A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps

Shuhui Wang[1] · Jiawei Xiang[1] (iD)

Abstract

Condition monitoring of piston pumps has great significance to ensure the reliability and security of hydraulic systems. However, the complex working conditions of the integrated electromechanical systems make the fault mechanism unclear which is difficult for fault diagnosis by feature matching techniques. In this paper, a novel minimum entropy deconvolution (MED)-based convolutional neural network (CNN) is presented to classify faults in axial piston pumps. Firstly, the collected raw signals are preprocessed using the MED technique. Then, the filtered signals are used to construct training samples and testing samples. Finally, the constructed samples are fed into the CNN to classify the multi-faults of axial piston pumps. With the convolution and subsampling operations, the present model can automatically obtain data features via iterative learning processes, which is suitable for the unknown fault mechanism problems. The learned features are visualized by t-distributed stochastic neighbor embedding technique. A benchmark data simulation of mechanical transmission systems and an experimental data investigation of an axial piston pump are performed to manifest the superiority of the present method by comparing with the traditional CNN.

Keywords Axial piston pumps · Minimum entropy deconvolution · Convolutional neural network · Fault classification

1 Introduction

Hydraulic transmission systems play an important role in modern industry. Condition monitoring of hydraulic systems has attracted increasing attention in recent decades (Jegadeeshwaran and Sugumaran 2015; Sepasi and Sassaniv 2010; Fu et al. 2014). Axial piston pumps are commonly used in hydraulic systems. The piston pumps defects may cause the machine breakdowns, lead to severe economic loss or even catastrophic casualty. As a matter of fact, pump defects take up a large percentage of electromechanical equipments faults (Ferdowsi et al. 2014; Lan et al. 2018; Wang et al. 2018a, b). Consequently, precisely and effectively detect faults in pumps has become an urgent task to ensure the safety and reliability of hydraulic systems.

Faults of axial piston pump mostly occurred in cylinder block, swash plate, bearings, pistons, etc. Pump health condition monitoring can be undertaken using different sensors, e.g., accelerometers, flowmeters, electric current, etc. Among them, vibration signals collected by accelerometers have enjoyed great success in condition monitoring and fault diagnosis of rotating machinery (Wang and Liang 2012; Xiang et al. 2015; Samanta et al. 2006; Pandya et al. 2014; Wang et al. 2009; Qiao et al. 2017). Many of the existing machinery diagnosis methods have been referred and introduced to detect faults in hydraulic pumps, including wavelet packet analysis (Gao and Zhang 2006), intermittent chaos (Zhao et al. 2009), local mean decomposition (LMD) (Jiang et al. 2014), empirical mode decomposition (EMD) (Lu et al. 2016, 2015), and minimum entropy deconvolution (MED) (He et al. 2016; Du et al. 2017; Dong et al. 2017), etc. Gao and Zhang (2006) performed wavelet packet transformation and wavelet coefficient residual analysis to discharge pressure signals to detect faults in hydraulic piston pumps. Zhao et al. (2009) introduced intermittent chaos combined with sliding window symbol sequence statistic to detect early faults in hydraulic pumps. Jiang et al. (2014)

✉ Jiawei Xiang
jwxiang@wzu.edu.cn

1 College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, People's Republic of China

proposed an improved adaptive multi-scale morphology analysis (IAMMA) in associated with LMD to detect faults of hydraulic pumps using vibration signals. Lu et al. (2016) developed a two-step fault diagnosis method based on EMD and fuzzy C-means clustering. EMD and Hilbert transform were employed to extract fault features from the discharge pressure signals collected by flowmeters. Fuzzy C-means clustering algorithm was applied to recognize hydraulic pump fault conditions. Meanwhile, Lu et al. (2015) also investigated fault severity recognition methods of hydraulic piston pumps, which would attain higher recognition accuracy. He et al. (2016) investigated multiple faults detection in rotating machinery by means of MED combined with SK analysis, which obtain a nice impulse restoration result of the vibration signals from a vacuum pump.

Generally speaking, once there are faults occurred to the components of the working piston pump, they are reflected in the measured signals with certain characteristics, which can be obtained through signal processing techniques. Most fault diagnosis methods are based on matching the extracted fault features with the corresponding analyzed fault characteristics. Unfortunately, many researches have reported that the frequencies of some different faults are identical (Gao and Zhang 2006; Lu et al. 2015). Besides, the fault mechanism of some faults is unclear, namely it is hard to point out such a fault with its corresponding characteristic. In such cases, the feature matching is failed. To tackle the challenge mentioned above, the main objective of the paper is to search for an automatic feature learning method for axial piston pump fault diagnosis.

Deep learning is new branch of machine learning, which has the great capacity in feature learning from raw data (Hinton and Salakhutdinov 2006; LeCun et al. 2015; LeCun and Bengio 1998). Although conventional shallow learning models, such as support vector machine (Suganyadevi et al. 2016; Liu et al. 2017) and artificial neural network (Trujillo et al. 2017; Huang et al. 2016), are usually employed to solve pattern recognition problems, they depend heavily on manually selected features by signal processing techniques. Compared with the shallow learning model, the deep learning model can realize very complicated transformation and abstraction of the raw data with its multilayer structures (LeCun et al. 2015). That is to say, deep learning models can decrease the dependence on various signal processing techniques. CNN (LeCun and Bengio 1998) is a kind of the deep learning model with powerful feature learning ability, which enjoyed great success in a variety of fields (Vu et al. 2018; Rafique et al. 2018; Liu et al. 2018). Nowadays, many researchers have exploited the excellent performance of CNN in mechanical fault diagnosis. Ince et al. (2016) constructed a motor condition monitoring and early fault detection system by employing one-dimensional CNNs. Appana et al. (2018) proposed an CNN-based feature learning and fault diagnosis method for the condition monitoring of bearings under varying rotational speeds. Janssens et al. (2016) introduced the CNN to autonomously learn useful features for bearing fault detection from the raw data. Wang et al. (2018a, b) proposed a CNN-based hidden Markov models for rolling element bearing fault identification. Ding and He (2017) presented an energy-fluctuated multi-scale feature mining approach based on wavelet packet energy and convolutional networks to classify faults in spindle bearings.

In the light of the feature learning ability of CNN models and great impulse restoration capacity of MED technique, a MED-enhanced CNN model is proposed in terms of detecting faults in axial piston pumps, and the contributions of this job is summarized as follows:

(1) To solve automatic feature learning problem in detecting faults of axial piston pumps, the CNN model is introduced, which could effectively avoid feature matching difficulty due to the unclear fault mechanism.

(2) To prompt the feature learning stability and classification accuracy, the MED technique is used to preprocess the raw vibration data. In fact, the preprocessing step here is something like a data cleaning method (Xu et al. 2019).

(3) The feature learning process is visualized using the t-distributed stochastic neighbor embedding (t-SNE) technique (van der Maaten and Hinton 2008), and the stability of the MED-enhanced CNN is demonstrated by the learning trials.

The remainder of this paper is organized as follows. Section 2 gives a brief review of the basic theory of MED and CNN. In Sect. 3, the fault detection method is introduced in details. A benchmark study is investigated in Sect. 4. In Sect. 5, the proposed method is applied to analyze the vibration signals from an axial piston pump. Conclusion remarks are drawn in Sect. 6.

2 A brief review of MED and CNN

2.1 Minimum entropy deconvolution

The minimum entropy deconvolution technique aims at modeling an inverse filtering process against the commonly signal transfer process (Sawalhi et al. 2007; Endo and Randall 2007). The main target is to find an optimum inverse filter coefficient vector f to recover the fault impact by

$$y = f * x. \tag{1}$$

As shown in Fig. 1, without any prior knowledge about the input x, the MED filter could adaptively adjust the filter coefficients by optimizing the objective function of the output y.

High-order statistics are well known for depicting the shape of the probability density function (PDF). For example, a high value of kurtosis (fourth-order statistics) is usually related to spikes in the PDF. Therefore, kurtosis is often employed as an objective function to quantify the characteristics of a signal, which is given as

$$K(y(n)) = \frac{\frac{1}{N}\sum_{n=1}^{N}|y(n)|^4}{\left[\frac{1}{N}\sum_{n=1}^{N}|y(n)|^2\right]^2} \tag{2}$$

where $y(n)$ $(n = 1, 2, \ldots, N)$ is the output sequence after data sampling. As mentioned by Endo and Randall (2007), kurtosis is an effective indicator that reflects the "peakiness" of a signal, therefore the property of impulses.

The main process of the MED filtering is summarized as follows:

Step 1 Construct the relationship between y and f. The relationship can be generalized as

$$y(n) = \sum_{l=1}^{L} f(l)\, x(n-l+1) \tag{3}$$

where $x(n)$ $(n = 1, 2, \ldots, N)$ is the input sequence after data sampling, $f(l)$ $(l = 1, 2, \ldots, L)$ is lth element in f, and finally get

$$\frac{\partial y(n)}{\partial f(l)} = x(n-l+1). \tag{4}$$

Step 2 Maximize the objective function with respect to $f(l)$

$$\frac{\partial\left(O_2^4(y(n))\right)}{\partial(f(l))} = 0 \tag{5}$$
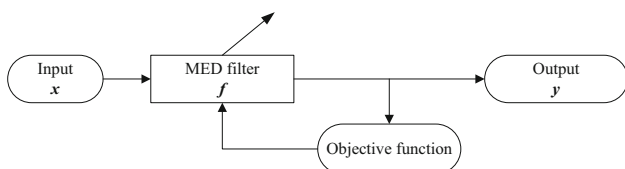
Step 3 Update f by adjusting $f(l)$ iteratively through

$$\sum_{n}^{N+L-1}\left\{\left[\sum_{k=1}^{L} f(k)\, x(n-k+1)\right] x(n-l+1)\right\}$$
$$= \frac{\sum_{n}^{N+L-1} (y(n))^2}{\sum_{n}^{N+L-1} (y(n))^4}\left[\sum_{n}^{N+L-1} (y(n))^3 x(n-l+1)\right] \tag{6}$$

Step 4 Obtain the final output y to approximate fault impact by the updated f using Eq. (3).

2.2 Convolutional neural network

A CNN is one kind of deep learning model with distinctive operation, namely convolution, subsampling and fully connection. There are three main traits of a CNN, that's local field, subsampling and weight sharing (LeCun and Bengio 1998).

Figure 2 demonstrates the main architecture of a CNN model with five hidden layers. The hidden layers include convolutional layer C_1 (first hidden layer), subsampling layer S_1 (second hidden layer), convolutional layer C_2 (third hidden layer), subsampling layer S_2 (fourth hidden layer) and fully connected layer FC (fifth hidden layer). In order to illustrate the transfer process concisely, we give the input layer "S_0."

Supposing an input map x^{S_0} to the CNN model, the kth $(k = 1, 2, \ldots, K$, where K is the number of feature maps) feature map $x_k^{C_1}$ in layer C_1 can be represented by

$$x_k^{C_1} = f\left(x^{S_0} * w_k^{C_1} + b_k^{C_1}\right) \tag{7}$$

where $f(\bullet)$ is the output activation function, $w_k^{C_1}$ denotes the kth kernel of C_1, and $b_k^{C_1}$ is the kth bias of C_1. Every kernel is like a small window observing local fields of x^{S_0} with a certain stride, and thus is the first trait of a CNN.

The precise location of data features is no more important once they are obtained by the convolution operation. Hence, subsampling operation is directly followed to decrease the computation complexity. As shown in Fig. 2, the K feature maps in C_1 is the input maps to layer S_1, and the transformation in this layer can be expressed by

$$x_k^{S_1} = f\left(s_k^{S_1}\mathrm{down}\left(x_k^{C_1}\right) + b_k^{S_1}\right) \tag{8}$$

where $x_k^{S_1}$ is the kth feature map in S_1, $s_k^{S_1}$ denotes the kth scale of S_1, down($\bullet$) represents the subsampling function and $b_k^{S_1}$ is the kth bias of S_1. This is the second trait of a CNN.

Different from the situation of one input map in Eq. (7), the kth feature map $x_k^{C_2}$ in layer C_2 is

$$x_k^{C_2} = f\left(\sum_{i \in M_k} x_i^{S_1} * w_{ik}^{C_2} + b_k^{C_2}\right) \tag{9}$$

Fig. 1 The main process of the MED filtering

Fig. 2 The common architecture of a CNN model
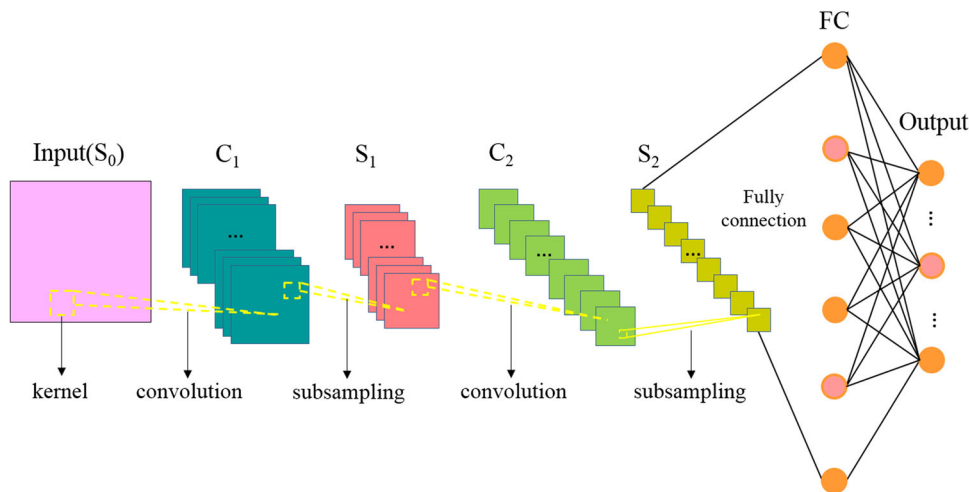
where M_k represents a selection of feature maps in layer S_1, i is the ith feature map of M_k, $w_{ik}^{C_2}$ is the kernel of C_2 and $b_k^{C_2}$ is the kth bias of C_2. Equation (9) indicates that each feature map $x_i^{S_1}$ shares the same weights $w_{ik}^{C_2}$, which illustrates the weight sharing trait of a CNN.

Similarly, the kth feature map $x_k^{S_2}$ in layer S_2 is

$$x_k^{S_2} = f\left(s_k^{S_2}\text{down}\left(x_k^{C_2}\right) + b_k^{S_2}\right). \tag{10}$$

The feature maps in S_2 are fully connected in layer FC. Input maps are transformed hierarchically by feedforward pass. The weights and bias in each hidden layers are updated by backpropagation pass. We can finally get the learned features by the whole training process, which can be referred in (Bouvrie 2006).

3 Automatic fault diagnosis using the MED based CNN

When faults occur, the vibration signals are different from those of the normal working pumps, which exhibit certain characteristics, such as impulses. However, the fault excitation impulses are usually too weak to be observed under strong noise environments (Antoni 2016; Li et al. 2017; Li and Zhao 2017; Xu et al. 2018). Meanwhile, due to the disturbances of transfer path in mechanical systems, the collected signals of a specific faulty component are inevitably distorted. Therefore, the MED technique is introduced with its great impulse restoration capacity.

In light of the strong feature learning ability of CNNs, this paper proposed a MED based CNN to the multi-fault detection of axial piston pumps. The MED technique is firstly employed to enhance the fault impulses in the raw vibration signals. To address the fault feature matching failure problem, a deep CNN model is then applied to

automatically learn fault features from the vibration signals.

The flowchart of the automatic fault diagnosis procedure is described in Fig. 3. The general procedure is summarized as follows:

Step 1 First, predefine the common fault patterns from 1 to R, which is the prerequisite of the next labeling work for collected samples.
Step 2 The raw vibration data are collected from the axial piston pump under different fault patterns by data acquisition system.
Step 3 The collected raw signals are filtered using the MED technique to largely get rid of disturbances of transfer paths and environments.
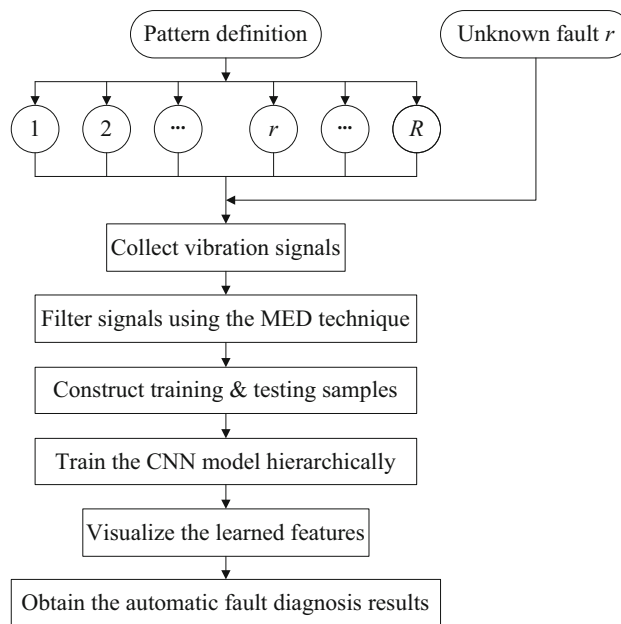

Fig. 3 The flowchart of the automatic fault diagnosis procedure

Step 4 The raw signals after MED filtering are directly separated into training sample set and testing sample set.
Step 5 The CNN model is trained hierarchically by alternate convolution and subsampling operations using the training sample set.
Step 6 The feature learning process is manifested via the two convolutional stage with t-SNE visualization.
Step 7 The testing sample set is treated as unknown faults to obtain the eventually fault classification results. In real applications, unknown fault pattern *r* must be included in the known fault patterns category.

It should be noted that, it is the lack of faulty training samples that prevents the intelligent classification based methods from practical applications. Numerical simulation (Xiang and Zhong 2016) of mechanical systems might be a probable tool to establish fault samples for machines under all kinds of working conditions.

4 Simulation verification with benchmark data

In this section, simulation verification is conducted with the benchmark data from Case Western Reserve University (CWRU) bearing data center (http://csegroups.case.edu/bearingdatacenter/home). The experimental bearing test setup contains an electrical driving motor, torque transducer and encoder, and a dynamometer.

4.1 Data source description

Table 1 gives detailed bearing data specifications in the case. Twelve bearing operating states are considered, as given in Table 2. The time waveform of the vibration signals (the first 12,000 points) and the corresponding spectrums are displayed in Fig. 4.

4.2 Results and analysis

The simulation verification is devoted to applying the CWRU data mentioned above to evaluate the feature learning performance of the proposed MED-based CNN model. Every bearing state contains 120,000 data points. There are 2400 (200 × 12) training samples and 1200 (100 × 12) testing samples, respectively. The size of the input map to CNN model is 20 × 20. The filter length of MED is chosen as 30, and the iteration number is 50. The main parameters of the CNN model are listed in Table 3.

The classification results using the MED-based CNN (MED–CNN) and the traditional CNN are shown in Fig. 5a, b, respectively. Confusion matrix is employed as an effective tool to describe classification results. The classification accuracies are listed in the diagonal line of the matrix. The rest elements describe the probability that a certain fault pattern is identified as another pattern. For example, as shown in Fig. 5a, the classification accuracy ratio of pattern 1 is 98% and 2% of the pattern 1 is identified as pattern 11. The average classification accuracy ratio using MED-based CNN is 97.33%, while that of the traditional CNN only remains 34.50%.

Table 1 Bearing specifications

Bearing type	Bearing location	Motor load (HP)	Sample frequency (Hz)	Motor speed (rpm)
SKF6205	Drive-end	0	12,000	1797

Table 2 Fault category and sample distribution in simulation verification

Rolling bearing state	Fault severity (in.)	Size of training samples/testing samples		Fault pattern
Normal state	No damage	200	100	1
Inner race fault	0.007	200	100	2
Ball fault	0.007	200	100	3
Outer race fault	0.007	200	100	4
Inner race fault	0.014	200	100	5
Ball fault	0.014	200	100	6
Outer race fault	0.014	200	100	7
Inner race fault	0.021	200	100	8
Ball fault	0.021	200	100	9
Outer race fault	0.021	200	100	10
Inner race fault	0.028	200	100	11
Ball fault	0.028	200	100	12

Fig. 4 Time waveform and the corresponding spectrums **a** normal state, **b** 0.007 in./inner race fault, **c** 0.007 in./ball fault, **d** 0.007 in./outer race fault, **e** 0.014 in./inner race fault, **f** 0.014 in./ball fault, **g** 0.014 in./outer race fault, **h** 0.021 in./inner race fault, **i** 0.021 in./ball fault, **j** 0.021 in./ outer race fault, **k** 0.028 in./ inner race fault, **l** 0.028 in./ball fault
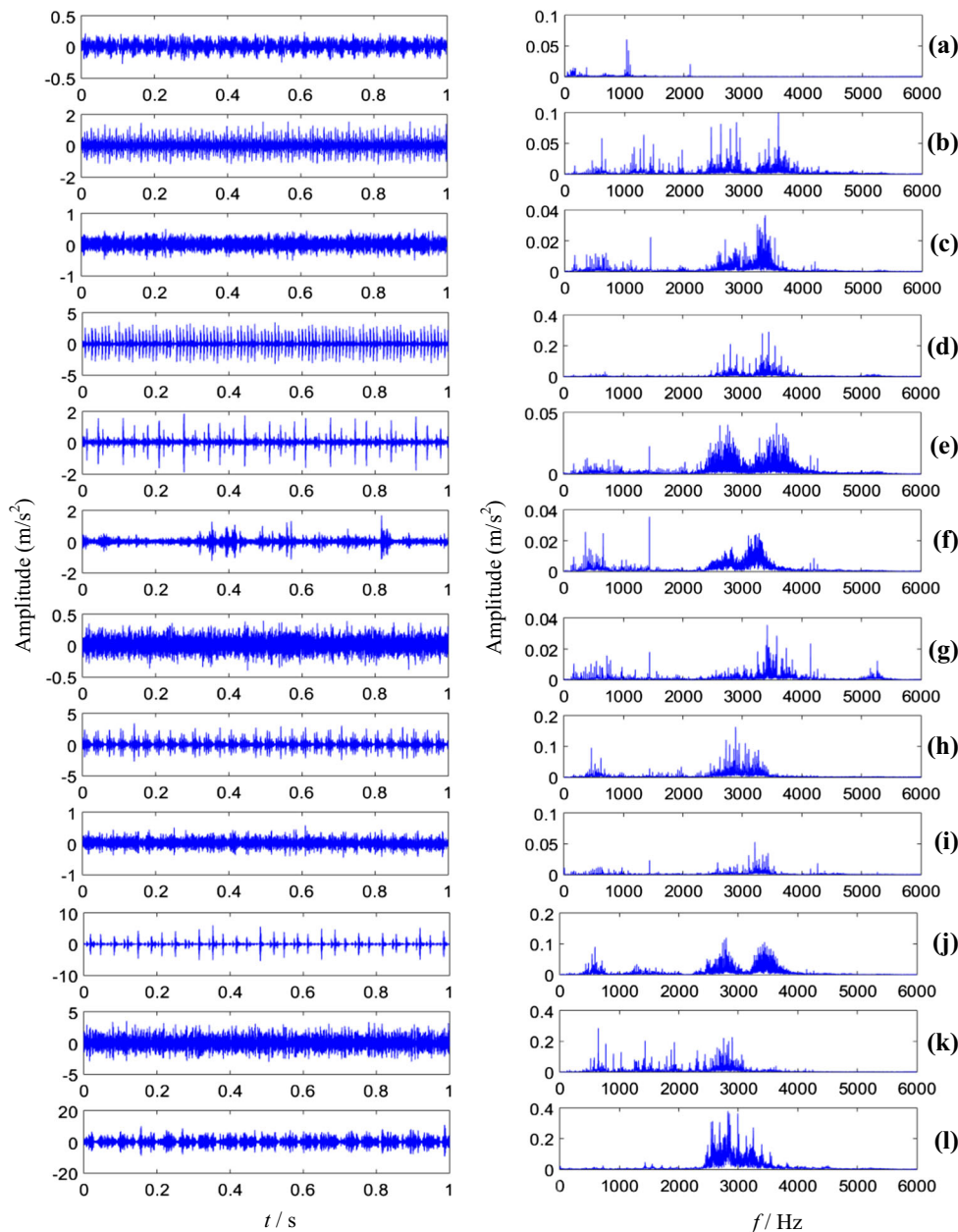
Table 3 The main parameters of the CNN model

CNN layer	Parameters
Input layer	Size of the input maps: 20×20
C_1	Kernel size: 5, number of output feature maps: 6
S_1	Scale: 2, number of output feature maps: 6
C_2	Kernel size: 5, number of output feature maps: 12
S_2	Scale: 2, number of output feature maps: 12
Other parameters	Maximum iterations: 100, learning rate: 1, batch size: 30

To validate the feature learning ability of the CNN model in diagnosing mechanical faults, the t-distributed stochastic neighbor embedding (t-SNE) technique is introduced for feature visualization. The feature learning process is described by three-dimensional scatter plot (3D plot) through the three representative stages, i.e., the input stage, the C_1-S_1 learning stage and the C_2-S_2 learning

(a) The MED based CNN
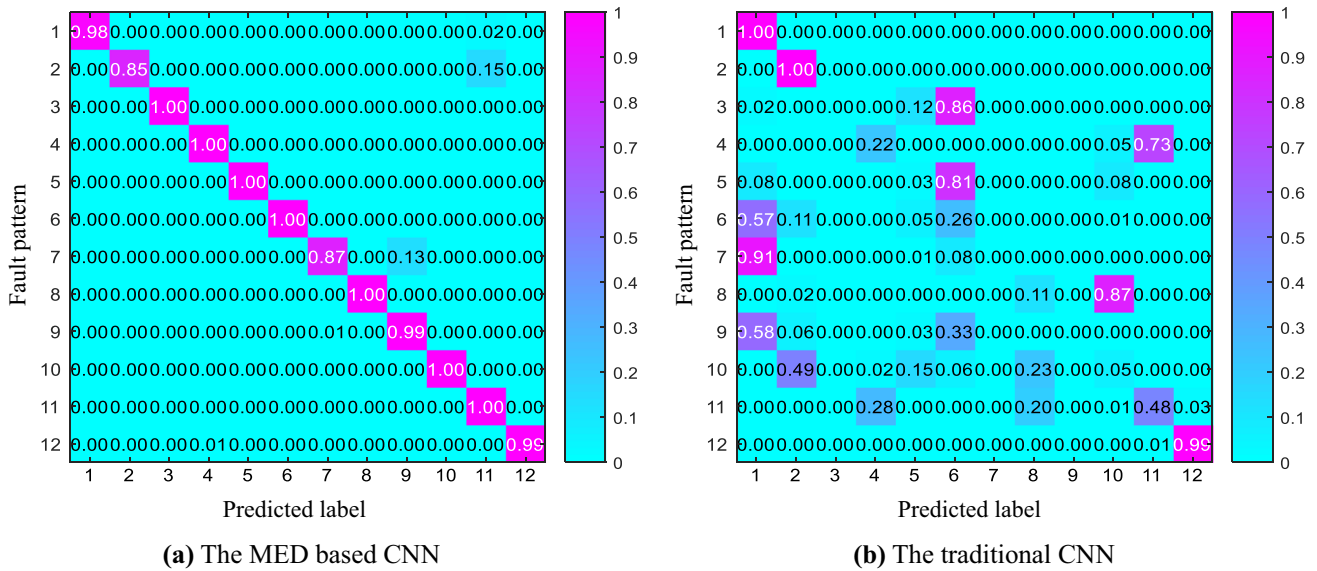
(b) The traditional CNN

Fig. 5 Multi-class confusion matrices in simulation verification

stage. In real world, we can get a complete visual experience by rotating the 3D plot to observe the learned features.

Figure 6a$_1$–a$_3$ gives vivid description of the feature learning processes of the MED–CNN (D1, D2 and D3 denote the three dimensions). Figure 6a$_1$ shows the features of the input data after the MED filtering. Pattern 8 and pattern 10 are well identified after the MED filtering process. However, the other ten patterns are still mixed together. Features of the twelve patterns are learned in progress through the second stage by C$_1$-S$_1$ operations. As shown in Fig. 6a$_2$, fault patterns can already be identified using the learned features. Compared with the second stage, features are further learned with smaller intra-category distances and larger inter-category distances, which can be seen in Fig. 6a$_3$.

The feature learning processes of the traditional CNN model are illustrated by the three stages. As shown in Fig. 6b$_1$–b$_3$, the features are learned constantly but still fail to describe for some individual pattern except for patterns 1, 2 and 4. From Fig. 5, we can conclude that the feature learning ability of the traditional CNN model is greatly enhanced by the advanced MED filtering.

Table 4 shows the comparison results among traditional stacked auto-encoder (SAE), MED-enhanced SAE (MED–SAE), traditional CNN and the present MED–CNN. We chose SAE here for comparison in that it is known as a feature mapping model which could also learn features automatically by encoder and decoder process. The architecture of the SAE here is 400-200-100-80-12, learning rate is 0.05, number of pre-training epochs is 10, and number of fine-tuning epochs is 500 (these are chosen by experiences). It can be seen that, the MED techniques indeed prompt the classification accuracy of SAE from 52.67% to

58.17%. However, it still cannot achieve a satisfactory result. Fortunately, the present MED-enhanced CNN shows superiority against others.

To further evaluate the effectiveness and robustness of the MED–CNN to detect multi-faults in mechanical systems, thirty trails are conducted. We train the model using 6 batch size under 5 iteration number conditions. Figure 7 shows the average accuracy ratio in the 30 trails. The maximum iteration number is chosen as 500, 400, 300, 200 and 100, respectively. The batch size is given successively as 30, 25, 20, 15, 10 and 5. The detailed accuracies in the 30 trails are listed in Tables 5 and 6, respectively.

As shown in Fig. 7, the MED–CNN gives a higher and steadier classification accuracy ratio compared with the traditional CNN. It can also be seen that, with the increase in iterations, the classification steadiness gets stronger. Moreover, in addition to trail 23, the batch size 30, 25, 15 and 10 give similar high classification accuracy while the batch size 5 gives lower classification accuracy instead. Deepening the network did not obviously improve the classification accuracy. MED–CNN with seven hidden layers (i.e., three convolutional layers, three subsampling layers, and one FC layer) are tested, and Table 7 lists the results.

5 Experimental investigation for axial piston pumps

5.1 Data source description

This section is devoted to the experimental investigations on the common faults diagnosis of axial piston pumps. Figure 8 shows the experimental platform. The tested axial

(a1)-(a3): The MED based CNN   **(b1)-(b3):** The traditional CNN

Fig. 6 The feature learning processes in simulation verification
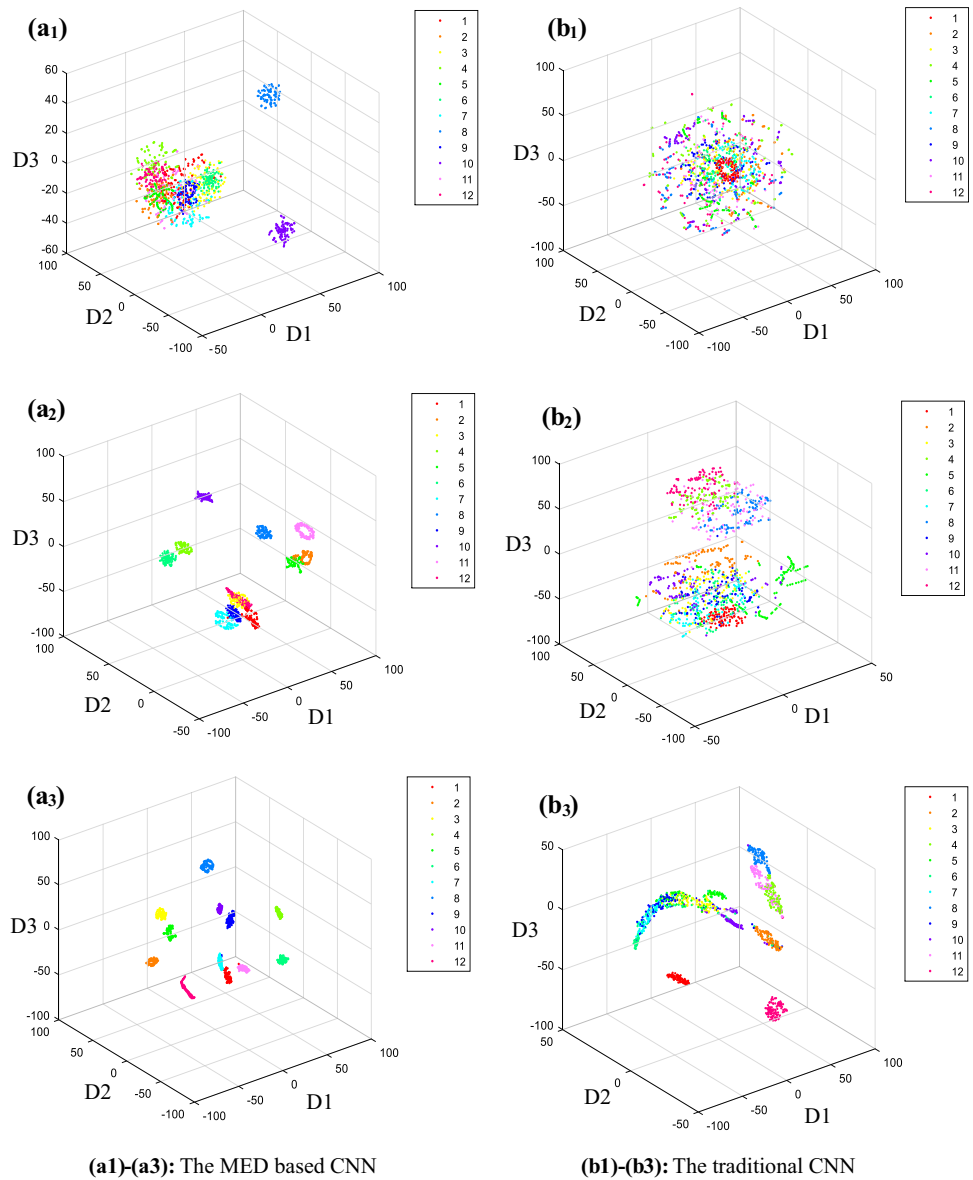
Table 4 Comparison results in simulation verification

Methods description	Classification accuracy (%)
Traditional SAE (400-200-100-80-12)	52.67
MED-SAE (the same as traditional SAE)	58.17
Traditional CNN	34.50
Present MED-CNN	97.33

piston pump (A in Fig. 8) is made by Ningbo Hilead Hydraulic Co., Ltd. (P. R. China), which is located at the end of the test rig. The data acquisition system includes a signal conditioner, a laptop with data acquisition software and multiple accelerometers. Some parameters of the tested axial piston pump are list in Table 8.

In the experimental investigation, one channel (accelerometer#3, marked with red circle in Fig. 8) vibration signals that sampled at 48 k Hz are applied. The pump running states and sample distributions are list in Table 9. Every kind operating state contains 160,000 data points. There are 1500 (300 × 5) training samples and 500 (100 × 5) testing samples, respectively.

The four common faults (shown in Fig. 9, marked with red circle) were as follows:

(a) Wear in three pistons, 0.03 mm wear amount in diameter to the tagged pistons.

(b) Blocked support hole in static pressure slippers.

(c) Wear in shaft shoulder, 0.03 mm wear amount in diameter.

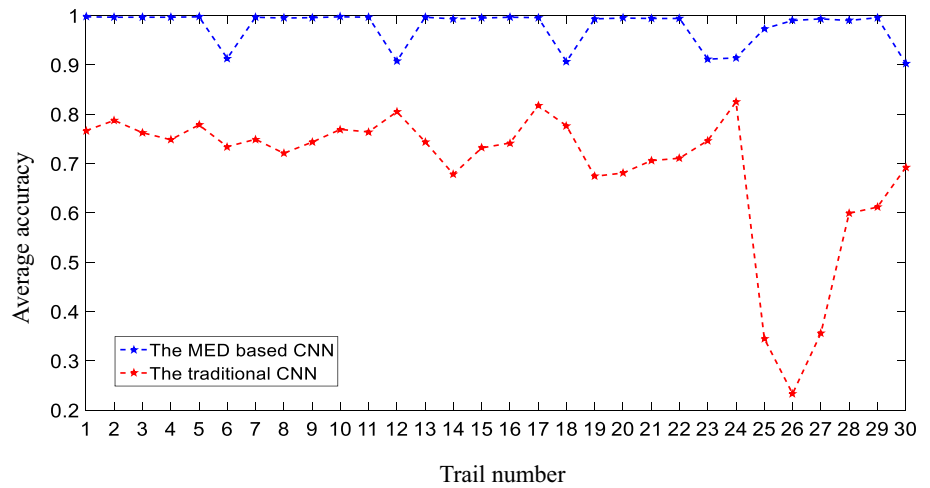**Fig. 7** The average accuracy ratio of the 30 trails in simulation verification

Table 5 The average accuracy ratio of the 30 trails based on MED-CNN in simulation verification

Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	0.9975	0.9967	0.9967	0.9933	0.9733
25	0.9967	0.9950	0.9933	0.9950	0.9900
20	0.9967	0.9958	0.9950	0.9942	0.9933
15	0.9967	0.9975	0.9967	0.9942	0.9900
10	0.9975	0.9967	0.9958	0.9117	0.9958
5	0.9133	0.9075	0.9067	0.9142	0.9033

Table 6 The average accuracy ratio of the 30 trails based on traditional CNN in simulation verification

Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	0.7667	0.7492	0.7442	0.6742	0.3450
25	0.7875	0.7208	0.6783	0.6808	0.2333
20	0.7625	0.7433	0.7317	0.7058	0.3558
15	0.7483	0.7692	0.7417	0.7108	0.5992
10	0.7783	0.7642	0.8175	0.7467	0.6117
5	0.7342	0.8050	0.7767	0.8250	0.6917

Table 7 Results of MED-CNN with 7 hidden layers in simulation verification

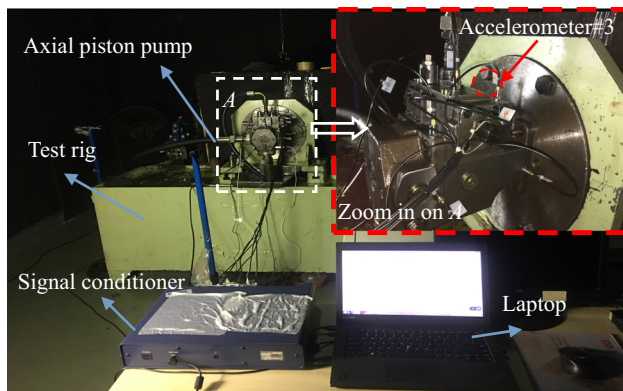| Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	0.9975	0.9967	0.9950	0.9958	0.9833
25	0.9975	0.9967	0.9958	0.9958	0.9408
20	0.9975	0.9967	0.9975	0.9125	0.9917
15	0.9967	0.9958	0.9967	0.9967	0.9908
10	0.9158	0.9975	0.9967	0.9983	0.7425
5	0.5825	0.9150	0.7483	0.8292	0.7442

Fig. 8 Experimental platform for the tested axial piston pump

(d) Cylinder block with a pitting defect, 0.5 mm in width, 0.3 mm in depth.

The time waveform of the vibration signals (the first 0.5 s) and the corresponding spectrums are displayed in Fig. 10.

5.2 Results and analysis

In order to realize automatic feature learning from the raw data, we adopt 400 data points as a sample. Therefore, the

Table 8 Some parameters of the tested pump

Parameter	Number of piston	Rated pressure (Mpa)	Rated pressure (rpm)
Value	9	35	1500

size of the input map is 20×20. In this case study, we adopt the same model parameters as listed in Table 3.

Figure 11 demonstrates the multi-class confusion matrices of the experimental data. Figure 11a, b gives the classification results using the MED–CNN and the traditional CNN. The average classification accuracy ratio using MED based CNN is 100%, and that of the traditional CNN remains 71.20%.

The fault patterns are totally identified as itself using the MED–CNN model, which is demonstrated by the diagonal line of confusion matrix in Fig. 11a. It is not so well of using the traditional CNN model. As shown in Fig. 11b, pattern 3 is identified as itself with 100%; 2% of pattern 1 is identified as pattern 4; 2% of pattern 2 is identified as pattern 1, 26% of pattern 2 is identified as pattern 4; 4% of pattern 4 is identified as pattern 1, 54% of pattern 4 is identified as pattern 2, and 6% of pattern 4 is identified as pattern 5; 18% of pattern 5 is identified as pattern 1, 9% of pattern 5 is identified as pattern 2 and 23% of pattern 5 is identified as pattern 4.

To validate the feature learning ability of the CNN model in diagnosing multi-faults of axial piston pumps, three-dimensional scatter plot is employed to describe the learning results through the three representative stages.

Figure $12a_1$–a_3 gives vivid description of the pump fault feature learning processes of the MED–CNN. Figure $12a_1$ shows the features of the pump data after the MED filtering. The fault features are comparably visible even with large intra-category distances after the MED filtering process. Features of the five patterns of axial piston pump are learned by C_1-S_1 operations. As shown in Fig. $12a_2$, fault patterns are explicitly observed using the learned features.

Table 9 Fault category and sample distribution in experimental investigation

Axial piston pump state	Size of training samples/testing samples		Fault pattern
Normal state	300	100	1
Wear in three pistons	300	100	2
Blocked support hole in static pressure slippers	300	100	3
Wear in shaft shoulder	300	100	4
Cylinder block with a pitting defect	300	100	5

(a) piston wear   **(b)** support hole blocked   **(c)** shaft shoulder wear   **(d)** cylinder block pitting
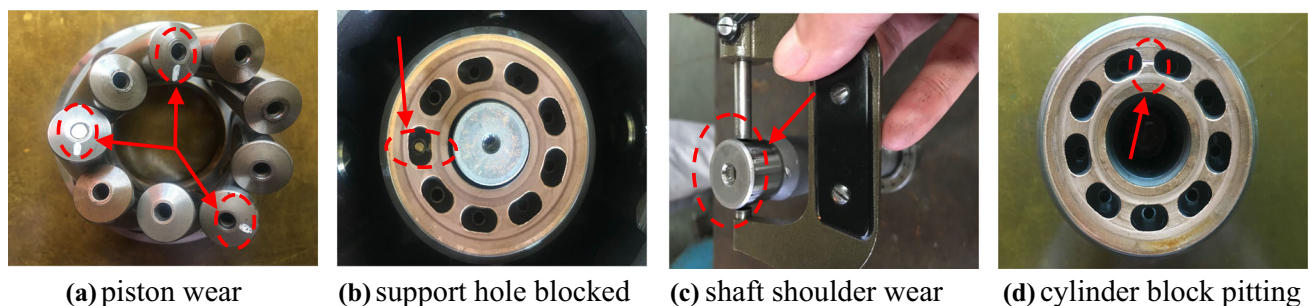
Fig. 9 The four common faults in the axial piston pump

Fig. 10 Time waveform and corresponding spectrums **a** normal state, **b** wear in three pistons, **c** blocked support hole in static pressure slippers, **d** wear in shaft shoulder, **e** Cylinder block with a pitting defect
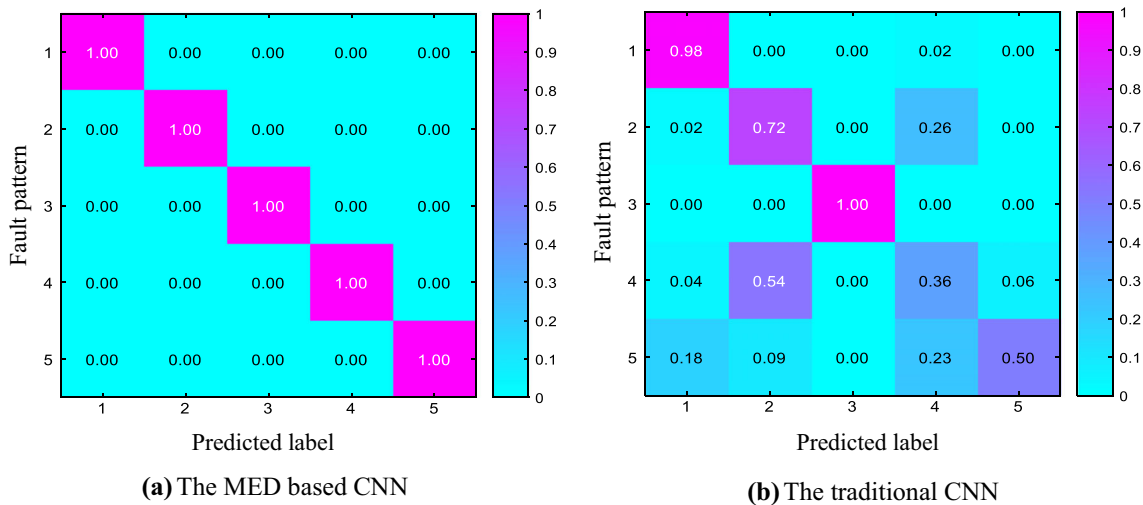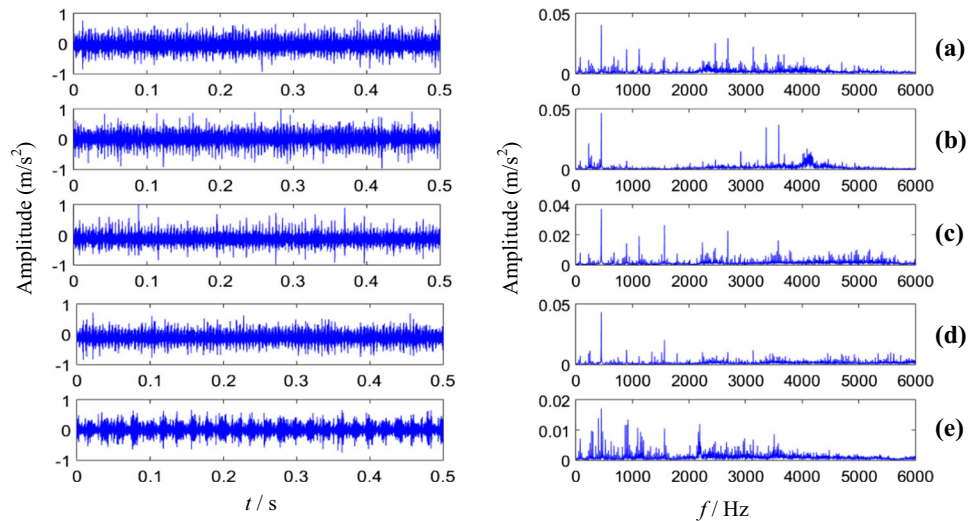

(a) The MED based CNN

(b) The traditional CNN

Fig. 11 Multi-class confusion matrices in experimental investigation

Different from the results of the second stage, features are further learned with smaller intra-category distances and larger inter-category distances, which can be seen in Fig. 12a$_3$.

Similarly, the pump fault feature learning processes of the traditional CNN model are also illustrated by the three stages. As shown in Fig. 12b$_1$–b$_3$, the features are learned constantly but still failed to describe for some individual pattern except for pattern 3. The traditional CNN has limited ability to describe the explicit features of the axial piston pump data.

Table 10 shows the comparison results in the case study. The architecture of the SAE here is 400-200-100-80-5, learning rate is 0.05, number of pre-training epochs is 10, and number of fine-tuning epochs is 500 (the same as in

simulation verification). The results also demonstrate the effectiveness of the present MED–CNN.

To further evaluate the effectiveness and robustness of the MED–CNN to detect multi-faults in the axial piston pump, thirty trails are conducted, which is the same as in the simulation verification section.

Figure 13 shows the average accuracy ratio of axial piston pump in the 30 trails. Similar to the simulation verification, the maximum iteration number is chosen as 500, 400, 300, 200 and 100, respectively. Meanwhile, the batch size is given as 30, 25, 20, 15, 10 and 5. The detailed accuracies in the 30 trails are listed in Tables 11 and 12, respectively.

As shown in Fig. 13, the MED–CNN gives a higher and steadier classification accuracy ratio compared with the

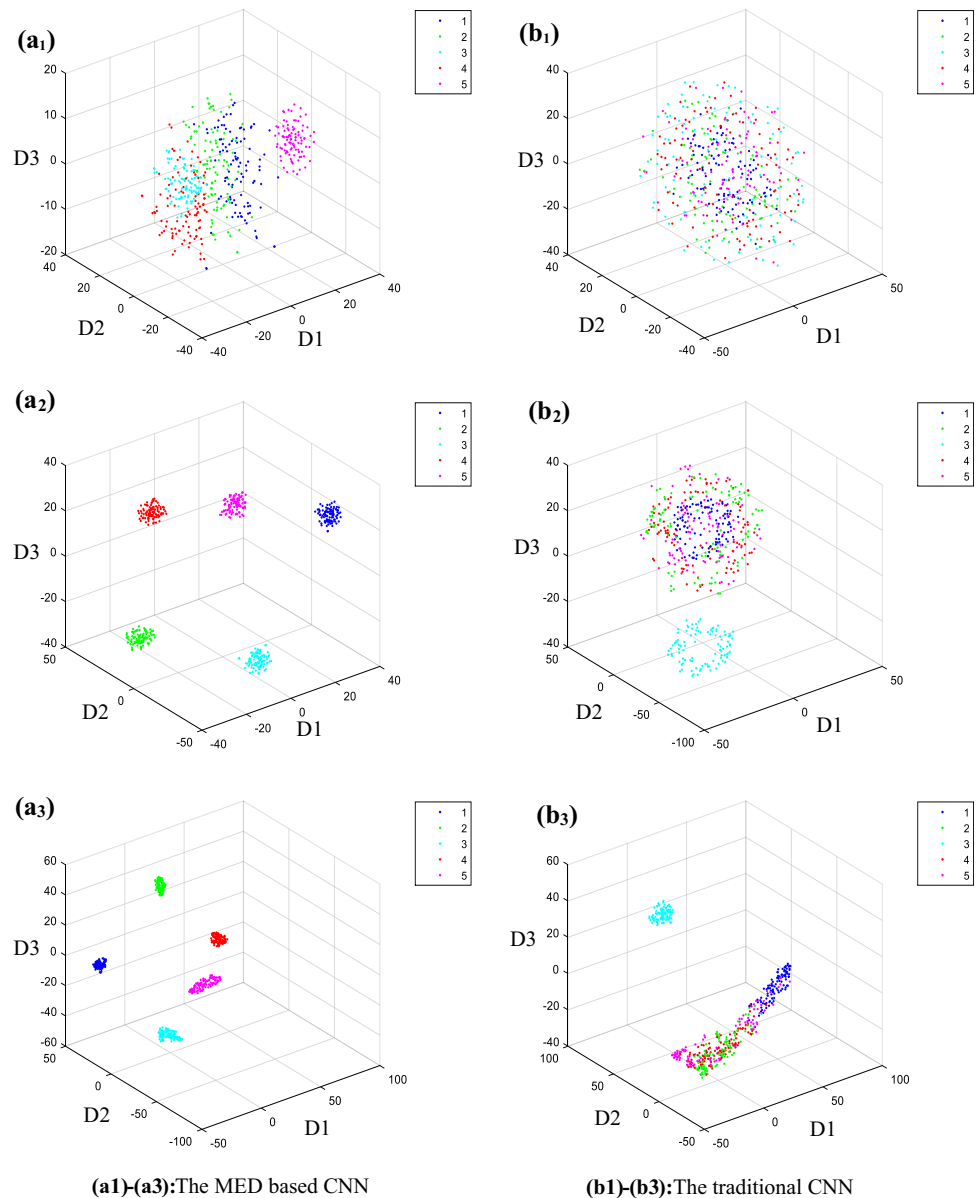**Fig. 12** The feature learning processes in experimental investigation

(a₁)

(b₁)

(a₂)

(b₂)

(a₃)

(b₃)

(a1)-(a3):The MED based CNN

(b1)-(b3):The traditional CNN

Table 10 Comparison results in experimental investigation

Methods description	Classification accuracy (%)
Traditional SAE (400-200-100-80-5)	61.00
MED-SAE (the same as traditional SAE)	79.00
Traditional CNN	71.20
Present MED-CNN	100.0

traditional CNN. With the increase in iterations, the classification accuracies tend to be steadier with high values, which illustrate the effectiveness and robustness of the MED–CNN to detect multi-faults in axial piston pumps.

The MED-CNN with seven hidden layers in the case study is also tested whose results are given in Table 13.

6 Conclusion

Faults occurred in the piston pumps are difficult to be detected due to the complex working environment in hydraulic systems. In order to get rid of feature selection expertise, this paper proposed a MED based CNN model to automatically detect faults in axial piston pump. Both simulations and experiments are conducted to investigate the performance of the present model by comparing with the traditional CNN model. Using the present model, the

Fig. 13 The average accuracy ratio of the 30 trails in experimental investigation
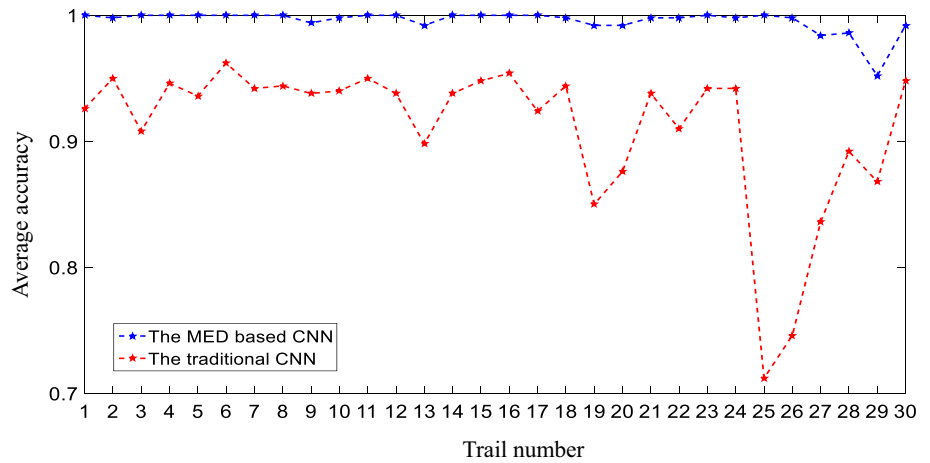

Table 11 The average accuracy ratio of the 30 trails based on MED-CNN in experimental investigation

Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	1	1	0.9920	0.9920	1
25	0.9980	1	1	0.9920	1
20	1	0.9940	1	0.9980	0.9840
15	1	0.9980	1	0.9980	0.9860
10	1	1	1	1	0.9520
5	1	1	0.9980	0.9980	0.9920

Table 12 The average accuracy ratio of the 30 trails based on traditional CNN in experimental investigation

Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	0.9260	0.9420	0.8980	0.8500	0.7120
25	0.9500	0.9440	0.9380	0.8760	0.7460
20	0.9080	0.9380	0.9480	0.9380	0.8360
15	0.9460	0.9400	0.9540	0.9100	0.8920
10	0.9360	0.9500	0.9240	0.9420	0.8680
5	0.9620	0.9380	0.9440	0.9420	0.9480

Table 13 Results of MED-CNN with 7 hidden layers in experimental investigation

Batch size	Trails 1–6 (500 iterations)	Trails 7–12 (400 iterations)	Trails 13–18 (300 iterations)	Trails 19–24 (200 iterations)	Trails 25–30 (100 iterations)
30	1	1	0.9980	0.8000	0.800
25	1	1	1	0.8000	0.9320
20	1	1	1	0.9920	0.8000
15	1	1	1	0.9980	0.8000
10	1	1	1	0.97	0.6320
5	1	0.9980	0.9980	0.7980	0.9900

average classification accuracy ratios for benchmark data with 12 classes of bearing operating states from the CWRU and for experimental data with five classes of commonly occurred faults from the experimental axial piston pump, are 97.33% and 100%, respectively. However, the average classification accuracy ratios for the traditional CNN model

only attain 34.50% and 71.20%, respectively. In addition, the superiority for classification robustness of the present model is illustrated by thirty trails under different iteration numbers and batch sizes. Therefore, the proposed model can learn effective fault features with comparable satisfactory results in multi-fault classifications and is expecting to classify faults in more complex mechanical systems.

Compliance with ethical standards

Conflict of interest The authors declare that there is conflict of interest regarding the publication of this paper.

References

Antoni J (2016) The infogram: entropic evidence of the signature of repetitive transients. Mech Syst Signal Process 74:73–94

Appana DK, Prosvirin A, Kim JM (2018) Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks. Soft Comput 22(20):6719–6729

Bouvrie J (2006) Notes on convolutional neural networks. http://cogprints.org/5869/1/cnn_tutorial.pdf. Accessed 22 Nov 2006

Ding XX, He QB (2017) Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. IEEE Trans Instrum Meas 66(8):1926–1935

Dong GM, Chen J, Zhao FG (2017) Incipient bearing fault feature extraction based on minimum entropy deconvolution and K-singular value decomposition. J Manuf Sci Eng 139(10):101006

Du ZH, Chen XF, Zhang H (2017) Convolutional sparse learning for blind deconvolution and application on impulsive feature detection. IEEE Trans Instrum Meas 67(2):338–349

Endo H, Randall RB (2007) Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. Mech Syst Signal Process 21(2):906–919

Ferdowsi H, Jagannathan S, Zawodniok M (2014) An online outlier identification and removal scheme for improving fault detection performance. IEEE Trans Neur Netw Learn Syst 25(5):908–919

Fu XB, Liu B, Zhang YC, Lian LN (2014) Fault diagnosis of hydraulic system in large forging hydraulic press. Measurement 49:390–396

Gao YJ, Zhang Q (2006) A wavelet packet and residual analysis based method for hydraulic pump health diagnosis. Proc Inst Mech Eng Part D J Automob Eng 220(6):735–745

He D, Wang XF, Li SC, Lin J, Zhao M (2016) Identification of multiple faults in rotating machinery based on minimum entropy deconvolution combined with spectral kurtosis. Mech Syst Signal Process 81:235–249

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

http://csegroups.case.edu/bearingdatacenter/home

Huang J, Wang YN, Liu ZL, Guan BY, Long D, Du XP (2016) On modeling microscopic vehicle fuel consumption using radial basis function neural network. Soft Comput 20(7):2771–2779

Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M (2016) Real-time motor fault detection by 1-d convolutional neural networks. IEEE Trans Ind Electron 63(11):7067–7075

Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, Van de Walle R, Van Hoecke S (2016) Convolutional neural network based fault detection for rotating machinery. J Sound Vib 377:331–345

Jegadeeshwaran R, Sugumaran V (2015) Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. Mech Syst Signal Process 52–53:436–446

Jiang WL, Zheng Z, Zhu Y, Li Y (2014) Demodulation for hydraulic pump fault signals based on local mean decomposition and improved adaptive multiscale morphology analysis. Mech Syst Signal Process 58–59:179–205

Lan Y, Hu JW, Huang JH, Niu LK, Zeng XH, Xiong XY, Wu B (2018) Fault diagnosis on slipper abrasion of axial piston pump based on extreme learning machine. Measurement 124:378–385

LeCun Y, Bengio Y (1998) Convolutional networks for images, speech, and time-series. In: Arbib MA (ed) The handbook of brain theory and neural networks. MIT Press, Cambridge, MA

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444

Li G, Zhao Q (2017) Minimum entropy deconvolution optimized sinusoidal synthesis and its application to vibration based fault detection. J Sound Vib 390:218–231

Li JM, Li M, Zhang JF (2017) Rolling bearing fault diagnosis based on time-delayed feedback monostable stochastic resonance and adaptive minimum entropy deconvolution. J Sound Vib 401:139–151

Liu P, Choo KKR, Wang LZ, Huang F (2017) SVM or deep learning? A comparative study on remote sensing image classification. Soft Comput 21(23):7053–7065

Liu YJ, Huang H, Cao JD, Huang TW (2018) Convolutional neural networks-based intelligent recognition of Chinese license plates. Soft Comput 22(7):2403–2419

Lu CQ, Wang SP, Tomovic M (2015) Fault severity recognition of hydraulic piston pumps based on EMD and feature energy entropy. In: Proceedings of IEEE 10th conference on industrial electronics and applications (ICIEA), pp 489–494

Lu CQ, Wang SP, Zhang C (2016) Fault diagnosis of hydraulic piston pumps based on a two-step EMD method and fuzzy C-means clustering. P I Mech Eng C-J Mec 230(16):2913–2928

Pandya DH, Upadhyay SH, Harsha SP (2014) Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. Soft Comput 18(2):255–266

Qiao ZJ, Lei YG, Lin J, Jia F (2017) An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis. Mech Syst Signal Process 84:731–746

Rafique MA, Pedrycz W, Jeon M (2018) Vehicle license plate detection using region-based convolutional neural networks. Soft Comput 22(19):6429–6440

Samanta B, Al-Balushi KR, Al-Araimi SA (2006) Artificial neural networks and genetic algorithm for bearing fault detection. Soft Comput 10(3):264–271

Sawalhi N, Randall RB, Endo H (2007) The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech Syst Signal Process 21(6):2616–2633

Sepasi M, Sassani F (2010) On-line fault diagnosis of hydraulic systems using unscented kalman filter. Int J Control Autom 8(1):149–156

Suganyadevi MV, Babulal CK, Kalyani S (2016) Assessment of voltage stability margin by comparing various support vector regression models. Soft Comput 20(2):807–818

Trujillo MCR, Alarcon TE, Dalmau OS, Ojeda AZ (2017) A Segmentation of carbon nanotube images through an artificial neural network. Soft Comput 21(3):611–625

van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

Vu TD, Ho NH, Yang HJ, Kim J, Song HC (2018) Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection. Soft Comput 22(20):6825–6833

Wang YX, Liang M (2012) Identification of multiple transient faults based on the adaptive spectral kurtosis method. J Sound Vib 331(2):470–486

Wang YX, He ZJ, Zi YY (2009) A demodulation method based on improved local mean decomposition and its application in rub-impact fault diagnosis. Meas Sci Technol 20(2):025704

Wang SH, Xiang JW, Zhong YT, Tang HS (2018a) A data indicator-based deep belief networks to detect multiple faults in axial piston pumps. Mech Syst Signal Process 112:154–170

Wang SH, Xiang JW, Zhong YT, Zhou YQ (2018b) Convolutional neural network-based hidden Markov models for rolling element bearing fault identification. Knowl-Based Syst 144:65–76

Xiang JW, Zhong YT (2016) A novel personalized diagnosis methodology using numerical simulation and an intelligent method to detect faults in a shaft. Appl Sci 6(12):414

Xiang JW, Zhong YT, Gao HF (2015) Rolling element bearing fault detection using PPCA and spectral kurtosis. Measurement 75:180–191

Xu XF, Qiao ZJ, Lei YG (2018) Repetitive transient extraction for machinery fault diagnosis using multiscale fractional order entropy infogram. Mech Syst Signal Process 23(5):1573–1585

Xu XF, Lei YG, Li ZD (2019) An incorrect data detection method for big data cleaning of machinery condition monitoring. IEEE Transactions on Industrial Electronics 1:2–3. https://doi.org/10.1109/tie.2019.2903774

Zhao Z, Jia MX, Wang FL, Wang S (2009) Intermittent chaos and sliding window symbol sequence statistics-based early fault diagnosis for hydraulic pump on hydraulic tube tester. Mech Syst Signal Process 103:312–326