



# Bearing fault diagnosis base on multi-scale CNN and LSTM model

Xiaohan Chen<sup>1</sup> · Beike Zhang<sup>1</sup> · Dong Gao<sup>1</sup>

Received: 23 December 2019 / Accepted: 23 May 2020 / Published online: 5 June 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Intelligent fault diagnosis methods based on signal analysis have been widely used for bearing fault diagnosis. These methods use a pre-determined transformation (such as empirical mode decomposition, fast Fourier transform, discrete wavelet transform) to convert time-series signals into frequency domain signals, the performance of diagnostic system is significantly rely on the extracted features. However, extracting signal characteristic is fairly time consuming and depends on specialized signal processing knowledge. Although some studies have developed highly accurate algorithms, the diagnostic results rely heavily on large data sets and unreliable human analysis. This study proposes an automatic feature learning neural network that utilizes raw vibration signals as inputs, and uses two convolutional neural networks with different kernel sizes to automatically extract different frequency signal characteristics from raw data. Then long short-term memory was used to identify the fault type according to learned features. The data is down-sampled before inputting into the network, greatly reducing the number of parameters. The experiment shows that the proposed method can not only achieve 98.46% average accuracy, exceeding some state-of-the-art intelligent algorithms based on prior knowledge and having better performance in noisy environments.

**Keywords** Bearing fault diagnosis · Convolutional neural network · Recurrent neural network · Deep learning · Feature fusion

## Introduction

Rolling-element bearings, also known as rolling bearings, are common parts in rotating machinery, and rolling bearing faults can affect normal rotating machine operation, causing serious equipment damage, and economic costs, sometimes even human casualties (Wu et al. 2012; Li et al. 2015). Many studies have indicated that 40–50% of rotating machinery failures were directly caused by rolling bearing faults (Thorsen and Dalva 1999; Donnell and Heising 1987; Garcia-Perez et al. 2011). Therefore, it is particular important to diagnose rolling bearing status effectively, promptly, and accurately, and this problem has drawn significant attention from researchers and engineers over previous years (He and He 1987).

Bearing fault diagnosis based on its vibration signal can be generally divided into two parts: feature extraction and classification (Tian and Liu 2019). Bearing vibration, collected by sensors placed on the equipment, contains considerable information regarding the bearing status as well as noise. Therefore, we first need to extract valuable features characterizing intrinsic machine information from the disorganizing signal. Different faults cause different vibrational characteristics, and fault assessment based on a single feature is insufficient (Sun and Tang 2002; Jin and Sanchez-Sinencio 2015). A good fault diagnosis model should be robust and correctly identify the fault type in any state. However, time domain signals alone cannot accurately represent all bearing fault information, and for most cases we need to transform time domain into frequency or time-frequency domains, e.g. amplitude spectrum, power spectrum analysis, envelope spectrum analysis, short-time Fourier transform, empirical mode decomposition distribution, and s-transform (Mojiri et al. 2007; Petsounis and Fassois 2001; Lucchese and Cortelazzo 2000; Wu et al. 2017; Jantunen and Vaajoensuu 2010; Stridh et al. 2004).

Not all statistical features will be useful for diagnosis, many features not only increase computational burden but

✉ Dong Gao  
gaodong@mail.buct.edu.cn

Xiaohan Chen  
cxh\_bb@outlook.com

<sup>1</sup> College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China

also decrease fault classification accuracy. Therefore, it is necessary to select suitable features from high dimensional features and input only these relevant features into the classifier. Current popular dimension reduction methods include independent component analysis (ICA) (Widodo and Yang 2007), principal component analysis (PCA) (Humberstone et al. 2012), and locally linear embedding (LLE) (Li et al. 2010). Once useful features are selected, these are input to train the classifier, e.g. support vector machine (SVM) (Saravanan et al. 2008), artificial neural network (ANN) (Zhang et al. 2013) naïve Bayes (NB) (Aydin et al. 2015), etc. These methods have achieved satisfactory results for diverse datasets, with some subsequently applied in industrial situations (Wu et al. 2012).

Artificial intelligence (AI) has been widely applied for pattern recognition, with remarkable achievements for image processing (Xu et al. 2016), natural language processing (Luong et al. 2015) and medical diagnosis (Chauhan et al. 2019). Fault diagnosis using AI was first attempted in the 1990s (Gao et al. 2015), and has subsequently developed rapidly. Shuang and Meng (2007) proposed a rolling bearing fault diagnosis method based on PCA and SVM. High-dimensional signal characteristics were reduced used PCA, and then SVM was applied to identify the failure pattern. Wang et al. (2018) found that if conduct sparse coding with discriminative dictionary, different residual will be generated, inspired this, they introduced sparse representation to extract feature from wavelet packet transform. Van and Kang (2015) introduced the particle swarm algorithm (PSO) to select positive features, and then followed a proposed decision fusion strategy to classify multiple faults using SVM. ANN have been widely used for classification because they can fit nonlinear models. Amar et al. (2015) used a two dimensional filter to suppress frequency amplitude caused by instantaneous vibration, and used spectrum images generated from Fourier transform as input for the ANN. Ramos (2019) used Kernel Fuzzy C-Means (KFCM) to identify the type of failures, and reducing the classification errors. Ant Colony Optimization optimization algorithm was used to train KFCM for improving the diagnosis results.

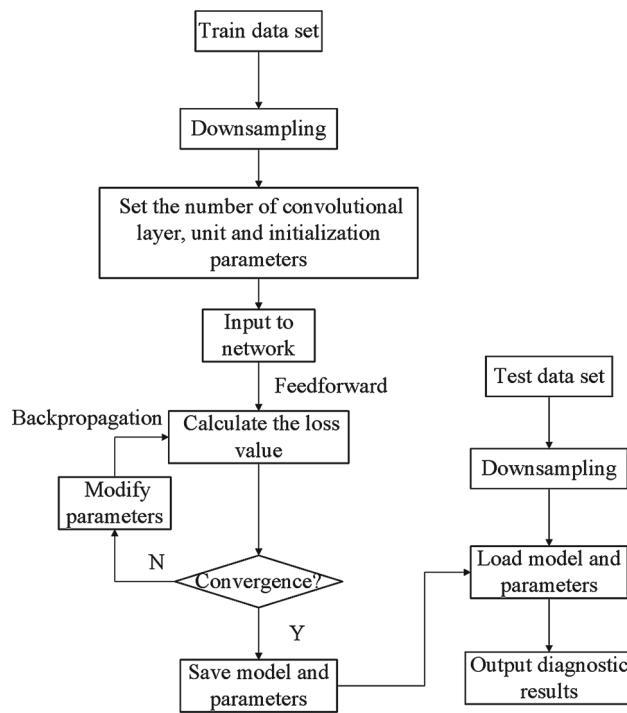
Recently, deep learning approaches have promoted rapid AI development, helped by dramatically increased computing power to allow processing large datasets very quickly. Yu et al. (2019) proposed a hierarchical deep learning algorithm with 3 long short-term memory (LSTM) modules and a dropout layer for each module. Tian and Liu (2019) proposed a deep convolutional neural networks (CNN) for real-time roller bearing damage detection, comprising 3 convolutional, 3 pooling, and 1 fully connected layer, with the softmax function for final fault prediction. Ali (2015) used an autoencoder to extract positive features. They used 4 datasets collected from 4 sensors placed in different positions

as autoencoder inputs, and then the extracted features were input to a deep belief network (DBP). The proposed network was superior to other intelligent algorithms. Narendiranath Thamba and Aravind (2018) used empirical mode decomposition (EMD) and Hilbert Huang transform (HHT) to calculate the instantaneous frequency peak, then applied an ANN with 10 hidden layers for classification. Yi et al. (2019) employed the wavelet packet transform (WPT) to dynamically decompose the vibration signal, and selected the bandpass signal with the largest negentropy as training data for dictionary learning. Rajeswaran and Madhu (2011) used genetic algorithm to select the characteristic parameters of the ANN to diagnose induction motor fault. Shao et al. (2016) proposed a deep belief network (DBN)-based method to automatically learn characteristics from vibration data that reflect the working condition of an induction motor.

These various algorithms and methods discussed above have achieved satisfactory results, and some of them have been practically applied in industrial situations. However, several limitations remain.

1. Mapping between vibration signal and fault type is complex. Most current studies use signal processing methods, e.g. EMD, HHT and WPT, to extract features. Hence, the fault diagnosis effectiveness relies heavily on manually extracted feature quality, and intelligent diagnostic method advantages have not been exploited.
2. Manual feature extraction is a very laborious and time-consuming task, and usually requires considerable core knowledge related to signal processing and mathematics.
3. Generally, ANN and SVM network structures are shallow, which limits their ability to fit complex nonlinear signals. Although increasing the number of hidden layers can extract more useful features, it also significantly increases the computational burden.
4. Most previous studies consider only a small subset of fault types, typically 3–6 types. Thus, these methods fail when the number of faults increase, or different fault types occur (as commonly happens for real failure cases).

This paper proposes the Multi-scale Convolutional Neural Network and Long Short-Term Memory (MCNN-LSTM) fault diagnosis model to address the problems discussed above. MCNN-LSTM comprises feature extractor and classifier, which allows raw data to be directly input into the model without preprocesses. The feature extractor includes two CNNs with different kernel size, to automatically extract characteristic representations of roller bearing fault vibrational signals. The extracted features are then fed into the classifier, a stacked LSTM network, for bearing fault assessment. We down-sample the sensor data before inputting the

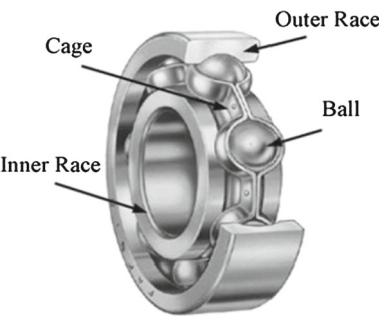
**Fig. 1** Block diagram of the proposed method

raw data to reduce computing power requirements and the number of required network parameters. The overall block diagram for the system is shown in Fig. 1. The proposed method offers the following advantages.

1. Compact network structure and raw data input, allowing bearing status detection in real time.
2. Automatically learns features from raw signal without preprocessing (such as EMD, HHT, etc.).
3. Provides effective training and classification methods with small datasets.
4. Accurately identifies 10 different fault type with high accuracy.

Experimental results confirm that the proposed model outperformed some current state-of-the-art intelligent fault diagnosis systems and having better performance in noisy environments.

The remainder of this paper is arranged as follows. “Theoretical background” section briefly introduces the rolling bearing fault diagnosis problem, CNN and LSTM networks. “Proposed multi-scale CNN and LSTM model” and “Validating the proposed model” sections describe the proposed MCNN-LSTM framework, and a series of experiments to verify the proposed method classification accuracy, respectively. We visualize key experimental results to explore the proposed system’s internal mechanisms. Finally, “Conclusions” section summarizes and concludes the paper.

**Fig. 2** Typical rolling bearing structure

## Theoretical background

### Rolling bearing fault signature

Figure 2 shows the typical roller bearing structure comprises an outer race mounted in the bearing housing, an inner race mounted on the rotating shaft, rolling elements, and a supporting cage (Zhang et al. 2020). Roller bearings are the most vulnerable rotating machine parts. Any damage will cause the bearing to degrade rapidly due wear and tear with the internal balls. Many underlying reasons cause bearing failure, including inappropriate size selection, electrical leakage, excessive load, etc. Periodic abnormal vibrations are produced once faults occur on a bearing, with magnitude determined by the fault type. Fault vibration-frequency characteristics depend on fault size, shaft speed, the load, and fault location (Yu et al. 2018), with some typical vibration frequencies for different mechanical fault characteristics as follows:

$$f_c = \frac{f_r}{2} \left( 1 - \frac{d}{D_m} \cos \alpha \right) = \frac{f_{bi}}{N_b} \quad (1)$$

$$f_{be} = \frac{f_r}{2} N_b \left( 1 - \frac{d}{D_m} \cos \alpha \right) \quad (2)$$

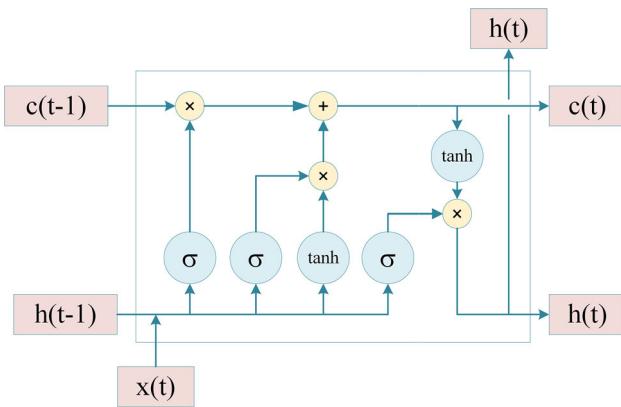
$$f_{bi} = \frac{f_r}{2} N_b \left( 1 + \frac{d}{D_m} \cos \alpha \right) \quad (3)$$

$$f_b = \frac{f_r D_m}{2d} \left( 1 - \left( \frac{d}{D_m} \cos \alpha \right)^2 \right) \quad (4)$$

where  $D_m$  is the pitch diameter;  $d$  is the roller diameter;  $N_b$  is the number of rollers;  $\alpha$  is the ball contact angle; and  $f_c$ ,  $f_{bi}$ ,  $f_{be}$ , and  $f_b$  are the cage, inner race, outer race and roller fault frequencies, respectively.

### Convolutional neural network

Convolutional Neural Network (CNN) is a kind of feed-forward neural network with convolution computation and deep structure and is widely used in image processing and natural language processing. Each CNN hidden layers includes a convolutional and pooling layer. The convolutional layer



**Fig. 3** LSTM neuron internal structure

extracts characteristics from input signal using a filter with shared weights to map the local signal of the previous layer to the next layer. Shared weights provide significant advantages, greatly reducing computational burden for complex nonlinear transformations. One dimensional convolution was used in this study since the vibration signal is one dimension time series,

$$y_i^{l+1}(j) = K_i^l \times x^l(j) + b_i^l \quad (5)$$

where  $k$  and  $b$  are the weight and bias for the  $i$ -th filter and  $l$ -th layer respectively, and  $x$  denotes the  $j$ -th local input in  $l$ -th layer.

We apply a pooling layer behind the convolutional layer to select and filter extracted features. Max-pooling is the most common pooling operation, which selects the largest statistic to obtain local features of this region and reduce the number of parameters. The result in  $(l+1)$ -th layer of  $i$ -th channel after pooling can be described as:

$$P_i^{l+1}(j) = \max_{(j-1)W+1 \leq t \leq jW} \{q_i^l(t)\} \quad (6)$$

where  $q$  denotes the  $t$ -th neuron in the  $l$ -th layer of the  $i$ -th channel,  $w$  is the width of pooling kernel.

### Long short-term memory

Long Short-Term Memory (LSTM) is the improved traditional recurrent neural network (RNN), which can capture the entire history information of input data (Yu et al. 2019). However, RNN has some shortcomings, including possible gradient disappearance or gradient explosion in backpropagation. LSTM solves these problems by adding input, output, and forget gates. Figure 3 shows the internal structure of LSTM neuron. The main idea of LSTM is that several gates control the update of the information flow along the time axis to obtain information at each time step. We introduce

the input state to determine whether the input  $x^t$  and hidden state of previous layer  $h^{t-1}$  should be added to current cell state. The forget gate is related to the previous hidden state and current input, which determines if the cell value should be retained or thrown away. The output of current layer are based on the input to this neuron and the output from previous time. LSTM neurons will output a list of state responses at each step, which contains the previously output information and the currently input information. The memory cell ensures the gradient can be passed to many time steps without gradient disappearing or exploding. Therefore, LSTM overcomes the difficulty caused by the vanishing gradient effect during training process. The input, forget, output, cell, and hidden gate updates can be expressed as follows.

$$i^t = \sigma(W^i x^t + V^i h^{t-1} + b^i) \quad (7)$$

$$f^t = \sigma(W^f x^t + V^f h^{t-1} + b^f) \quad (8)$$

$$o^t = \sigma(W^o x^t + V^o h^{t-1} + b^o) \quad (9)$$

$$c^t = f^t \odot c^{t-1} + i^t \odot \tanh(W^c x^t + V^c h^{t-1} + b^c) \quad (10)$$

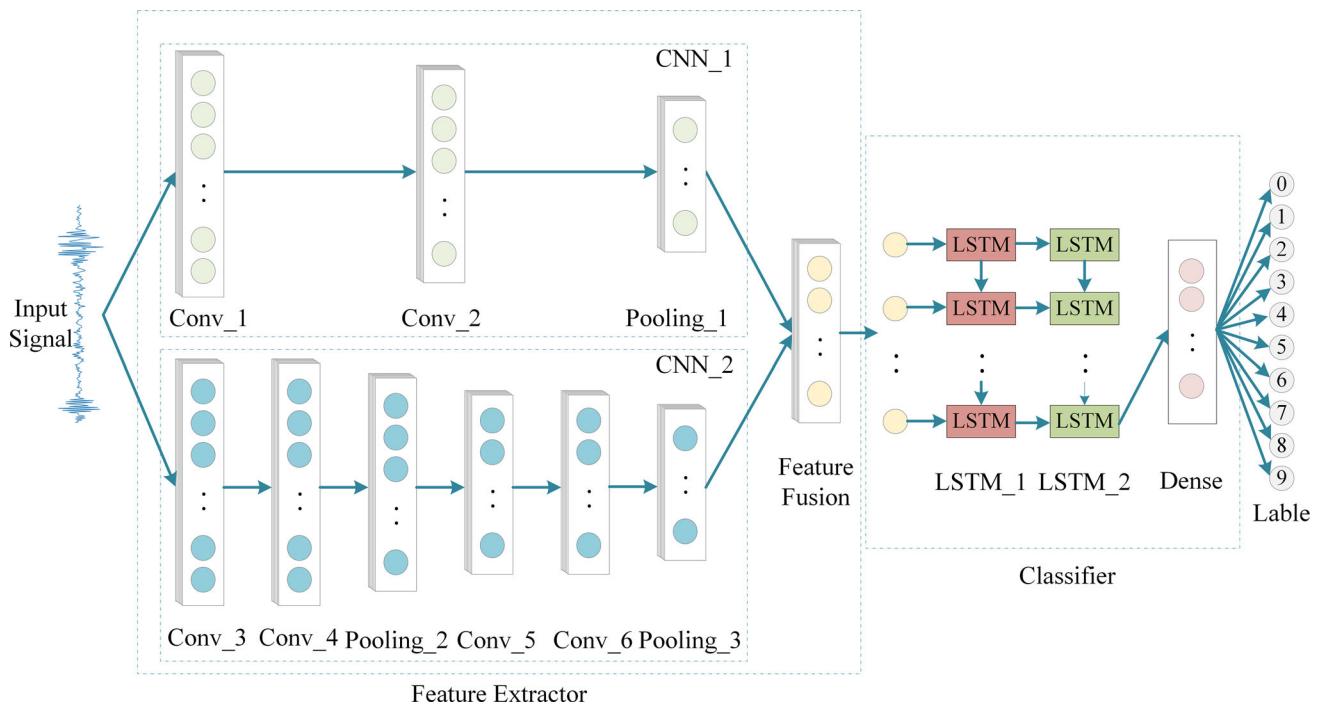
$$h^t = o^t \odot \tanh(c^t) \quad (11)$$

where  $b$  denotes the biases; and  $W$  and  $V$  denote the input and hidden state weights, respectively. In  $t$ -th update step, input gate,  $i$ , forget gate,  $f$ , output gate,  $o$ , and cell state,  $c$  are updated by input  $x$  and hidden state of step  $n-1$ .

### Proposed multi-scale CNN and LSTM model

This study aimed at input raw vibration signals directly into the neural network allowing the neural network to extract inner characteristic representations from the raw signal and hence detect the roller bearing health state without pre-processing. Figure 4 shows that the MCNN-LSTM model includes feature extractor and classifier modules. The proposed model was inspired by the traditional CNN-LSTM model (Sainath et al. 2015) with some critical improvements.

We down-sample the signal before submitting it to the proposed model to improve calculation speed and performance. Raw signals are directly fed into the model for automatic feature extraction and failures classification. The first module (feature extractor) comprises two one-dimensional CNNs with different kernel sizes and depths. Raw signals are simultaneously fed into CNNs to extract features in different frequency domains. We employed a large receptive field ( $20 \times 20$  and  $10 \times 10$ ) for CNN\_1 to automatically extract low frequency features, whereas CNN\_2 extracts features from high frequency signals, hence uses a smaller receptive field ( $6 \times 6$ ). Feature vectors from CNN\_1 and CNN\_2 are fused by element-wise product. The 1D CNN offer several advantages as feature extractors: they automatically learn underlying meanings for different vibration signals; high



**Fig. 4** Proposed multi-scale CNN and LSTM model

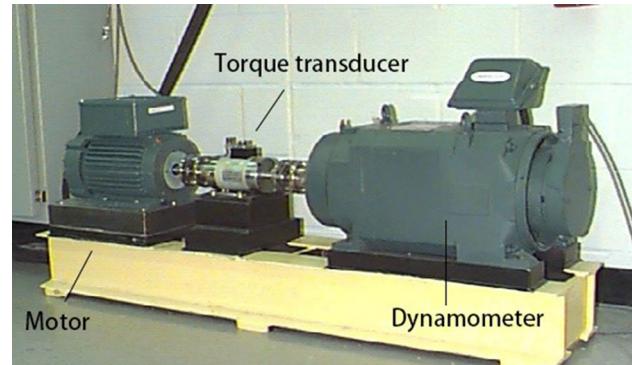
dimensional input vectors are considerably reduced using the shared weight strategy, also reducing the number of parameters; and outputs from previous layers and inputs for the next layer are limited by the kernel size.

The second module, classifier, comprises a hierarchical LSTM and fully connected layer, to construct complex non-linear models between the input and output. Figure 4 shows how LSTM<sub>1</sub> hidden states provide input for LSTM<sub>2</sub>, and LSTM<sub>2</sub> outputs are fed into a fully connected layer for classification. The input is divided into several steps to the LSTM network. The output of the next step is affected by the output of the previous step. Therefore, the LSTM network makes good use of the characteristics of the timing signal and fully extracts the internal characteristics of the vibration signal. Finally, the softmax function converts neuron outputs into a probability distribution over 10 roller bearing health states. The softmax function is described as follow. where  $z$  denotes the output from the  $j$ -th neuron.

$$q(z_j) = \text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{10} e^{z_k}} \quad (12)$$

## Validating the proposed model

As discussed above, roller bearings are important rotating machinery components. Therefore it is important to diagnose roller bearing condition in a timely and accurate manner. This

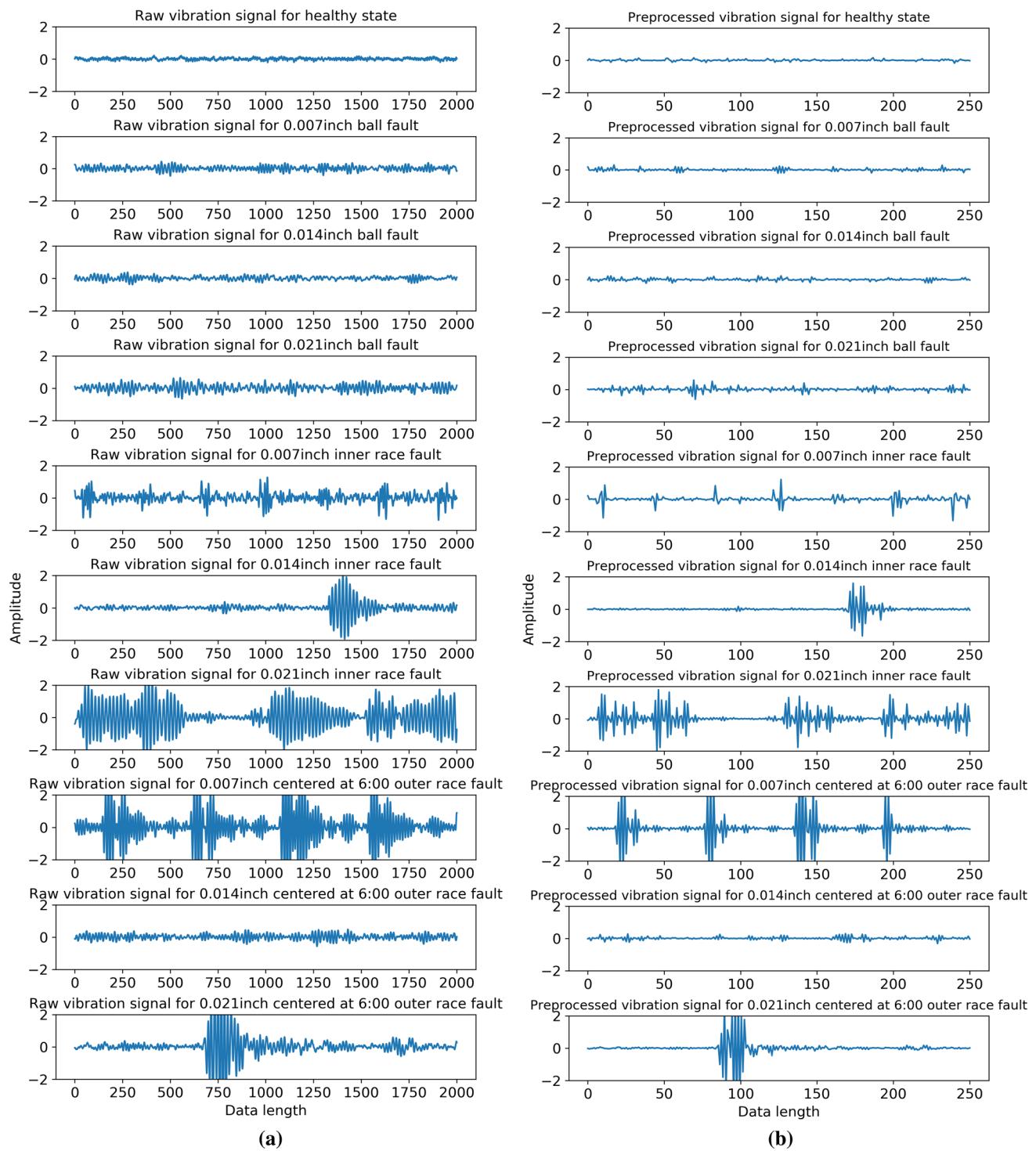


**Fig. 5** Experiment platform for rolling bearing fault sensing

study used the Case Western Reserve University (CWRU) roller bearing data set to validate the proposed MCNN-LSTM. We first trained the network using a training dataset comprising 9 different faults and 1 normal signal, i.e., 10 different patterns in total. Test results show that the MCNN-LSTM achieved 98.46% accuracy. We then visualized the hidden layers from the model to help understand the neural network reasoning, and finally compared the proposed network outcomes with current state-of-the-art models.

## Datasets

Figure 5 shows the experimental platform employed to generate rolling bearing fault data. Deep groove ball bearings (6205-2RS JEM SKF) were placed on the shaft of a 1.5 kW



**Fig. 6** Raw vibration signals for 10 bearing states **a** before down-sampling and **b** after down-sampling

**Table 1** Generated dataset details

| Bearing fault    | Label | Defect size (inch) | No. samples | Data length | Load zone | Accelerometer position | Shaft load (hp) | Shaft speed (rpm) |
|------------------|-------|--------------------|-------------|-------------|-----------|------------------------|-----------------|-------------------|
| Ball fault       | 0     | 0.007              | 240         | 2000        | —         | Drive end              | 3               | 1730              |
| Ball fault       | 1     | 0.014              | 240         | 2000        | —         | —                      | —               | —                 |
| Ball fault       | 2     | 0.021              | 240         | 2000        | —         | —                      | —               | —                 |
| Healthy state    | 3     | —                  | 240         | 2000        | —         | —                      | —               | —                 |
| Inner race fault | 4     | 0.007              | 240         | 2000        | —         | —                      | —               | —                 |
| Inner race fault | 5     | 0.014              | 240         | 2000        | —         | —                      | —               | —                 |
| Inner race fault | 6     | 0.021              | 240         | 2000        | —         | —                      | —               | —                 |
| Outer race fault | 7     | 0.007              | 240         | 2000        | 6:00      | —                      | —               | —                 |
| Outer race fault | 8     | 0.014              | 240         | 2000        | 6:00      | —                      | —               | —                 |
| Outer race fault | 9     | 0.021              | 240         | 2000        | 6:00      | —                      | —               | —                 |

**Table 2** MCNN-LSTM model parameters applied

| Name      | Filters | Kernel size/stride | Units | Input size | Output size | Activation function | No. of trainable parameters |
|-----------|---------|--------------------|-------|------------|-------------|---------------------|-----------------------------|
| Conv_1    | 50      | 20/2               |       | 1 × 250    | 50 × 116    | tanh                | 1050                        |
| Conv_2    | 30      | 10/2               |       | 50 × 116   | 30 × 54     | tanh                | 15,030                      |
| Pooling_1 |         | 2/2                |       | 30 × 54    | 30 × 27     |                     | 0                           |
| Conv_3    | 50      | 6/1                |       | 30 × 54    | 50 × 245    | tanh                | 350                         |
| Conv_4    | 40      | 6/1                |       | 60 × 245   | 40 × 240    | tanh                | 12,040                      |
| Pooling_2 |         | 2/2                |       | 40 × 240   | 40 × 120    |                     | 0                           |
| Conv_5    | 30      | 6/1                |       | 40 × 120   | 30 × 115    | tanh                | 7230                        |
| Conv_6    | 30      | 6/2                |       | 30 × 115   | 30 × 55     | tanh                | 5430                        |
| Pooling_3 |         | 2/2                |       | 30 × 55    | 30 × 27     |                     | 0                           |
| LSTM_1    |         |                    | 60    | 30 × 27    | 30 × 60     | relu                | 21,120                      |
| LSTM_2    |         |                    | 30    | 30 × 60    | 1 × 30      | relu                | 10,920                      |
| Dense     |         |                    | 10    | 1 × 30     | 1 × 10      | softmax             | 310                         |
|           |         |                    |       |            |             | Total               | 73,480                      |

motor. The middle module was a torque sensor to collect motor speed and horsepower data. The final module was a dynamometer to measure power. There were also various control electronics not specifically shown in Fig. 5.

Training and test data were derived from accelerometers placed at drive end (DE) sampled at 48 kHz. Motor power and rotation were fixed at 3 hp and 1730 rpm, respectively (Yu et al. 2018). Three bearing failure types were considered: rolling, inner race, and outer race faults. Each fault type was induced in the rolling bearings by electro-discharge with different fault diameters (0.007, 0.014, and 0.021 inch). Thus, there were 10 faults in total: 3 fault types × 3 defect sizes + 1 normal signal; each with different and characteristic vibration signals.

Figure 6a shows typical raw vibration signals for the 9 fault (ball fault, inner race fault and outer race fault, each fault type contains three kinds of fault size, 0.007, 0.014, and 0.021 inch) and normal states. We selected partial vibration acceleration signals of 10 bearing states and plotted them. It can be seen from the figures that it is difficult to identify the type of failure from the original data without the help of data processing knowledge. To solve this problem, an automatic feature learning method based on neural networks will be introduced. We down-sampled the signals before input into MCNN-LSTM to improve computation efficiency and reduce the number of network parameters, selecting one datapoint of 8 for the input signal, 250 datapoints per sample, as shown in Fig. 6b. It can be seen from the figure that the down-sampled data still reflects the characteristics of the input data, But the amount of input data is greatly reduced. To ensure effective and robust of proposed model, each bearing state (including healthy) included 240 samples acquired from accelerometers in the time domain, split into 70% for training and 30% for testing. Thus, we provided 1680 training

and 720 test samples in total, where each sample contained 2000 consecutive vibration signals. Table 1 shows the data set details.

We employed accuracy, precision, recall, and F1-score to verify the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

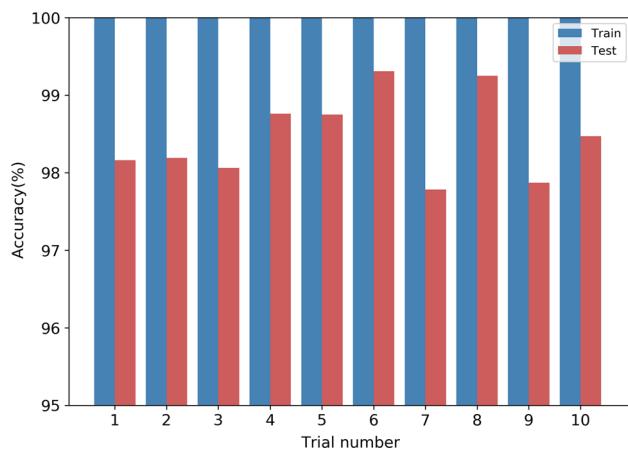
$$\text{Recall} = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

where TP, FP, FN, and TN representing the number of true positive, false positive, false negative, and true negative outcomes, respectively.

## Experiment setup

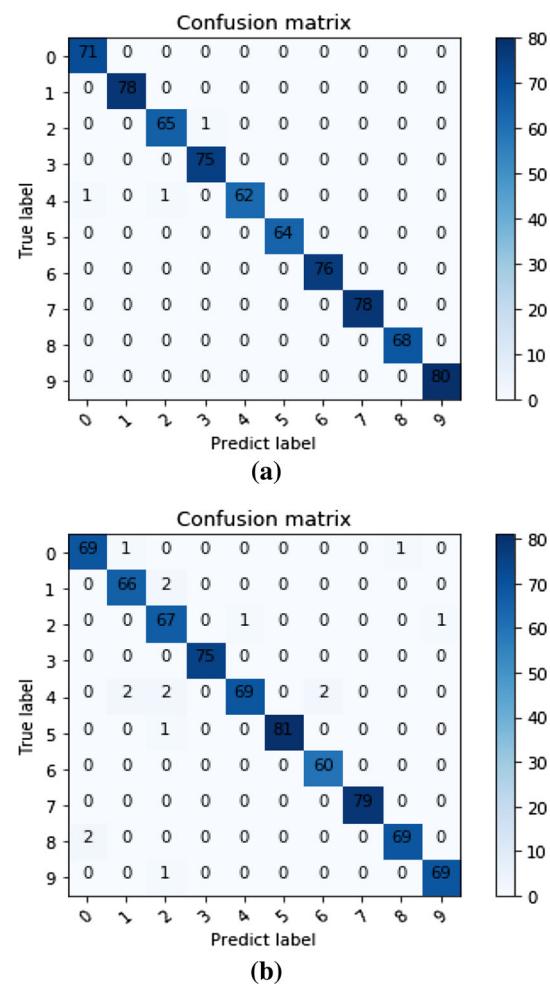
Table 2 shows the CNN and LSTM parameters. Nonlinear activation functions tanh and Relu were applied for CNN convolutional layers and LSTM, respectively. We used dropout during training to avoid overfitting after LSTM\_2, mean square error (MSE) as the loss function to evaluate network performance, and back propagation to update the weights. Since the initial weight and bias settings can strongly impact ANN results, we performed 10 independent experiments using the same parameters to reduce initial choice influence on classification accuracy. Experiments were run on an i7 8565 CPU, 8G RAM, using the keras 2.2.4 framework with tensorflow backend, training for 400 epochs and learning rate = 0.0006.

**Fig. 7** Accuracy of 10 trails**Table 3** Rolling bearing fault diagnosis results using MCNN-LSTM

| Fault label | Precision (%) | Recall (%) | F1-score (%) |
|-------------|---------------|------------|--------------|
| 0           | 97.81         | 98.21      | 98.10        |
| 1           | 98.72         | 96.80      | 97.75        |
| 2           | 96.42         | 94.75      | 95.10        |
| 3           | 99.23         | 99.72      | 99.45        |
| 4           | 96.43         | 94.28      | 95.34        |
| 5           | 98.64         | 99.86      | 99.24        |
| 6           | 99.15         | 99.43      | 99.28        |
| 7           | 99.19         | 99.24      | 98.69        |
| 8           | 97.95         | 97.81      | 97.89        |
| 9           | 98.03         | 98.42      | 97.72        |
| Accuracy    | 98.46         |            |              |

## Experimental results

Figure 7 shows the accuracy of 10 trials, the results show that the training accuracy achieve 100% and the test accuracy is above 97%. Table 3 shows average accuracy, precision, recall and F1-score over the 10 repeated runs for the proposed MCNN-LSTM. The MCNN-LSTM achieved 98.46% average and 99.31% maximum accuracy. F1-score exceeds 95% for all detection results and labels 3, 5, and 6 (healthy; inner race fault defect size = 0.014 mm; and inner race fault defect size = 0.021 inch) exceed 99%. Figure 8 show the confusion matrix for the highest and lowest accuracy outcomes (99.31% and 97.78%), respectively. The main reason for the lower accuracy is misclassification of label 4 to labels 1, 2, and 6. This may due to similarities between these fault signals, which makes classification more difficult when noise is present. However, the overall fault diagnosis results verify that the proposed MCNN-LSTM framework can effectively detect various roller bearing faults.

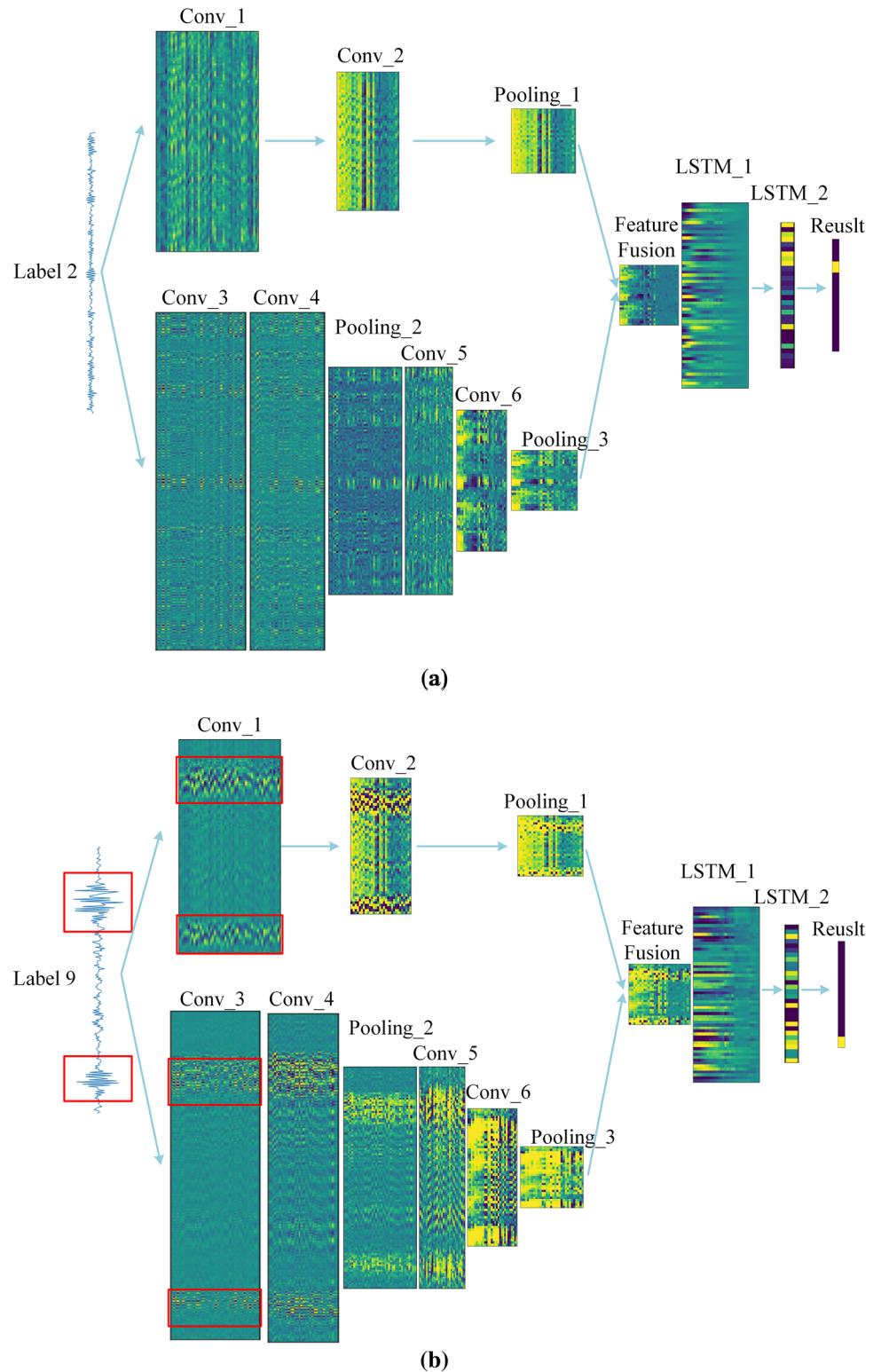
**Fig. 8** Confusion matrix of the classification for MCNN-LSTM: **a** highest accuracy = 99.31%, **b** lowest accuracy = 97.78%

## Network visualization

Generally, ANN models are black boxes, and it is difficult to understand their internal mechanisms. This section visualizes various CNN internal structural aspects to explore the underlying reasoning process. Figure 9 visualizes the activation functions for each hidden layer for specific cases of ball fault and outer race fault centered at 6:00 ( both with defect size = 0.021 inch) network inputs. Yellow indicates a region is activated and green indicates it is not activated. More neurons are activated in Fig. 9b than a. It is difficult to distinguish fault type from the first hidden layer, but the feature map becomes clearer as the network deepens.

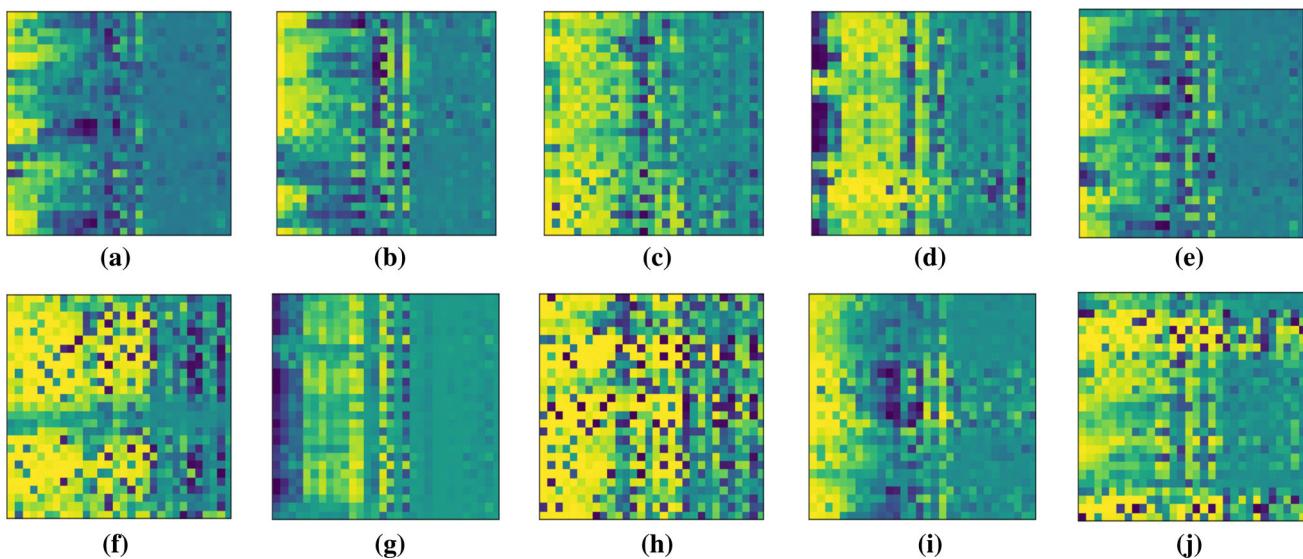
We then investigated feature extractor outputs after inputting 10 different signal segments, with fault signals and healthy signal feed as shown in Table 1. Feature extractors converted the  $1 \times 250$  input signal into a  $30 \times 27$  matrix. Figure 10 shows that different signals types generated different feature matrix types. Yellow indicates neurons are activated and green indicates that they are not activated. It

**Fig. 9** MCNN-LSTM hidden layer visualizations with signal segment inputs: **a** ball fault, defect size = 0.021 inch; and **b** outer race fault, defect size = 0.021 inch, centered at 6:00. Yellow indicates neurons are activated and green that they are not activated. (Color figure online)



can be seen from the figures that different fault types have different neuron activation areas, which confirms that the

proposed method automatically extracts features from raw signals without manual processing.



**Fig. 10** Feature map for ball fault with **a** 0.007 inch, **b** 0.014 inch, and **c** 0.021 inch defect sizes; **d** healthy state; **e** inner race fault with 0.007 inch, **f** 0.014 inch, and **g** 0.021 inch defect size; and outer race fault centered at 6:00 with **h** 0.007 inch, **i** 0.014 inch, and **j** 0.021 inch defect size

**Table 4** Comparison of diagnosis accuracy with different features and algorithms

| Classifier     | Features   | Test accuracy(%) |
|----------------|--|------------------|
| SVM            | Time domain features and frequency features (TD&FD)            | 83.23            |
| KNN            | Time domain features and frequency features (TD&FD)            | 82.26            |
| ANN            | Time domain features and frequency features (TD&FD)            | 79.72            |
| LSTM           | Time domain features and frequency features (TD&FD)            | 86.81            |
| SVM            | Time domain features and empirical mode decomposition (TD&EMD) | 89.72            |
| KNN            | Time domain features and empirical mode decomposition (TD&EMD) | 80.14            |
| ANN            | Time domain features and empirical mode decomposition (TD&EMD) | 85.16            |
| LSTM           | Time domain features and empirical mode decomposition (TD&EMD) | 86.81            |
| LSTM           | —  | 66.39            |
| Proposed model | —  | 98.46            |

## Comparisons with other experiments

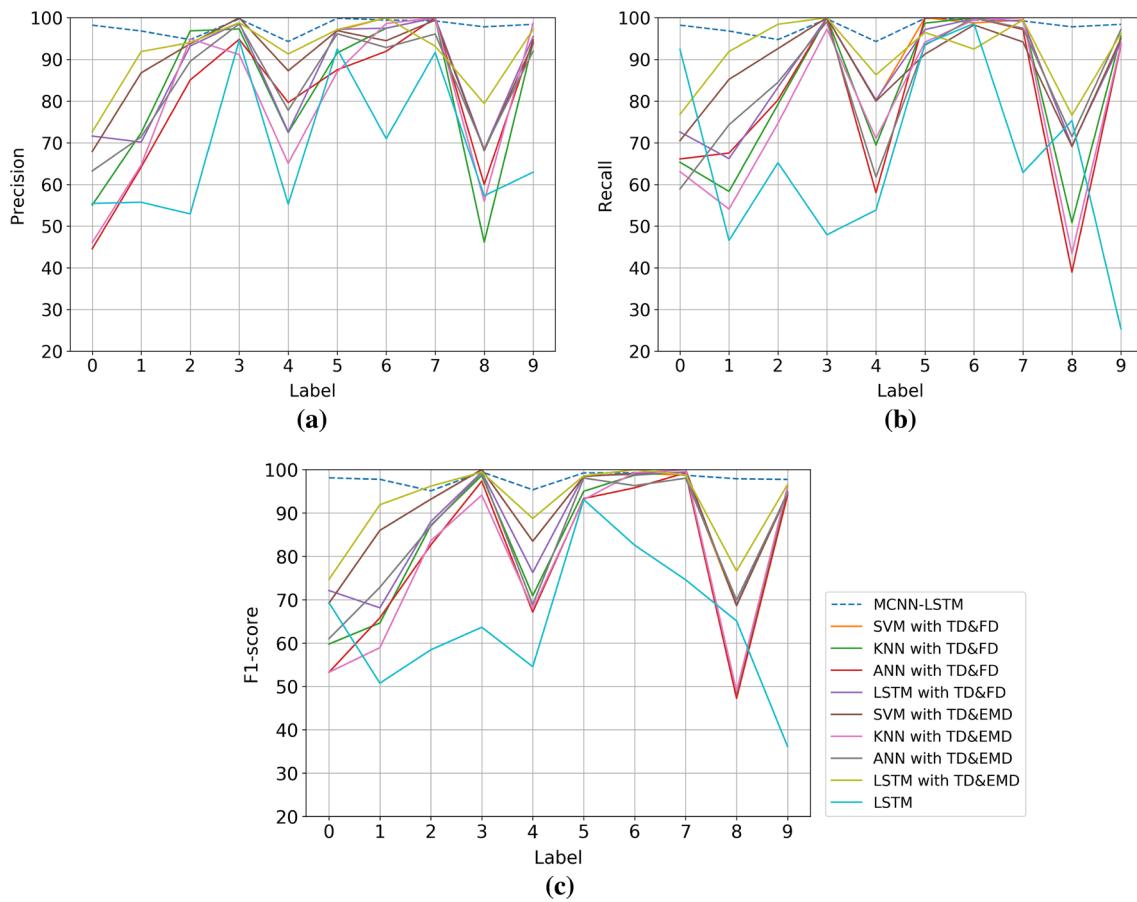
### Case study I: performance of 10 intelligent diagnostic models

The MCNN-LSTM proposed model provides an end-to-end fault diagnosis method for rolling bearings, with directly input unprocessed raw signals to detect fault types. We compared MCNN-LSTM outcomes with current state-of-the-art intelligent diagnosis methods using the same experiment data (Table 1).

Traditional fault diagnosis method performance relies heavily on extracted feature parameter selection, such as wavelet kernel size, and specific classification algorithms. However, there is not much difference between the classification methods with classification performance depending on negative features. Commonly used features include time

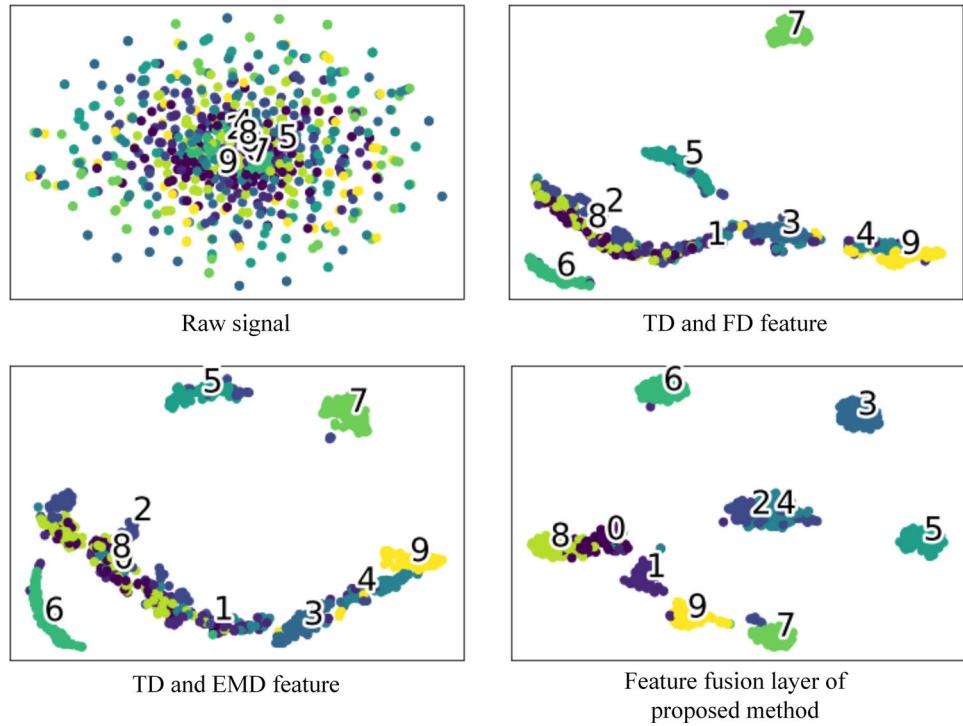
domain and frequency domain features, and Fourier transform (FT) and empirical mode decomposition (EMD) are also commonly applied to extract features. After manually extracting features, we need to fine tune the parameters depending on different issues, which is time consuming and laborious. Redundant features also increase computational cost and reduce accuracy. Therefore, bearing fault diagnosis methods based on feature extraction limits developing a suitable fault diagnosis system. In contrast, the proposed MCNN-LSTM framework automatically extracts bearing vibration signal characteristics, and subsequently accurately classifies the extracted features.

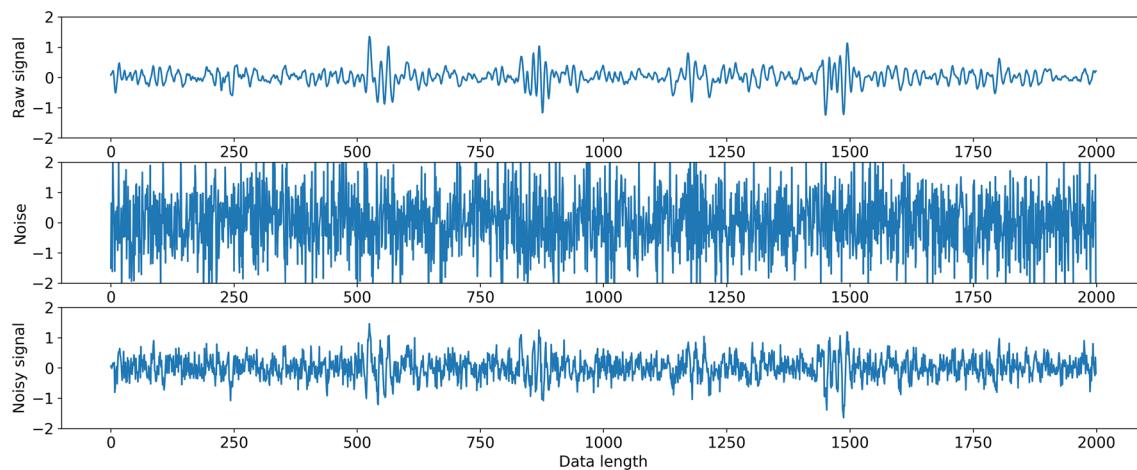
We used two feature sets for comparative experiments between the proposed and current approaches. One set was the 15 time domain (TD) and 3 frequency domain (FD) features extracted from the vibration spectrum, the other was 10 time domain and 4 energy features generated from EMD



**Fig. 11** **a** Precision of 10 fault diagnose methods, **b** recall of 10 fault diagnose methods, **c** F1-score of 10 fault diagnose methods

**Fig. 12** Feature visualization based on T-SNE



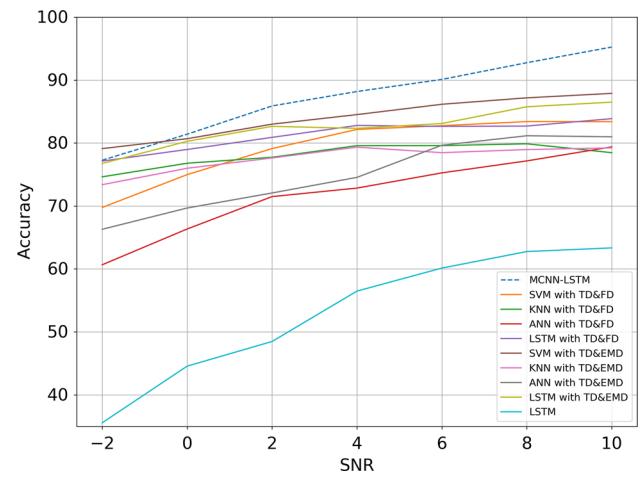


**Fig. 13** Waveform of 0.014 mm ball fault, white Gaussian noise, and composite noise signal with SNR = 0

and Hilbert transform. Detailed information regarding these features could be found in Chen and Li (2017) and Thamba and Aravind (2018), respectively.

We performed 10 trials bearing fault diagnosis experiments, as shown in Table 4. The proposed MCNN-LSTM framework achieved average diagnostic accuracy 98.46%, considerably higher than feature based fault diagnosis methods (highest accuracy = 89.72%). We also tested the LSTM network alone with raw data inputs, achieving average accuracy = 66.39%. We plot precision, recall and F1-score of 10 fault labels as shown in Fig. 11. Prior knowledge based intelligent diagnose methods only achieve great performance on label 3, 5, 6, 7. LSTM with raw data fail to classify 10 faults. But MCNN-LSTM method can correctly distinguish 10 different signal states. Thus, roller bearing fault diagnosis methods based on prior knowledge cannot completely mine raw signal internal features to correctly identify faults when the test data has similarities.

To investigate differences between automatic and artificial feature extraction methods, we used t-distributed stochastic neighbor embedding (T-SNE) (Der Maaten and Hinton 2008) to visualize time and frequency domain features, time domain and energy features, and auto-learning features in two-dimensional feature space, as shown in Fig. 12. Feature expression distribution automatically learned by the proposed MCNN-LSTM method exhibited considerably cleaner boundaries. It is clear that the features extracted by proposed method are much more divisible, which means the classifier will more easily distinguish fault categories. It can be seen from the visualization of proposed method, label 3, 5, and 6 are clearly distinguished from the overall feature distribution, and labels 0, 1, 7, and 9, they are also clearly clustered. In contrast, T-SNE fails to distinguish the 10 different signal states from raw data inputs.



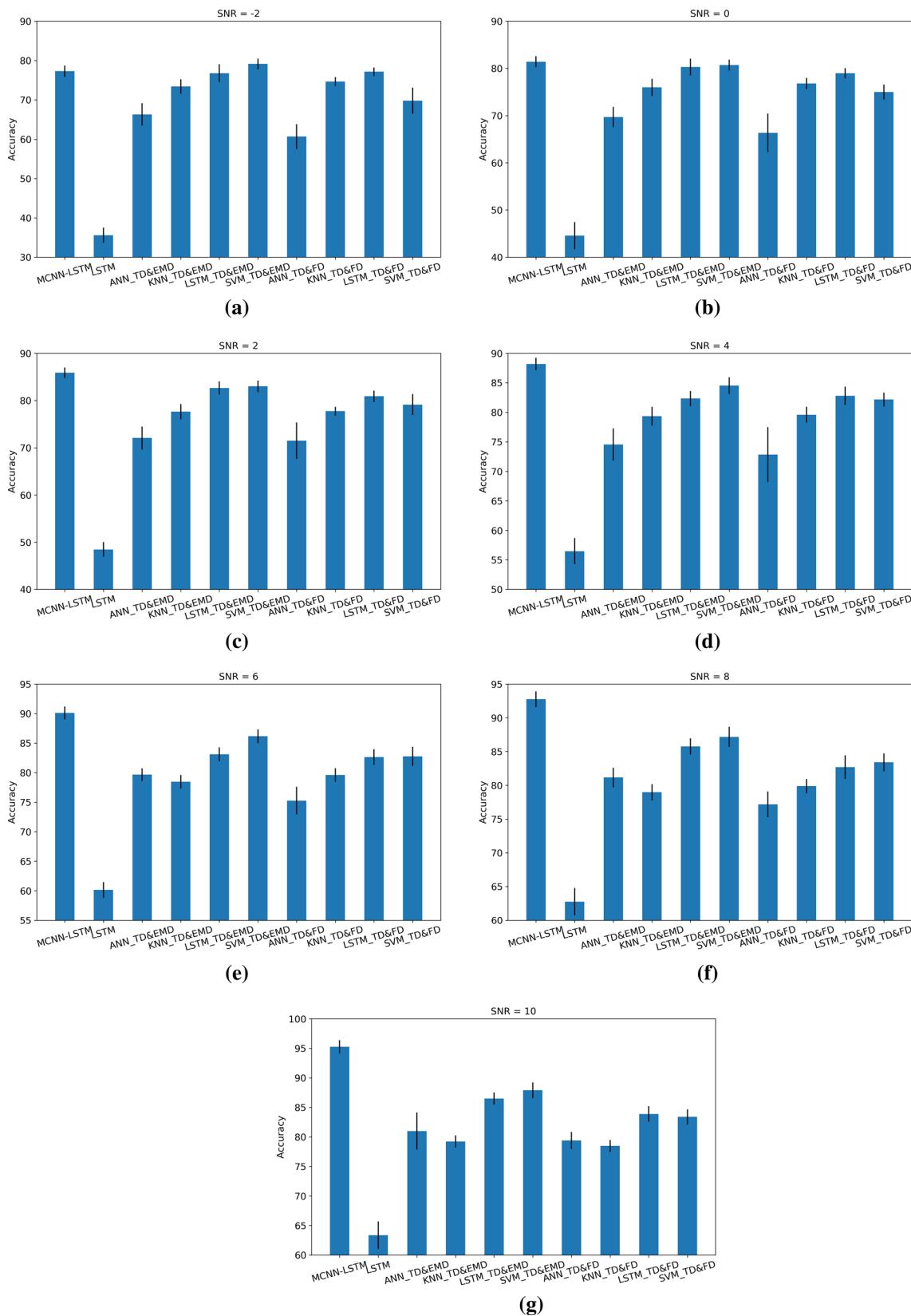
**Fig. 14** Result of 10 different models under different noisy environment

#### Case study II: performance of 10 intelligent diagnostic models under noise environment

In realistic industrial environment, raw signal collected by sensors contains strong background noise. In order to test the robustness of proposed model under strong background noise, white Gaussian noise was added to the raw signals. Figure 13 show the 0.014 inch ball fault signal is added with white Gaussian noise. The signal to noise ratio (SNR) of composite signal is 0, which means the noisy signal has the same power with original signal. From Fig. 13, we can clearly see that the original vibration acceleration signal has been completely destroyed when SNR = 0, extracting fault information from complex waveform diagrams under such a noise background is difficult. The robustness of the proposed system is verified by testing the performance of MCNN-LSTM in different noise environments. The equation of SNR can be

**Table 5** Result of 10 different models under different noisy environment

| Model            | SNR                 |                     |                     |                     |                     | 10                  |                     |
|------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                  | -2                  | 0                   | 2                   | 4                   | 6                   |                     |                     |
| ANN with TD&FD   | 60.67 ± 3.15        | 66.35 ± 4.08        | 71.50 ± 3.84        | 72.85 ± 4.65        | 75.26 ± 2.35        | 77.17 ± 1.91        | 79.39 ± 1.44        |
| KNN with TD&FD   | 74.64 ± 1.13        | 76.79 ± 1.18        | 77.74 ± 0.93        | 79.58 ± 1.34        | 79.60 ± 1.16        | 79.88 ± 1.06        | 78.47 ± 1.01        |
| LSTM with TD&FD  | 77.14 ± 1.10        | 78.98 ± 1.05        | 80.90 ± 1.20        | 82.80 ± 1.54        | 82.65 ± 1.31        | 82.70 ± 1.74        | 83.88 ± 1.30        |
| SVM with TD&FD   | 69.77 ± 3.32        | 74.99 ± 1.60        | 79.13 ± 2.22        | 82.16 ± 1.17        | 82.76 ± 1.63        | 83.41 ± 1.31        | 83.39 ± 1.28        |
| ANN with TD&EMD  | 66.30 ± 2.86        | 69.68 ± 2.16        | 72.07 ± 2.40        | 74.55 ± 2.74        | 79.66 ± 1.07        | 81.16 ± 1.45        | 81.00 ± 3.13        |
| KNN with TD&EMD  | 73.39 ± 1.81        | 75.99 ± 1.80        | 77.63 ± 1.62        | 79.35 ± 1.60        | 78.47 ± 1.13        | 78.97 ± 1.19        | 79.23 ± 1.02        |
| LSTM with TD&EMD | 76.75 ± 2.30        | 80.28 ± 1.78        | 82.65 ± 1.39        | 82.34 ± 1.29        | 83.10 ± 1.20        | 85.77 ± 1.19        | 86.50 ± 1.02        |
| SVM with TD&EMD  | <b>79.12 ± 1.35</b> | 80.69 ± 1.15        | 83.00 ± 1.21        | 84.52 ± 1.42        | 86.16 ± 1.15        | 87.18 ± 1.48        | 87.89 ± 1.34        |
| LSTM             | 35.57 ± 1.93        | 44.56 ± 2.86        | 48.48 ± 1.56        | 56.49 ± 2.22        | 60.14 ± 1.33        | 87.18 ± 1.48        | 63.34 ± 2.33        |
| MCNN-LSTM        | 77.27 ± 1.43        | <b>81.41 ± 1.15</b> | <b>85.90 ± 1.10</b> | <b>88.19 ± 1.05</b> | <b>90.11 ± 1.09</b> | <b>92.77 ± 1.15</b> | <b>95.25 ± 1.14</b> |



**Fig. 15** Result of 10 intelligent diagnosis models under different background noise **a** SNR = -2, **b** SNR = 0, **c** SNR = 2, **d** SNR = 4, **e** SNR = 6, **f** SNR = 8, **g** SNR = 10

described as follow:

$$SNR = 10 \log_{10} \frac{P_s}{P_n} \quad (17)$$

where  $P_s$  represents signal power and  $P_n$  represents noise power. When SNR is less than 0, it indicates that the energy of the noise signal is higher than the original signal, in this case it is quite difficult to extract useful features from the composite noise signal. We test the proposed MCNN-LSTM and other 9 state-of-the-art intelligent diagnosis model with different SNR ranging from –2 to 10 db. Experiments were performed 10 trials and the structure and parameters are not changed. The average accuracy and standard deviation are shown in Table 5, and Fig. 14 shows the average accuracy of the 10 methods under different noise environments. The accuracy of proposed model is 77.27% lower than the SVM model based on time feature and EMD and Hilbert transform when  $SNR = -2$ . It's clear that the accuracy becomes higher as the SNR becomes larger. But when  $SNR = 10$ , the accuracy of proposed model surges to 95.25%, much higher than other intelligent diagnosis model. That is, when background noise is strong, MCNN-LSTM model fails to distinguish different signal states, and the bearing fault characteristics are drowned in background noise. This means that none of these models above are suitable for working in strong background noise. In addition, when SNR is above 6db, the accuracy of proposed model is easy to achieve 90% and outperform other intelligent diagnosis method based on manual feature extraction. Figure 15 show more detailed result comparison under different noise intensities. It can be seen form the figure that the proposed model doesn't have too much difference from the model based on manual feature extraction when  $SNR = -2$  or 0. When SNR is higher than 2, MCNN-LSTM is obviously has a better performance than other models. At the same time, the standard deviation of MCNN-LSTM is smaller than other methods, which means the rolling bearing fault diagnosis method based on automato learning features has better performance than the method based on expert knowledge under noisy environment. That is, the proposed method is more robust than other methods. We can conclude that the proposed MCNN-LSTM framework can realize end-to-end roller bearing fault diagnosis from direct raw data input; automatically learning features from the raw signal; and achieve higher accuracy than traditional intelligent diagnosis methods based on prior knowledge.

## Conclusions

This paper proposed a novel end-to-end intelligent diagnosis method for rolling bearing faults. The proposed model includes feature extractor and classifier modules. The feature extractor can learn signal features from raw data, and then

provide the extracted features for the classification module, which comprises a stacked LSTM neural network. Finally, a softmax function was used to convert neuron outputs into a probability distribution of rolling bearing fault (including healthy). The proposed MCNN-LSTM framework offers several advantages, as follows: (1) the compact network structure and accepts raw data input, which make it possible to detect roller bearing status in real time; (2) the proposed algorithm can automatically learn features from raw signals without preprocessing, such as EMD, or HHT; (3) the proposed method provides effective training and classification methods for small datasets; (4) the final trained network accurately identified 10 different fault conditions. We used the CWRU bearing data set to verify the proposed fault diagnosis method feasibility, achieving average accuracy = 98.46%. Compared with two current state-of-the-art diagnosis methods based on prior knowledge, the proposed model successfully learned features from raw data and achieved higher accuracy, whereas the prior knowledge methods achieved 89.72%, and the LSTM neural network alone with raw signal input only achieved 66.39%. Feature visualization based on T-SNE also supported the proposed MCNN-LSTM superiority.

**Acknowledgements** This work was supported by National Natural Science Foundation (NNSF) of China (Grants 61703026 and 61873022).

## Compliance with ethical standards

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

## References

- Ali, J. B., et al. (2015). Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, 89, 16–27.
- Amar, M., Gondal, I., & Wilson, C. (2015). Vibration spectrum imaging: A novel bearing fault classification approach. *IEEE Transactions on Industrial Electronics*, 62(1), 494–502.
- Aydin, I., Karaköse, M., & Akin, E. (2015). Combined intelligent methods based on wireless sensor networks for condition monitoring and fault diagnosis. *Journal of Intelligent Manufacturing*, 26(4), 717–29.
- Chauhan, S., Vig, L., & Ahmad, S. (2019). ECG anomaly class identification using LSTM and error profile modeling. *Computers in Biology and Medicine*, 109, 14–21.
- Chen, Z., & Li, W. (2017). Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Transactions on Instrumentation and Measurement*, 66(7), 1693–1702.
- Der Maaten, L. V., & Hinton, G. E. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
- Donnell, P. O., Heising, C., et al. (1987). Report of large motor reliability survey of industrial and commercial installations: Part 3. *IEEE Transactions on Industry Applications*, IA-23(1), 153–58.

- Gao, Z., Cecati, C., & Ding, S. (2015). A survey of fault diagnosis and fault-tolerant techniques part II: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics*, 62, 3768–3774.
- Garcia-Perez, A., de Jesus Romero-Troncoso, R., Cabal-Yepez, E., & Osornio-Rios, R. A. (2011). The application of high-resolution spectral analysis for identifying multiple combined faults in induction motors. *IEEE Transactions on Industrial Electronics*, 58(5), 2002–10.
- He, M., & He, D. (1987). Report of large motor reliability survey of industrial and commercial installations: Part 3. *IEEE Transactions on Industry Applications*, IA-23(1), 153–58.
- Humberstone, M., Wood, B., Henkel, J., & Hines, J. W. (2012). Differentiating between expanded and fault conditions using principal component analysis. *Journal of Intelligent Manufacturing*, 23(2), 179–88.
- Jantunen, E., & Vaajoensuu, E. (2010). Self adaptive diagnosis of tool wear with a microcontroller. *Journal of Intelligent Manufacturing*, 21(2), 223–30.
- Jin, J., & Sanchez-Sinencio, E. (2015). A home sleep apnea screening device with time-domain signal processing and autonomous scoring capability. *IEEE Transactions on Biomedical Circuits and Systems*, 9(1), 96–104.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Li, H., Lian, X., Guo, C., & Zhao, P. (2015). Investigation on early fault classification for rolling element bearing based on the optimal frequency band determination. *Journal of Intelligent Manufacturing*, 26(1), 189–98.
- Li, Z. W., Nie, R., & Han, Y. F. (2010). Supervised locally linear embedding for fault diagnosis. *Advanced Materials Research*, 139–141, 2599–2602.
- Lucchese, L., & Cortelazzo, G. M. (2000). A noise-robust frequency domain technique for estimating planar roto-translations. *IEEE Transactions on Signal Processing*, 48(6), 1769–86.
- Luong, T., et al. (2015). Addressing the rare word problem in neural machine translation. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing* (Volume 1: Long papers) (pp. 11–19). Beijing, China: Association for Computational Linguistics.
- Mojiri, M., Karimi-Ghartemani, M., & Bakhshai, A. (2007). Time-domain signal analysis using adaptive notch filter. *IEEE Transactions on Signal Processing*, 55(1), 85–93.
- Petsounis, K. A., & Fassois, S. D. (2001). Parametric time-domain methods for the identification of vibrating structures—a critical comparison and assessment. *Mechanical Systems and Signal Processing*, 15(6), 1031–60.
- Rajeswaran, N., Madhu, T., et al. (2011). Analysis and study of fault diagnosis of induction motor drives using hybrid artificial intelligence. *Journal of Foshan University*, 29, 313–318.
- Ramos, A. R., et al. (2019). An approach to robust fault diagnosis in mechanical systems using computational intelligence. *Journal of Intelligent Manufacturing*, 30(4), 1601–15.
- Sainath, T. N., Vinyals, O., Senior, A., & Sak, H. (2015). Convolutional, long short-term memory, fully connected deep neural networks. In *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, South Brisbane, Queensland (pp. 4580–84). IEEE.
- Saravanan, N., Kumar Siddabattuni, V. N. S., & Ramachandran, K. I. (2008). A comparative study on classification of features by SVM and PSVM extracted using morlet wavelet for fault diagnosis of spur bevel gear box. *Expert Systems with Applications*, 35(3), 1351–66.
- Shao, S., et al. (2016). Learning features from vibration signals for induction motor fault diagnosis. In: *2016 International symposium on flexible automation (ISFA)*, Cleveland, OH (pp. 71–76). IEEE.
- Shuang, L., & Meng, L. (2007). Bearing fault diagnosis based on PCA and SVM. In *2007 International Conference on Mechatronics and Automation*, Harbin (pp. 3503–3507). IEEE.
- Stridh, M., Sornmo, L., Meurling, C. J., & Olsson, S. B. (2004). Sequential characterization of atrial tachyarrhythmias based on ECG time-frequency analysis. *IEEE Transactions on Biomedical Engineering*, 51(1), 100–114.
- Sun, Q., & Tang, Y. (2002). Singularity analysis using continuous wavelet transform for bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 16(6), 1025–41.
- Thamba, N. B., Aravind, A., et al. (2018). Application of EMD, ANN and DNN for self-aligning bearing fault diagnosis. *Archives of Acoustics*, 43(2), 163–175.
- Thorsen, O. V., & Dalva, M. (1999). Failure identification and analysis for high-voltage induction motors in the petrochemical industry. *IEEE Transactions on Industry Applications*, 35(4), 810–18.
- Tian, Y., & Liu, X. (2019). A deep adaptive learning method for rolling bearing fault diagnosis using immunity. *Tsinghua Science and Technology*, 24(6), 750–62.
- Van, M., & Kang, H.-J. (2015). Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization. *IEEE Transactions on Industrial Informatics*, 12, 124–135.
- Wang, C., Gan, M., & Zhu, C. (2018). Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *Journal of Intelligent Manufacturing*, 29(4), 937–51.
- Widodo, A., & Yang, B.-S. (2007). Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors. *Expert Systems with Applications*, 33(1), 241–50.
- Wu, C., et al. (2017). A novel approach to wavelet selection and tree kernel construction for diagnosis of rolling element bearing fault. *Journal of Intelligent Manufacturing*, 28(8), 1847–58.
- Wu, S.-D., et al. (2012). Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy*, 14(8), 1343–56.
- Xu, Y., Wen, J., Fei, L., & Zhang, Z. (2016). Review of video and image defogging algorithms and related studies on image restoration and enhancement. *IEEE Access*, 4, 165–88.
- Yi, Lu, Xie, R., & Liang, S. Y. (2019). Bearing fault diagnosis with nonlinear adaptive dictionary learning. *The International Journal of Advanced Manufacturing Technology*, 102(9–12), 4227–39.
- Yu, L., Qu, J., Gao, F., & Tian, Y. (2019). A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM. *Shock and Vibration*, 2019, 1–10.
- Yu, X., et al. (2018). Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection. *IEEE Access*, 6, 3715–30.
- Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2020). Deep learning algorithms for bearing fault diagnostics—a comprehensive review. *IEEE Access*, 8, 29857–81.
- 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), 1213–27.