang C (2020) Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning. *J Intell Manuf* 31:1467–1481
- Sousa R, Antunes J, Coutinho F, Silva E, Santos J, Ferreira H (2019) Robust cepstral-based features for anomaly detection in ball bearings. *Int J Adv Manuf Technol* 103:2377–2390
- Singh J, Darpe AK, Singh SP (2018) Rolling element bearing fault diagnosis based on over-complete rational dilation wavelet transform and auto-correlation of analytic energy operator. *Mech Syst Signal Process* 100:662–693
- Yeap, Y.M., Ukil, A., 2016. Fault detection in HVDC system using Short Time Fourier Transform, In: 2016 IEEE Power and Energy Society General Meeting.
- Manjurul Islam MM, Kim J-M (2019) Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Comput Ind* 106: 142–153
- Djebala A, Babouri M, Ouelaa N (2015) Rolling bearing fault detection using a hybrid method based on empirical mode decomposition and optimized wavelet multi-resolution analysis. *Int J Adv Manuf Technol* 79:2093–2105
- Zhang D, Yu D (2017) Multi-fault diagnosis of gearbox based on resonance-based signal sparse decomposition and comb filter. *Measurement* 103:361–369
- Moumene I, Ouelaa N (2016) Application of the wavelets multiresolution analysis and the high-frequency resonance technique for gears and bearings faults diagnosis. *Int J Adv Manuf Technol* 83:1315–1339
- Liang P, Deng C, Wu J (2019) Compound fault diagnosis of gear-boxes via multi-label convolutional neural network and wavelet transform. *Comput Ind* 113:103–132
- Vincent P, Larochelle H, Lajoie I (2010) Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 11:3371–3408
- Qi Y, Shen C, Wang D (2017) Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access* 5:15066–15079
- de Moura EP, Souto CR, Silva AA (2011) Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. *Mech Syst Signal Process* 25(5):1765–1772
- Zhu K, Song X, Xue D (2014) A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm. *Measurement* 47:669–675
- Chen Z, Li W (2017) Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE T Instrum Meas* 66(7):1693–1702
- Gan M, Wang C, Zhu C (2018) Fault feature enhancement for rotating machinery based on quality factor analysis and manifold learning. *J Intell Manuf* 29(2):463–480
- Lu Y, Xie R, Liang S (2019) Adaptive online dictionary learning for bearing fault diagnosis. *Int J Adv Manuf Technol* 101:195–202
- Tsironi E, Barros P, Weber C (2017) An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition. *Neurocomputing* 268:76–86
- Weimer D, Scholz-Reiter S, B. (2016) Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Ann Manuf Technol* 65(1):417–420
- Ince T, Kiranyaz S, Eren L (2016) Real-time motor fault detection by 1 - D convolutional neural networks. *IEEE Trans Ind Electron* 63(11):7067–7075
- Abdeljaber O, Avci O, Kiranyaz S (2017) Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J Sound Vib* 388:154–170
- Pan J, Zi Y, Chen J (2017) LiftingNet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification. *IEEE T Ind. Electron.* 65(6):4973–4982
- Duan Z, Wu T, Guo S, Shao T, Malekian R, Li Z (2018) Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: a review. *Int J Adv Manuf Technol* 96:803–819

25. LeCun Y, Boser BE, Denker J (1989) Back propagation applied to handwritten zip code recognition. *Neural Comput* 1(4):541–551
26. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, In: the Advances in neural information processing systems, 1097–1105
27. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition, In: the International Conference on Learning Representations, 1–14
28. Szegedy C, Vanhoucke V, Ioffe S (2016) Rethinking the inception architecture for computer vision, In: the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2818–2826
29. Wen, L., Gao, L., Li, X., 2018. A new data-driven intelligent fault diagnosis by using convolutional neural network, In: the IEEE International Conference on Industrial Engineering and Engineering Management, 813–817.
30. Zhang W, Li C, Peng G (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech Syst Signal Process* 100:439–453
31. Zhang R, Tao H, Wu L (2017) Transfer learning with neural networks for bearing fault diagnosis in changing working conditions. *IEEE Access* 5:14347–14357
32. Cao P, Zhang S, Tang J (2018) Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning. *IEEE Access* 6:26241–26253
33. Manjurul Islam MM, Kim JM (2018) Motor bearing fault diagnosis using deep convolutional neural networks with 2D analysis of vibration signal, In: the 31st Canadian Conference on Artificial Intelligence, 8–11
34. Li X, Zhang W, Ding Q (2018) A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. *Neurocomputing* 310:77–95
35. Janssens O, Slavkovikj V, Vervisch B (2016) Convolutional neural network based fault detection for rotating machinery. *J Sound Vib* 377:331–345
36. Ma S, Chu F (2019) Ensemble deep learning-based fault diagnosis of rotor bearing systems. *Comput Ind* 105:143–152
37. Zan T, Liu Z, Wang H, Wang M, Gao X (2020) Control chart pattern recognition using the convolutional neural network. *J Intell Manuf* 31:703–716
38. Ragab A, Yacout S, Ouali M, Osman H (2019) Prognostics of multiple failure modes in rotating machinery using a pattern based classifier and cumulative incidence functions. *J Intell Manuf* 30(1): 255–274
39. Yan R, Gao RX, Chen X (2014) Wavelets for fault diagnosis of rotary machines: a review with applications. *Signal Process* 96:1–15
40. Miao R, Gao Y, Ge L (2019) Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network. *Comput Ind* 112:103–115
41. Khan A, Sung J, Kang J (2019) Multi-channel fusion convolutional neural network to classify syntactic anomaly from language-related ERP components, *Inform. Fusion* 52:53–61
42. Liu Y, Chen X, Wang Z (2018) Deep learning for pixel-level image fusion: Recent advances and future prospects, *Inform. Fusion* 42: 158–173
43. Long J, Shelhamer E, Darrell T (2014) Fully convolutional networks for semantic segmentation, *IEEE T. Pattern Anal* 39(4): 640–650
44. Cui Z, Chen W, Chen Y (2016) Multi-scale convolutional neural networks for time series classification, *arXiv preprint arXiv: 1603.06995*
45. Zeiler MD, Taylor GW, Fergus R (2011) Adaptive deconvolutional networks for mid and high level feature learning, In: the Proceedings of the IEEE International Conference on Computer Vision, 2018–2025
46. Srivastava N, Hinton G, Krizhevsky A (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15(1):1929–1958
47. Smith WA, Randall RB (2015) Rolling element bearing diagnostics using the case western reserve university data: a benchmark study. *Mech Syst Signal Process* 64:100–131
48. Zhao, W., Wang, Z., Lu, C., 2016. Fault diagnosis for centrifugal pumps using deep learning and softmax regression, In: the 12th world congress on intelligent control and automation (WCICA).
49. Feng Y, Lu B, Zhang D (2015) Fault feature extraction for rolling bearing based on dual impulse Morlet Wavelet, In: ASME. International design engineering technical conferences and computers and information in engineering conference, V008T13A064

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.