Journal of Intelligent Manufacturing
https://doi.org/10.1007/s10845-020-01671-1

An end-to-end fault diagnostics method based on convolutional neural network for rotating machinery with multiple case studies

Yiwei Wang[1] · Jian Zhou[1] · Lianyu Zheng[1] · Christian Gogu[2]

Received: 29 November 2019 / Accepted: 15 September 2020

Abstract
The fault diagnostics of rotating components are crucial for most mechanical systems since the rotating components faults are the main form of failures of many mechanical systems. In traditional diagnostics approaches, extracting features from raw input is an important prerequisite and normally requires manual extraction based on signal processing techniques. This suffers of some drawbacks such as the strong dependence on domain expertise, the high sensitivity to different mechanical systems, the poor flexibility and generalization ability, and the limitations of mining new features, etc. In this paper, we proposed an end-to-end fault diagnostics model based on a convolutional neural network for rotating machinery using vibration signals. The model learns features directly from the one-dimensional raw vibration signals without any manual feature extraction. To fully validate its effectiveness and robustness, the proposed model is tested on four datasets, including two public ones and two datasets of our own, covering the applications of ball screw, bearing and gearbox. The method of manual, signal processing based feature extraction combined with a classifier is also explored for comparison. The results show that the manually extracted features are sensitive to the various applications, thus needing fine-tuning, while the proposed framework has a good robustness for rotating machinery fault diagnostics with high accuracies for all the four applications, without any application-specific manual fine-tuning.

Keywords Fault diagnostics · Rotating machinery · Vibration signals · Convolutional neural network

Introduction

Rotating machinery is the essential equipment playing a crucial character in the modern industry. As indispensable key transmission devices of rotating machinery, the typical rotating components such as ball screws, bearings, and gears, are the leading cause of failure in essential industrial equipment such as induction motors, wheelset of high-speed railway bogie, aero-engines, wind-turbine, etc. According to statistics, 30–51% of rotating machinery failure are caused by these key components (Islam and Kim 2019a; Zhao et al. 2020). Failure of the rotating components results in machine performance degradation, unwanted downtime, economic losses and even human casualties. Normally, the rotating components are installed deep inside the machine and undergo a long degradation process from healthy to failure. It is not practical to frequently shut down and disassemble the machines to examine their health state. If damaged rotating components are left unattended, it may cause secondary damage for the machines. On the other hand, due to different working conditions and other uncertainties, even the same type of rotating components may exhibit their own degradation process individually, making it difficult to accurately estimate the health states based on statistics of large samples. Therefore, online monitoring and real-time fault diagnostics of individual rotating components based on manufacturing big data is an urgent demand.

Smart manufacturing, which is characterized by the integration of Artificial Intelligence (AI) with recent emerging technologies (Lee et al. 2018), enables online monitoring and massive manufacturing data acquisition from sensors and terminals installed in equipment. However, the data must be converted into useful information before it can be of value to the industry. Prognostics and health management (PHM)

✉ Lianyu Zheng
lyzheng@buaa.edu.cn

[1] School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

[2] Institut Clément Ader (UMR CNRS 5312) INSA/UPS/ISAE/Mines Albi, Université de Toulouse, 31400 Toulouse, France

Published online: 16 October 2020

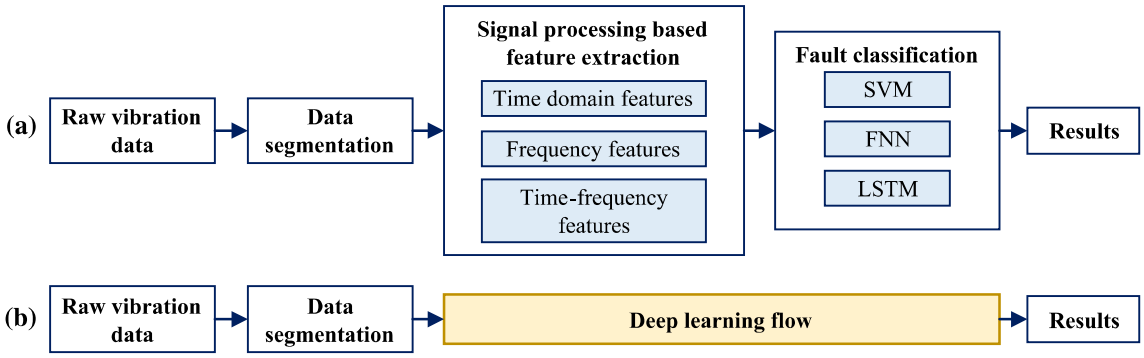

Fig. 1 Diagnostics methods: traditional versus deep learning based

is such a bridge converting manufacturing big data to useful information. As an emerging discipline receiving great attention from both academia and various industries, PHM has been listed as a part of the "standard architecture of smart manufacturing" proposed by China. PHM deeply fuses AI into manufacturing industries through a complete architecture containing functions such as intelligent fault diagnostics, prognostics, predictive maintenance, etc. (Vogl et al. 2019; Xia and Xi 2019). This fusion enables timely online fault diagnostics of devices as well as their future state prediction, and consequently, results in the improvement of the maintainability, supportability, reliability and safety of essential industrial equipment. As an important part consisting of PHM, intelligent fault diagnostics provide solutions for real-time fault diagnostics of individual rotating components.

For rotating machinery, the vibration signal is widely used for fault diagnostics due to various advantages, such as continuous monitoring without stopping the machines, ease of use, sensitivity towarding faults etc. Traditional intelligent fault diagnostics normally contains two sequential steps of manually extracting features from raw vibration signals followed by establishing the mapping between the extracted features and the corresponding states based on classification techniques such as support vector machine (SVM) (Goyal et al. 2019) or feedforward neural network (FNN), as shown in Fig. 1a. Whether a fault sensitive feature can be extracted affects the performance of the diagnostics model significantly, and hence lots of effort are devoted to extracting suitable features before a classification algorithm can be employed. The features are normally extracted from time domain (Park et al. 2018), frequency domain, or time-frequency domain using various signal processing techniques such as fast Fourier transform, Hilbert-Huang transform (Feng and Pan 2012), empirical mode decomposition (Liu et al. 2018), variation mode decomposition (Yan and Jia 2018), wavelet transform (Dhamande and Chaudhari 2018 ; Wang et al. 2018a), intrinsic time scale decomposition (Feng et al. 2016), local mean decomposition (Wang et al. 2018b), etc. Manually extracting features, while having led to satisfying results in the past, also exhibits some drawbacks. The complex signal processing techniques required by feature extraction highly depend on the expertise and prior knowledge, and also require lots of human labour. In addition, manually extracted features are normally empirical and thus sensitive to changes. These empirical features reduce the flexibility and the generalization ability of the diagnostics model, i.e., the model performs highly accurately for one particular diagnostics task while much less accurately for another task. Therefore, significant human labour and expertise are required for exploring and designing suitable features for different diagnostics tasks (Jing et al. 2017). These difficulties in feature extraction seriously hinder fault diagnostics evolving into a mature technology that can be widely deployed in industry.

The strong feature-learning ability of deep learning such as auto encoder and convolutional neural network (CNN) provides a potential solution to the aforementioned drawbacks (Hamadache et al. 2019; Zhao et al. 2019; Li et al. 2019a; Jia et al. 2018). The hierarchical structures of multiple neural layers enable deep learning networks to directly mine information from raw data layer by layer (Fig. 1b). Compared with other deep learning methods, CNN significantly reduces the number of parameters to be optimized by the strategies of weight sharing and sub-sampling. CNN also has strong anti-noise ability because of its insensitivity to the local change due to the convolution process. Inspired by the successful employment of CNN in image classification area, it is easy to think of converting waveform signal into images and then using CNN for fault diagnostics. Hoang and Kang (2019) converted vibration signal into grayscale images through a simple method proposed by Nguyen et al. (2013), and then fed the images into CNN for bearing diagnostics. Chen et al. (2019) proposed a scheme combining discrete wavelet transformation (DWT) with CNN for planetary gearboxes fault diagnostics. A series of sets of wavelet coefficients of DWT were used as the input of CNN. Wang et al. (2019) proposed a conversion method converting vibration signals from multiple sensors to images. A bottleneck layer optimized CNN

was used for rotating machinery diagnostics. Islam and Kim (2019b) used 2D representation of acoustic emission signal processed by wavelet packet transform as the input of an adaptive deep CNN for bearing fault diagnostics. Wang et al. (2017) converted time sequences signal of gear box into time-frequency images using continuous wavelet analysis and then fed the images into a deep CNN. Zhu et al. (2019a) transformed multiple vibration signals of a rotor into symmetrized dot pattern (SDP) images before classified by CNN. Zhu et al. (2019b) employed short-time Fourier transform to convert one-dimensional signals of bearing into a time-frequency graph and then a novel capsule network was proposed for diagnosing. Liang et al. (2020) employed wavelet transform to extract time-frequency image features from raw signals. Generative Adversarial Networks (GANs) were used to generate additional fake training images for data augmentation purposes. A CNN model was built for fault modes classification. The proposed method was validated on a gearbox application. Chen et al. (2020) used cyclic spectral analysis to obtain the two-dimensional Cyclic Spectral Coherence maps of vibration signals and a CNN model was constructed to learn high-level feature representations and conduct fault classification. The method was validated on a public dataset of bearing faults published by the Case Western Reserve University (CWRU). Zhang et al. (2020) processed the raw vibration signals to gray-scale images without any predetermined parameters and then fed into a CNN with two dropout layers and two fully-connected layers for fault classification. CWRU bearing dataset were used for validation.

It can be seen that most studies require one additional step that converts 1D vibration signal into 2D representations before using the CNN model, which circumvent some drawbacks of manually feature extraction but still need application-specific adaptation. Recently, directly extracting features from one-dimensional raw vibrational data without any signal processing techniques has begun to be proposed by researchers. This provides an end-to-end solution for fault diagnostics, which reduces the dependencies on expertise and prior knowledge, and hence facilitates the use and deployment of diagnostics model. Wu et al. (2019) optimized the 2D CNN to be a one-dimensional CNN that is suitable for processing vibration signals, and validated the proposed model on gearbox application. Li et al. (2018a) proposed a 1D CNN model with the residual learning algorithm for bearing fault diagnostics, and the raw data without any pre-processing were fed into the built model. Li et al. (2020) developed an adaptive 1D separable convolution with residual connection network for diagnosing gear pitting. Peng et al. (2019) proposed a deeper 1D CNN based on a 1D residual block for the fault diagnostics of wheelset bearings in high-speed trains. Wide convolution kernel and dropout technology were used in the CNN to enhance the network's generalization performance. The traditional fault diagnostics, 1D CNN and 2D CNN methods employed in the literature reviewed above are summarized in Table 1.

The above studies focus only on one specific application. Specifically, bearings and gearboxes are more extensively studied than ball screws, whose fault diagnostics studies are very limited due to lack of public dataset. The generalizability of CNN to fault diagnostics of rotating components is necessary to be fully investigated. In this paper, we propose an end-to-end fault diagnostics method based on CNN using raw vibration signal. A CNN model consisting of three stacks of convolutional and pooling layer, dropout layer and fully connected layer is proposed. The alternating convolution and pooling layers of the CNN model automatically extract feature maps from raw data layer by layer. The softmax function is used as the activation function of the last fully connected layer for dealing with multi-class classification problems. No manually extracted feature is necessary. To fully validate the effectiveness and the generalizability of the proposed model for fault diagnostics of rotating components, we tested on four datasets, including two public ones and two of our own, covering the applications of ball screw, bearing and gearbox. These three types of rotating components are the typical ones that are widely used as the key components in essential industrial equipment such as machine tools, high-speed trains, aero-engines, wind-turbine, gas-turbine, etc. To our best knowledge, our work firstly validates the CNN model for fault diagnostics in such wide applications. Moreover, the signal processing based feature extraction combined with long short-term memory (LSTM) network (the combination method is referred to as traditional method here) is also explored and compared with the proposed CNN model. Specifically, three typical engineered features, i.e., (a) wavelet packet energy (WPE) based on wavelet packet decomposition, (b) instantaneous frequency (IF), and (c) instantaneous spectral entropy (ISE) based on power spectrogram, are constructed from the raw vibration data and then used as the input of an LSTM network. The proposed CNN model is compared with the traditional method in terms of accuracy and robustness in various applications

The remainder of the paper is organized as follows. "The CNN-based diagnostics framework for rotating machinery" section details the structure and the feature learning mechanism of the proposed model. In "Case studies and discussions" section, the generalization of the proposed model is verified by four case studies covering the commonly used rotating components of ball screw, bearing and gear. The generalizability and robustness of the proposed model is further discussed by the comparison with traditional methods. Finally, conclusions and perspectives are given in "Conclusions and future work" section.

Table 1 Summary of different categories of fault diagnostics

Category	Method	Advantage/disadvantage	References
Traditional fault diagnostics	First manually extract features based on signal processing techniques such as Fourier transform, Hilbert-Huang transform, empirical mode decomposition, wavelet transform, etc. Then feed the features to classifier such as support vector machine, shallow neural networks, etc.	Signal processing techniques highly depends on expertise and prior knowledge Manually extracted features are application-specific and quite sensitive to environment or working conditions. Require lots of skilled labour to explore and design suitable features for new diagnostics task.	Goyal et al. (2019), Feng and Pan (2012), Liu et al. (2018), Yan and Jia (2018), Dhamande and Chaudhari (2018), Wang et al. (2018a, b), Feng et al. (2016), and Jing et al. (2017)
2D convolutional neural network	First convert raw vibration signal to 2D representations such as grayscale images, time-frequency images, symmetrized dot pattern images, cyclic spectral coherence maps, etc. Then feed the 2D representations to 2D convolutional neural network for classification.	Circumvent some drawbacks of manually feature extraction but still need application-specific adaptation.	Hoang and Kang (2019), Nguyen et al. (2013), Wang et al. (2017, 2019), Islam and Kim (2019b), Zhu et al. (2019a, b); Chen et al. (2020), Zhu et al. (2019), Chen et al. (2019), Liang et al. (2020) and Zhang et al. 2020
1D convolutional neural network	Use 1D convolutional neural network to accomplish direct feature extraction from raw vibration signal and classification.	Provide end-to-end solutions for fault diagnostics Reduce the dependencies on expertise and prior knowledge Reduce the sensitivities to environment or working conditions Facilitate the use and deployment of diagnostics models.	Wu et al. (2019), Li et al. (2018b, 2020) and Peng et al. (2019)

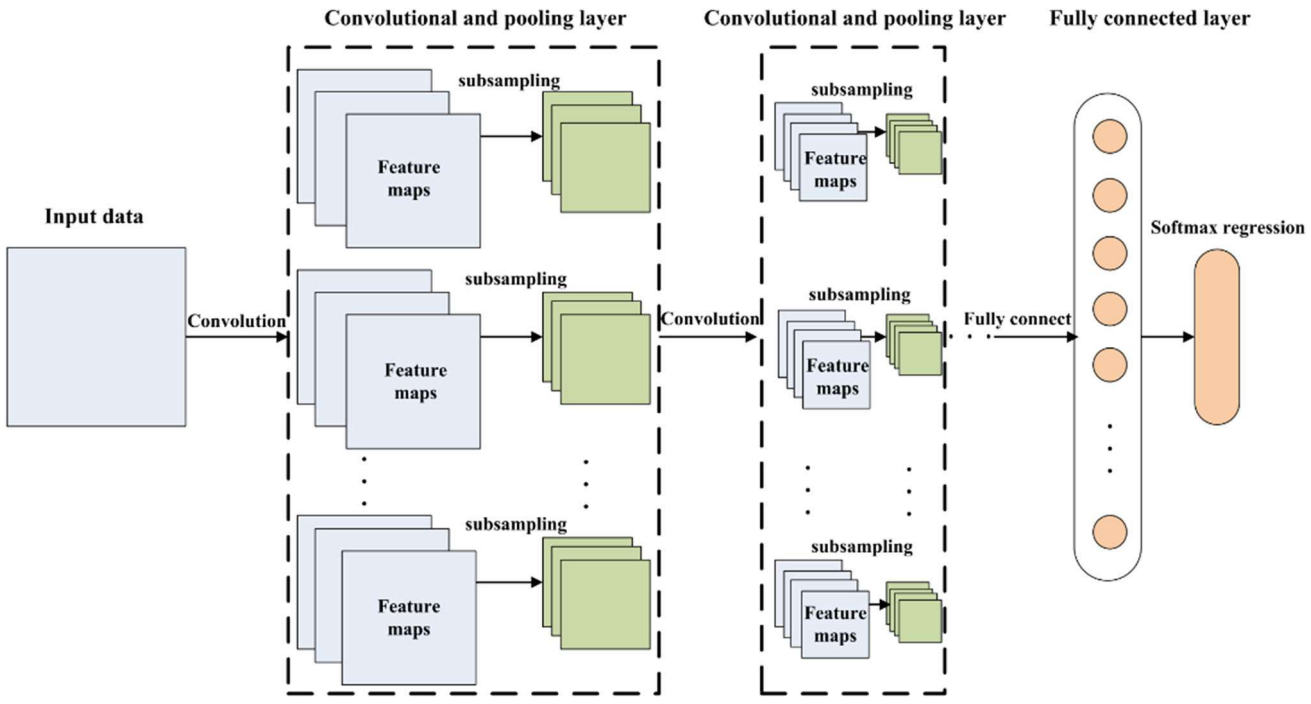

Fig. 2 A typical architecture of CNN (Jing et al. 2017)

The CNN-based diagnostics framework for rotating machinery

Convolutional neural networks (CNNs), first proposed by LeCun for image processing, has two characteristics, i.e., spatially shared weights and spatial pooling (Goodfellow et al. 2019). The architecture of a typical CNN is illustrated in Fig. 2, which is structured by series of stages (Jing et al. 2017). The convolutional layer convolves multiple filters with raw input data and generate feature maps. Pooling layer often follows the convolutional layer to reduce the size of feature map and extract the most significant local features (Li et al. 2019b). The last stage of the architecture consists of a fully-connected layer, which is normally a multi-class classification model.

The schematic diagram of the proposed framework is illustrated in Fig. 3. The sliding window method was used to segment the raw time sequence vibration data of each health state and then reshape to a matrix before feeding into the neural network. The one-hot encoding method is used to manually create the labels of samples, which serve as the output of the network. For example, if there are three classes of data, the first class is encoded as (1, 0, 0), the second (0, 1, 0), and the third (0, 0, 1). The self feature learning ability is realized by the hidden layers, which is comprised of stacks of alternated convolutional layers and pooling layers. One-dimensional convolution kernels and pooling kernels are used in the network since the input is a one-dimensional time series signal. The structure of the CNN model and the feature learning process are detailed below.

Structure of the proposed CNN model

The structure of the proposed model is illustrated in Fig. 4, including three stacks of convolution-pooling layers and a fully connected layer. In the convolutional layer, multiple filters are convolved with raw input data and generate translation invariant features. In the subsequent pooling layer, the feature is compressed by sliding a fixed-length window following several rules such as average, max and so on. In the first two stacks maxpooling layer is used while in the last stack the average pooling layer is used. The data flow from the input of the network to the final output in Fig. 4 is detailed by explaining the entities (denoted by the Greek letters) and the actions (denoted by arrows).

1 $\boldsymbol{\alpha}$ is the input matrix of the network, which has the shape (m_1,n_1). Note that we use the form (m,n) to represent a m-by-n matrix. The subscript of m and n as well as f and s that will be introduced later represents the index of the layer.

2 $\boldsymbol{\beta}_i$ is a filter with shape (h,n), in which, i = 1,2,…,f_1. f_1 is the number of filters in the 1st layer. h is the kernel size of the convolution.

3 $\boldsymbol{\gamma}$ is the output matrix of the 1st convolution layer, having the shape (m_1, f_1).

4 From $\boldsymbol{\alpha}$ to $\boldsymbol{\gamma}$, the convolution operation is carried out, which is detailed as follows. The dot product between filter $\boldsymbol{\beta}_i$ and a concatenation vector $\boldsymbol{\alpha}_{k:k+h-1}$ defines the convolution operation.

$$c_j = \varphi(\boldsymbol{\beta}_i \cdot \boldsymbol{\alpha}_{k:k+h-1} + \boldsymbol{b}) \tag{1}$$

in which, · represents the dot product, b the bias term and φ the non-linear activation function. $\boldsymbol{\alpha}_{k:k+h-1}$ is a h-length window starting from the k-*th* row to the (k + h-1)-*th* row, which is defined as:

$$\boldsymbol{\alpha}_{k:k+h-1} = \boldsymbol{\alpha}_k \oplus \boldsymbol{\alpha}_{k+1} \oplus \cdots \oplus \boldsymbol{\alpha}_{k+h-1} \tag{2}$$

where $\oplus$ is the concatenation operation of two vectors. As defined in Eq. 1, the output scalar c_j can be regarded as the activation of the filter $\boldsymbol{\beta}_i$ on the corresponding concatenation vector $\boldsymbol{\alpha}_{k:k+h-1}$. By sliding the filter $\boldsymbol{\beta}_i$ through $\boldsymbol{\alpha}$ and applying zero padding technique, m_1 output scalar c_j can be obtained, forming a column vector $\mathbf{c}_i$, also known as a feature map:

$$\mathbf{c}_i = [c_1, c_2, ..., c_j, ..., c_{m_1}] \tag{3}$$

One filter corresponds to one column vector. Since there are f_1 filters in the first layer, the output matrix $\boldsymbol{\gamma}$ is thus (m_1, f_1) matrix. From the above operation it can be seen that one filter performs multiple convolution operations, during which the weights of the filter are shared. The feature map $\mathbf{c}_i$, obtained by convolving one filter $\boldsymbol{\beta}_i$ over the input data, represents the feature of the input data extracted from a certain level. By convolving the input data with multiple filters, a high-dimensional feature map containing multiple column vectors that reflect the input data from different perspectives are extracted.

5. $\boldsymbol{\mu}$ is the output matrix of the 2nd layer, having the shape $(m_2/s_2, n_2)$, where s_2 is the pooling length of the 2nd layer. Note that m_2 and n_2 denote input size of the 2nd layer. Since the output of the current layer is the input of the next layer, $m_2 = m_1$ and $n_2 = f_1$.

6 From $\boldsymbol{\gamma}$ to $\boldsymbol{\mu}$, max pooling operation is carried out, which is detailed as follows. The max operation is taken over the s_2 consecutive values in $\mathbf{c}_i$. Then the compressed column vector h is obtained as:

$$\mathbf{h}_i = [h_1, h_2, ..., h_l, ..., h_{m/s}] \tag{4}$$

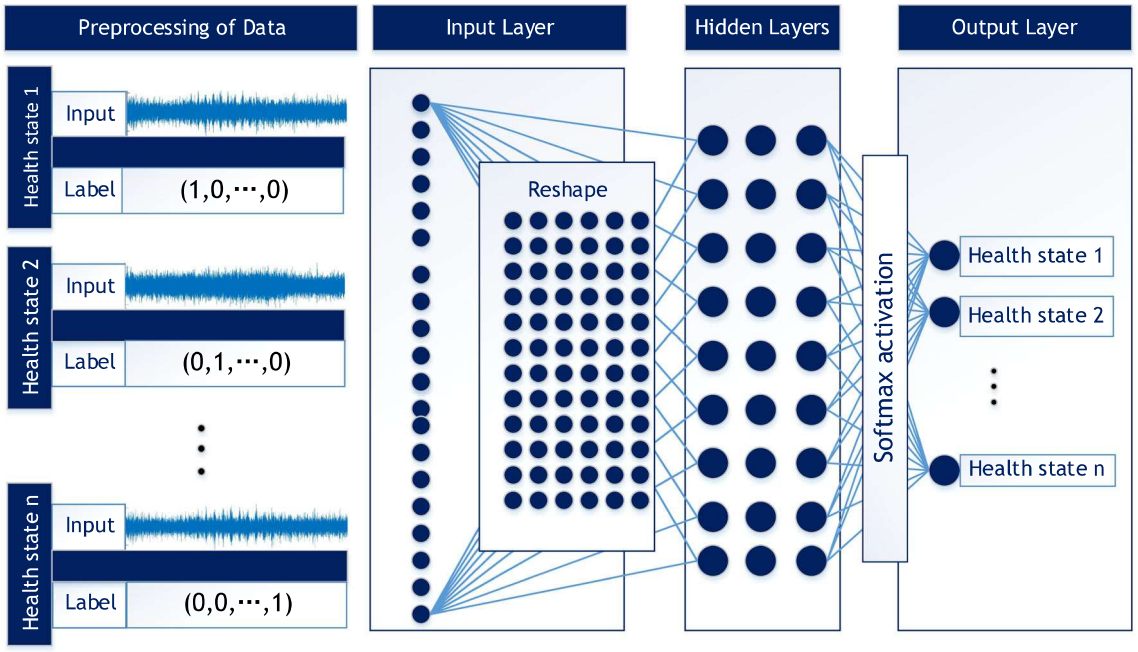

Fig. 3 Framework of proposed diagnostics model

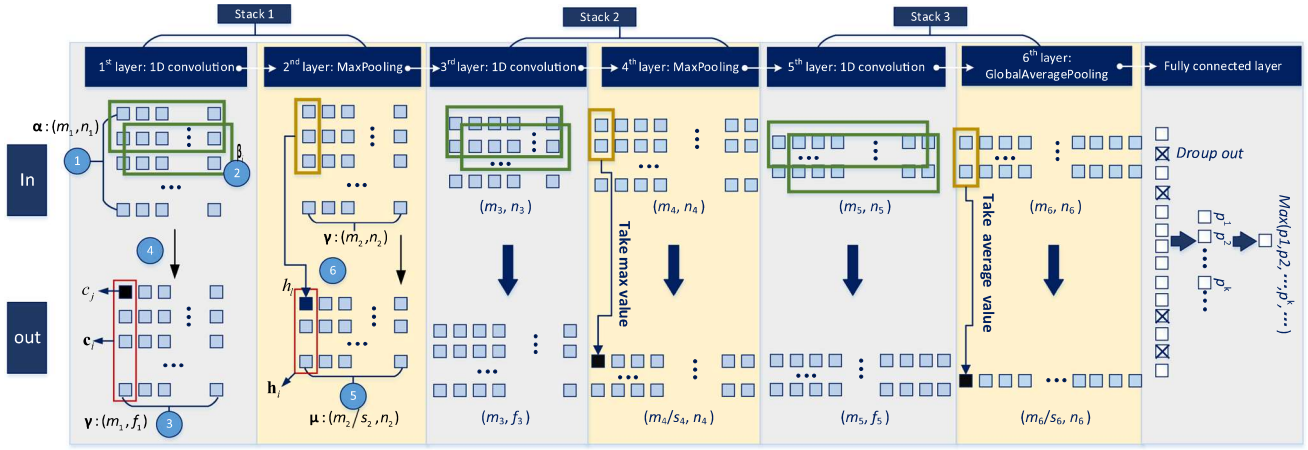

Fig. 4 Structure of proposed CNN network

where h_l=max[$c_{(l-1)s+1}, c_{(l-1)s+2}, \cdots, c_{ls}$]

From above we see that when a matrix goes through one convolution layer, its number of rows keep unchanged and the number of column equals to the number of filters. In the case of pooling layer, the number of columns keeps unchanged while the number of rows is compressed depending on the pooling length.

7. In the 2nd and 3rd stacks, the convolution and pooling propagate. The only difference is that the number of filters and the pooling length varies.
8. The output of the 7th layer is flattened and connected with a fully connected layer, which is similar to a traditional multilayer neural network and can be applied through different classification. The dropout technique is employed to prevent overfitting. The softmax function (Behley et al. 2013) is used as the last layer, which gives the probability of being each label. Specifically, assuming a K-label classification task, the output of the softmax function can be calculated as Eq. 5, in which $\boldsymbol{W}_k$ and $\boldsymbol{b}_k$ are the weight matrix and bias, $P(y = k|\mathbf{x}; \boldsymbol{W}_k, \boldsymbol{b}_k)$ is the probability of being the k-th label (denoted as p^k in Fig. 4) given the input $\mathbf{x}$ and the corresponding weight and bias. Here $\mathbf{x}$ is the vector after drop out in the fully connected layer. The final output of the

network is the health sate label with the highest probability.

$$\begin{bmatrix} P(y=1|\mathbf{x};\mathbf{W_1},\mathbf{b_1}) \\ \cdots \\ P(y=k|\mathbf{x};\mathbf{W}_k,\mathbf{b}_k) \\ P(y=K|\mathbf{x};\mathbf{W}_K,\mathbf{b}_K) \end{bmatrix} = \frac{1}{\sum_{k=1}^{K}\exp(\mathbf{W}_k\mathbf{x}+\mathbf{b}_k)} \begin{bmatrix} \exp(\mathbf{W_1}\mathbf{x}+\mathbf{b_1}) \\ \cdots \\ \exp(\mathbf{W}_k\mathbf{x}+\mathbf{b}_k) \\ \exp(\mathbf{W}_K\mathbf{x}+\mathbf{b}_K) \end{bmatrix} \tag{5}$$

Hyperparameters

The activation function of all the convolutional layers is the Relu function due to its ability to avoid gradient vanishing and to its fast convergence. The loss function of the CNN model is cross-entropy and the precision function is categorical accuracy. L2 regularization term is set for the first and third convolution layers to reduce overfitting. The parameter of L2 term is a trade-off between the effectiveness of training and overfitting, i.e., a too-large value will lead to inadequate training and a too-small value is not enough to reduce the risk of overfitting. We set this value to 0.001 based on the prior study (Ng 2004). Dropout is set for the fully connected layer to reduce overfitting by directly setting the neurons of the network to zero in a given proportion. We referred to the study of the founder of the dropout technique (Srivastava et al. 2014) and set this proportion to 0.5, which is a typical value in deep learning.

The initial weight of the network is set by the glorot uniform function, and the bias are set to 0. The weight is optimized by the adaptive moment estimation (ADAM) solver with initial learning rate 0.001 and exponentially decayed rate 0.1. Adam solver is a combination of the Momentum and RMSProp optimization algorithms. It designs an independent adaptive learning rate for different parameters by calculating the first-order moment estimation and second-order moment estimation of the gradient, which typically gives better optimization performance than the alternative stochastic gradient descent with momentum (SGDM) solver (Kingma and Ba 2015). Adam algorithm is currently the most widely used optimization algorithm embedded in the field of machine learning and deep learning.

The mini batch training strategy is adopted here. Specifically, the training examples are divided into small batches. The model parameters will be updated after each batch passing through the network. The passing through of one batch is called one iteration. When the entire training example is passed through the network once and each example has the opportunity to update the model parameters, it is one epoch. The execution environment is an Intel e5-2620v4 CPU and a GeForce RTX2080Ti GPU. The above network setting and the execution environment will be used in all the following cases.

Case studies and discussions

Case 1: Ball screw lubrication states diagnostics

Experiment and data preparation

In this case study, the proposed model is validated for diagnosing the lubrication states of the ball screw. Ball screws are crucial mechanical components being intensively used in many engineering systems that requires precise positioning such as the feed system in machine tool, and in high precision leveling systems for aircrafts and missiles (Li et al. 2018a). The growing demand for high speed and large lead for ball screws makes it increasingly important to keep good lubrication in order to reduce the friction. Indeed, correct lubrication is vital to ball screws since the lubrication affects significantly its performance. Poor lubrication may increase the friction and impairs the positioning accuracy of ball screws. In addition, abnormal vibration caused by poor lubrication accelerates the damage of the machine tool and affects the quality of machining. Therefore, monitoring and online diagnosing of the lubrication state of the ball screw is important for improving position accuracy and lifetime of ball screws.

Very few reports are available regarding ball screw lubrication state diagnostics. Motivated by this, we design an experiment that simulates the different lubrication states of ball screws. The experiment is carried out in the test bench which was originally designed for measuring the friction torque of a ball screw, as shown in Fig. 5. The drive system drives the nut moving along the screw back and forth. Three states labeled as “Grease”, “Oil”, and “Absent” are simulated by (1) lubricating the ball screw using grease, (2) lubricating using oil and (3) removing the original lubricant, respectively. These three health states simulate the typical lubrication states that ball screws may encounter in real working environment. The vibration signals corresponding to the three states are acquired at a sampling rate 5 kHz with the data acquisition system Prosig P8020, as shown in Fig. 6. 128-s data are acquired for each lubrication states.

The raw time domain signals of one round trip (forward and reverse motion) of the nut under “Absent” lubrication is shown in Fig. 7. Two parts can be clearly seen, which correspond to the signals of forward and reverse motions of the nut, respectively. The abrupt “peaks” are due to the sharp slowdown and stop of the nut near the end of the motions. The data near the beginning and end of the motions are discarded. Only the “steady state” data in the middle stage of the motions are retained. For conciseness, the full raw signals under “Oil” and “Grease” lubrications are not presented. Instead, the retained segment of the forward motion of the three lubrication conditions are given in Fig. 8a. It can be seen that the differences among the three states are quite small, thus we further transform the signal into frequency domain

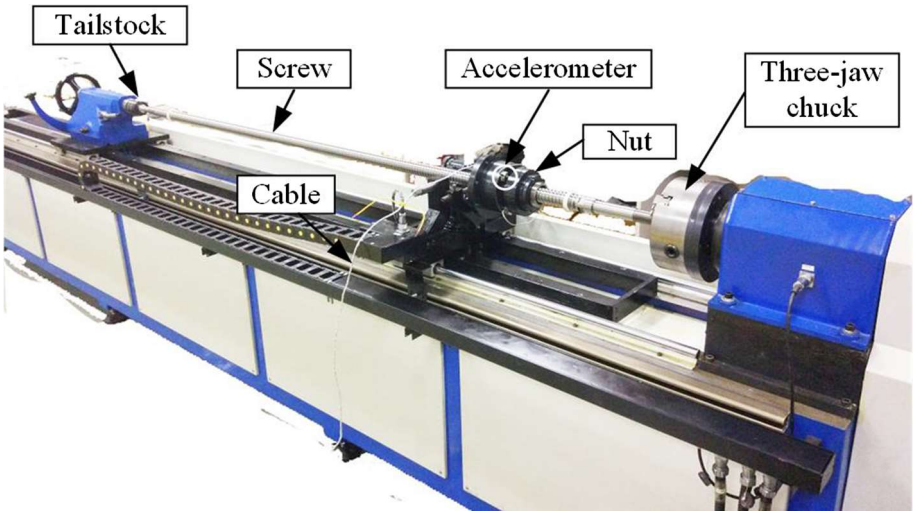

Fig. 5 Ball screw test bench

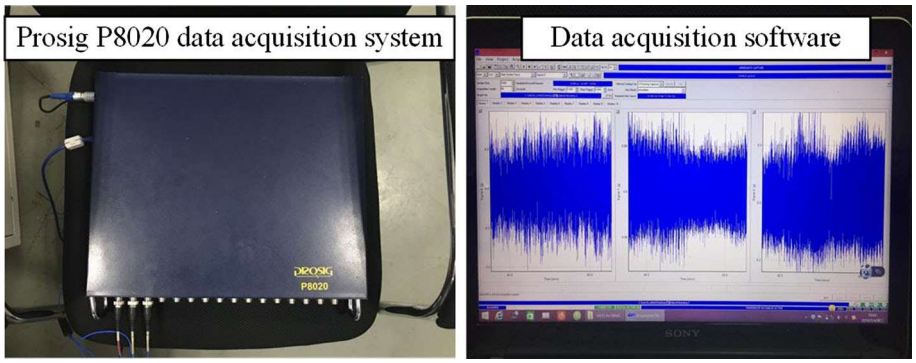

Fig. 6 Data acquisition set-up

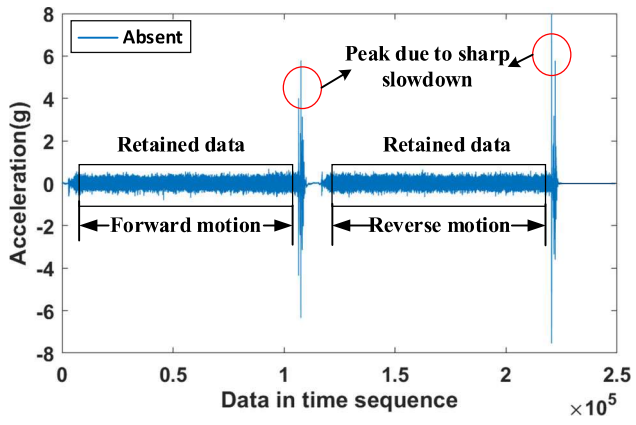

Fig. 7 Raw signal under "Absent" lubrication

using FFT, as shown in Fig. 8b. The differences among the three cases are not obvious and it is hard to see appropriate patterns, making it more challenging to correctly distinguish different lubrication conditions.

The raw vibration signal is divided into segments to form the input samples of the network. For each state, there are $128 \times 5000 = 6.4 \times 10^5$ data point. 6400 samples are selected as one segment and is further reshaped to a (64, 100) matrix. It is worth pointing out that the sample length should be traded off between the number of samples and the feature information that one sample contains. A too-short length of time window may carry incomplete feature information, leading to the difficulty of diagnostics, while a long length of time window will result in insufficient training data. Based on the sampling rate of data used in this paper as well as other related research works, we take 6400 data points as one sample. 80% data (80 samples) are reserved for training and the rest 20% (20 samples) for testing. Finally, the training/testing samples taken from each lubrication state form the overall training sets ($80 \times 3 = 240$ samples) and the testing sets ($20 \times 3 = 60$ samples). The input/output shape, the kernel size, stride and number of filters of each layer during the training process are reported in Table 2. Note that the above hyperparameters (the kernel size, stride and number of filters of each layer) remain unchanged in all the following case studies.

Results and discussions

The diagnostics accuracy on the test set is 100% and all the three states are correctly classified (thus the confusion matrix is not given). In order to better illustrate the feature learning process of the CNN model, the t-distributed

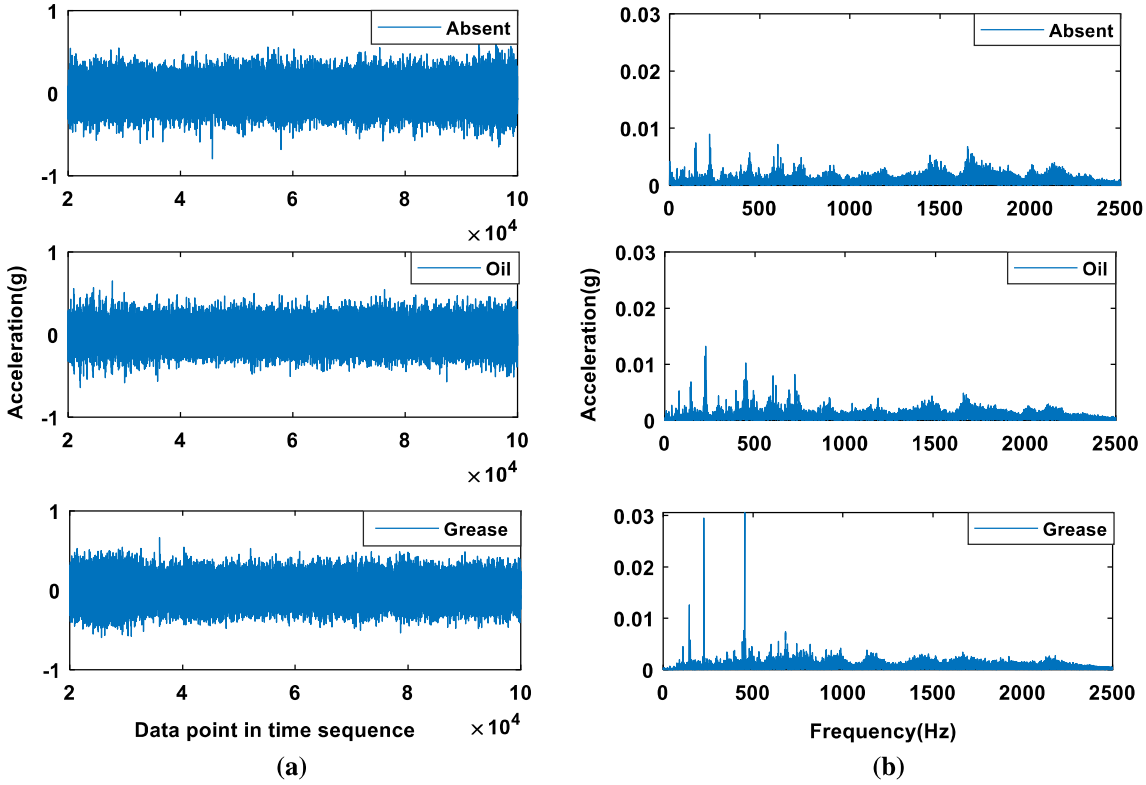

Fig. 8 Retained segment of the forward motion of the three lubrication conditions: (a) in the time domain, (b) in the frequency domain

Table 2 Parameters of the proposed model

No. of layer	Layer	Input shape	Kernel size/stride/number of filters	Output shape
1	Convolution	(240, 64, 100)	(3, 100)/1/64	(240, 64, 64)
2	Maxpooling	(240, 64, 64)	3/3	(240, 21, 64)
3	Convolution	(240, 21, 64)	3 × 64/1/128	(240, 21, 128)
4	Maxpooling	(240, 21, 128)	3/3	(240, 7, 128)
5	Convolution	(240, 7, 128)	3 × 128/1/128	(240, 7, 128)
6	Average pooling	(240, 7, 128)	/	(240, 128)
7	Dropout	(240, 128)	/	(240, 64)
8	Fully connected	(240, 64)	/	(240, 3)

stochastic neighbour embedding (t-SNE) technique (Maaten and Hinton 2008) is used to illustrate the output of each layer. t-SNE is a machine learning algorithm for high dimensional data visualization using nonlinear dimensionality reduction technique. For the current case study, the feature outputted by each layer is high-dimensional, whose shape is given in Table 2 (e.g., after dropout layer, the feature that is fed into the fully connected layer for classification is a 1-by-64 vector). We use the t-SNE technique to reduce the feature after each layer to two-dimensional space in order to show how the data "flow" from input to output, and thus to see how the features belonging to the same state aggregate. Figure 9 shows this process during testing, in which the distance between points represents the similarity of different samples. The symbols "1" (red), "2" (blue) and "3" (green) in the figures represents the "Absent", "Oil", and "Grease" lubrication states, respectively. We see that the in the input layer, the dots of the three states are completely mixed and no pattern can be observed to distinguish different fault modes. With the convolutional and pooling operations implemented, the mixed dots gradually separated, and in the output layer, dots belonging to the same state are clustered and dots belongs to different state are completely separated. In the output layer of Fig. 9, it can be seen that all the features belonging to the same state are clustered and the features belonging to different state are completely separated. There is no confusion, corresponding to 100% accuracy.

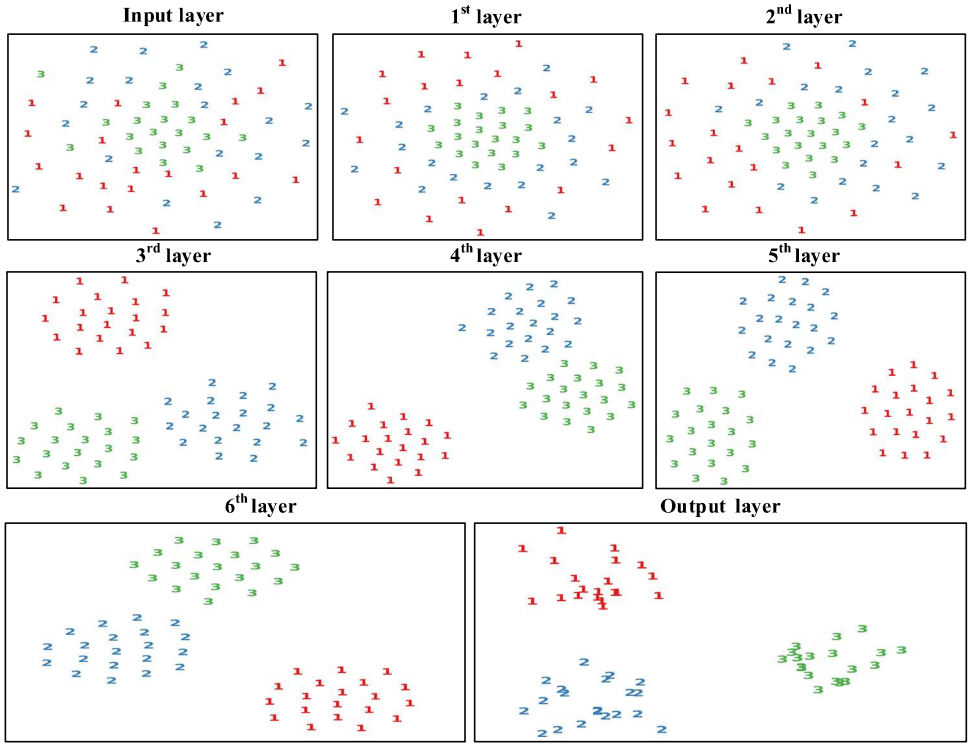

Fig. 9 Model testing process visualization by t-SNE, case study 1

Two well-known and widely used machine learning methods, i.e., feedforward backpropagation (BP) network and support vector machine (SVM) are used here as comparison methods. The sample length remains the same as that used by the CNN model, and accordingly, the number of training samples and testing samples are unchanged. Since these two methods normally accept low-dimensional or moderate-dimensional data as input, each raw sample is pre-processed by wavelet packet decomposition (WPD) to extract a feature vector. Specifically, based on our prior knowledge on the study of fault diagnostics using the vibration signal, a five-level WPD is applied on each raw sample and accordingly $2^5 = 32$ frequency sub-bands of the raw sample are obtained. The energy of each sub-band is calculated and concatenated to form a 1-by-32 feature vector, which is the input of the BP neural network and SVM.

A typical three-layer BP neural network is used. The number of neurons in the input layer and output layer are 32 and 3, which equals to the length of the feature vector and to the three lubrication states, respectively. A hidden layer containing 10 neurons is adopted. Note that the number of neurons in the hidden layer is typically an empirical value. Too many or too few neurons may reduce the classification accuracy of the network. We gradually increased the number of neurons from 5 to 20, and finally set this value to 10, where the network achieved its highest accuracy.

The basic SVM for binary classification is employed and is turned into multi-classes classifiers by the strategy of "one-vs-all". The strategy involves training a single SVM classifier for each class, with the samples of that class as positive samples and all other samples as negatives. Specifically, for the current case study of ball screw, we trained three basic SVM classifiers that are able to diagnose "Absent", "Oil", and "Grease", respectively. The Radial Basis Function (RBF) is used as the kernel function.

The diagnostics accuracies of BP neural network and SVM are 95%, and 90%. The corresponding confusion matrices are given in Fig. 10a, b, respectively. The last column in the matrix shows the percentages of examples predicted to belong to each label that are correctly classified (also called precision, or positive predictive value). For example, in the 1st row of Fig. 10a, 19 samples are classified by the proposed model as lubrication state "Absent", while 18 out of these 19 are correctly classified. One sample that should belong to label "Oil" are incorrectly classified as "Absent". The precision for label "Absent" hence equals to 18/19 = 94.7%. The row at the bottom of the matrix shows the percentages of all the samples belonging to each class that are correctly classified (also called sensitivity or true positive rate, recall, probability of detection, etc.). For example, in the 1st column of Fig. 10a there are 20 samples of "Absent". 18 out of 20 are correctly classified and one sample is incorrectly identified. The sensitivity for label "Absent" is thus 18/20 = 90%. The value in the bottom right corner is the overall classification accuracy of the network, which equals to the number of correctly classified samples divided by the total number of testing samples, in this case, 95%.

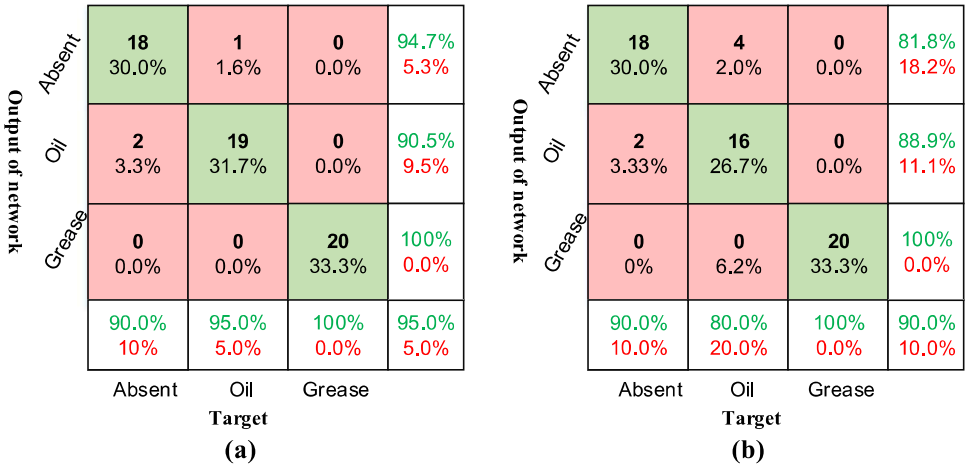

Fig. 10 Confusion matrices of case study 1, given by **a** BP neural network and **b** SVM

Compared with the proposed CNN model, the overall accuracies of the BP neural network and SVM are lower. Figure 10 indicates that confusions occur between the lubrication states of "Absent" and "Oil". Objectively, in real working environment, the differences between "Absent" and "Oil" are indeed small. This is also reflected in Fig. 8 that the time-domain and frequency-domain signals of these two states are very similar and hard to be visually distinguished. In contrast to the BP neural network and SVM, the CNN can well classify these two states.

Case study 2: Bearing fault diagnostics using CWRU dataset

Data description and preparation

The public bearing fault dataset from Case Western Reserve University (CWRU) (Case Western Reserve University Bearing Data Center Website, https://csegroups.case.edu/bearingdatacenter/home) is used to validate the proposed model in this section. A benchmark study of CWRU dataset was studied by (Smith and Randall 2015). As shown in Fig. 11, the test bench consists of a 2 hp motor (1 hp = 735 W), a torque transducer/encoder, a dynamometer, and control electronics. SKF-6202 deep groove ball bearings are used as the test bearings and support the motor shaft. The experiments were performed under four working conditions, as reported in Table 3. Four fault modes: "Outer race fault", "Inner race fault", "Ball fault" and "Normal" are introduced. For each fault mode, a single fault point with three severities levels, i.e., fault diameters of 0.007 mil, 0.014 mil, and 0.021 mil were seeded, which is regarded as a different fault mode. Therefore, there are 10 fault modes. Vibration data of each fault type under each working condition were collected using accelerometers, which were attached to the housing with magnetic bases. The sampling rate was 48 kHz.

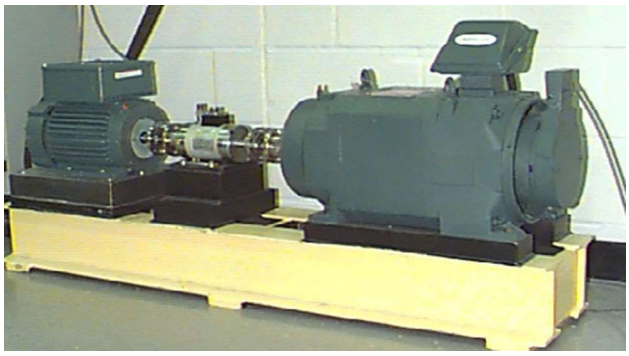

Fig. 11 Test bench of case study 2 [Case Western Reserve University Bearing Data Center Website, https://csegroups.case.edu/bearingdatacenter/home]

Table 3 Description of working conditions

Working condition	Motor load (hp)	Motor speed (rpm)
1	0	1797
2	1	1772
3	2	1750
4	3	1730

The number of data acquired of each fault type under each working condition are reported in Table 4. The amount of data provided is different for each working condition. To adapt this variation, the data preparation is adjusted. The training/testing ratio is set to be 4:1. For conditions 2–4, the number of data for each label is truncated to 4.8×10^5. 6.4×10^3 data points are segmented and further reshaped to (64, 100) matrix as one sample. Thus $4.8 \times 10^5 / 6.4 \times 10^3 = 75$ samples for each fault mode are obtained. For condition 1, the mode of the inner race damage with 0.014 mil is left out due to insufficient data. Indeed, we are aware of some data augmentation techniques such as adding noise or Generative Adversarial Network that may mitigate the problem of

Table 4 Description of CWRU bearing fault data

Fault mode	Fault label	Number of data points available in each working condition			
		Condition 1	Condition 2	Condition 3	Condition 4
Normal	1	243,938	483,903	483,903	485,643
Inner race fault with fault diameter 0.007 mil	2	243,938	486,224	485,643	485,643
Inner race fault with fault diameter 0.014 mil	3	63,788	489,125	487,964	485,063
Inner race fault with fault diameter 0.021 mil	4	244,339	485,063	491,446	491,446
Outer race fault with fault diameter 0.007 mil	5	243,538	486,804	486,804	487,964
Outer race fault with fault diameter 0.014 mil	6	245,140	484,483	486,804	488,545
Outer race fault with fault diameter 0.021 mil	7	246,342	489,125	487,964	489,125
Ball fault with fault diameter 0.007 mil	8	243,938	487,384	486,804	488,545
Ball fault with fault diameter 0.014 mil	9	249,146	486,224	487,384	486,804
Ball fault with fault diameter 0.021 mil	10	243,938	486,804	487,384	486,804

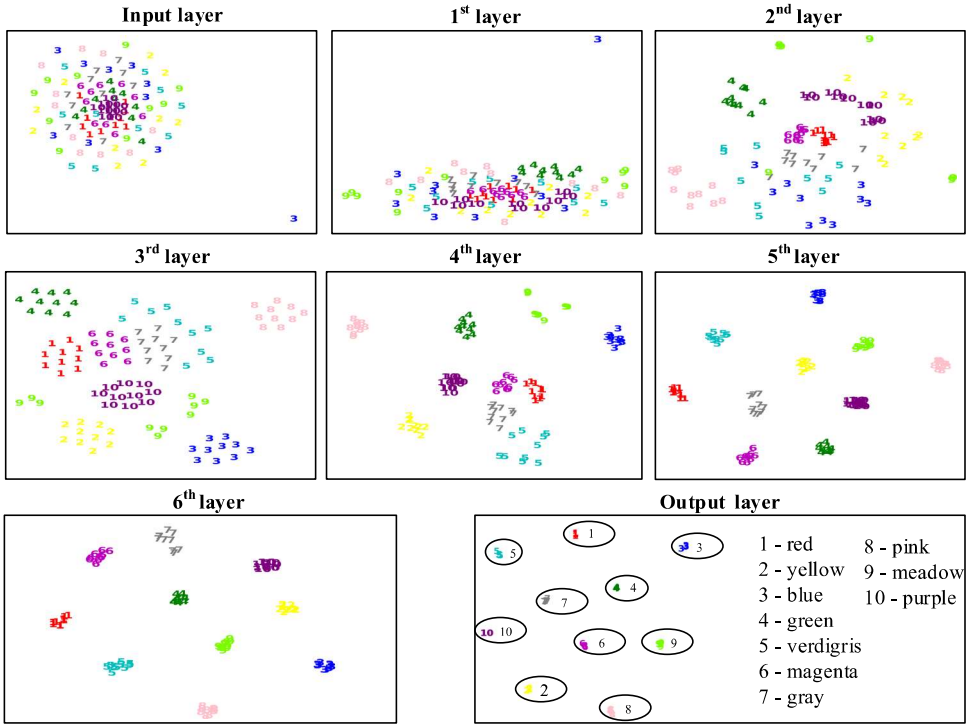

Fig. 12 Model testing process visualization under condition 2, case study 2

insufficient data, but such an investigation will be left for our coming work. The data of other modes is truncated to 2.4×10^5. 2.4×10^3 data points are segmented and further reshaped to (24, 100) as one sample. 100 samples are obtained for each fault label, and there are hence $80 \times 9 = 720$ samples for training and $20 \times 9 = 180$ samples for testing.

Results and discussions

For each condition the diagnostics accuracy on the test set is 100%. Due to the space limitation, only the feature learning process during testing of condition 2 are visualized by t-SNE, as shown in Fig. 12. The 10 symbols in different colours represents the 10 fault labels of condition 2. It can be seen that from the 5th layer, features of same fault mode have already been well aggregated and the features belonging to different modes have been well separated.

Case study 3: Bearing fault diagnostics with private dataset

Experiment and data preparation

In this case study, we validate the proposed model with the bearing fault dataset acquired from our own test bench, as shown in Fig. 13. Seven health states are considered, includ-

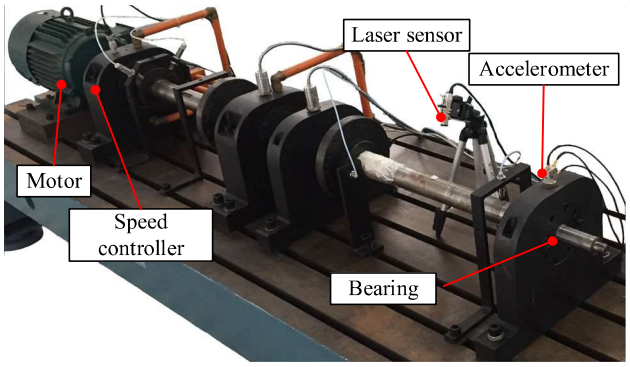

Fig. 13 Private test bench for bearing fault

ing the normal state, four types of single-point faults (i.e., inner race, outer race and ball), and three types of compound faults (i.e., inner race and ball, inner race and outer race, outer race and ball). Based on the literature and based on our observations and experiences, these faults are the most frequently occurs. The vibration data are collected from an NSK-6308 deep groove ball bearing in the experiment performed under three motor speeds 1500 rpm, 2000 rpm and 2500 rpm at the sampling rate 20 kHz. For each health state under each motor speed, the data acquisition lasts for 256 seconds, thus 5.12×10^6 data points are acquired. 6.4×10^3 data points are segmented and further reshaped to (64, 100) matrix as one sample. Thus 800 samples are obtained for each health state. The train/test ratio is set to 4:1. Figure 14 illustrates the vibration signal recorded in one second corresponding to the eight health states. For confidentiality reasons, the raw data is normalized to (− 1, 1).

Results and discussions

The diagnostics confusion matrices obtained on the test set for all three motor speeds are shown in Fig. 15, where the accuracies are nearly 100%. Labels 0–6 representes the following fault modes, i.e., **0**-ball, **1**-inner race, **2**-outer race, **3**-compound fault of inner race and ball, **4**-compound fault of outer race and ball, **5**-compound fault of outer race and inner race, **6**-normal.

Due to the space limitation, only the feature learning process during testing under the motor speed 1500 rpm is visualized by t-SNE, as shown in Fig. 16. It can be seen that in the output layer the features of same fault mode have been well aggregated and the features belonging to different modes have been well separated. Note that in the output layer, the samples belonging to "6" are totally aggregate. Very few confusions occur between "3" and "4", and between "4" and "5", which is consistent with the confusion matrix of Fig. 15a.

Case study 4: PHM 2009 spur gearbox challenge data

Data description and preparation

The 2009 PHM data challenge of gearbox fault data is used in this case study. Readers are referred to "PHM data challenge 2009, https://www.phmsociety.org/competition/PHM/09" for more information about the experiment setting. The overview of the apparatus is shown in Fig. 17a, including the drive system, a tachometer for providing zero-crossing information, the testing gearbox, and two accelerometers for collecting data. Two sets of gears, i.e., spur gears and helical gears were tested. We used the data of the spur gear since it contains more fault modes than that of the helical gears. The spur gearbox is a generic industrial one containing 3 shafts, 4 gears and 6 bearing, as shown in Fig. 17b. The teeth of the input gear, 1st idler gear, 2nd idler gear and the output gear are 32, 96, 48, and 80, respectively. Therefore, from input to output the gear reduction ratio is (16/48) × (24/40), or 5 to 1 reduction. For the gearbox, instead of single-point fault, eight types of compound faults caused by gear chipped, gear eccentric, bearing ball fault, shaft imbalance, shaft keyway fault, etc. are considered. The detail descriptions of the eight fault types are reported in Table 5. The faults were seeded in the experiments. These faults covered the common failures of gearboxes in real cases.

The experiments were carried out under 10 working conditions, i.e., 1800, 2100, 2400, 2700, 3000 rpm (revolutions per minute) shaft speed under high and low loading, respectively. For each fault type under each working condition, vibration signals were sampled synchronously from accelerometers mounted on both the input and output shaft retaining plates, as shown in Fig. 18. Data were acquired with a sampling rate 66.67 kHz and sampling time 4 s, and thus 266,655 data points are obtained and further truncated to 2.56×10^5 for each fault type. Additionally, for each working condition, the experiment was repeated twice.

For data preparation, 6.4×10^3 data points are segmented and reshaped to (64, 100) as one sample. Therefore, 80 samples are obtained for each label. We randomly draw 80% data (64 samples) for training and the 20% data (16 samples) for testing. Finally, the training/testing data taken from each label form the training sets (512 samples) and the testing sets (128 samples).

Results and discussions

For all the 10 working conditions, the testing accuracy of the eight types of faults are 100%. Due to space limitation, we only show the feature learning process during testing at working condition 2700 rpm under low loading as an example, as given in Fig. 19. The labels 1–8 represent the fault label as listed in Table 6.

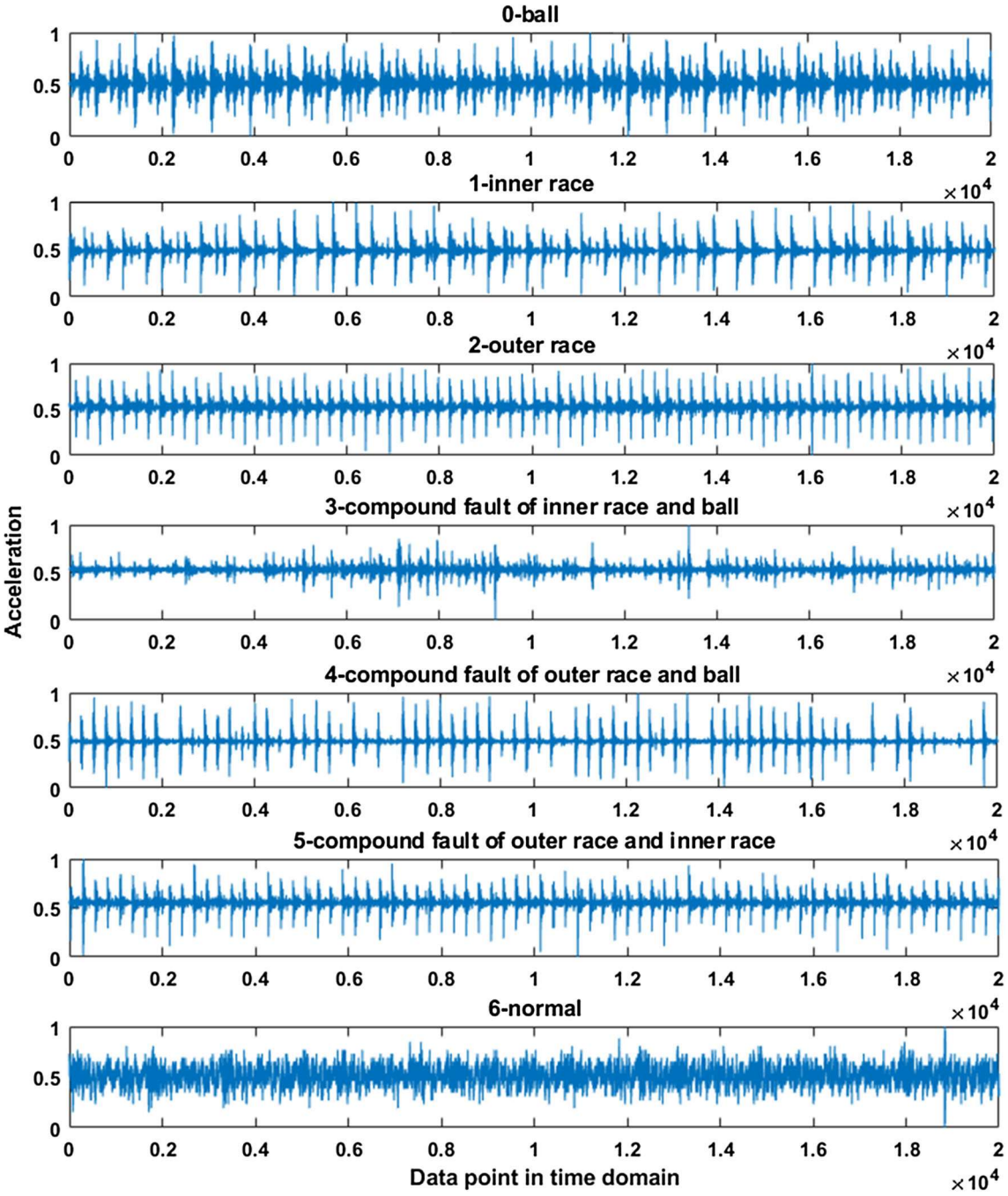

Fig. 14 Visualization of the raw vibration data of the eight heath states

The fivefold cross validation is used to evaluate the model. All samples and corresponding labels are randomly divided into five groups (the total number of samples in each group is the same). Each round four out of the five groups are used for training the model and the remaining is used for testing. By cross validation, the model has been tested five times and all the samples have the chance of being training/testing data. Since the samples are totally divided randomly, in each group the number of samples belonging to each label may imbalanced. For all the ten working conditions, the testing accuracies are 100%. We take working condition 2700 rpm shaft speed and low loading as an example and show the confusion matrices of the five testing results in Fig. 20.

Comparison with traditional diagnostics methods

The method of signal processing based feature extraction combined with a long short-term memory (LSTM) network

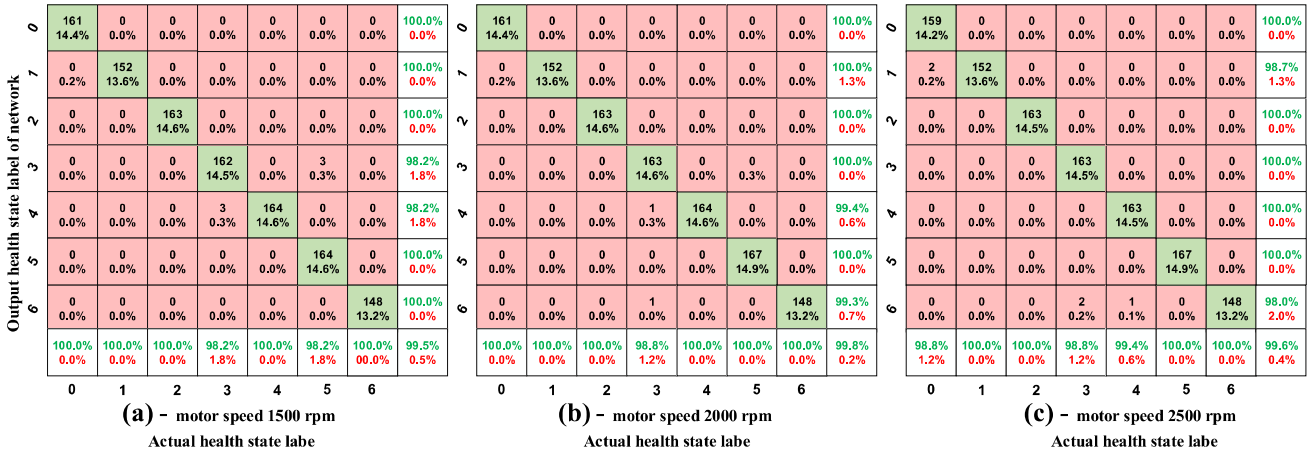

Fig. 15 Confusion matrix of case study 3 under three motor speed, given by proposed CNN model

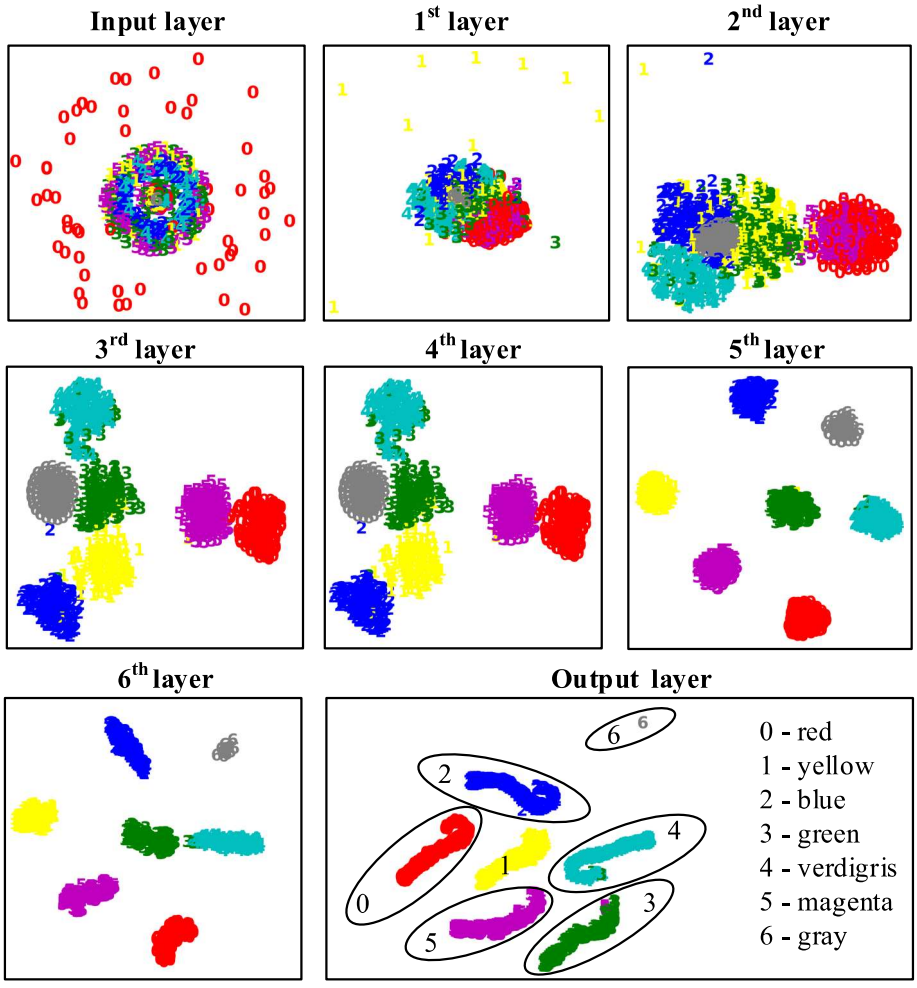

Fig. 16 Model testing process visualization under motor speed 1500 rpm, case study 3

as a classifier (which is referred to as traditional method from now on) is utilized to compare with the proposed.

CNN model in the above four case studies. The flowchart of the traditional method is shown in Fig. 21. The time domain signal of different health states is firstly divided into data segments. Then three manually extracted features, i.e., wavelet packet energy (WPE) based on wavelet packet decomposition (Zhang et al. 2013), instantaneous frequency (IF) (Boashash 1992a, 1992b) and instantaneous spectral entropy (ISE) (Pan et al. 2008) based on power spectrogram,

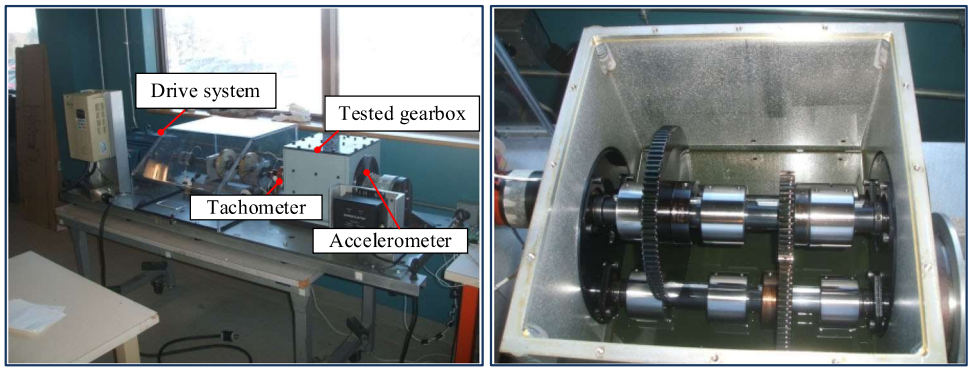

Fig. 17 Gearbox used in 2009 PHM data challenge [PHM data challenge 2009, https://www.phmsociety.org/competition/PHM/09]

Table 5 Fault modes description of 2009 PHM spur gears

Fault label	Fault description											
	Gear				Bearing						Shaft	
	32T	96T	48T	80T	IS:IS	ID:IS	OS:IS	IS:OS	ID:OS	OS:OS	Input	Output
1	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good	Good
2	Chipped	Good	Eccentric	Good	Good	Good	Good	Good	Good	Good	Good	Good
3	Good	Good	Eccentric	Good	Good	Good	Good	Good	Good	Good	Good	Good
4	Good	Good	Eccentric	Broken	Ball	Good	Good	Good	Good	Good	Good	Good
5	Chipped	Good	Eccentric	Broken	Inner	Ball	Outer	Good	Good	Good	Good	Good
6	Good	Good	Good	Broken	Inner	Ball	Outer	Good	Good	Good	Imbalance	Good
7	Good	Good	Good	Good	Inner	Good	Good	Good	Good	Good	Good	Keyway Sheared
8	Good	Good	Good	Good	Good	Ball	Outer	Good	Good	Good	Imbalance	Good

T teeth of the gear, *IS* input shaft, *ID* idler shaft, *OS* output side, *OS* output side, *IS* input side

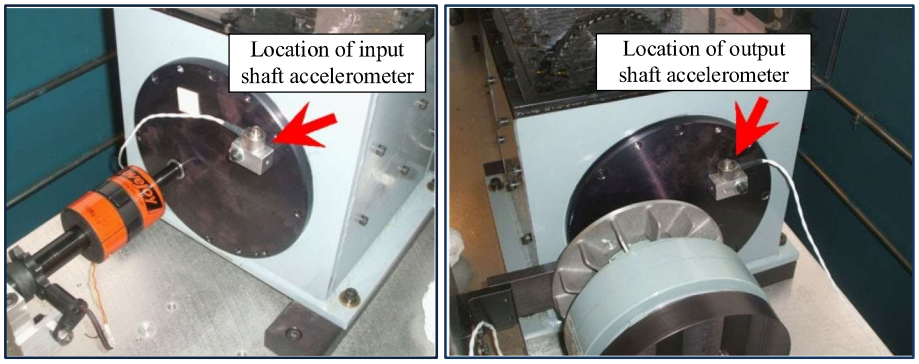

Fig. 18 The location of input and output shaft accelerometers [PHM data challenge 2009, https://www.phmsociety.org/competition/PHM/09]

are constructed from each data segment. The LSTM serves as the classifier. Therefore, three traditional methods, i.e., WPD-LSTM, IF-LSTM, ISE-LSTM are compared.

The architecture of the LSTM is composed of a sequence input, one LSTM layer, a fully connected layer and a softmax layer. The fully connected layer multiplies the input by the weight matrix and adds a bias vector. The output is finally calculated by a softmax transfer function. For the hyperparameters of LSTM, through initial trials we found that the number of LSTM units and batch size are two obvious parameters that affect the accuracy, given an appropriate learning rate. Specifically, we changed the number of LSTM units consecutively from 2^2 to 2^9 for the four case studies and found that a large amount of LSTM units are normally required when the number of training samples is large, and verse vice. For instance, in case study 1, where 240 training samples are available, a LSTM network with 128 units (or even fewer) performs better than that with 256 units, while in case study 2 in which 720 training samples are available, a LSTM network with 256 units are better than that with fewer units. In terms of batch size, we find that a smaller batch size tends to result in a higher accuracy but the training oscillation increases accordingly. In addition, too small batch sizes suffer the risk of non-convergence.

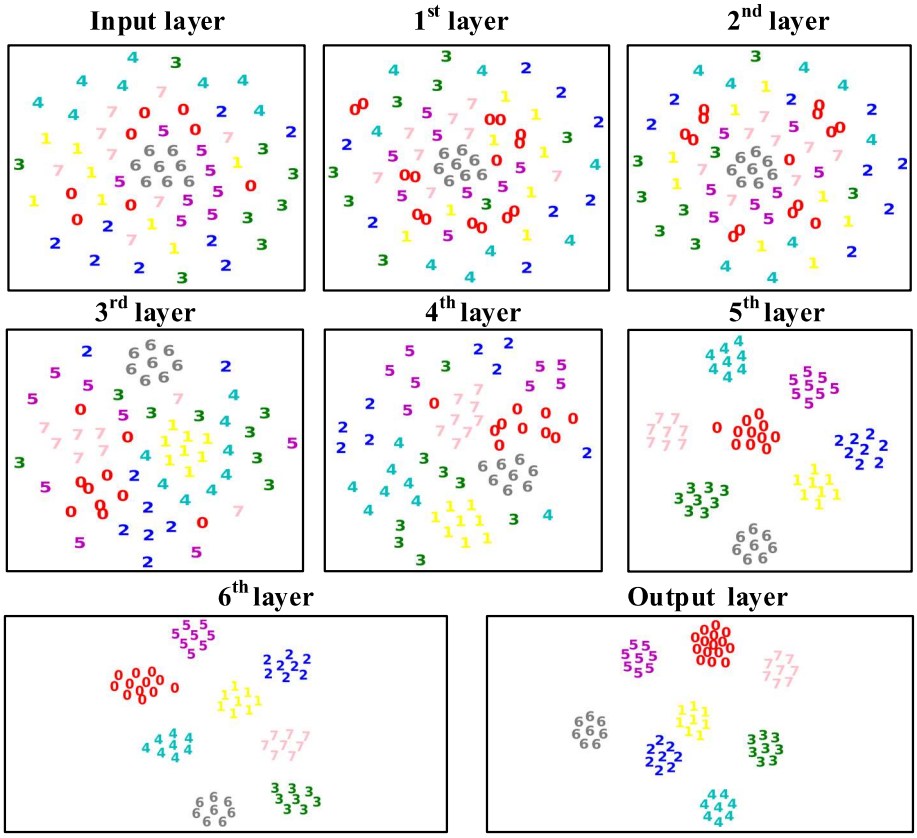

Fig. 19 Model testing process visualization under 2700 rpm and low loading, case study 4

Specifically, the accuracies with all three features of WPE, IF and ISE in case study 3 are high. Figure 22 illustrates the confusion matrices in case study 3 under motor speed 1500 rpm as an example. The accuracies of IF-LSTM and ISE-LSTM are acceptable in the application of gearbox under the low loading condition, but dramatically decrease under the high loading as well as in the case of ball screw. WPE, which performs the best among the three manually extracted features in many cases but suffers the risk of non-convergence in some working conditions of gearbox application. The confusion matrices given by the traditional methods for gearbox application under speed 2700 rpm and low loading are shown in Fig. 23 as an example. The accuracy given by WPE-LSTM is very low due to non-convergence. In contrast, the proposed CNN model can well identify the eight health states under this working condition, which can be clearly visualized in Fig. 19.

Through the comparison among the proposed CNN model and the traditional methods in various applications under various working conditions, it can be seen that the proposed CNN model exhibits much more robustness, giving consistently high accuracies in all four case studies. Moreover, the end-to-end structure of the CNN model requires less reliance on empirical expertise and advanced signal processing techniques, which enables the proposed model to be easily adapted to different diagnostics tasks.

Conclusions and future work

Manual feature extraction based on signal processing techniques is normally required in traditional diagnostics for rotating machinery, which has the drawbacks such as strong dependencies on the expertise and prior knowledge, the requirement for lots of skilled human labour, the sensitivity to changes, etc., and thus requires extensive fine-tuning. Some recent works based on deep learning convert the vibration signal to images based on some time-frequency methods, which can circumvent some of the previous drawbacks but still need application-specific adaptation. In this paper, we proposed an end-to-end health state diagnostics model based on convolutional neural network (CNN), which can directly learn feature representation from the raw vibration signal and no manually extracted feature is required. In addition, to fully validate the effectiveness and the generalizability of the proposed model for fault diagnostics of the rotating component, we carried out tests on four datasets, including two public ones and two datasets of our own, covering the applications of ball screw, bearing and gearbox. The results show high diagnostics accuracies for all the four tasks. To our best knowledge, our work firstly validates the CNN model in such wide applications.

Moreover, the signal processing based feature extraction combined with long short-term memory (LSTM) network

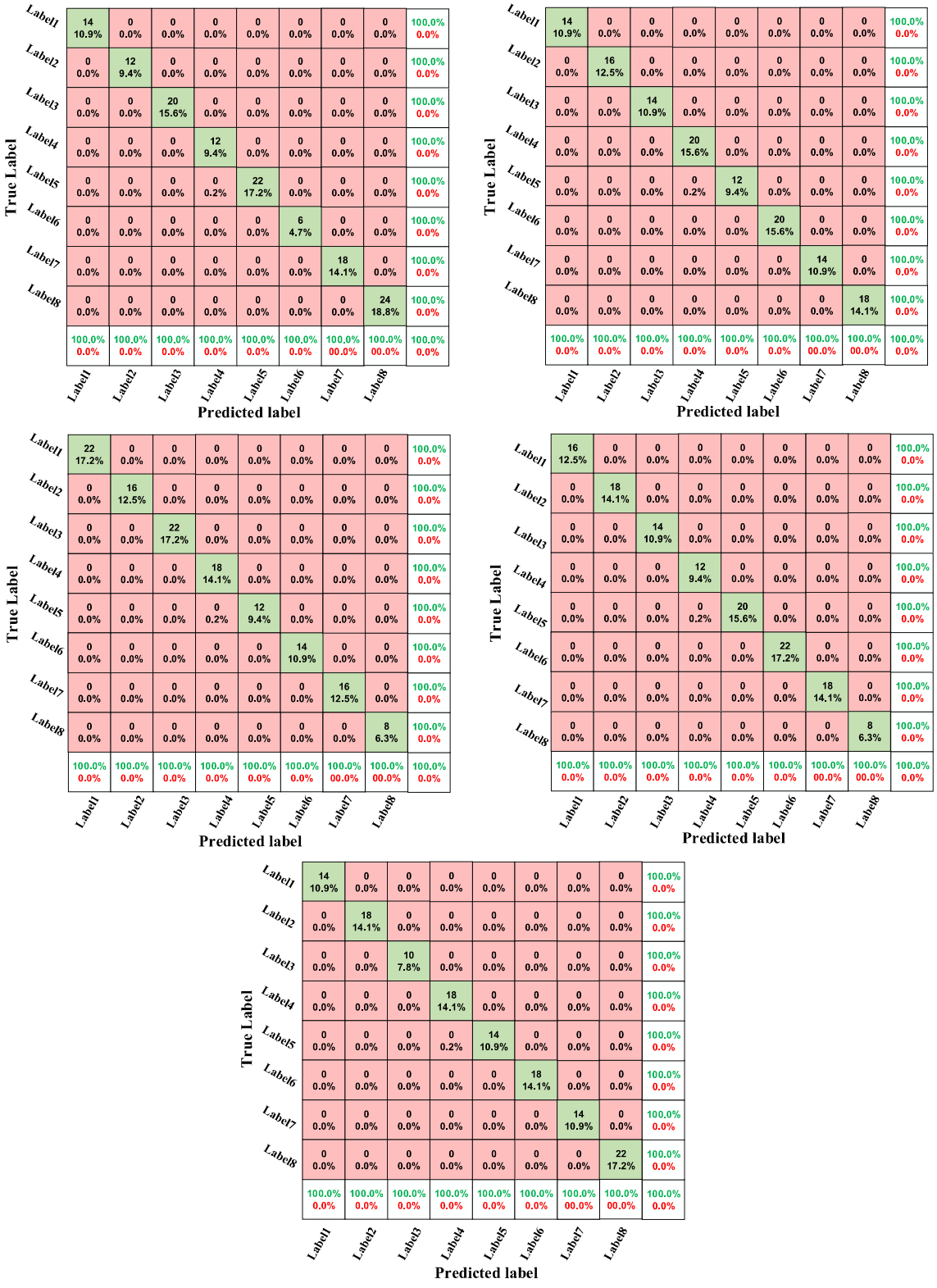

Fig. 20 Testing accuracies of fivefold cross validation under the working condition 2700 rpm shaft speed and low loading

(here is referred to as traditional method) is also explored and compared with the proposed CNN model. Specifically, three typical engineered features, i.e., (a) wavelet packet energy (WPE) based on wavelet packet decomposition, (b) instanta-

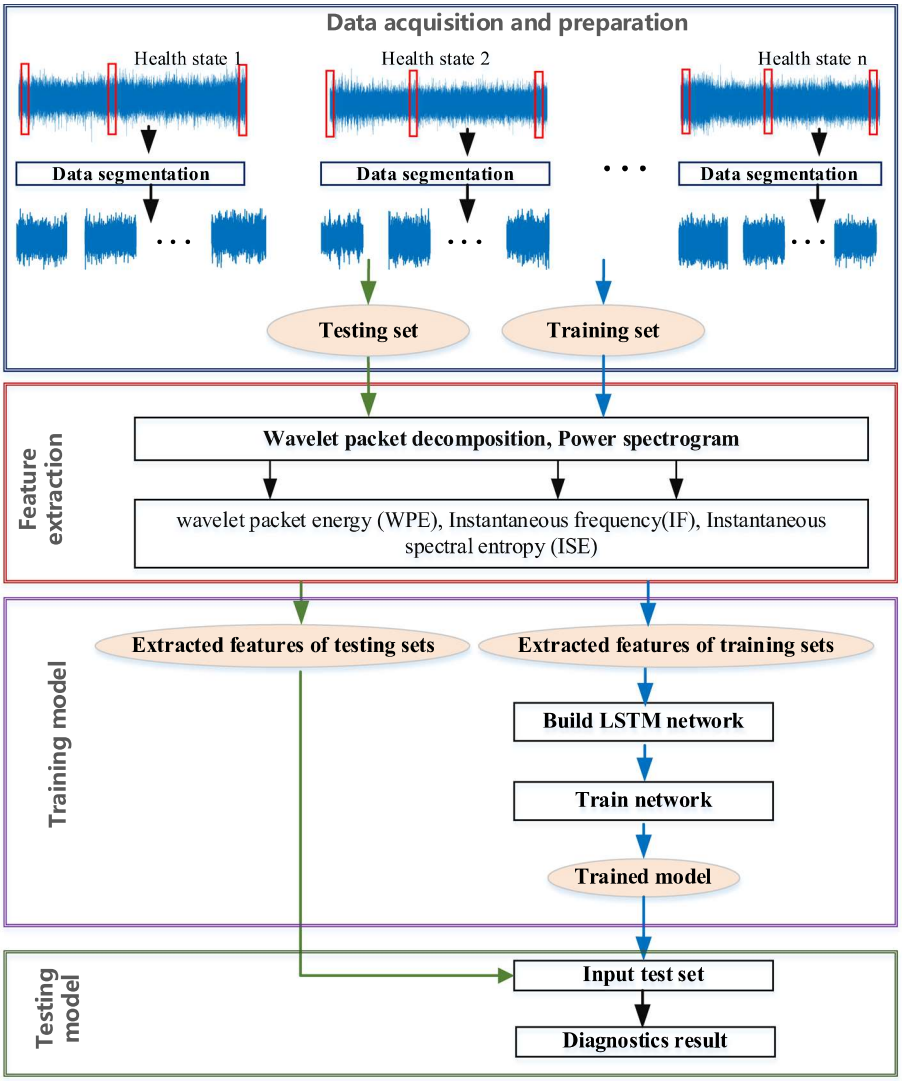

Fig. 21 Flowchart of the implementation of the traditional methods

Fig. 22 Confusion matrix of case study 3 under motor speed 1500, given by the traditional method

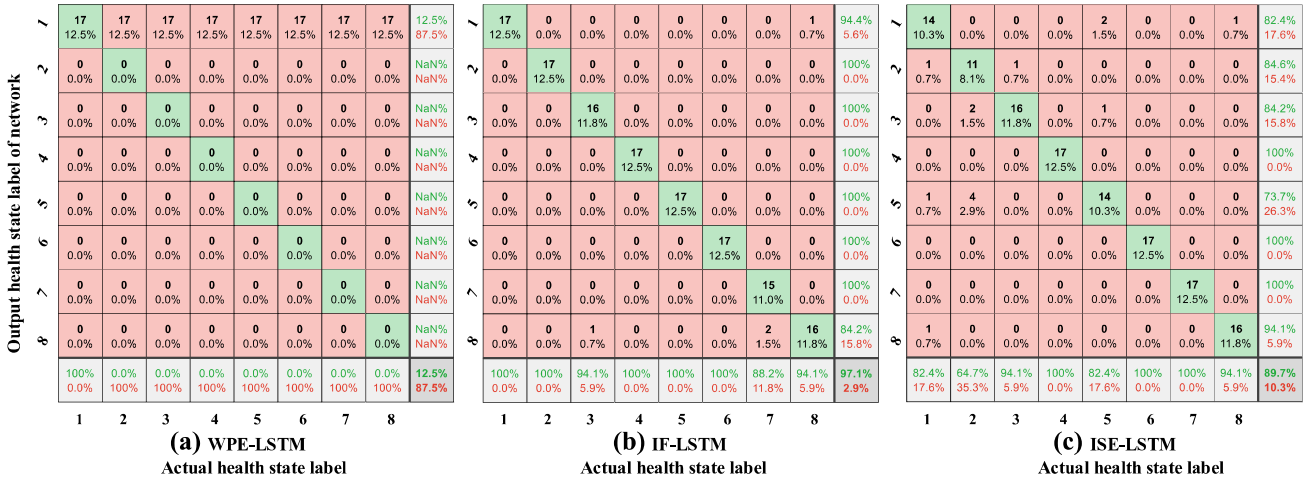

Fig. 23 Confusion matrix of case study 4 under speed 2700 rpm and low loading, given by the traditional method

Table 6 Accuracies of proposed CNN and traditional methods in all case studies

Case study	Working conditions	Number of labels	Proposed CNN model (%)	WPE-LSTM (%)	IF-LSTM (%)	ISE-LSTM (%)
1	500 rpm, 0 load	3	100	97.3 ± 3.2	86.8 ± 4.3	69.6 ± 5.6
2	1797 rpm, 0 load	9	100	79.4 ± 5.0	89.2 ± 2.5	90.3 ± 3.5
	1772 rpm, 1hp load	10	100	92.7 ± 5.2	86.4 ± 6.0	82.2 ± 4.5
	1750 rpm, 2hp load	10	100	82.2 ± 7.4	86.8 ± 1.8	88.7 ± 3.6
	1730 rpm, 3hp load	10	100	92.4 ± 7.0	86.8 ± 3.6	79.4 ± 4.6
3	1500 rpm, 0 load	7	99.5 ± 0.5	96.3 ± 2.1	97.5 ± 2.4	92.2 ± 2.5
	2000 rpm, 0 load	7	99.8 ± 0.5	95.4 ± 5.9	99.1 ± 2.3	92.3 ± 2.1
	2500 rpm, 0 load	7	99.5 ± 0.5	98.5 ± 3.4	94.6 ± 2.4	99.4 ± 2.1
4	1800 rpm, low load	8	100	93.4 ± 3.6	90.1 ± 2.3	81.9 ± 4.1
	1800 rpm, high load	8	100	91.0 ± 1.7	82.1 ± 7.3	84.5 ± 2.3
	2100 rpm, low load	8	100	99.5 ± 0.5	88.0 ± 8.4	89.4 ± 5.4
	2100 rpm, high load	8	100	–	83.5 ± 3.4	84.0 ± 2.1
	2400 rpm, low load	8	100	98.6 ± 1.7	89.9 ± 2.5	90.7 ± 1.8
	2400 rpm, high load	8	100	86.8 ± 2.4	82.2 ± 2.9	81.1 ± 2.7
	2700 rpm, low load	8	100	–	97.1 ± 1.9	89.7 ± 2.4
	2700 rpm, high load	8	100	92.2 ± 4.6	84.0 ± 1.6	76.9 ± 2.7
	3000 rpm, low load	8	100	95.2 ± 5.9	90.9 ± 4.0	91.0 ± 1.7
	3000 rpm, high load	8	100	96.1 ± 2.7	82.2 ± 1.6	76.8 ± 2.1

"–" represents non-convergence

neous frequency (IF), and (c) instantaneous spectral entropy (ISE) based on power spectrogram, are constructed from the raw vibration data and then used as the input of a classifier (LSTM network). The results indicate that manually extracted features based on signal processing techniques are indeed sensitive to diagnostics tasks. One feature performs well in one task but may fail to give satisfactory accuracy or lead to non-convergence in another task. The comparison shows that the proposed CNN based model has indeed good robustness and ability of generalization that is easy to adapt to different diagnostics task without any manual tuning.

The limits of the current work, and the corresponding future work are summarized as follows. The current work used the data acquired from the test benches of the laboratory. Next, we will investigate the performance of the proposed model in real industry environment. In the current work, the high diagnostics accuracy of each application is based on the assumptions that sufficient labeled data are available, and that the training and testing data are from the same distribution, which may be a limiting factor in industrial applications. To release these assumptions, our future work will focus on transfer learning methods, which are able to transfer

vibration-based diagnostics capabilities to new working conditions, experimental protocols and instrumented devices while avoiding the requirement for new labeled fault data. By this way, the diagnostics models trained with laboratory data have the potential of being used in the real industry environment. In the current work, the fault data of each label are balanced. In our future work, we will focus on building the diagnostics model when the fault data are unbalanced, i.e., only small fault data or even no fault data are available for some specific fault labels, since in practice, faults of high-stakes industrial devices are rare. In addition to the single fault type considered in the current work, we will study the fault diagnostics of compound faults. The issue of low signal to noise ratio in the acquired vibration signal caused by the strong coupling of different components is also our interests for future work.

Acknowledgements The present work was funded by the National Natural Science Foundation of China (No.51805262) and the Graduate Student Innovation Fund of Beihang University (YCSJ-03-2019-06). The authors gratefully acknowledge the Key Laboratory of Performance Test for CNC Machine Tool Components affiliated of Ministry of Industry and Information Technology of China for providing the ball screw test bench and experiment materials.

References

Behley, J., Steinhage, V., & Cremers, A. B. (2013). Laser-based segment classification using a mixture of bag-of-words. In *2013 IEEE/RSJ international conference on intelligent robots and systems* (pp. 4195–4200). https://doi.org/10.1109/IROS.2013.6696957.

Boashash, B. (1992a). Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals. Proceedings of the IEEE, 80(4), 520–538, doi:https://doi.org/10.1109/5.135376.

Boashash B (1992b). Estimating and interpreting the instantaneous frequency of a signal. Proceedings of the IEEE 80(4), 540–568, doi:https://doi.org/10.1109/5.135378.

Case Western Reserve University Bearing Data Center Website, Available: https://csegroups.case.edu/bearingdatacenter/home.

Chen, R., Huang, X., Yang, L., Xu, X., Zhang, X., & Zhang, Y. (2019). Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Computers in Industry, 106*, 48–59. doi:https://doi.org/10.1016/j.compind.2018.11.003.

Chen, Z., Mauricio, A., Li, W., & Gryllias, K. (2020). A deep learning method for bearing fault diagnosis based on Cyclic Spectral Coherence and Convolutional Neural Networks. *Mechanical Systems and Signal Processing, 140*, 106683. doi:https://doi.org/10.1016/j.ymssp.2020.106683.

Dhamande, L. S., & Chaudhari, M. B. (2018). Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. *Measurement, 125*, 63–77. doi:https://doi.org/10.1016/j.measurement.2018.04.059.

Feng, Z., Lin, X., & Zuo, M. J. (2016). Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis. *Mechanical Systems and Signal Processing, 72–73*, 223–240. doi:https://doi.org/10.1016/j.ymssp.2015.11.024.

Feng, G., & Pan, Y. (2012). Establishing a cost-effective sensing system and signal processing method to diagnose preload levels of ball screws. *Mechanical Systems and Signal Processing, 28*, 78–88. doi:https://doi.org/10.1016/j.ymssp.2011.10.004.

Goodfellow, I., Bengio, Y., & Courville, A. (2019). Deep learning. Cambridge, MIT Press.

Goyal, D., Choudhary, A., Pabla, B. S., & Dhami, S. S. (2019). Support vector machines based non-contact fault diagnosis system for bearings. *Journal of Intelligent Manufacturing*. doi:https://doi.org/10.1007/s10845-019-01511-x.

Hamadache, M., Jung, J. H., Park, J., & Youn, B. D. (2019). A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning. *JMST Advances, 1*(1), 125–151. doi:https://doi.org/10.1007/s42791-019-0016-y.

Hoang, D. T., & Kang, H. J. (2019). Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research, 53*, 42–50. doi:https://doi.org/10.1016/j.cogsys.2018.03.002.

Islam, M. M. M., & Kim, J. M. (2019a). Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines. *Reliability Engineering & System Safety, 184*, 55–66. doi:https://doi.org/10.1016/j.ress.2018.02.012.

Islam, M. M. M., & Kim, J. M. (2019b). Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Computers in Industry, 106*, 142–153. doi:https://doi.org/10.1016/j.compind.2019.01.008.

Jia, F., Lei, Y., Guo, L., Lin, J., & Xing, S. (2018). A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines. *Neurocomputing, 272*, 619–628. doi:https://doi.org/10.1016/j.neucom.2017.07.032.

Jing, L., Zhao, M., Li, P., & Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement, 111*, 1–10. doi:https://doi.org/10.1016/j.measurement.2017.07.017.

Kingma, D. P., & Ba, J. (2015). Adam: A method for Stochastic Optimization. the 3rd International Conference for Learning Representations, San Diego, 2015, arXiv preprint arXiv:1412.6980.

Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters, 18*, 20–23. doi:https://doi.org/10.1016/j.mfglet.2018.09.002.

Li, P., Jia, X., Feng, J., Davari, H., Qiao, G., Hwang, Y., et al. (2018a). Prognosability study of ball screw degradation using systematic methodology. *Mechanical Systems and Signal Processing, 109*, 45–57. doi:https://doi.org/10.1016/j.ymssp.2018.02.046.

Li, X., Li, J., Qu, Y., & He, D. (2019a). Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning. *Chinese Journal of Aeronautics*. doi:https://doi.org/10.1016/j.cja.2019.04.018.

Li, X., Li, J., Zhao, C., Qu, Y., & He, D. (2020). Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection. *Mechanical Systems and Signal Processing, 142*, 106740. doi:https://doi.org/10.1016/j.ymssp.2020.106740.

Li, X., Zhang, W., & Ding, Q. (2019b). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability Engineering & System Safety, 182*, 208–218. doi:https://doi.org/10.1016/j.ress.2018.11.011.

Li, X., Zhang, W., Ding, Q., & Sun, J. Q. (2018b). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. *Journal of Intelligent Manufacturing*. doi:https://doi.org/10.1007/s10845-018-1456-1.

Liang, P., Deng, C., Wu, J., & Yang, Z. (2020). Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network. *Measurement, 159*, 107768. doi:https://doi.org/10.1016/j.measurement.2020.107768.

Liu, L., Liang, X., & Zuo, M. J. (2018). A dependence-based feature vector and its application on planetary gearbox fault classification. *Journal of Sound and Vibration, 431*, 192–211. doi:https://doi.org/10.1016/j.jsv.2018.06.015.

Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning research, 9*(Nov), 2579–2605.

Ng, A. Y. (2004). Feature selection, L1 vs L2 regularization, and rotational invariance. In *Proceedings of the 21th international conference on machine learning*.

Nguyen, D., Kang, M., Kim, C. H., & Kim, J.-M. (2013). Highly reliable state monitoring system for induction motors using dominant features in a two-dimension vibration signal. *New Review of Hypermedia and Multimedia, 19*(3–4), 248–258. doi:https://doi.org/10.1080/13614568.2013.832407.

PHM data challenge. (2009). Available from https://www.phmsociety.org/competition/PHM/09.

Pan, Y. N., Chen, J., & Li, X. L. (2008). Spectral entropy: A complementary index for rolling element bearing performance degradation assessment. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 223(5), 1223–1231, doi:https://doi.org/10.1243/09544062JMES1224.

Park, S., Kim, S., & Choi, J. H. (2018). Gear fault diagnosis using transmission error and ensemble empirical mode decomposition. *Mechanical Systems and Signal Processing, 108*, 262–275. doi:https://doi.org/10.1016/j.ymssp.2018.02.028.

Peng, D., Liu, Z., Wang, H., Qin, Y., & Jia, L. (2019). A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access, 7*, 10278–10293. doi:https://doi.org/10.1109/ACCESS.2018.2888842.

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing, 64–65*, 100–131. doi:https://doi.org/10.1016/j.ymssp.2015.04.021.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutsjever, I., & Salakhutdinov, R. (2014). DropOut: A simple way to prevent neural network from overfitting. *Journal of Machine Learning research, 15*, 1929–1958.

Vogl, G. W., Weiss, B. A., & Helu, M. (2019). A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing, 30*(1), 79–95. doi:https://doi.org/10.1007/s10845-016-1228-8.

Wang, P., Ananya, Yan, R., & Gao, R. X. (2017). Virtualization and deep recognition for system fault classification. *Journal of Manufacturing Systems, 44*, 310–316. doi:https://doi.org/10.1016/j.jmsy.2017.04.012.

Wang, C., Gan, M., & Zhu, C. a. (2018a). Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *Journal of Intelligent Manufacturing, 29*(4), 937–951. doi:https://doi.org/10.1007/s10845-015-1153-2.

Wang, H., Li, S., Song, L., & Cui, L. (2019). A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. *Computers in Industry, 105*, 182–190. doi:https://doi.org/10.1016/j.compind.2018.12.013.

Wang, L., Liu, Z., Miao, Q., & Zhang, X. (2018b). Complete ensemble local mean decomposition with adaptive noise and its application to fault diagnosis for rolling bearings. *Mechanical Systems and Signal Processing, 106*, 24–39. doi:https://doi.org/10.1016/j.ymssp.2017.12.031.

Wu, C., Jiang, P., Ding, C., Feng, F., & Chen, T. (2019). Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Computers in Industry, 108*, 53–61. doi:https://doi.org/10.1016/j.compind.2018.12.001.

Xia, T., & Xi, L. (2019). Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems. *Journal of Intelligent Manufacturing, 30*(4), 1659–1672. doi:https://doi.org/10.1007/s10845-017-1342-2.

Yan, X., & Jia, M. (2018). A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing*. doi:https://doi.org/10.1016/j.neucom.2018.05.002.

Zhang, J., Sun, Y., Guo, L., Gao, H., Hong, X., & Song, H. (2020). A new bearing fault diagnosis method based on modified convolutional neural networks. *Chinese Journal of Aeronautics, 33*(2), 439–447. doi:https://doi.org/10.1016/j.cja.2019.07.011.

Zhang, Z., Wang, Y., & Wang, K. (2013). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *Journal of Intelligent Manufacturing, 24*(6), 1213–1227. doi:https://doi.org/10.1007/s10845-012-0657-2.

Zhao, X., Jia, M., & Lin, M. (2020). Deep Laplacian Auto-encoder and its application into imbalanced fault diagnosis of rotating machinery. *Measurement, 152*, 107320. doi:https://doi.org/10.1016/j.measurement.2019.107320.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237. doi:https://doi.org/10.1016/j.ymssp.2018.05.050.

Zhu, X., Hou, D., Zhou, P., Han, Z., Yuan, Y., Zhou, W., et al. (2019a). Rotor fault diagnosis using a convolutional neural network with symmetrized dot pattern images. *Measurement, 138*, 526–535. doi:https://doi.org/10.1016/j.measurement.2019.02.022.

Zhu, Z., Peng, G., Chen, Y., & Gao, H. (2019b). A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. *Neurocomputing, 323*, 62–75. doi:https://doi.org/10.1016/j.neucom.2018.09.050.