

Fault diagnosis of axial piston pumps with multi-sensor data and convolutional neural network

Qun CHAO^{a,b,c}, Haohan GAO^a, Jianfeng TAO (✉)^{a,c}, Chengliang LIU^{a,c}, Yuanhang WANG^d, Jian ZHOU^d

^a State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

^b State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China

^c MoE Key Laboratory of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai 200240, China

^d China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 510610, China

✉ Corresponding author. E-mail: jftao@sjtu.edu.cn (Jianfeng TAO)

© Higher Education Press 2022

ABSTRACT Axial piston pumps have wide applications in hydraulic systems for power transmission. Their condition monitoring and fault diagnosis are essential in ensuring the safety and reliability of the entire hydraulic system. Vibration and discharge pressure signals are two common signals used for the fault diagnosis of axial piston pumps because of their sensitivity to pump health conditions. However, most of the previous fault diagnosis methods only used vibration or pressure signal, and literatures related to multi-sensor data fusion for the pump fault diagnosis are limited. This paper presents an end-to-end multi-sensor data fusion method for the fault diagnosis of axial piston pumps. The vibration and pressure signals under different pump health conditions are fused into RGB images and then recognized by a convolutional neural network. Experiments were performed on an axial piston pump to confirm the effectiveness of the proposed method. Results show that the proposed multi-sensor data fusion method greatly improves the fault diagnosis of axial piston pumps in terms of accuracy and robustness and has better diagnostic performance than other existing diagnosis methods.

KEYWORDS axial piston pump, fault diagnosis, convolutional neural network, multi-sensor data fusion

1 Introduction

Axial piston pumps play the role of “heart” in hydraulic systems with a wide variety of applications in the fields of construction, agriculture, aerospace, and robotics. The pump delivers pressurized fluid to other hydraulic components by converting rotating mechanical energy into fluid power. Axial piston pumps often need to operate under harsh working conditions, such as high pressure, high speed, and high temperature [1,2]. The pump failure will cause the breakdown of hydraulic systems, increase in maintenance time and cost, and even catastrophic accidents. Therefore, monitoring and recognizing the pump health conditions in real time are necessary and urgent.

Over the last several decades, various machine learning methods, including artificial neural network [3], support

vector machine [4], extreme learning machine [5], XGBoost [6], and k -nearest neighbor, and decision trees, have been applied for the fault diagnosis of axial piston pumps [7]. A common drawback of these traditional machine learning methods lies in the manual feature extraction. Therefore, in recent years, deep learning has become popular in the field of fault diagnosis of rotating machinery because of its powerful end-to-end ability. Many researchers have also found applications of deep learning methods in the fault diagnosis of axial piston pumps, such as deep belief network [8] and one-dimensional (1D) and two-dimensional (2D) convolutional neural networks (CNNs) [9–15]. Among these previous studies, the vibration signal is most frequently selected to monitor pump health conditions [4–14] because it contains abundant fault information. Meanwhile, the discharge pressure signal appears to be suitable for monitoring the performance of axial piston pumps [3,16–18].

Machine learning, especially recent deep learning, has great potential for the fault diagnosis of axial piston pumps. Most previous studies used single-sensor data for the fault diagnosis of axial piston pumps. However, an axial piston pump is a complicated fluid–solid–thermal coupling system and single-sensor data are insufficient for the accurate and reliable fault diagnosis in noisy environment [19,20]. For example, even though an accelerometer is installed at the same location of the pump housing, the vibration signals differ significantly from each other in three orthogonal directions [9,10].

By contrast, multi-sensor data contain redundant and complementary information for an accurate and reliable fault diagnosis. Multi-sensor data fusion can be fulfilled at three levels, namely, data, feature, and decision levels [21]. Among the three fusion levels, the most popular data-level fusion requires less expert knowledge and loses less information than the two other fusion levels [22]. Therefore, data-level fusion is attracting increasing attention in the field of fault diagnosis. Recently, data-level fusion and machine learning have been combined for the fault diagnosis of rotating machinery, such as bearing [23,24], electric motor [25], and gearbox [22,26,27]. The common signals to be fused include vibration signal, current signal, and acoustic signal, among which multiple vibration signals are typically fused for the fault diagnosis of rotating machinery [23,24,26,28,29].

Although deep learning-based intelligent methods have achieved state-of-the-art performance in the fault diagnosis of rotating machinery, some limitations are still observed in previous studies:

(1) Most CNN-based studies on the fault diagnosis of axial piston pumps focus on single-sensor data, especially vibration signal. Only few studies have used discharge pressure and vibration signals simultaneously for the fault diagnosis of axial piston pumps.

(2) Although data fusion has been used in the fault

diagnosis of bearing, electric motor, and gearbox, only few research reported the application of data fusion to the fault diagnosis of axial piston pumps.

Therefore, the main contributions of this work are presented as follows:

(1) Vibration and discharge pressure signals are used simultaneously for the fault diagnosis of axial piston pumps to obtain a diagnostic performance that is better than that of single-sensor data.

(2) The proposed method, which can accurately recognize the four different degradation states of an axial piston pump in noisy environment, realizes multi-sensor data fusion at the data and feature levels.

The remainder of this paper is structured as follows. Section 2 describes the experimental setup and data acquisition for an axial piston pump. Section 3 outlines the theoretical basis of CNN and short-time Fourier transform (STFT). Section 4 details the proposed multi-sensor data fusion method based on CNN. Section 5 presents and discusses the comparative results to validate the proposed method. Finally, Section 6 provides the concluding remarks.

2 Experiments on an axial piston pump

2.1 Machine description

Figure 1 schematically shows a typical construction of the swash-plate type axial piston pump. Two bearings at the shaft ends support the whole rotating group, including a driving shaft, a cylinder block, and piston–slipper assemblies. Each pair of piston and slipper is connected by a ball-and-socket joint that allows rotation motions of the slipper around the piston ball. The cylinder block accommodates an odd number of pistons at equal angular intervals and has a slight contact with the stationary valve plate via a built-in compressed spring. The valve plate

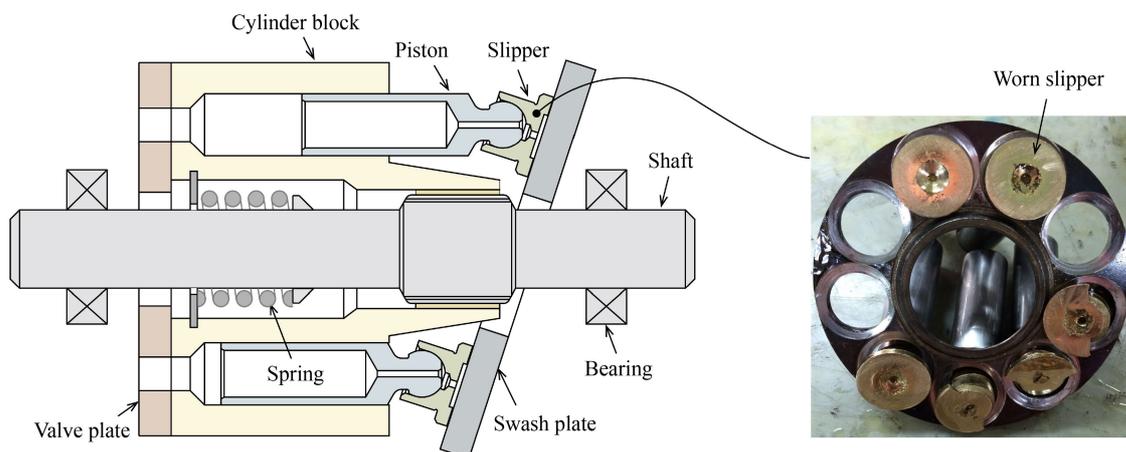


Fig. 1 Illustration of an axial piston pump and worn slippers.

communicates with the pump inlet and outlet ports through two kidney-shaped slots. Once the shaft drives the cylinder block to rotate, the pistons reciprocate within the cylinder bores because of the action of the sliding slippers on the swash plate. As a result, low-pressure fluid enters the cylinder block through the pump inlet port and then high-pressure fluid discharges from the cylinder block to the pump outlet port. The above motions repeat themselves to generate a continuous delivery flow for each shaft revolution.

A large number of pump maintenance cases suggest that the slipper is a common failure part in axial piston machines [30]. The sliding slippers are subjected to heavy loads from the pistons, thereby leading to contact wear at their bottom faces [1,31] (Fig. 1). The slipper wear increases the clearance between the slippers and the swash plate and hence causes severe high-frequency impact between them. The metal-to-metal impact is transmitted to the pump housing and end cover, which will generate abnormal vibration. In addition, large clearance between the slippers and the swash plate enhances the discharge pressure fluctuation because of increased leakage flow from the slipper pairs [32]. The transient and periodic vibration and discharge pressure signals contain rich information about the health state of axial piston pumps and thus can be used for the condition monitoring of axial piston pumps.

2.2 Experimental setup and data acquisition

A test rig was constructed to collect the vibration and discharge pressure signals from a swash-plate type axial piston pump under healthy and faulty conditions. Figure 2(a) [2] depicts the hydraulic circuit diagram of the test rig. The electric motor (M1) provided power for the test pump (Fig. 2(b) [33]) that operated at a discharge pressure of 21 MPa, a rotational speed of 10000 r/min, and a full displacement of 1.3 mL/r. The test pump received low-pressure hydraulic oil at its inlet port and discharged high-pressure hydraulic oil at its outlet port. The test pump communicated with the pressurized reservoir (PR1) through the inlet port when the ball valve (B1) opened. The proportional relief valve (R1) downstream the test pump served as an adjustable load on the test pump. During the pump operation, the oil temperature at the inlet port of the test pump was maintained at $(60 \pm 2) ^\circ\text{C}$ to avoid temperature effects on the pump performance.

Several transducers were used to monitor the performance of the test pump. A tri-axial accelerometer was fixed to the pump end cover to record the vibration signals in three orthogonal directions. A high-frequency dynamic pressure transducer (P2) was installed near the outlet port of the test pump to measure the dynamic discharge pressure. The vibration and discharge pressure

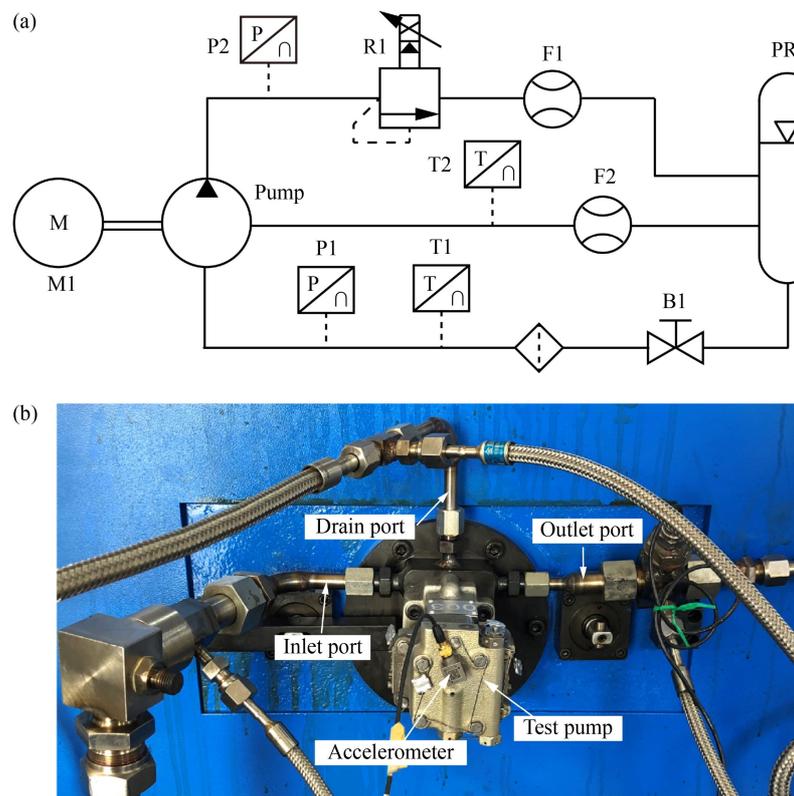


Fig. 2 (a) Hydraulic circuit diagram of the test rig [2] and (b) the test pump [33]. Reproduced with permission from Springer Nature, copyright 2021.

signals were acquired synchronously at a sampling frequency of 10240 Hz by a data acquisition card from National Instruments (NI). In addition, a flow meter (F2) was located downstream the drain port of the test pump to record the case drain flow rate (also called external leakage flow rate), which was a good indicator of the pump health conditions.

The experiments were carried out on the test pump under healthy and faulty conditions. To reflect the degradation process of worn friction pairs, four different degradation states were simulated by adjusting the clearance between the slippers and the swash plate. Table 1 lists four different levels of leakage flow rates to represent four pump degradation states, where the case drain flow increased with the clearance to simulate the gradual wear process of the friction pairs in actual applications.

Figure 3 shows the vibration signals under different degradation states. The vibration in each direction becomes increasingly intense with the increasing clearance and leakage flow rate of the slipper pairs. Each health condition exhibits different vibration behaviors for

Table 1 Different degradation states of the axial piston pump

Degradation state	Increased clearance/mm	Leakage flow rate/(L·min ⁻¹)
Normal	0.00	0.40
Slight leakage	0.05	0.57
Medium leakage	0.15	0.65
Severe leakage	0.20	1.10

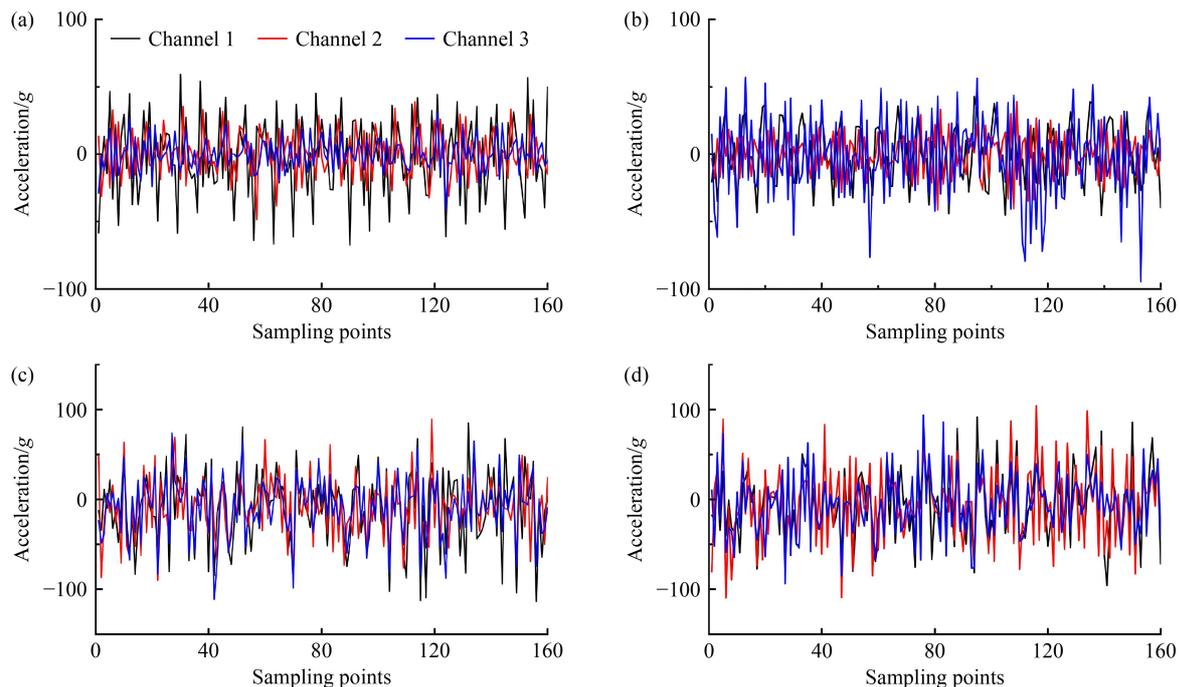


Fig. 3 Raw vibration signals for four different pump degradation states: (a) normal; (b) slight leakage; (c) medium leakage; and (d) severe leakage.

three vibration signal channels because the periodic shock and vibration have exclusive transmission paths from the slipper pairs to the pump housing and end cover in different directions [20]. One can hardly distinguish the waveform signals between two neighboring pump health conditions. Moreover, it is difficult to identify the most sensitive channel of the vibration signals to the pump health conditions.

Figure 4 presents the dynamic discharge pressure of the test pump, where the pressure ripples at individual degradation states differ from one another under the same working conditions. The difference in pressure ripples arises from the different contributions of the leakage flow to the displacement chamber pressure of the axial piston pumps [32,34].

3 Theoretical basis

3.1 Convolutional neural network

As a popular deep learning architecture, the CNN is a multi-layer feed-forward neural network. CNN has a strong ability in automatic feature extraction and nonlinear mapping from 2D data because of its characteristics of local connection, weight sharing, and sub-sampling. A typical CNN mainly contains an input layer, a convolutional layer, a pooling layer, and a full-connected layer.

The convolutional layer plays an important role in automatic feature extraction. It applies a group of

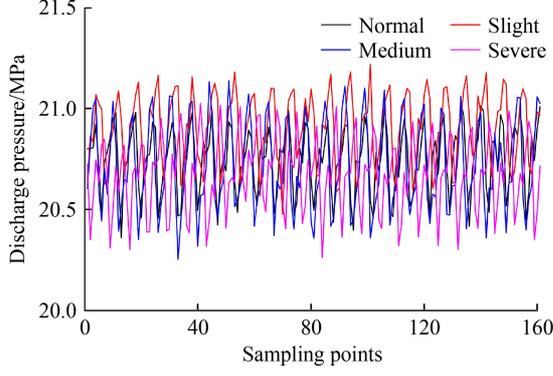


Fig. 4 Raw discharge pressure signals for four different pump degradation states.

convolution kernels (also known as filters) to extract features from the input image separately. The filter size determines the local receptive field to be convolved. The feature maps are generated by the following convolution and nonlinear mapping operations [22].

$$\mathbf{X}_k^l = f \left(\sum_{c=1}^C (\mathbf{W}_{c,k}^l * \mathbf{X}_c^{l-1}) + \mathbf{B}_k^l \right), \quad (1)$$

where symbol $*$ represents a 2D convolution operation, superscript l denotes the index of network layers, subscript k denotes the index of group filters or output feature maps, subscript c denotes the index of channels for input feature maps or the group filters, C is the total number of filter channels, \mathbf{X}_k^l is the k th output feature map at layer l , \mathbf{X}_c^{l-1} is the c th-channel component of the input feature map at layer $(l-1)$, $\mathbf{W}_{c,k}^l$ is the c th-channel component of the k th group filter weight at layer l , and \mathbf{B}_k^l is the bias of the k th group filter at layer l . $f(\cdot)$ is an activation function applied to each output feature map. The rectified linear unit $f(x) = \max(0, x)$ is often selected as the activation function due to its superiority to other activation functions in terms of over-fitting avoidance and training process acceleration [35].

The pooling layer (also called subsampling layer) usually follows the convolutional layer to reduce the dimension of feature maps and the number of trainable parameters. Max-pooling and average-pooling are two common pooling operations for the pooling layer, and they output the maximum and average values from the windowed elements, respectively. Max-pooling is often adopted at the pooling layer in the image recognition task because of its advantageous texture feature extraction. The max-pooling operation is expressed as

$$p_k^{l+1} = \max \{ a_{hw} \in A_k^l \mid (i'-1)H+1 \leq h \leq i'H, \\ (j'-1)W+1 \leq w \leq j'W \}, \quad (2)$$

where A_k^l is the k th feature map at layer l and a_{hw} is its element at pixel (h, w) within the pooling window with height H and width W , i' and j' are the height and width indices of element pixels, respectively, and p_k^{l+1} is the maximum element within the pooling window.

After a group of convolutional and pooling layers, the input image is flattened into a 1D array. The fully connected layer receives the 1D array to further reduce the array dimension. Finally, a softmax classifier is used to calculate the multi-class probability as follows:

$$p \{ y^{(s)} = q \mid \mathbf{x}^{(s)}; \boldsymbol{\theta}^L \} = \begin{bmatrix} p(y^{(s)} = 1 \mid \mathbf{x}^{(s)}; \boldsymbol{\theta}^L) \\ p(y^{(s)} = 2 \mid \mathbf{x}^{(s)}; \boldsymbol{\theta}^L) \\ \vdots \\ p(y^{(s)} = Q \mid \mathbf{x}^{(s)}; \boldsymbol{\theta}^L) \end{bmatrix} \\ = \frac{1}{\sum_{q=1}^Q \exp \{ (\boldsymbol{\theta}_q^L) \cdot \mathbf{x}^{(s)} \}} \begin{bmatrix} \exp \{ (\boldsymbol{\theta}_1^L) \cdot \mathbf{x}^{(s)} \} \\ \exp \{ (\boldsymbol{\theta}_2^L) \cdot \mathbf{x}^{(s)} \} \\ \vdots \\ \exp \{ (\boldsymbol{\theta}_Q^L) \cdot \mathbf{x}^{(s)} \} \end{bmatrix}, \quad (3)$$

where $\mathbf{x}^{(s)}$ is the s th sample, $y^{(s)}$ is the predicted label, $\boldsymbol{\theta}^L = [\boldsymbol{\theta}_1^L \ \boldsymbol{\theta}_2^L \ \dots \ \boldsymbol{\theta}_Q^L]$ represents the trainable parameters including weight and bias at the last layer L , and lowercase letter q and uppercase letter Q denote the q th class and total classification number, respectively. The

sum term $1 / \left\{ \sum_{q=1}^Q \exp \{ (\boldsymbol{\theta}_q^L) \cdot \mathbf{x}^{(s)} \} \right\}$ aims to normalize the probability distribution so that the probability summation of all classes is equal to unity.

The classification task usually adopts cross-entropy to be a loss function in machine learning models. The cross-entropy function measures the “distance” between real and predicted probability distributions. For a multi-classification task, the loss function J based on cross-entropy is defined as

$$J = -\frac{1}{S} \sum_{s=1}^S \sum_{q=1}^Q 1 \{ y^{(s)} = q \} \log \frac{\exp \{ (\boldsymbol{\theta}_q^L) \cdot \mathbf{x}^{(s)} \}}{\sum_{q=1}^Q \exp \{ (\boldsymbol{\theta}_q^L) \cdot \mathbf{x}^{(s)} \}}, \quad (4)$$

where S denotes the total number of samples, and the indicator function $1 \{ \cdot \}$ outputs 1 for true condition and 0 for false condition.

The training process minimizes the loss function by optimizing the trainable parameters. The gradient descent method updates the trainable parameters of CNN models by back-propagation algorithm.

$$\boldsymbol{\theta}^{\text{new}} = \boldsymbol{\theta}^{\text{old}} - \eta \frac{\partial J}{\partial \boldsymbol{\theta}^{\text{old}}}, \quad (5)$$

where η is the learning rate, and $\boldsymbol{\theta}^{\text{old}}$ and $\boldsymbol{\theta}^{\text{new}}$ are the trainable parameters before and after update, respectively.

3.2 Short-time Fourier transform

STFT is a common and simple signal processing method that can transform 1D time-series data into 2D spectrogram images. To take advantage of CNNs in the

task of image recognition, STFT is used to transform the collected raw signals into images.

The continuous-time form of the STFT for the vibration signal $x(\tau)$ is expressed as [36]

$$\text{STFT}(t, \omega) = \int_{-\infty}^{\infty} x(t + \tau) w^*(\tau) \exp(-j\omega\tau) d\tau, \quad (6)$$

where t is time, τ is the time variable of integration, ω is the angular frequency, j is the imaginary unit, and $w^*(\tau)$ is the conjugated form of window function $w(\tau)$. Equation (6) suggests that the STFT actually represents a Fourier transform of the signal $x(\tau)$ truncated by the window function at instant t .

Numerical calculation needs a discretized STFT instead of a continuous-time one. The discretized form of STFT is given by

$$\text{STFT}\{x[n]\} = \sum_{-\infty}^{\infty} x[n] w[n-m] \exp(-j\omega n), \quad (7)$$

where m and n are the indices of discrete sampling points.

In this work, Hanning window is chosen to be the window function, which is expressed as

$$w[n] = \begin{cases} 0.5 \left(1 - \cos \frac{2\pi n}{N-1} \right), & 0 \leq n \leq N-1, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where N is the size of Hanning window.

4 Proposed method

The axial piston pump is a typical fluid–solid–thermal coupling system, and its comprehensive degradation state is difficult to capture using single-sensor signals. Multi-sensor data fusion can provide abundant and complementary information and avoid possible measurement errors from single sensor. Vibration and discharge pressure are two common signals for axial piston pumps. They are suitable to be used for monitoring the pump performance.

Figure 5 illustrates the process of multi-sensor data fusion for the fault diagnosis of axial piston pumps. Taking two vibration signals and one discharge pressure signal for example, the multi-sensor data fusion includes three main steps. First, the raw data of each signal are divided into equal data segments, of which each data segment contains 256 sampling points. These data segments are transformed into grayscale spectrograms with a size of 128×128 pixels by STFT. Second, the grayscale spectrograms of each signal are discretized to become temporary matrices, and the magnitude of matrix elements are normalized to 0–255. The temporary matrices of individual signals act as elements of R, G, and B channels. A group of three temporary matrices at the same period of time are composited into an RGB image.

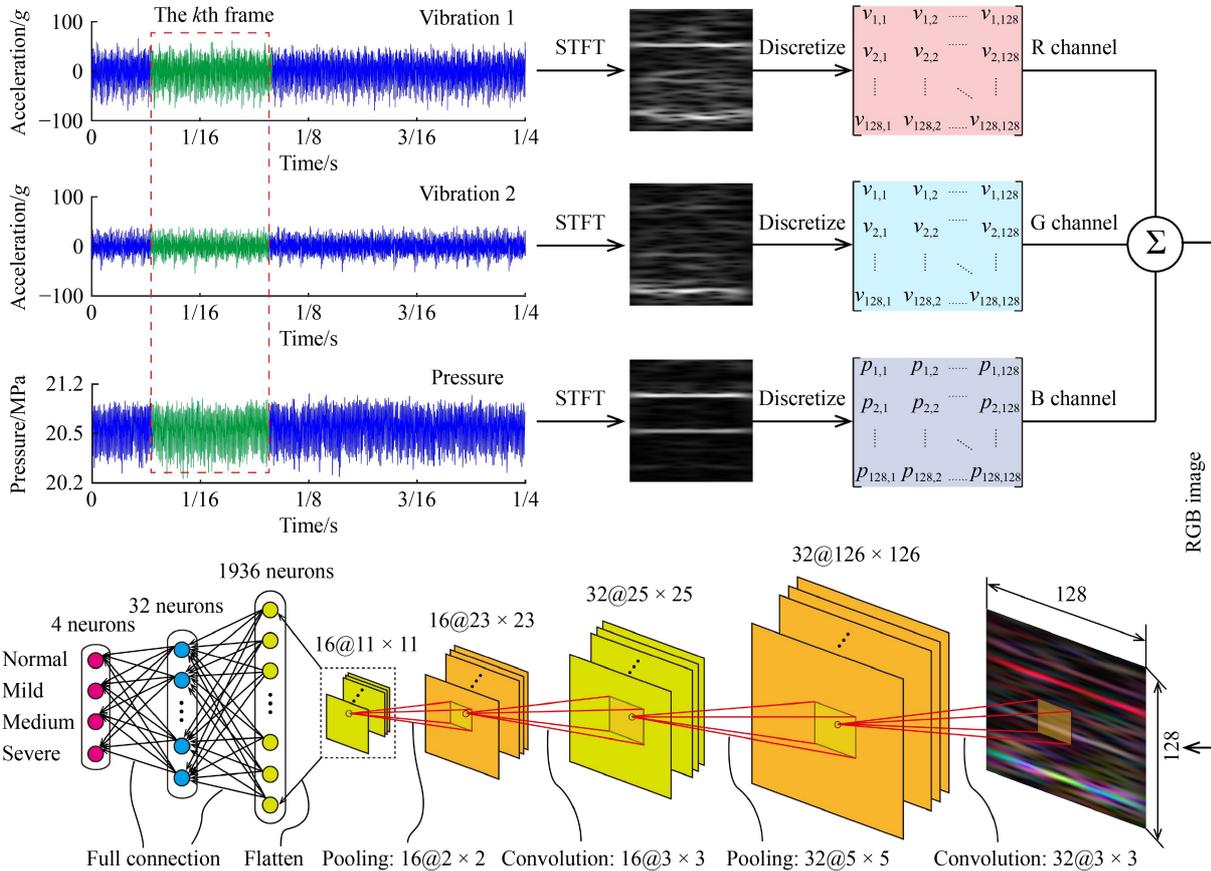


Fig. 5 Illustration of the proposed multi-sensor data fusion for the fault diagnosis of axial piston pumps.

Finally, a 2D CNN model accomplishes the fault diagnosis task of axial piston pumps by recognizing the RGB images that represent corresponding pump health conditions. The proposed method actually integrates information fusion at the data and feature levels to realize an end-to-end multi-sensor data fusion.

The first stage fulfills data-level fusion by compositing an RGB image from a group of three grayscale spectrograms. The grayscale spectrograms are converted from the raw data segments of two vibration signals and one pressure signal by STFT. The grayscale spectrograms contain information in time and frequency domains and can reflect the pump health state better than the raw time-series data. In fact, the time–frequency representations can be obtained by other time–frequency analysis approaches, such as continuous wave transform [14,15]. In addition, the multi-sensor data can either be from heterogeneous or homogeneous signals. For example, the RGB images are composited from a group of three vibration signals instead of two vibration signals and one pressure signal. Notably, the first stage can fuse more input signals. For example, if four signals are to be fused, the corresponding grayscale spectrograms will be composited into a four-dimensional tensor rather than a three-dimensional RGB image. In this case, the magnitude of each discretized grayscale spectrogram can be normalized to other uniform ranges, not just between 0 and 255.

The second stage of the proposed multi-sensor data fusion fulfills feature-level fusion by a CNN model. The previous stage only extends the raw data from different sensors to 2D representations and fuses them at data level. However, unlike conventional feature engineering methods, it has not extracted features from the extended data. The second stage uses a CNN model to receive the RGB images of the previous stage and automatically extracts features from the RGB images further layer by layer. The feature extraction by the CNN model is actually a feature fusion process.

The CNN model in this work is a modified version of the famous LeNet-5 developed by LeCun et al. [37]. Figure 5 illustrates the architecture of the CNN model used in this work, where the network contains one input layer, two convolutional layers, two max-pooling layers, two full-connected layers, and one output layer in sequence. The input layer receives $128 \times 128 \times 3$ -pixel RGB images, where the first two numbers represent the image height and weight and the last number represents the image channels. The first convolutional layer applies 32 filters with a size of 3×3 pixels to extract features from the input image separately at a stride of 1 pixel, generating 32 feature maps with a size of 126×126 pixels. The first max-pooling layer has a 5×5 pooling window to slide across each feature map at a stride of 5 pixels. This means that the output feature maps are scaled down by five times in the height and width dimensions.

Consequently, 32 feature maps have a size of 25×25 pixels after the max-pooling operation.

Similarly, another group of convolutional and max-pooling layers is stacked behind the previous one. The second convolutional layer has 16 filters with the same size as the first convolutional layer. It slides across the input feature maps at a stride of 1 pixel to generate 16 new feature maps with a size of 23×23 pixels. The second max-pooling layer has a pooling window with a size of 2×2 pixels, and the new feature maps are further reduced by half in height and width after max-pooling at a stride of 2 pixels.

After two alternating convolutional and pooling layers, each input image is flattened into a 1936×1 vector and then fed into the first full-connected layer with 32 hidden neurons. The feature maps are further reduced to a 32×1 vector. The second full-connected layer contains four hidden neuron nodes and a soft-max activation function to accomplish the classification task (i.e., the classification of four different pump degradation states, including normal state, mild leakage, medium leakage, and severe leakage).

5 Results and discussion

Figure 6 shows examples of the composited RGB images under different pump degradation states. The top row of the RGB images is converted from three vibration signals, whereas the three other rows are converted from two vibration signals and one discharge pressure signal. The energy of all RGB images is dominated at 1500, 3000, and 4500 Hz, which actually represent the piston pass frequency and its second-order and third-order harmonic frequencies. The comparison of RGB images at the same row shows that the image fusion makes it possible to classify the pump degradation states intuitively. By contrast, the grayscale images converted from a single signal offer less discriminative information to distinguish them (Fig. 7).

Figure 8 compares the diagnostic performance of the CNN model with different input signals. For the single-sensor signals, the pressure signal achieves the highest average accuracy rate of 99.9% and the lowest standard deviation of 0.2% over five trials. By contrast, the same CNN model has lower accuracy rates of 96.5%, 93.4%, and 95.6% when it receives a vibration signal. The various accuracy rates among the vibration signals in different directions mainly result from the sensitivity of vibration signal to the transmission paths. That is, obtaining an accurate fault diagnosis using single-sensor data is difficult due to limited information.

Multi-sensor data fusion can capture comprehensive fault information effectively to improve the diagnostic performance. This work investigates two types of multi-sensor data fusion: homogeneous data fusion and

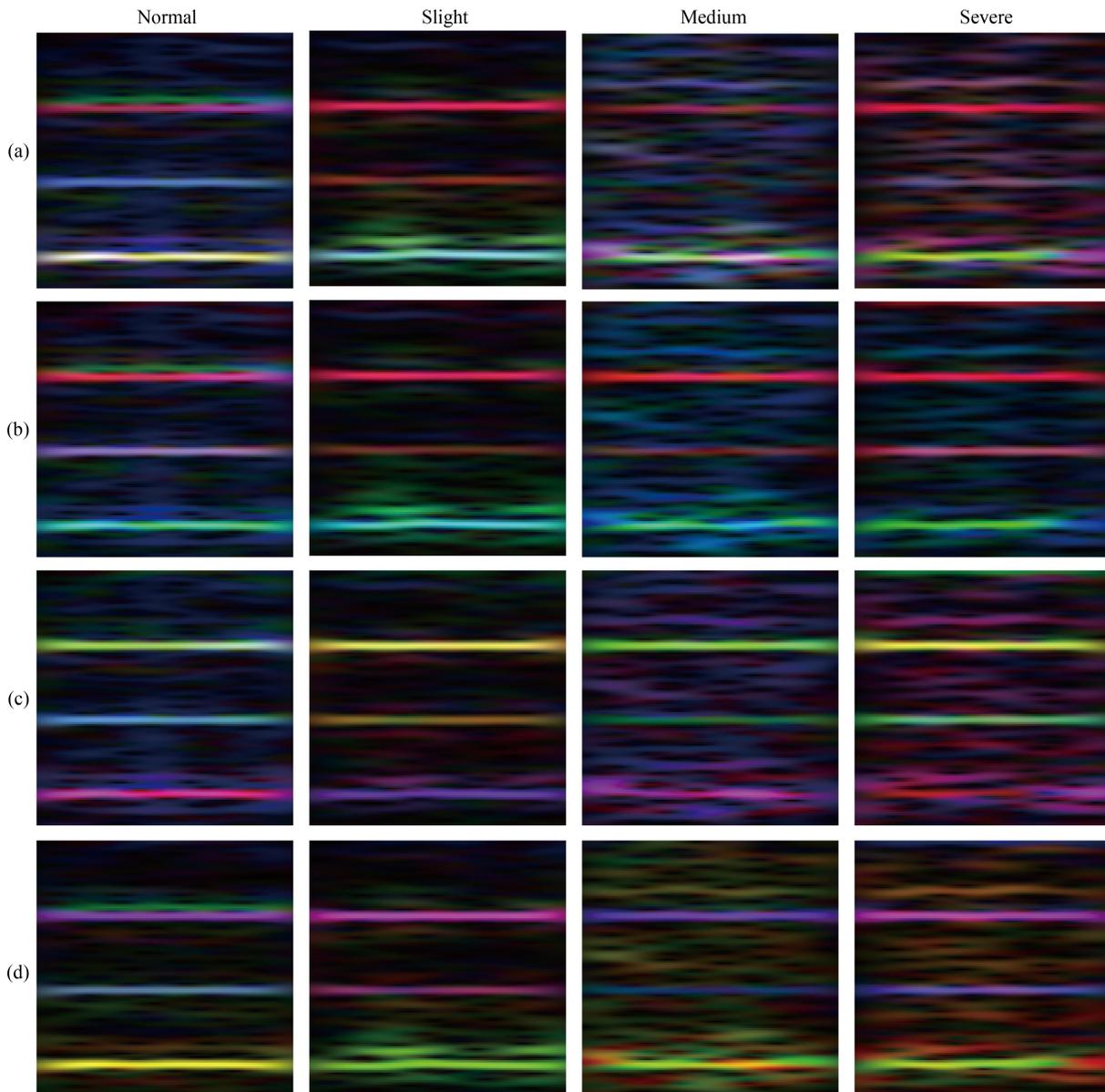


Fig. 6 RGB images under different pump degradation states: (a) three vibration signals; (b) vibration signals 2 and 3, and pressure signal; (c) vibration signals 1 and 3, and pressure signal; and (d) vibration signals 1 and 2, and pressure signal.

heterogeneous data fusion. The former integrates three vibration signals, while the latter integrates two vibration signals and one pressure signal. Figure 8 further compares the average accuracy rate and standard deviation before and after multi-sensor data fusion. The average accuracy rate can be increased up to 100% when the CNN model receives multi-sensor data in spite of being homogeneous or heterogeneous data. The multi-sensor data fusion has a higher average accuracy rate of 4%–6% than a single vibration signal. These comparison results confirm the advantages of multi-sensor data fusion in the fault diagnosis of axial piston pumps.

In industrial applications, the collected vibration and pressure signals often contain background noise, and it is difficult to extract weak fault features from the

contaminated signals. In the meantime, the entrained noise may have negative effects on the overall diagnostic performance of the multi-sensor data fusion. To evaluate the anti-noise ability of the fusion algorithm, white Gaussian noise is added intentionally to the original 1D time-series testing dataset [38,39]. Figure 9 shows examples of RGB images converted from contaminated signals under different pump degradation states. Compared with the “clean” RGB images in Fig. 6, the contaminated RGB images in Fig. 9 seem to be less intuitively identifiable among different pump degradation states.

Figure 10 compares the diagnostic accuracy between single-sensor data and multi-sensor data at different signal-to-noise ratio (SNR) levels. Among the

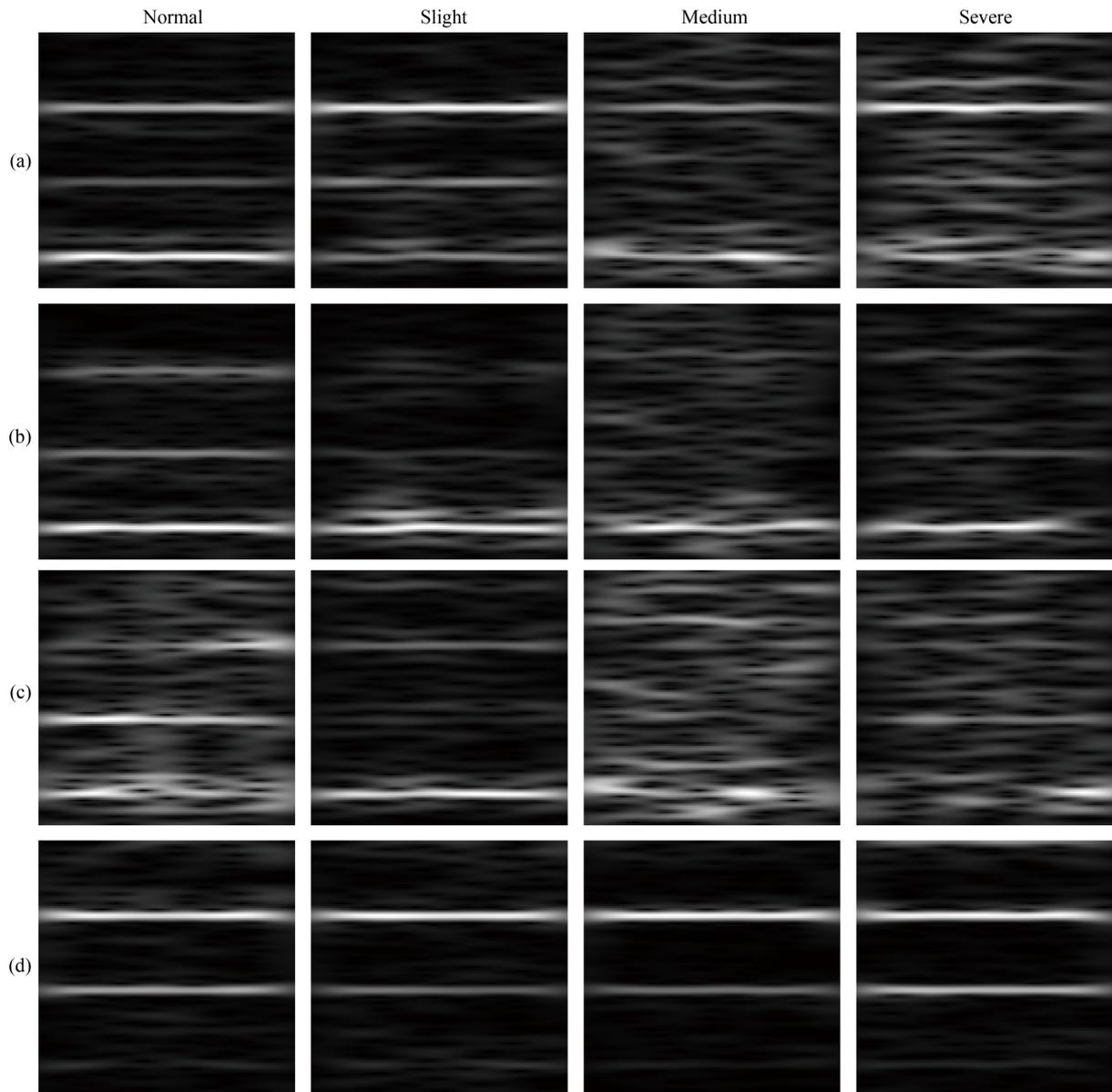


Fig. 7 Grayscale images under different pump degradation states: (a) vibration signal 1; (b) vibration signal 2; (c) vibration signal 3; and (d) pressure signal.

single-sensor data, the pressure signal and the second vibration signal have the worst robustness against noise, whereas the third vibration signal has the best robustness against noise. The CNN model with an input of single-sensor data has relatively low diagnostic accuracy in a noisy environment. For the first or second vibration signal, the classification accuracy drops below 90% at SNRs below 10 dB. The classification accuracy with only pressure signals even remains below 50%. Although the classification accuracy with the third vibration signal achieves an acceptable classification accuracy above 90% at high SNRs, it will drop dramatically below 80% at SNRs below 6 dB.

The multi-sensor data fusion improves the anti-noise

ability of the CNN model significantly in spite of being homogeneous or heterogeneous data. For example, the classification accuracy is only 56.4%, 76.3%, and 50.0% for the first vibration signal, the third vibration signal, and the pressure signal, respectively, at an SNR of 4 dB, while it can achieve up to 96.7% when these signals are fused at the same SNR.

The improvement of diagnostic performance depends on the type of fused signals. For the axial piston pump in this work, the optimal multi-sensor data fusion strategy, which outperforms other types of fused signals by a large margin, especially at low SNRs, is to combine the pressure signal and the first and third vibration signals. For instance, the optimal multi-sensor data fusion can

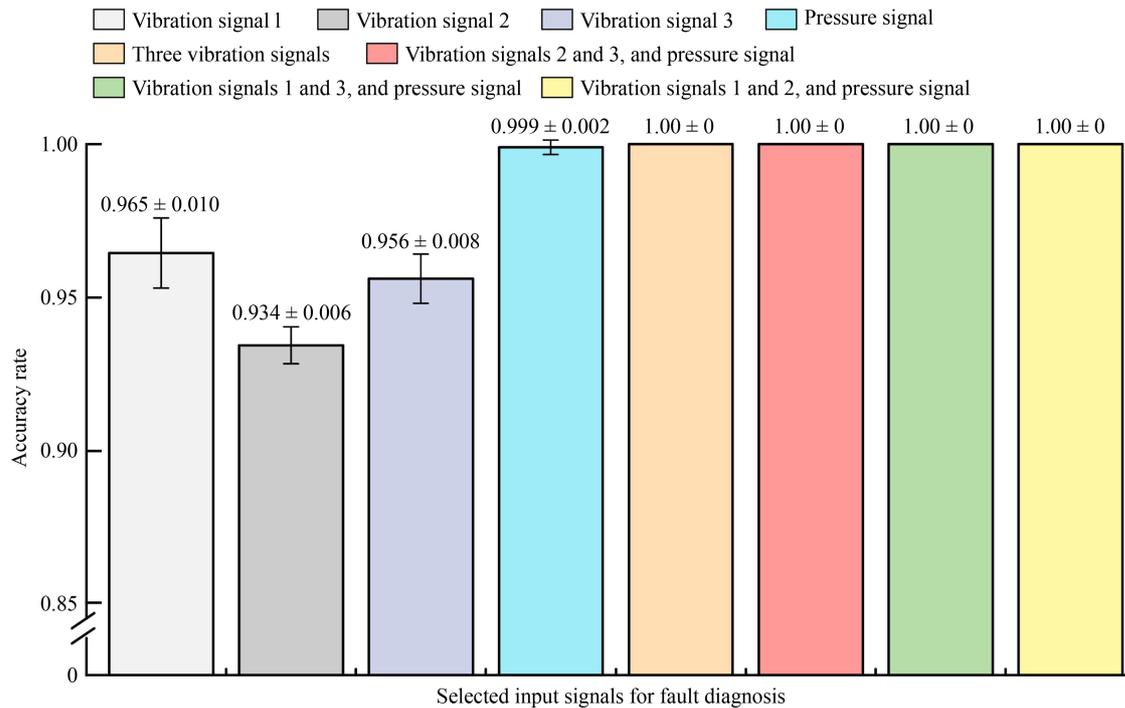


Fig. 8 Comparison of classification accuracy between single-sensor data and multi-sensor data.

achieve a classification accuracy of 99.9% at an SNR of 6 dB, which is higher than the three other types of data fusion by 6.7%, 2.7%, and 3.4% (Fig. 10). This finding provides guidelines to select the most suitable single-sensor data to be fused.

Figure 11 further compares the confusion matrices of diagnostic accuracy for one trial with and without multi-sensor data fusion. The confusion matrices reflect the contribution of each signal to the classification results and how the multi-sensor data fusion improves the diagnostic performance. The horizontal and vertical coordinates of each confusion matrix represent the predicted and real labels, respectively. The confusion matrix presents the accuracy rates on the diagonal elements and error rates on the other elements. Figures 11(a)–11(d) indicate that the single-sensor signal leads to unsatisfactory classification results in noisy environment. The third vibration signal appears to have stronger anti-noise ability than other signals and hence should be involved during data fusion.

As expected, the CNN model has better robustness against noise when multiple sensor signals are integrated. The class accuracy of each degradation state becomes more than 95% after multi-sensor data fusion, except for two cases: the first case is three vibration signals (Fig. 11(e)); and the second case is one pressure signal plus the first two vibration signals (Fig. 11(h)). The inferior anti-noise ability of the above two cases may arise from the following factors. First, the homogeneous data fusion is inferior to the heterogeneous one for the fault diagnosis of the investigated axial piston pump. Second, the first two vibration signals are less sensitive to pump health

conditions but more sensitive to noise than the third vibration signal.

Table 2 [5,15,19,40,41] compares the diagnostic performance of the slipper faults in axial piston pumps among different methods and input signals. The first three diagnosis methods all adopt CNN, but they only use single-sensor data (i.e., acoustic or vibration signal). The methods of CNN-Bayesian optimization (BO) [15] and 2D CNN [40] also recognize 2D spectrogram images converted from 1D time-series data by continuous wavelet transform, whereas the method of 1D CNN [41] directly handles 1D time-series data. The comparison of diagnostic performance among the first three methods and the proposed method suggests that the multi-sensor data fusion is helpful in improving the diagnostic performance of the CNN model. The last three diagnosis methods all integrates multi-sensor data fusion, but they use different types of input signals. The extreme learning machine method [5] uses nine vibration signals and one discharge flow signal, whereas the empirical wavelet transform (EWT) and variance contribution rate method [19] uses three vibration signals. Compared with the proposed method, these two methods adopt conventional machine learning rather than deep learning, and their multi-sensor data fusion only occurs at the feature level. The multi-sensor data fusion method proposed in this work exhibits superior diagnostic performance to the ones in previous studies [5,19]. This may be explained by two reasons. First, the proposed method integrates data-level multi-sensor data fusion in addition to feature-level multi-sensor data fusion. Second, heterogeneous signals

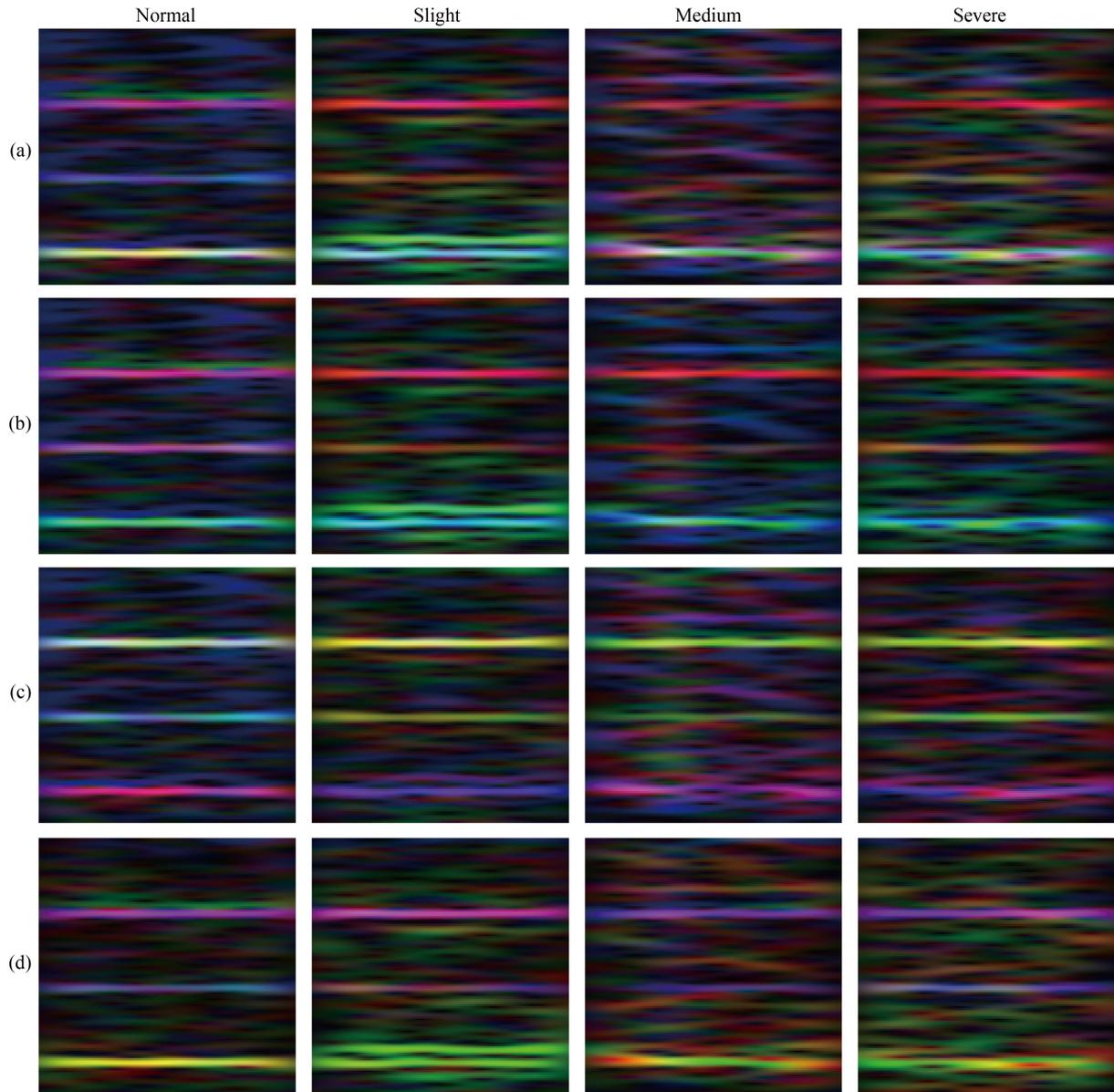


Fig. 9 Comparison of contaminated spectrograms among different pump degradation states at an SNR of 6 dB: (a) three vibration signals; (b) vibration signals 2 and 3, and pressure signal; (c) vibration signals 1 and 3, and pressure signal; and (d) vibration signals 1 and 2, and pressure signal.

(especially including discharge pressure signal) are more sensitive to the performance degradation of axial piston pumps than homogeneous signals.

6 Conclusions

This paper proposes an end-to-end multi-sensor data fusion method for the fault diagnosis of axial piston pump. The proposed method integrates two stages of multi-sensor data fusion at data and feature levels through composited RGB image and CNN. The vibration signals and discharge pressure signal are fused to recognize the four different degradation states of an axial piston pump.

The following conclusions can be drawn from the results and discussion:

(1) The diagnostic performance of the CNN model depends on the monitoring signals of axial piston pumps. Compared with the vibration signal, the CNN model with discharge pressure signal has a higher recognition accuracy of 3.4%–6.5%.

(2) The multi-sensor data fusion significantly improves the diagnostic accuracy and reliability. Compared with each single vibration signal, the data fusion of the three vibration signals increases the recognition accuracy from 96.5%, 93.4%, and 95.6% to 100.0%.

(3) The diagnostic performance the CNN model becomes unsatisfactory in noisy environment when only a

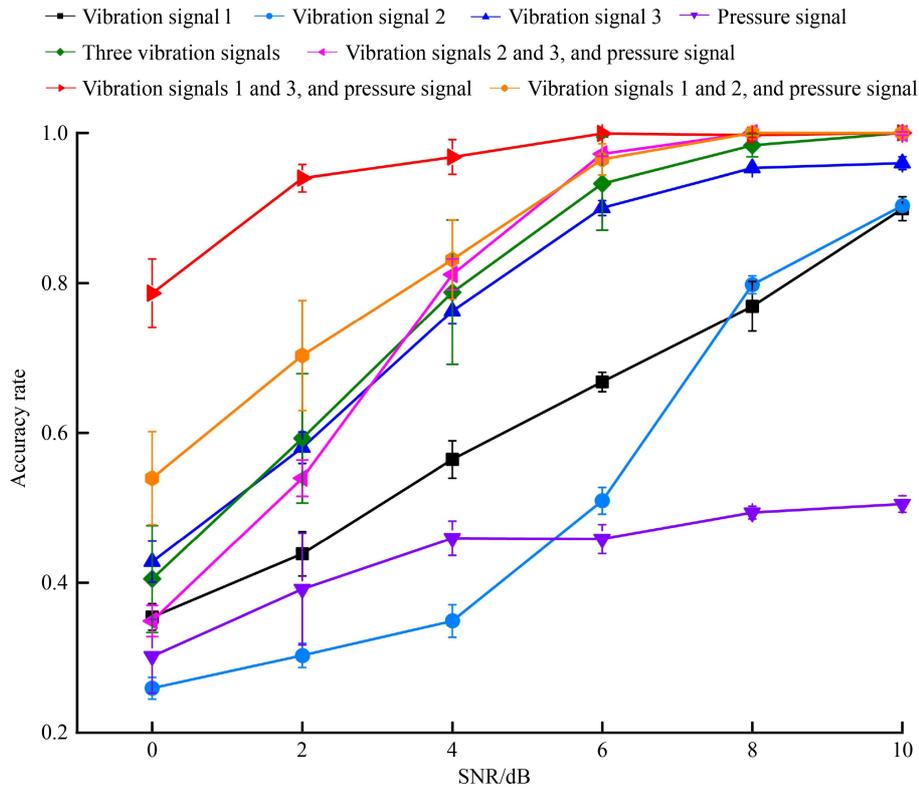


Fig. 10 Comparison of anti-noise performance between single-sensor data and multi-sensor data.

Normal	0.000	0.000	0.875	0.125	0.458	0.000	0.542	0.000	1.000	0.000	0.000	0.000	0.896	0.083	0.021	0.000
Slight	0.000	1.000	0.000	0.000	0.000	0.188	0.729	0.083	0.000	0.917	0.062	0.021	0.062	0.917	0.021	0.000
Medium	0.000	0.000	0.958	0.042	0.000	0.000	0.917	0.083	0.000	0.000	0.958	0.042	0.021	0.917	0.062	0.000
Severe	0.000	0.000	0.271	0.729	0.000	0.000	0.583	0.417	0.000	0.000	0.229	0.771	0.979	0.021	0.000	0.000
	Normal	Slight	Medium	Severe												
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)								

Fig. 11 Comparison of confusion matrix before and after multi-sensor data fusion at an SNR of 6 dB: (a) vibration signal 1; (b) vibration signal 2; (c) vibration signal 3; (d) pressure signal; (e) three vibration signals; (f) vibration signals 2 and 3, and pressure signal; (g) vibration signals 1 and 3, and pressure signal; and (h) vibration signals 1 and 2 and pressure signal.

single-sensor signal is available. The multi-sensor data fusion, especially heterogeneous data fusion, can improve the diagnostic accuracy and reliability in noisy environment significantly. For example, the recognition accuracy of the pump degradation states at an SNR of 6 dB is increased by 9.9%–54.1% when two vibration

signals and one pressure signal are fused and by 3.2%–47.4% when three vibration signals are fused.

(4) For the slipper fault diagnosis of axial piston pumps, the proposed method has a better diagnostic performance than other existing methods because of two reasons. First, the proposed method integrates data-level

Table 2 Classification accuracy of slipper faults in axial piston pumps for different methods and input signals

Method	Input signal	Classification accuracy/%
CNN-BO [15]	One acoustic signal	97.8
2D CNN [40]	One vibration signal	96.1
1D CNN [41]	One vibration signal	98.7
Extreme learning machine [5]	Nine vibration signals + one discharge flow signal	84.1
EWT and variance contribution rate [19]	Three vibration signals	66.5
Proposed method	Two vibration signals + one discharge pressure signal	100.0

and feature-level multi-sensor data fusion. Second, the multi-sensor data fusion integrates heterogeneous data, including vibration and pressure signals.

The present method only considers vibration and discharge pressure signals and treats each signal equally. The future research will improve the multi-sensor data fusion further by integrating other types of available signals and emphasizing the data weights of different signals.

Nomenclature

Abbreviations

1D	One-dimensional
2D	Two-dimensional
CNN	Convolutional neural network
SNR	Signal-to-noise ratio
STFT	Short-time Fourier transform

Variables

a_{hw}	Feature map element at pixel (h, w) in the pooling window
A_k^l	The k th feature map at layer l
B_k^l	Bias of the k th group filter at layer l
c	Index of channels for input feature maps or the group filters
C	Total number of filter channels
$f(\cdot)$	Activation function
H	Pooling window height
i'	Height index of element pixels
j	Imaginary unit
j'	Width index of element pixels
J	Loss function
k	Index of group filters or output feature maps
l	Index of network layers
L	Total layer number
m, n	Indices of discrete sampling points
N	Size of Hanning window
p_k^{l+1}	Maximum element in the pooling window
q	The q th class

Q	Total classification number
s	Index of samples
S	Total number of samples
t	Time
$x(\tau)$	Vibration signal
$x^{(s)}$	The s th sample
X_c^{l-1}	The c th-channel component of the input feature map at layer $(l-1)$
X_k^l	The k th output feature map at layer l
$y^{(s)}$	Predicted label
$w(\tau), w^*(\tau)$	Window function and its conjugated form
W	Pooling window width
$W_{c,k}^l$	The c th-channel component of the k th group filter weight at layer l
η	Learning rate
θ^L	Trainable parameters at the last layer L
$\theta^{\text{new}}, \theta^{\text{old}}$	Trainable parameters after and before update, respectively
τ	Time variable of integration
ω	Angular frequency

Acknowledgements This study was supported by the National Key R&D Program of China (Grant No. 2018YFB1702503), the Open Foundation of the State Key Laboratory of Fluid Power and Mechatronic Systems, China (Grant No. GZKF-202108), the National Postdoctoral Program for Innovative Talents, China (Grant No. BX20200210), the China Postdoctoral Science Foundation (Grant No. 2019M660086), and Shanghai Municipal Science and Technology Major Project, China (Grant No. 2021SHZDZX0102).

References

- Chao Q, Zhang J H, Xu B, Wang Q N, Lyu F, Li K. Integrated slipper retainer mechanism to eliminate slipper wear in high-speed axial piston pumps. *Frontiers of Mechanical Engineering*, 2022, 17(1): 1–13
- Chao Q, Xu Z, Tao J F, Liu C L, Zhai J. Cavitation in a high-speed aviation axial piston pump over a wide range of fluid temperatures. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 2022, 236(4): 727–737
- Maradey Lázaro J G, Borrás Pinilla C. Detection and classification of wear fault in axial piston pumps: using ANNs and pressure signals. In: Burgos D A T, Vejar M A, Pozo F, eds. *Pattern*

- Recognition Applications in Engineering. Hershey: IGI Global, 2020, 286–316
4. Xia S Q, Zhang J H, Ye S G, Xu B, Huang W D, Xiang J W. A sparse support vector machine based fault detection strategy on key lubricating interfaces of axial piston pumps. *IEEE Access*, 2019, 7: 178177–178186
 5. Lan Y, Hu J W, Huang J H, Niu L K, Zeng X H, Xiong X Y, Wu B. Fault diagnosis on slipper abrasion of axial piston pump based on extreme learning machine. *Measurement*, 2018, 124: 378–385
 6. Guo R, Zhao Z Q, Wang T, Liu G H, Zhao J Y, Gao D R. Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. *Applied Sciences*, 2020, 10(18): 6593
 7. Keller N, Sciancalepore A, Vacca A. Demonstrating a condition monitoring process for axial piston pumps with damaged valve plates. *International Journal of Fluid Power*, 2022, 23(2): 205–236
 8. Wang S H, Xiang J W, Zhong Y T, Tang H S. A data indicator-based deep belief networks to detect multiple faults in axial piston pumps. *Mechanical Systems and Signal Processing*, 2018, 112: 154–170
 9. Chao Q, Tao J F, Wei X L, Wang Y H, Meng L H, Liu C L. Cavitation intensity recognition for high-speed axial piston pumps using 1-D convolutional neural networks with multi-channel inputs of vibration signals. *Alexandria Engineering Journal*, 2020, 59(6): 4463–4473
 10. Chao Q, Tao J F, Wei X L, Liu C L. Identification of cavitation intensity for high-speed aviation hydraulic pumps using 2D convolutional neural networks with an input of RGB-based vibration data. *Measurement Science and Technology*, 2020, 31(10): 105102
 11. Wang S H, Xiang J W. A minimum entropy deconvolution-enhanced convolutional neural networks for fault diagnosis of axial piston pumps. *Soft Computing*, 2020, 24(4): 2983–2997
 12. Tang S N, Yuan S Q, Zhu Y, Li G P. An integrated deep learning method towards fault diagnosis of hydraulic axial piston pump. *Sensors*, 2020, 20(22): 6576
 13. Chao Q, Gao H H, Tao J F, Wang Y H, Zhou J, Liu C L. Adaptive decision-level fusion strategy for the fault diagnosis of axial piston pumps using multiple channels of vibration signals. *Science China Technological Sciences*, 2022, 65(2): 470–480
 14. Tang S N, Zhu Y, Yuan S Q. Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization. *ISA Transactions*, 2022 (in press)
 15. Tang S N, Zhu Y, Yuan S Q. A novel adaptive convolutional neural network for fault diagnosis of hydraulic piston pump with acoustic images. *Advanced Engineering Informatics*, 2022, 52: 101554
 16. Lu C Q, Wang S P, Makis V. Fault severity recognition of aviation piston pump based on feature extraction of EEMD paving and optimized support vector regression model. *Aerospace Science and Technology*, 2017, 67: 105–117
 17. Wang Y D, Zhu Y, Wang Q L, Yuan S Q, Tang S N, Zheng Z J. Effective component extraction for hydraulic pump pressure signal based on fast empirical mode decomposition and relative entropy. *AIP Advances*, 2020, 10(7): 075103
 18. Lu C Q, Wang S P, Zhang C. Fault diagnosis of hydraulic piston pumps based on a two-step EMD method and fuzzy C-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2016, 230(16): 2913–2928
 19. Yu H, Li H R, Li Y L. Vibration signal fusion using improved empirical wavelet transform and variance contribution rate for weak fault detection of hydraulic pumps. *ISA Transactions*, 2020, 107: 385–401
 20. Yu H, Li H R, Li Y L, Li Y F. A novel improved full vector spectrum algorithm and its application in multi-sensor data fusion for hydraulic pumps. *Measurement*, 2019, 133: 145–161
 21. Safizadeh M S, Latifi S K. Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. *Information Fusion*, 2014, 18: 1–8
 22. Xia M, Li T, Xu L, Liu L Z, de Silva C W. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *IEEE/ASME Transactions on Mechatronics*, 2018, 23(1): 101–110
 23. Wang H Q, Li S, Song L Y, Cui L L. A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals. *Computers in Industry*, 2019, 105: 182–190
 24. Gong W F, Chen H, Zhang Z H, Zhang M L, Wang R H, Guan C, Wang Q. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors*, 2019, 19(7): 1693
 25. Wang J J, Fu P L, Zhang L B, Gao R X, Zhao R. Multilevel information fusion for induction motor fault diagnosis. *IEEE/ASME Transactions on Mechatronics*, 2019, 24(5): 2139–2150
 26. Chen H P, Hu N Q, Cheng Z, Zhang L, Zhang Y. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. *Measurement*, 2019, 146: 268–278
 27. Azamfar M, Singh J, Bravo-Imaz I, Lee J. Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. *Mechanical Systems and Signal Processing*, 2020, 144: 106861
 28. Kolar D, Lisjak D, Pająk M, Pavković D. Fault diagnosis of rotary machines using deep convolutional neural network with wide three axis vibration signal input. *Sensors*, 2020, 20(14): 4017
 29. Yan X S, Sun Z, Zhao J J, Shi Z G, Zhang C A. Fault diagnosis of rotating machinery equipped with multiple sensors using space-time fragments. *Journal of Sound and Vibration*, 2019, 456: 49–64
 30. Chao Q, Zhang J H, Xu B, Huang H P, Pan M. A review of high-speed electro-hydrostatic actuator pumps in aerospace applications: challenges and solutions. *Journal of Mechanical Design*, 2019, 141(5): 050801
 31. Ma J M, Chen J, Li J, Li Q L, Ren C Y. Wear analysis of swash plate/slipper pair of axis piston hydraulic pump. *Tribology International*, 2015, 90: 467–472
 32. Huang J H, Yan Z, Quan L, Lan Y, Gao Y S. Characteristics of delivery pressure in the axial piston pump with combination of variable displacement and variable speed. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 2015, 229(7): 599–613
 33. Chao Q, Tao J F, Lei J B, Wei X L, Liu C L, Wang Y H, Meng L H. Fast scaling approach based on cavitation conditions to estimate

- the speed limitation for axial piston pump design. *Frontiers of Mechanical Engineering*, 2021, 16(1): 176–185
34. Chacon R, Ivantysynova M. Virtual prototyping of axial piston machines: numerical method and experimental validation. *Energies*, 2019, 12(9): 1674
 35. Dahl G E, Sainath T N, Hinton G E. Improving deep neural networks for LVCSR using rectified linear units and dropout. In: *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. Vancouver: IEEE, 2013, 8609–8613
 36. Stankovic L, Daković M, Thayaparan T. *Time–Frequency Signal Analysis with Applications*. Boston: Artech House, 2013
 37. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324
 38. Zhang W, Li C H, Peng G L, Chen Y H, Zhang Z J. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, 2018, 100: 439–453
 39. Liu X C, Zhou Q C, Zhao J, Shen H H, Xiong X L. Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network. *Sensors*, 2019, 19(4): 972
 40. Tang S N, Zhu Y, Yuan S Q, Li G P. Intelligent diagnosis towards hydraulic axial piston pump using a novel integrated CNN model. *Sensors*, 2020, 20(24): 7152
 41. Jiang W L, Wang C Y, Zou J Y, Zhang S Q. Application of deep learning in fault diagnosis of rotating machinery. *Processes*, 2021, 9(6): 919