APPLIED RESEARCH

Fault Diagnosis of Industrial Robot Based on Multi-Source Data Fusion and Channel Attention Convolutional Neural Networks

YIWEN ZHANG[1,2], ZIHAO ZANG[1], XINMING ZHANG[1,2], LINSEN SONG[1], ZHENGLEI YU[3], YITIAN WANG[1], YAN GAO[1], AND LEI WANG[1]

[1]School of Mechanical and Electrical Engineering, Changchun University of Science and Technology, Changchun 130022, China
[2]School of Mechatronic Engineering and Automation, Foshan University, Foshan 528225, China
[3]College of Biological and Agricultural Engineering, Jilin University, Changchun 130025, China

Corresponding authors: Xinming Zhang (zxm@fosu.edu.cn), Linsen Song (songlinsen@cust.edu.cn), and Zhenglei Yu (zlyu@jlu.edu.cn)

ABSTRACT Industrial robots are prone to failure due to harsh working environments, which affects movement accuracy. The fault diagnosis of industrial robots has become an indispensable part of robot collaborative maintenance in intelligent manufacturing. Most existing diagnostic methods only use a single data source, and the diagnostic accuracy will be affected due to signal acquisition errors and noise interference. This paper proposes a multi-source data fusion and channel attention convolutional neural network (MD-CA-CNN) for fault diagnosis of multi-joint industrial robots. The network takes the time domain data and time-frequency domain data of the vibration signal, torque signal, and current signal of the six joints of the robot as input. Then, we realize the diagnosis of the faults by using a Softmax Classifier layer after the two parts of feature extraction and feature fusion. In addition, a channel attention mechanism is developed. It acts on the two parts of feature extraction and feature fusion, respectively. It assigns weights to different source data and weights to time-domain and time-frequency domain features. Finally, we established a test bench to compare the proposed method with the deep learning algorithm that only uses multi-data source fusion, the deep learning algorithm that only uses a single data source, and the commonly used machine learning algorithm. The results show that the MD-CA-CNN model proposed in this paper has the highest accuracy and stability, reaching $95.8\% \pm 0.39\%$, which verifies the method's effectiveness.

INDEX TERMS Industrial robots, fault diagnosis, multi-source data fusion, convolutional neural networks, channel attention mechanisms.

I. INTRODUCTION

Industrial robots have long been used to automate manufacturing processes to improve productivity, quality, and safety [1]. Due to their advantages of high flexibility, high-cost efficiency, ample working space, high repeatability, and multiple functions, they have been widely used in modern manufacturing industries such as automobile assembly, chemical industry, and aerospace [2]. The transmission

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Shu.

system is the core of an industrial robot. In actual industrial production, the motor, reducer, and other components in the industrial robot transmission system may fail due to overload, long-term work, or long-term non-maintenance, affecting the joint stability of the industrial robot. If it is severe, the robot will not be able to continue the task, interrupting the production task. Condition monitoring and fault diagnosis of industrial robots can reduce unplanned downtime on highly automated production lines, saving associated costs [3]. Therefore, it is meaningful to study a highly accurate fault diagnosis method for industrial robots to arrange

appropriate maintenance strategies and reduce maintenance costs.

In the industrial field, there are three main ways to diagnose equipment faults: knowledge-based, model-based, and data-driven [4]. Knowledge-based diagnostics require expert experience to identify the status of the device. Chen et al. realized the composite fault identification of rolling bearings by adaptive extraction of resonant frequency bands [5]. Anouar et al. applied wavelet analysis to rotating machinery monitoring [6]. However, fault detection is a real-time process, and this approach requires one or more specialized personnel to assess machine performance, which is impractical in modern industrial production [7]. Model-based diagnostic methods typically analyze the correct behavior of each component in a robotic system, build a mathematical or physical model, and then identify faults by comparing the model's expected output with the actual output of the robotic system [4]. Kim et al. proposed a new phase-based time-domain averaging (PTDA) method to detect industrial robot reducer faults [8]. Sun et al. realized the fault diagnosis of actuators based on constructing six free-space motion equations mathematical models of autonomous underwater vehicles [9]. Sabry et al. propose an energy model-based method for monitoring industrial robots [10]. Muradore and Fiorini proposed a fault detection and isolation method based on industrial robot models and signal processing [11]. Although model-based diagnostic methods have been widely researched in fault diagnosis, complex nonlinear characteristics and extremely variable operating conditions limit the fault diagnosis of industrial robots with complex structures. With the development of sensing technology, data-driven robot fault diagnosis methods have emerged in large numbers. No detailed mechanistic knowledge is required only fault diagnosis by analyzing and extracting different features of sensor signals collected from different health states [12]. Machine learning (ML) has been widely used to diagnose equipment failures. Guo et al. propose an improved random forest (IRF) algorithm to reselect and weight highly accurate and heterogeneous decision trees through hierarchical clustering. They used digital twin and transfer learning technology to diagnose faults on the physical production line. They verified the algorithm through a case study of an automobile rear axle assembly line [13]. Xu et al. proposed a bolt-loosening detection method for industrial robot joints based on electromechanical modeling and motor current signature analysis (MCSA). They established the kinetic equations coupling the motor current and bolt loosening, extracted the time-frequency characteristics of the motor current, and then used SVM to recognize the bolt loosening [14]. Izagirre et al. proposed a method and practical implementation of a network architecture for industrial robot data acquisition and predictive maintenance. They use SVR or ELM to detect degraded states of Robot joints. The architecture is implemented by extracting torque signals from a PLC on a real automotive assembly line [1]. Raouf et al.

used statistical analysis of three-phase current signals for feature engineering to detect and diagnose six faults in the transmission mechanisms of industrial robots. They investigated a variety of algorithms. The accuracy assessment showed that the SVM model was the most accurate [15]. Pan et al. use vibration signals to detect gap failures in industrial robot joints. The Wigner-Vile distribution (WVD) of the vibration signal is used to extract healthy features and use the features as inputs to an artificial neural network (ANN) algorithm [16].

As a particular data-driven method, the deep learning (DL) model has demonstrated superior capabilities in adaptive feature extraction and fault classification for multilayer nonlinear transformations [17]. With the advent of the artificial intelligence era and powerful computing power support, deep learning methods have become a research hotspot in industrial robot fault diagnosis [18]. Zhi et al. designed a convolutional neural network and long short-term memory (CNN-LSTM) to diagnose harmonic reducer faults in industrial robots [19]. Yildirim et al. captured sound signals of various faults in an industrial robot. They used wavelet analysis to process and remove noise from the captured sound signals, extract relevant features, and train five neural networks for noise analysis and classification [20]. He et al. proposed the multi-scale hybrid convolutional neural network (MSMCNN). They used the vibration signals of robot joints for adaptive feature extraction, effectively extracted comprehensive and complementary weak fault features, and realized the diagnosis of the harmonic reducer of multiple robot joints [21]. Park et al. proposed a fault detection model for edge computing. They built an edge device that integrates data collection, processing, storage, and analysis. Then, they detected six types of faults in IR based on the LSTM network [22].

In recent years, multi-source information fusion has emerged in fault diagnosis. Due to signal acquisition errors and noise interference, we may need more than a single data source in some cases. In addition, since each different type of signal can perceive the environment from different aspects, multi-source data fusion can provide more useful and accurate information compared with a single data source approach, especially for complex systems [23]. For example, Miao et al. proposed a new channel-based convolutional neural network (CWNN-FA) with feature enhancement for fault diagnosis of wheeled mobile robots. This structure uses multi-heterogeneous sensor data such as IMUs and encoders as input [24]. Gültekin et al. proposed a data fusion method based on a convolutional neural network, using the sound and vibration signals of the motor to detect operational faults occurring in automatic vehicles (ATVs) [25]. Liu et al. proposed a fault diagnosis method for a chain jack hydraulic system based on multi-source sensor data fusion. The proposed method utilized pressure, temperature, and flow data under different operating conditions [26]. Cui et al. proposed a new multi-task multi-sensor fusion network (M2FN) to

improve fault diagnosis performance. The method used convolutional neural networks to extract and fuse features from raw vibration and current signals [27]. The above methods have all been proved through experiments that using multiple data sources has higher diagnostic accuracy than using a single data source.

Attention mechanism is a method of imitating the human visual and cognitive system. In recent years, the application of attention mechanisms in deep learning, especially in convolutional neural networks, has been proposed in many works. For example, Du et al. used an efficient channel-attentive deep dense convolutional neural network to achieve automatic classification of esophageal diseases in gastroscopy images [28]. In the field of fault diagnosis, Huang et al. proposed a Multi-scale Convolutional Neural Network with Channel Attention (CA-MCNN). In CA-MCNN, the maximum pooling and average pooling layers are used to extract the multi-scale information of the bearing signals, which increases the dimensions of input. The channel attention mechanism is introduced to increase the convolutional layer feature learning ability by adaptively scoring and assigning weights to the learned features [29]. Di et al. proposed a machine tool fault diagnosis method based on Multiscale-Channel Attention Network (MSCANet). MSCANet effectively integrates the vibration signal characteristics of machine tool spindles in different directions and extracts different hierarchical features of vibration signals using multi-scale structures. Adaptive fusion of features at different scales through Channel Attention (CA) mechanism. Thus improving the accuracy of tool wear status diagnosis [30]. Huang et al. proposed a multi-scale convolutional neural network for bearing fault diagnosis. This network structure will obtain time-frequency representations from vibration time-domain signals as inputs to the model. And use convolution kernels of different sizes to extract multi-scale information from time-frequency images. In addition, an attention mechanism has been established to more effectively adaptively select features of different scales for classification, emphasizing key features and weakening redundant features [31]. Tong et al. also proposed a coordinated attention (CA) model suitable for one-dimensional vibration signals and established a lightweight coordinated attention convolutional neural network model ACNN, which takes data from multiple vibration sensors as input. Secondly, a kurtosis weighted fusion strategy was designed. Then, based on ACNN and weighted fusion strategy, a rolling bearing fault diagnosis method based on Multi sensor ACNN was proposed [32]. The above methods all demonstrate that the application of attention mechanism can improve the performance and generalization ability of convolutional neural networks. However, these fault diagnosis methods still only utilize the time-domain or time-frequency domain signals of a single vibration data, and the attention mechanism is limited to different scales of a single data source.

Although various industrial robot fault diagnosis methods have improved the diagnosis accuracy, they only use a single data source, such as vibration, torque, or current. Extracting discriminative features from multi-source data to provide accurate and reliable diagnosis is still challenging [27]. Therefore, this paper proposes an industrial robot fault diagnosis method based on multi-source data fusion and channel-attention convolutional neural networks. It can detect and diagnose industrial robot drive train faults. The time-domain signals contain original feature information, and the time-frequency domain signals contain time-domain and frequency-domain information. The industrial robot is a coupled system, so each joint's change causes the other joints' response. Thus, we input the time-domain and time-frequency domain signals of each joint's original vibration, current, and torque signals. 1D-CNN and 2D-CNN extract deep features, and we realize multi-source data fusion through channel connection. At the same time, we add a channel attention module to adaptively assign weights to different source data and time domain and time-frequency domain features. Finally, we concatenate together the extracted features of each joint, and then the diagnosis result is obtained through a fully connected layer. Because this approach leverages multiple data sources, it can more accurately and robustly diagnose the condition of industrial robots. The innovation of using the channel attention mechanism in convolutional neural networks in this paper is to combine it with multi-source information fusion, which is used twice consecutively to assign different weights to different source data as well as to assign weights to time domain and time-frequency domain data, respectively. This study uses a 6-axis industrial robot as the experimental object. Comparison with other advanced fault diagnosis methods demonstrates the superiority of the MD-CA-CNN proposed in this paper. In summary, the main contribution of this paper:

1) We realize the fusion of robot joints' vibration, current, and torque data by adopting CNN channel superposition. We obtain the time-frequency domain image of the signal using CWT and extract the depth features of the original time-domain signal and the time-frequency domain signal using 1D-CNN and 2D-CNN simultaneously. It avoids the problem of a single data source being susceptible to noise and can provide more comprehensive information about the fault characteristics.

2) We propose a CNN structure based on multi-source data and channel attention mechanism, namely MD-CA-CNN. A channel attention mechanism is developed and added to the CNN input port and deep feature channel fusion. It can adaptively assign weights to different source data and time-domain and time-frequency-domain features to improve the contribution of essential channels to fault diagnosis.

The rest of the paper is organized as follows: Section II details the overall framework of the fault diagnosis system for industrial robots and the proposed diagnostic approach. Section III describes the details of the experimental setup and the collected experimental data used to evaluate the proposed methodology, and Section IV gives the experimental results

and comprehensive analysis. Finally, Section V summarizes the study.

II. PROPOSED METHODOLOGY

A. FRAMEWORK OF THE DIAGNOSTIC SYSTEM

This study first proposes a general architecture of an industrial robot fault diagnosis method. As shown in Fig. 1, the proposed method has four phases. In the fault data acquisition phase, vibration, current, and torque signals are collected from each joint of the six-axis industrial robot.
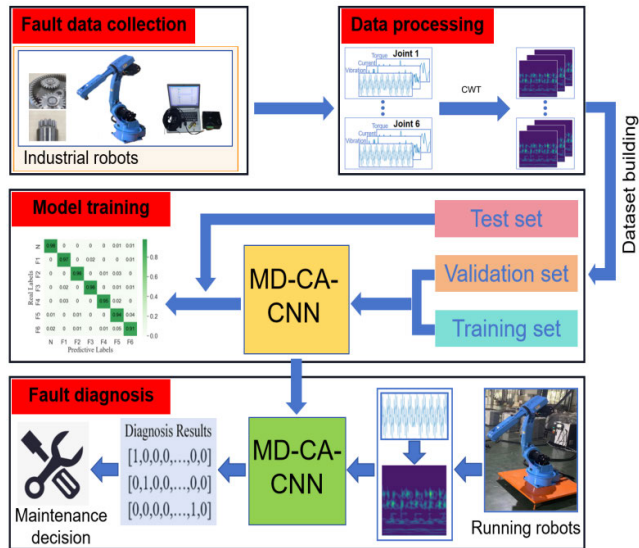

FIGURE 1. The overall diagnostic framework of MD-CA-CNN.

In the data preprocessing stage, a sliding window method segments the data along the time dimension. The sliding window method is a commonly used data enhancement method in fault diagnosis, which can significantly increase the number of training samples. As shown in Fig. 2, a time window of length w is used to segment the data, the step size is set to s, and the total number of data points is T. The formula for the number of training samples N is shown in (1), and the result is rounded down to the nearest integer.

$$N = \frac{T - w}{s} + 1 \tag{1}$$

The signals acquired in each time window are then converted into time-frequency images using the continuous wavelet transform (CWT), which reduces the noise effect and reveals fault-related features. The time-frequency images generated by CWT contain rich time-domain and frequency-domain information. The original time-domain signals also contain the most essential fault characterization information. Both are helpful for diagnosing robot faults, so both the time-domain and time-frequency-domain signals are used as input samples for model training.

In the model training stage, the processed data are randomly divided into three independent datasets, i.e., training set, validation set, and testing set, in the ratio of 7:2:1. The

structure of the MD-CA-CNN model was determined by pre-experimentation. The proposed MD-CA-CNN model is trained, evaluated, and optimized offline using the training and validation set samples and their corresponding labels. The loss function is computed, and the model weights are updated according to the backpropagation principle. The test set samples are used to test the accuracy of the MD-CA-CNN model after training.

In the fault diagnosis stage, the trained MD-CA-CNN model is deployed, and the robot field data is collected as input to the model using the same time window length as the data preprocessing stage to obtain the output results. Then, the diagnostic results are used to provide decision support for maintenance.
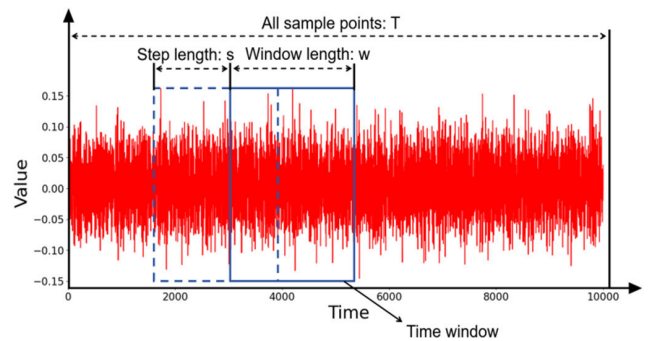

FIGURE 2. The method of sliding window.

B. CONTINUOUS WAVELET TRANSFORM

In the field of signal analysis, the commonly used STFT can locate the time. However, it has the same time and frequency resolution for different frequencies because the window size is fixed. Moreover, the length of the window function is also challenging to determine. Therefore, STFT is more suitable for smooth signals with small frequency fluctuations than non-smooth signals with large frequency fluctuations. CWT is an adaptive time-frequency analysis method due to the introduction of wavelet function as the basis function. It can automatically adjust the window size according to the height of the frequency and can more obviously detect the mutation point and oscillation part of the signal. Therefore, It is more conducive to the analysis of faulty signals. The CWT formula of the signal $x(t)$ is shown below:

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

where $\psi\left(\frac{t-b}{a}\right)$ is the mother wavelet and $\bar{\psi}\left(\frac{t-b}{a}\right)$ is the complex conjugate of $\psi\left(\frac{t-b}{a}\right)$; a is the scale factor, b is the translation factor, and both a and b are arbitrary real numbers; $X_\omega(a, b)$ is the wavelet coefficient.

C. CONVOLUTIONAL NEURAL NETWORK

In recent years, 2D-CNN has occupied an essential position in computer vision and image recognition fields. However,

2D-CNN is only a viable option for some applications on 1D signals. To solve this problem, 1D-CNN has been proposed. It has achieved state-of-the-art performance levels in several applications, such as biomedical data categorization and early diagnosis, as well as in power electronics and motor fault detection [33]. The principle of 1D-CNN and 2D-CNN is the same. The main difference is that the convolution kernel moves in different dimensions. The convolution kernels of 1D-CNN will only be convolved along the time-step order. In contrast, the convolution kernels of 2D-CNN will be convolved along the image's horizontal and vertical axes. The MD-CA-CNN proposed in this paper is based on 1D-CNN and 2D-CNN together.

The structure of CNN mainly consists of alternating convolutional layers, nonlinear layers, pooling layers, and fully connected output layers.

Convolutional layers are the most important part of CNN. Each convolutional layer contains multiple convolutional kernels. Unlike the traditional fully connected layer, the convolutional layer performs convolutional operations through convolutional kernels. So, the input to each node is only a part of the neurons of the previous neural network layer. Therefore, the convolution operation can extract the input data's feature information and significantly reduce the model computing burden and the noise interference on the features. The formula for the convolution operation is shown below:

$$y_i^l = \sum_{r=1}^{M} x_r^{l-1} \otimes k_{i,r}^l + b_i^l, \quad i = 1, 2, \cdots, N \quad (3)$$

where N represents the number of convolution kernels in the lth convolutional layer; M represents the number of convolution kernels in the l-1th convolutional layer; x_r^{l-1} denotes the rth feature map in the l-1th convolutional layer; k_i^l is the rth channel in the ith convolution kernel in the lth convolutional layer; b_i^l denotes the bias of the ith convolution kernel in the lth convolutional layer; and y_i^l is the output corresponding to the ith convolution kernel in the lth convolutional layer. $\otimes$ denotes the convolution operation.

The nonlinear layer is an integral part of the neural network. Generally, after the convolutional layer, the activation function is utilized to make a nonlinear mapping of the convolutional layer output. There are many activation functions, but in recent years, the ReLU function has been popular because of its computational simplicity, ability to prevent gradient vanishing, and other advantages. The formula for the ReLU function is shown below:

$$y^l = \max\left(0, x^l\right) \quad (4)$$

where x^l denotes the input of the lth ReLU nonlinear layer and y^l denotes the output of the lth ReLU nonlinear layer.

The pooling layer is generally sandwiched between consecutive convolutional layers and compresses the data while preserving the data feature information. It can reduce the number of parameters, effectively control the overfitting, and improve the network operation speed. Pooling methods include maximum pooling and average pooling, while the maximum pooling method is actually used more often. The formula for the maximum pooling method is shown below:

$$y_{k,i}^l = \max\left(x_{k,i,1}^l, \cdots, x_{k,i,j}^l, \cdots, x_{k,i,j^2}^l\right), \ k = 1, 2, \cdots, N$$
$$(5)$$

where N represents the number of output feature maps, which is the same as the number of output feature maps; represents the jth element of the ith image block of the kth input feature map, each image block contains j^2 elements and represents the ith element on the kth output feature map of the lth layer.

The final part of the CNN consists of a fully connected layer and an output layer for classification. First, the features obtained from the previous convolutional and pooling layers are flattened and fully connected and then passed through the fully connected layer to the output layer to output the results. The output layer usually uses the softmax function due to its high efficiency in classification tasks. The operations performed in the fully connected output layer are shown below:

$$y^l = f\left(w^l \bullet x^{l-1} + b^l\right) \quad (6)$$

$$s^l = \text{softmax}\left(y^l\right) \quad (7)$$

where x^{l-1} is the output of layer $l - 1$, and w^l is the weight coefficient of the fully connected layer l; b^l is the bias of the fully connected layer l. y^l is the output of the fully connected layer l. s^l denotes the final classification result obtained after the softmax activation function.

D. ATTENTION MECHANISM

Attention mechanism is an approach in deep learning based on the human visual system and cognitive mechanisms. When humans are confronted with a large amount of information, they will efficiently allocate their limited attentional resources to selectively focus on the information most valuable to them from a large amount of information [34]. The attention mechanism's core goal is to selectively focus on the information that is more critical to the current task goal from a multitude of inputs and assign higher weights to it to enhance its contribution to the outcome. The attention mechanism can improve the expressiveness of deep learning models and increase the model's sensitivity to key features [35].

Channel Attention Mechanism, a type of attention mechanism, can associate different channels of the input data to calculate each channel's degree of importance and assign weights to different channel features. It improves the expressive power of the network in the feature representation, which improves the model's performance. Therefore, it is often used inside convolutional neural networks [35]. The computational procedure of the channel attention mechanism constructed in this paper is as follows:

The input sample is denoted as $S = \{S_1, S_2, \cdots, S_i, \cdots, S_h\}$ and h is the number of input sample channels.
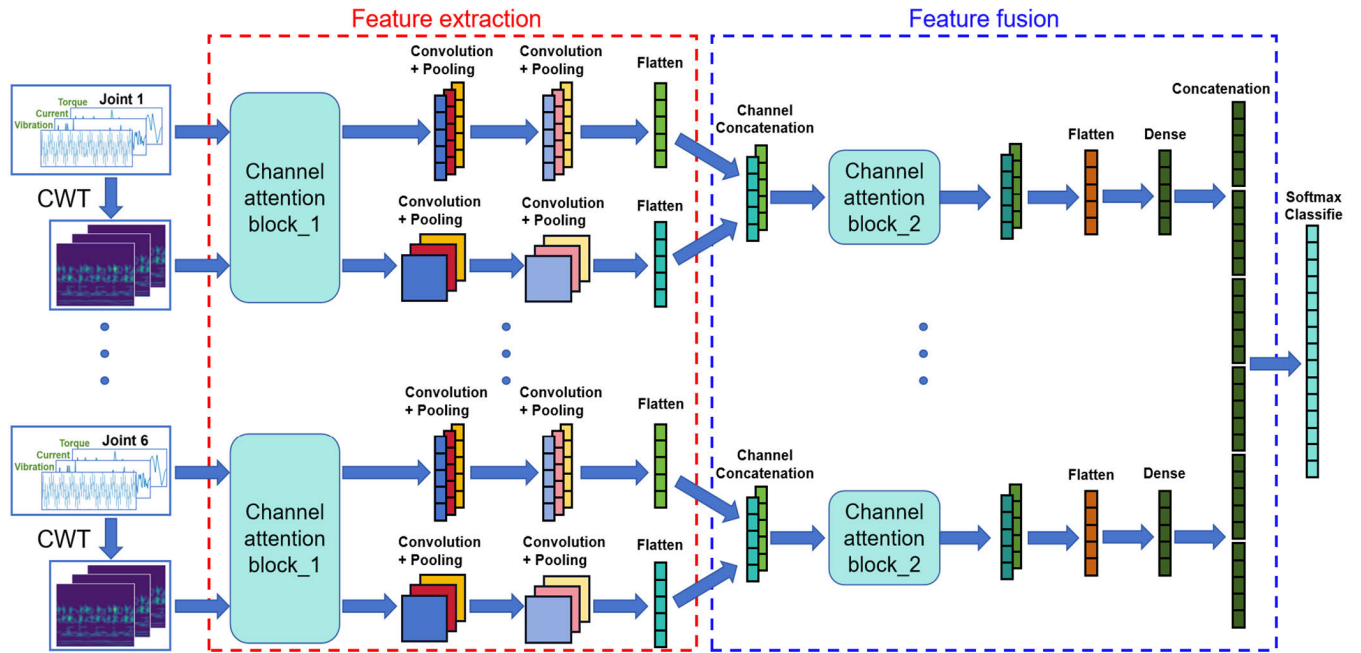
FIGURE 3. The structure of MD-CA-CNN.

1) Global average pooling is performed on the feature data for each channel:

$$m_i = \text{Avgpool}(S_i) = \frac{1}{l}\sum_{j=1}^{l} S_i^j \qquad (8)$$

where S_i denotes the feature sequence of the ith channel, and l denotes the length of the feature sequence. S_i^j denotes the jth data of the feature sequence of the ith channel. m_i is the result of the global average pooling of the ith channel.

2) Find the score of the ith channel feature:

$$F = \sigma(W \bullet M + b) \qquad (9)$$

where σ is a scoring function, such as a sigmoid function or a ReLU function. $M = \{m_1, m_2, \cdots, m_i, \cdots, m_h\}$, $F = \{f_1, f_2, \cdots, f_i, \cdots, f_h\}$, f_i is the score of the ith channel feature.

3) Normalization of scores:

$$\alpha_i = \text{softmax}(f_i) = \frac{\exp(f_i)}{\sum_i f} \qquad (10)$$

where α_i is the weight assigned to the ith channel data S_i by the channel attention mechanism.

4) Obtains the final output M of the attention mechanism:

$$M = S \otimes A = \{s_1\alpha_1, s_2\alpha_2, \cdots, s_h\alpha_h\} \qquad (11)$$

where $A = \{\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots, \alpha_h\}$, $\otimes$ is the multiplication of the corresponding elements of two sets.

E. PROPOSED MD-CA-CNN STRUCTURE

In this section, the specific steps of the proposed method are described. Fig. 3 shows the system structure of the proposed MD-CA-CNN. The proposed method uses the initial timing signals and time-frequency images of vibration, current, and torque of six joints as the input to the system. The feature extraction model is constructed for each of the six joints. Finally, the extracted features of the six joints are connected in series to determine the final diagnosis result. The proposed structure consists of two main parts. In the feature extraction part, the weights of the feature data of different channels are firstly determined by the channel attention mechanism and by combining the time domain information with the time-frequency domain information. Then, we extract the time-domain and time-frequency domain features using 1DCNN and 2DCNN, respectively. In the feature fusion part, the time domain and time-frequency domain features are first fused, and then the weights of the time domain features and time-frequency domain features are determined using the channel attention mechanism. Afterward, the data of the two channels are fused in series, and the fused features of the joint are obtained through a fully connected layer. After that, the fusion features obtained for each joint are again fused in series to obtain the final fusion features of the robot. Finally, the Softmax layer is utilized to obtain the fault diagnosis results.

Since the fusion feature extraction models of the six joints are constructed in the same way, we use one of the joints as an example to describe the specific steps of the proposed method.

1) FEATURE EXTRACTION

Before feature extraction of input data using CNN, the weights of different channel data are first determined by Channel attention block_1. The structure of Channel attention block_1 is shown in Fig. 4. Because the input data contains both time-domain and time-frequency domain
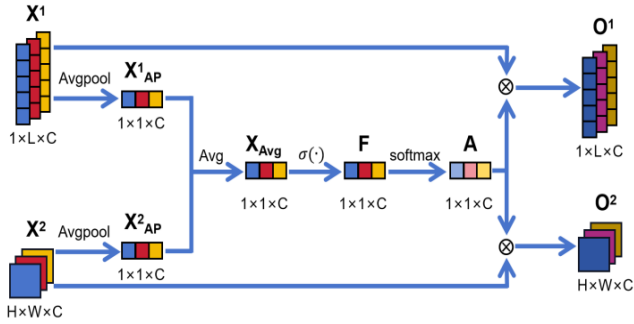
FIGURE 4. The structure of channel attention block_1.

representations of the same data, the effects of time-domain data and time-frequency domain data on the channel weights are considered simultaneously.

Global average pooling is first performed in Channel attention block_1 for each channel of the two types of input data X1 and X2, respectively. The formulas are as follows:

$$X_{AP,i}^1 = \text{Avgpool}(X_i^1) = \frac{1}{L}\sum_{j=1}^{L} X_{i,j}^1 \qquad (12)$$

$$X_{AP,i}^2 = \text{Avgpool}(X_i^2) = \frac{1}{H \times W}\sum_{j=1}^{H \times W} X_{i,j}^2 \qquad (13)$$

where $X_{i,j}^1$ and $X_{i,j}^2$ are the jth data of the ith channel of X^1 and X^2, respectively; $X_{AP,i}^1$ and $X_{AP,i}^2$ are the ith data of X_{AP}^1 and X_{AP}^2, respectively.

Then find the average of the corresponding elements of X_{AP}^1 and X_{AP}^2 to obtain X_{Avg}. The equation is as follows:

$$X_{Avg,i} = \frac{1}{2}(X_{AP,i}^1 + X_{AP,i}^2) \qquad (14)$$

After that find the score for each channel feature. The formula is as follows:

$$F = \sigma(W \bullet X_{Avg} + b) \qquad (15)$$

In this paper the scoring function σ uses the Sigmoid function; $F = \{f_1, f_2, \cdots, f_C\}$.

Next the score obtained for the ith channel is normalized using softmax function with the following formula:

$$\alpha_i = \text{softmax}(f_i) = \frac{\exp(f_i)}{\sum_i f} \qquad (16)$$

where α_i is the weight assigned to the ith channel.

Finally, each obtained channel weight is multiplied by the corresponding channel data of the initial timing signal and the time-frequency domain signal, respectively. The final output of Channel attention block_1 is obtained. The formula is as follows:

$$O^1 = X^1 \otimes A = \{X_1^1\alpha_1, X_2^1\alpha_2, \cdots, X_C^1\alpha_C\} \qquad (17)$$

$$O^2 = X^2 \otimes A = \{X_1^2\alpha_1, X_2^2\alpha_2, \cdots, X_C^2\alpha_C\} \qquad (18)$$

where $A = \{\alpha_1, \alpha_2, \cdots, \alpha_i, \cdots, \alpha_h\}$. O_1 and O_2 are the time domain data and time-frequency domain data after Channel attention block_1 assigns channel weights, respectively.

After Channel attention block_1, CNN is utilized to extract deep learning features. The 1D-CNN and 2D-CNN used in this paper extract the deep features of the time and time-frequency domain signals, respectively. Both 1D-CNN and 2D-CNN used in this paper consist of two convolutional modules and a flattened layer. Each convolutional module, in turn, consists of a convolutional layer, a nonlinear layer, a convolutional layer, a nonlinear layer, and a pooling layer. Finally, the Flatten layer performs a spreading operation on the extracted feature data. The parameters of the 1D-CNN and 2D-CNN convolutional modules used in this paper are determined experimentally, and the details are shown in Table 1.

TABLE 1. CNN parameters.

	Layer	Parameters
1D-CNN	1D convolution layer	filters=10; kernel size=32; Padding="same"; Stride=1
	Nonlinear layers	ReLU
	1D MaxPooling layer	kernel size=4
	1D convolution layer	filters=10; kernel size=16; Padding="same"; Stride=1
	Nonlinear layers	ReLU
	1D MaxPooling layer	kernel size=4
2D-CNN	2D convolution layer	filters=10; kernel size=32; Padding="same"; Stride=1
	Nonlinear layers	ReLU
	2D MaxPooling layer	kernel size=4
	2D convolution layer	filters=10; kernel size=16; Padding="same"; Stride=1
	Nonlinear layers	ReLU
	2D MaxPooling layer	kernel size=4

2) FEATURE FUSION

Firstly, the channels of the two 1D sequence data from the feature extraction part are spliced to obtain two-channel sequence data. Then, the weights of the data of two channels (i.e., the weights of the time-domain features and the time-frequency-domain features) are determined by Channel attention block_2. The structure of Channel attention block_2 is shown in Fig. 5.
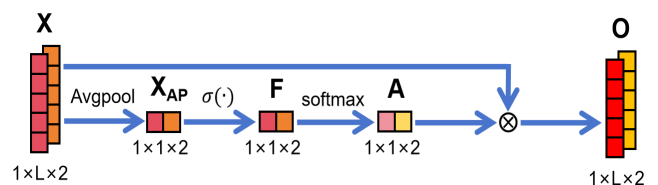

FIGURE 5. The structure of channel attention block_2.

A global average pooling operation is first performed on the dual channel data X in Channel attention block_2. The formula is as follows:

$$X_{AP,i} = Avgpool(X_i) = \frac{1}{L}\sum_{j=1}^{L} X_{i,j}, \quad i = 1, 2 \quad (19)$$

where $X_{i,j}$ is the jth data of the ith channel of data X; $X_{AP,i}$ is the ith data of X_{AP}.

After that find the score of each channel. The formula is as follows:

$$F = \sigma(W \bullet X_{AP} + b) \quad (20)$$

Here the scoring function σ still uses the Sigmoid function; $F = \{f_1, f_2\}$.

The score obtained for the ith channel is normalized. The formula is as follows:

$$\alpha_i = \text{softmax}(f_i) = \frac{\exp(f_i)}{\sum_i f}, \quad i = 1, 2 \quad (21)$$

where α_i is the weight assigned to the ith channel.

Finally, each of the obtained channel weights is multiplied with the corresponding channel data of the initial signal X respectively to obtain the final output of Channel attention block_2. The formula is given below:

$$O = X \otimes A = \{X_1\alpha_1, X_2\alpha_2\} \quad (22)$$

where $A = \{\alpha_1, \alpha_2\}$. O is the feature data after Channel attention block_2 gives channel weights.

After Channel attention block_2, the two channels data of feature data O are fused in series through the Flatten layer to form a one-dimensional sequence data. After that, the fused feature T of the joint is obtained through a fully connected layer. The formula is as follows:

$$T = W \bullet O + b \quad (23)$$

After that, the obtained fused features for each joint are again concatenated in tandem to obtain the final fused features of the robot. Finally, the Softmax layer is utilized to obtain the fault diagnosis results. The Softmax classifier is the most common classifier in deep learning, and its output reflects the probability distribution of features over the label space, which can handle multicategorization problems well [36]. Compared with some other classifiers, the Softmax function is computationally more stable and can be easily parallelized on GPUs, especially when the number of categories is not very large, with relatively low computational complexity. In contrast, some other classifiers (e.g., support vector machines) may require more complex computations, resulting in longer computation times.

III. EXPERIMENT
A. EXPERIMENTAL PLATFORM CONSTRUCTION
We construct a testbed to verify the performance of the proposed fault diagnosis method for industrial robots. As shown in Fig. 6, the testbed consists of a 6-axis multi-joint
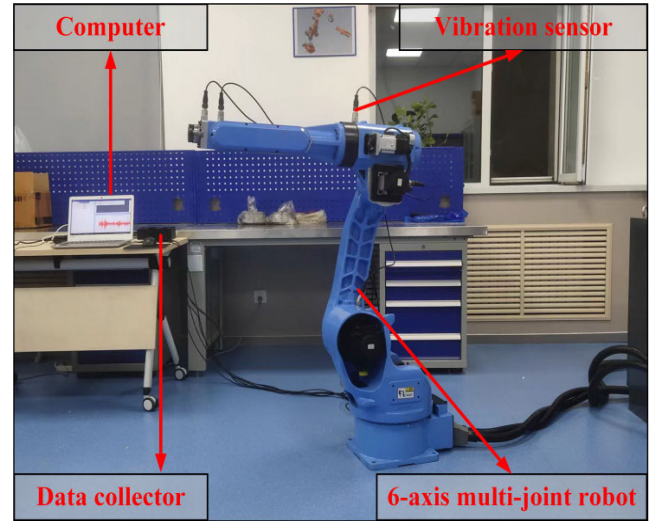

FIGURE 6. Robot experiment platform.

robot, a laptop computer (MECHREVO), a data collector (IN-SDG), and six vibration sensors. We set the task of the robot to carry a 5kg weight.

B. EXPERIMENTAL DATA ACQUISITION AND PROCESSING
Robot reducers are an essential power transmission device in the drive system of 6-axis multi-joint industrial robots. It can help the robot realize precise motion and command transmission and ensure that each part works closely together to complete complex work. In this experiment, four kinds of single faults and two kinds of compound faults of the reducers are simulated through human fault injection. As shown in Fig. 7, which are pitting of the sun gear of the second joint reducer (F1), cracking of the planetary gear of the second joint reducer (F2), cracking of the sun gear of the third joint reducer (F3), pitting of the planetary gear of the third joint
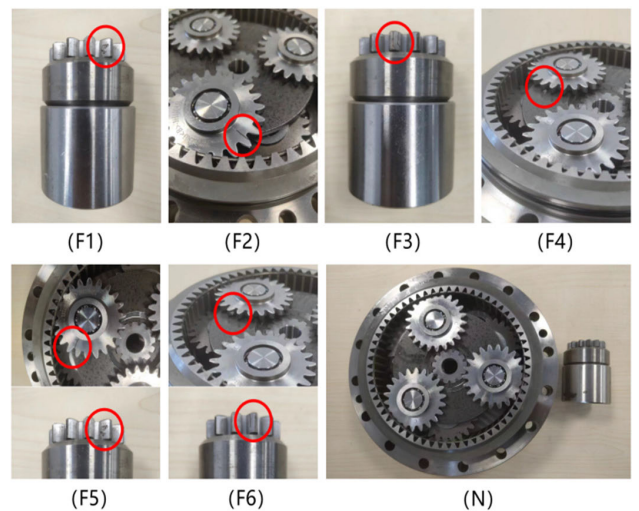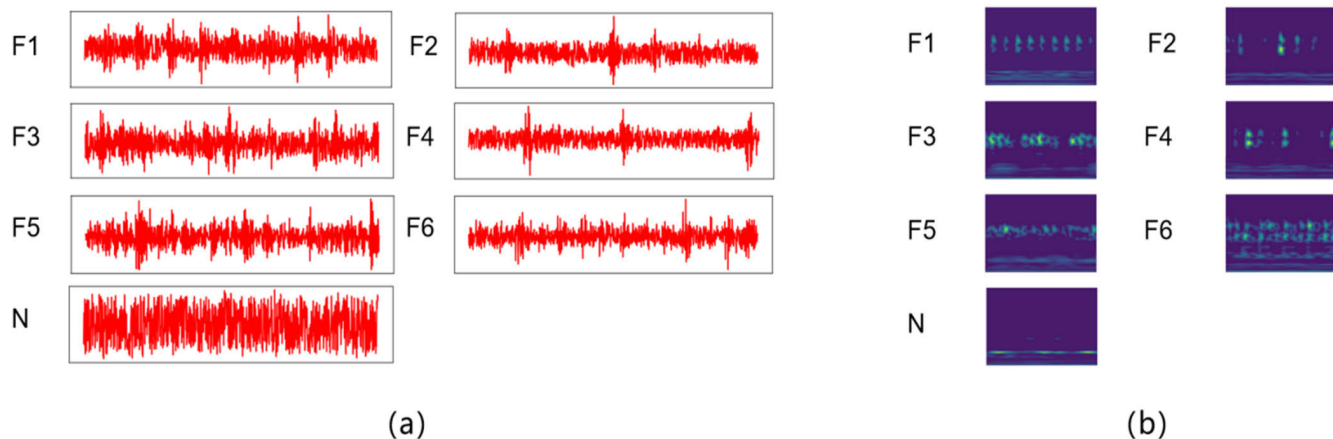

FIGURE 7. Failure modes.

FIGURE 8. Vibration data of third joint.

reducer (F4), pitting of the sun gear and cracking of the planetary gear of the second joint reducer (F5), cracking of the sun gear of the second joint reducer and pitting of the planetary gear of the third joint reducer (F6). In addition, the normal reducer mode (N) is added, realizing a total of seven different types of data to be simulated, as shown in Figure 7.

The vibration signals of each joint were acquired through vibration sensors, and the current and torque signals were acquired through the robot controller. The sampling frequency is set to 1024 Hz, and 750 data segments are obtained in each mode. Each segment contains 5120 data points. The window length is set to 512, and the step size is set to 200. Twenty-four samples are obtained for each segment of data by (1). So, a total of 18,000 samples are obtained in each mode. Each sample includes signals from three channels: vibration, current, and torque. Then, the time-frequency image of the sample data is obtained using CWT. The size of the time-frequency image is 512×512, as shown in Fig. 8, which shows the vibration data with a time-frequency image for each mode of the third joint. Finally, all is divided into three parts: 70 % is the training dataset, 20 % is the validation dataset, and 10 % is the test dataset, as shown in Table 2.

TABLE 2. Description of the data set.

Type	Training/Validation/Test	Sample unit size
N	12600/3600/1800	(6, (3,512), (3,512,512))
F1	12600/3600/1800	(6, (3,512), (3,512,512))
F2	12600/3600/1800	(6, (3,512), (3,512,512))
F3	12600/3600/1800	(6, (3,512), (3,512,512))
F4	12600/3600/1800	(6, (3,512), (3,512,512))
F5	12600/3600/1800	(6, (3,512), (3,512,512))
F6	12600/3600/1800	(6, (3,512), (3,512,512))

C. DESIGN OF COMPARATIVE EXPERIMENTS

In order to analyze the performance of the MD-CA-CNN model based on multi-source data, experiments were designed to compare it with other commonly used machine learning and deep learning models. The model used in this experiment was written based on Python 3.8 utilizing Pytorch framework and Scikit-learn machine learning library. Moreover, the model was trained on a laptop with an NVIDIA RTX 3060 GPU.

Firstly, the models proposed in this paper are compared with the models that only utilize 1D-CNN or 2D-CNN for feature extraction (MD-CA-1D-CNN, MD-CA-2D-CNN). The MD-CA-1D-CNN model only utilizes the original vibration signals, the original current signals, and the original torque signals as inputs. The MD-CA-2D-CNN model only utilizes the time-frequency images of the original vibration signals, the current signals, and the torque signals as inputs. Second, the model proposed in this paper is compared with a model that does not contain channel attention but utilizes multiple sources of data as inputs (MD-CNN) and CNNs that utilize a single source of data as inputs, including a convolutional neural network that uses only raw vibration signals (V-CNN), a convolutional neural network that uses only raw current signals (C-CNN), and a convolutional neural network that uses only raw torque signals (T-CNN). Finally, the model proposed in this paper is compared with other commonly used deep learning models and machine learning models. Deep learning models used as comparisons include Long Short-Term Memory Network (LSTM), Gated Recurrent Unit (GRU). Machine learning models used as comparisons include Random Forest (RF) [37], K-Nearest Neighbor Algorithm (KNN), and Support Vector Machines (SVM) [38]. LSTM and GRU are commonly used models for time-series tasks, and are generally only suitable for using time-series data as inputs to the model. Since traditional machine learning algorithms are unsuitable for directly taking the raw time series data as input, some statistical features of the raw time series data are used as algorithmic inputs

after standardization. The selected statistical features include maximum value, mean value, root mean square value, peak-to-peak value, crag, skewness, margin indicator, and impulse indicator. All these features can well characterize the hidden faults of mechanical equipment.

IV. RESULTS AND DISCUSSION

In order to reduce the effect of random factors, we conducted five trials for each algorithm model. Each experiment randomly shuffles the data order in the dataset. And in order to further reduce randomness and improve the model's generalization ability, a ten-fold cross-validation method was used in each experiment, and the average value obtained was taken as the result of each experiment. Figure 9 shows the test accuracies for each trial for each algorithm. Figure 10 shows the violin plots of the test results of different algorithms. Table 3 shows in more detail the detailed accuracies for each experiment, the average test results for each algorithm, and the corresponding standard deviations.
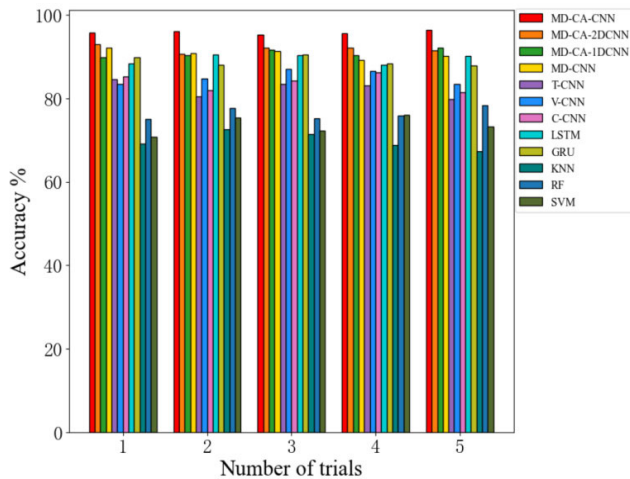

FIGURE 9. Diagnostic results for each algorithm in five trials.

The results show that the diagnostic accuracy of the proposed MD-CA-CNN method is 95.8%, with a standard deviation value of 0.39%. The diagnostic accuracy of the MD-CA-2D-CNN method is 91.8%, and the standard deviation value is 0.80%. The diagnostic accuracy of the MD-CA-1D-CNN method is 90.9%, and the standard deviation value is 0.83%. Through comparison, it can be found that the MD-CA-CNN method can simultaneously utilize 1D-CNN and 2D-CNN to extract time-domain features and time-frequency-domain features from the time series data, obtaining more comprehensive feature information, which improves the performance of model fault diagnosis. The diagnostic accuracy of the MD-CNN method is 90.7%, with a standard deviation value of 0.99%. By comparison, it can be found that the channel attention mechanism improves the performance of the model fault diagnosis by adaptively assigning weights to different source data as well as to time-domain and time-frequency-domain features. The diagnostic accuracy of
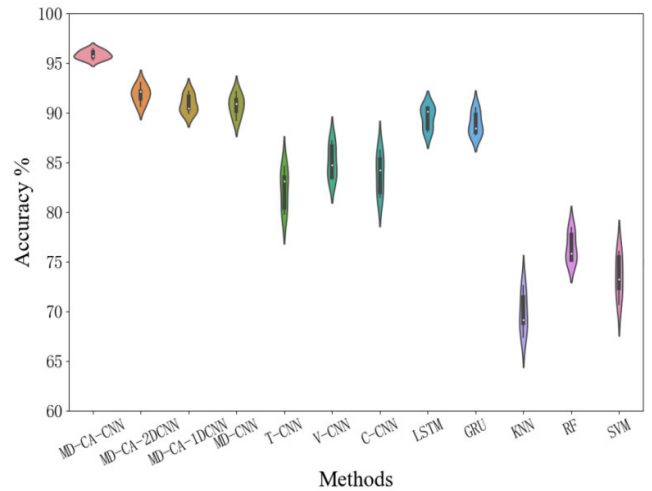

FIGURE 10. Violin plots of the diagnostic results for each algorithm.

TABLE 3. Detailed accuracy of experiments.

Methods	Single experiment accuracy (%)					Average test accuracy (%)
	1	2	3	4	5	
MD-CA-CNN	**95.7**	**96.1**	**95.3**	**95.6**	**96.4**	**95.8 (±0.39)**
MD-CA-2D-CNN	93.0	90.6	92.1	92.1	91.4	91.8 (±0.80)
MD-CA-1D-CNN	89.9	90.4	91.6	90.4	92.1	90.9 (±0.83)
MD-CNN	92.1	90.9	91.3	89.2	90.2	90.7 (±0.99)
T-CNN	84.6	80.4	83.5	83.1	79.8	82.3 (±1.86)
V-CNN	83.5	84.7	87.1	86.6	83.4	85.1 (±1.54)
C-CNN	85.3	82.0	84.2	86.2	81.5	83.8 (±1.83)
LSTM	88.4	90.5	90.4	88.1	90.1	89.5 (±1.03)
GRU	89.8	88.0	90.5	88.4	87.8	88.9 (±1.06)
KNN	69.1	72.6	71.4	68.8	67.4	69.9 (±1.88)
RF	75.1	77.7	75.2	75.8	78.4	76.4 (±1.35)
SVM	70.7	75.4	72.3	76.0	73.2	73.5 (±1.96)

T-CNN is 82.3%, with a standard deviation value of 1.86%. The diagnostic accuracy of V-CNN is 85.1%, with a standard deviation value of 1.54%. The diagnostic accuracy of C-CNN is 83.8%, with a standard deviation value of 1.83%. By comparison, it can be found that MD-CA-CNN and MD-CNN utilizing multi-source data fusion have higher diagnostic accuracy and more stable performance than T-CNN, V-CNN, and C-CNN utilizing only a single data source. The diagnostic accuracy of the LSTM algorithm is 89.5%, and the standard deviation value is 1.03%. The diagnostic accuracy of the GRU algorithm is 88.9%, and the standard deviation value is 1.06%. By comparison, it can be found that the MD-CA-CNN method proposed in this paper has a higher performance of fault diagnosis than deep learning models such as LSTM and GRU. Among the machine learning algorithms, the diagnostic accuracy of KNN is 69.9%, with a standard deviation value of 1.88%. The diagnostic accuracy of RF is 76.4%,
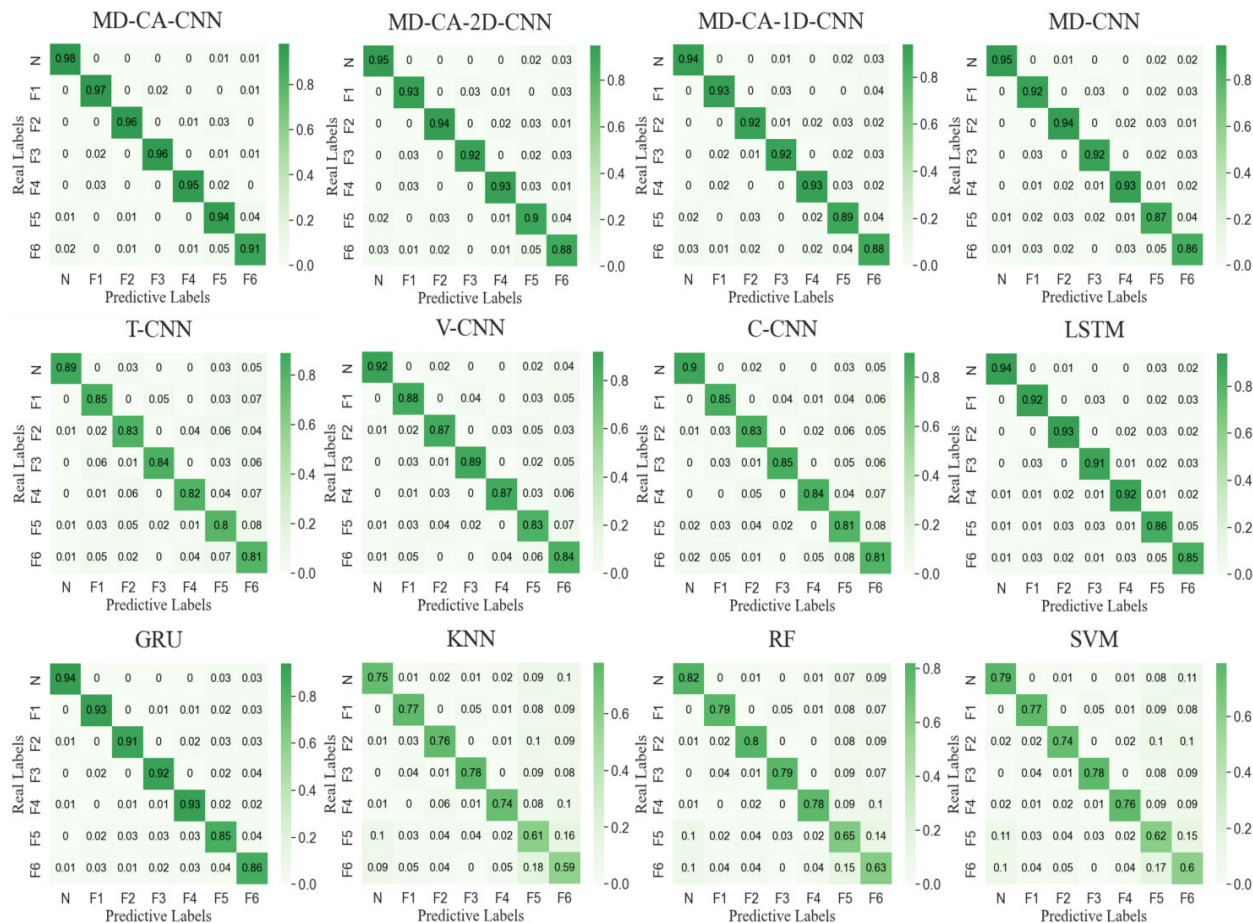
FIGURE 11. Confusion matrix for the 3rd trial of different algorithms.

with a standard deviation value of 1.35%. The diagnostic accuracy of SVM is 73.5%, with a standard deviation value of 1.96%. By comparing with the previous deep learning algorithms, it can be found that the diagnostic accuracy and stability performance of deep learning algorithms are much higher than that of machine learning algorithms. It reflects the advantages of deep learning algorithms in dealing with long-time series problems.

In addition, Fig. 11 shows the confusion matrix for the third trial of the different algorithms, where the classification and actual results are presented. All three machine learning algorithms (KNN, RF, SVM) have low diagnostic accuracy for all types of faults, especially for composite faults F5 and F6, which may be because machine learning is unsuitable for handling long time series. The three deep learning algorithms (T-CNN, V-CNN, C-CNN) utilizing a single data source significantly improve diagnostic accuracy for all types of faults over all three machine learning algorithms. However, the results are still less satisfactory, which may be due to the lack of robustness of the single data source. The MD-CNN, LSTM, and GRU have better diagnostic accuracy for a single fault type. However, they are also prone to misclassification for composite fault modes F5 and F6, which

may be due to confusion about the importance of multi-source data. MD-CA-CNN algorithm has higher diagnostic accuracy for various fault modes and normal modes, possibly because the multi-source data provides more comprehensive information and improves robustness. As well as the channel attention mechanism adaptively assigns weights to the multi-source data and to the time-domain and time-frequency-domain features, which improves the performance of the model's fault diagnosis, especially the diagnostic accuracy for F5 and F6.

V. CONCLUSION

In this paper, we proposed a multi-joint industrial robot fault diagnosis model based on multi-source data fusion and channel-attention convolutional neural network, namely MD-CA-CNN. The model takes the vibration data, torque data and current data of the robot as inputs. The proposed method differs from existing multi-source data fusion techniques in the field of fault diagnosis in the following ways: firstly, the fusion of vibration, current and torque data of robot joints is realized by using CNN channel superposition. Secondly, multiple dimensions of information such as original time-domain signals and time-frequency-domain signals are taken into account, and both 1D-CNN and 2D-CNN are

utilized to extract the deep features in both time-domain and time-frequency-domain signals dimensions to obtain more comprehensive fault feature information. Finally, a channel attention mechanism is developed to adaptively assign weights to different source data as well as to time-domain and time-frequency-domain features, thus improving the contribution of important channels to fault diagnosis.

We built a testbed to test the performance of the proposed method by diagnosing a variety of six-axis robot RV reducer failures. The performance of the proposed method is tested by comparing it with a model that utilizes only 1D-CNN or 2D-CNN for feature extraction (MD-CA-1D-CNN, MD-CA-2D-CNN), a deep learning algorithm that utilizes only the fusion of multiple data sources (MD-CNN), three deep learning algorithms that utilize only a single data source (T-CNN, V-CNN, and C-CNN), two commonly used deep learning algorithms (LSTM, GRU), and three commonly used machine learning algorithms (KNN, RF, SVM) are compared and analyzed, and the MD-CA-CNN algorithm has the highest accuracy and stability (95.8% ± 0.39%). Also, the following conclusions can be obtained:

1) Multi-source data can provide more comprehensive feature information. The combination of redundant or complementary information from multiple sources in space or time can reduce the disadvantage of a single signal susceptible to noise. In addition, simultaneously extracting the depth features of the original time-domain signal and the time-frequency domain signal can obtain more comprehensive fault feature information and improve the robustness of the algorithm;

2) The channel attention mechanism developed in this paper can model the importance of each feature channel and enhance or suppress different channels, which is beneficial to improve the performance of the model;

3) In the field of fault diagnosis, deep learning algorithms are more advantageous than machine learning algorithms in dealing with high-complexity, long-time series problems.

The model proposed in this paper takes a long time to train due to its large structure and many parameters, but the model after training meets the time requirements for real-time diagnosis in the field when performing fault diagnosis tasks. In future research, we will focus on reducing the complexity of the proposed method.

REFERENCES

[1] U. Izagirre, I. Andonegui, I. Landa-Torres, and U. Zurutuza, "A practical and synchronized data acquisition network architecture for industrial robot predictive maintenance in manufacturing assembly lines," *Robot. Comput.-Integr. Manuf.*, vol. 74, Apr. 2022, Art. no. 102287.

[2] G. Cui, B. Li, W. Tian, W. Liao, and W. Zhao, "Dynamic modeling and vibration prediction of an industrial robot in manufacturing," *Appl. Math. Model.*, vol. 105, pp. 114–136, May 2022.

[3] C. Nentwich and R. Daub, "Cost-benefit analysis of industrial robot gear condition monitoring," *Proc. CIRP*, vol. 107, pp. 143–148, Jan. 2022.

[4] E. Khalastchi and M. Kalech, "On fault detection and diagnosis in robotic systems," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 1–24, Jan. 2018.

[5] B. Chen, F. Peng, H. Wang, and Y. Yu, "Compound fault identification of rolling element bearing based on adaptive resonant frequency band extraction," *Mechanism Mach. Theory*, vol. 154, Dec. 2020, Art. no. 104051.

[6] B. A. E. Anouar, "Application of wavelet analysis and its interpretation in rotating machines monitoring and fault diagnosis," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 3465–3471, Jan. 2018.

[7] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.

[8] Y. Kim, J. Park, K. Na, H. Yuan, B. D. Youn, and C.-S. Kang, "Phase-based time domain averaging (PTDA) for fault detection of a gearbox in an industrial robot using vibration signals," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106544.

[9] Y.-S. Sun, X.-R. Ran, Y.-M. Li, G.-C. Zhang, and Y.-H. Zhang, "Thruster fault diagnosis method based on Gaussian particle filter for autonomous underwater vehicles," *Int. J. Nav. Archit. Ocean Eng.*, vol. 8, no. 3, pp. 243–251, May 2016.

[10] A. H. Sabry, F. H. Nordin, A. H. Sabry, and M. Z. A. A. Kadir, "Fault detection and diagnosis of industrial robot based on power consumption modeling," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7929–7940, Sep. 2020.

[11] R. Muradore and P. Fiorini, "A PLS-based statistical approach for fault detection and isolation of robotic manipulators," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3167–3175, Aug. 2012.

[12] J. Long, Y. Qin, Z. Yang, Y. Huang, and C. Li, "Discriminative feature learning using a multiscale convolutional capsule network from attitude data for fault diagnosis of industrial robots," *Mech. Syst. Signal Process.*, vol. 182, Jan. 2023, Art. no. 109569.

[13] K. Guo, X. Wan, L. Liu, Z. Gao, and M. Yang, "Fault diagnosis of intelligent production line based on digital twin and improved random forest," *Appl. Sci.*, vol. 11, no. 16, p. 7733, Aug. 2021.

[14] K. Xu, X. Wu, D. Wang, and X. Liu, "Electromechanical coupling modeling and motor current signature analysis of bolt loosening of industrial robot joint," *Mech. Syst. Signal Process.*, vol. 184, Feb. 2023, Art. no. 109681.

[15] I. Raouf, H. Lee, and H. S. Kim, "Mechanical fault detection based on machine learning for robotic RV reducer using electrical current signature analysis: A data-driven approach," *J. Comput. Des. Eng.*, vol. 9, no. 2, pp. 417–433, Feb. 2022.

[16] M.-C. Pan, H. Van Brussel, and P. Sas, "Intelligent joint fault diagnosis of industrial robots," *Mech. Syst. Signal Process.*, vol. 12, no. 4, pp. 571–588, Jul. 1998.

[17] Y. He, C. Zhao, X. Zhou, and W. Shen, "MJAR: A novel joint generalization-based diagnosis method for industrial robots with compound faults," *Robot. Comput.-Integr. Manuf.*, vol. 86, Apr. 2024, Art. no. 102668.

[18] Á. Nagy and I. Vajk, "Sequential time-optimal path-tracking algorithm for robots," *IEEE Trans. Robot.*, vol. 35, no. 5, pp. 1253–1259, Oct. 2019.

[19] Z. Zhi, L. Liu, D. Liu, and C. Hu, "Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2572–2581, Feb. 2022.

[20] S. Yildirim and I. Eski, "Noise analysis of robot manipulator using neural networks," *Robot. Comput.-Integr. Manuf.*, vol. 26, no. 4, pp. 282–290, Aug. 2010.

[21] Y. He, J. Chen, X. Zhou, and S. Huang, "In-situ fault diagnosis for the harmonic reducer of industrial robots via multi-scale mixed convolutional neural networks," *J. Manuf. Syst.*, vol. 66, pp. 233–247, Feb. 2023.

[22] D. Park, S. Kim, Y. An, and J.-Y. Jung, "LiReD: A light-weight real-time fault detection system for edge computing using LSTM recurrent neural networks," *Sensors*, vol. 18, no. 7, p. 2110, Jun. 2018.

[23] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network," *Measurement*, vol. 173, Mar. 2021, Art. no. 108518.

[24] Z. Miao, F. Zhou, X. Yuan, Y. Xia, and K. Chen, "Multi-heterogeneous sensor data fusion method via convolutional neural network for fault diagnosis of wheeled mobile robot," *Appl. Soft Comput.*, vol. 129, Nov. 2022, Art. no. 109554.

[25] Ö. Gültekin, E. Cinar, K. Özkan, and A. Yazıcı, "Multisensory data fusion-based deep learning approach for fault diagnosis of an industrial autonomous transfer vehicle," *Expert Syst. Appl.*, vol. 200, Aug. 2022, Art. no. 117055.

[26] Y. Liu, W. Li, S. Lin, X. Zhou, and Y. Ge, "Hydraulic system fault diagnosis of the chain jacks based on multi-source data fusion," *Measurement*, vol. 217, Aug. 2023, Art. no. 113116.

[27] J. Cui, P. Xie, X. Wang, J. Wang, Q. He, and G. Jiang, "M2FN: An end-to-end multi-task and multi-sensor fusion network for intelligent fault diagnosis," *Measurement*, vol. 204, Nov. 2022, Art. no. 112085.

[28] W. Du, N. Rao, C. Dong, Y. Wang, D. Hu, L. Zhu, B. Zeng, and T. Gan, "Automatic classification of esophageal disease in gastroscopic images using an efficient channel attention deep dense convolutional neural network," *Biomed. Opt. Exp.*, vol. 12, no. 6, p. 3066, May 2021.

[29] Y.-J. Huang, A.-H. Liao, D.-Y. Hu, W. Shi, and S.-B. Zheng, "Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis," *Measurement*, vol. 203, Nov. 2022, Art. no. 111935.

[30] Z. Di, D. Yuan, D. Li, D. Liang, X. Zhou, M. Xin, F. Cao, and T. Lei, "Tool fault diagnosis method based on multiscale-channel attention network," in *Proc. IEEE Int. Conf. Sens., Diag., Prognostics, Control (SDPC)*, Aug. 2021, pp. 129–136.

[31] T. Huang, S. Fu, H. Feng, and J. Kuang, "Bearing fault diagnosis based on shallow multi-scale convolutional neural network with attention," *Energies*, vol. 12, no. 20, p. 3937, Oct. 2019.

[32] J. Tong, C. Liu, J. Zheng, and H. Pan, "Multi-sensor information fusion and coordinate attention-based fault diagnosis method and its interpretability research," *Eng. Appl. Artif. Intell.*, vol. 124, Sep. 2023, Art. no. 106614.

[33] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, Apr. 2021, Art. no. 107398.

[34] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021.

[35] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[36] J. Li, S. Lü, and Z. Li, "Unsupervised domain adaptation via softmax-based prototype construction and adaptation," *Inf. Sci.*, vol. 609, pp. 257–275, Sep. 2022.

[37] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[38] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

ZIHAO ZANG was born in Tai'an, Shandong, China, in 1997. He received the B.S. degree in mechanical design, manufacturing, and automation from Inner Mongolia University for Nationalities, in 2020. He is currently pursuing the M.S. degree with the School of Mechanical and Electrical Engineering, Changchun University of Science and Technology. His research interests include industrial equipment fault diagnosis, predictive maintenance, and remaining useful life prediction.

XINMING ZHANG was born in Yuncheng, Shandong, China, in 1967. He received the M.S. degree from the Mechanical Engineering Department, Changchun Institute of Optics and Mechanics, in 1990, and the Ph.D. degree in mechanical manufacturing and automation from Changchun University of Science and Technology, in 2009.

From 1990 to 1993, he was a Mechanic with the Equipment Department, Second Engine Plant, FAW Group. From 1993 to 1995, he taught at the Science and Technology Development Center, Changchun Institute of Optics and Precision Mechanics. From 1999 to 2020, he taught successively at the Science and Technology Development Center and the School of Mechanical and Electrical Engineering, Changchun University of Science and Technology. In 2021, he was the Vice Chairperson of the Academic Committee of the School of Mechatronic Engineering and Automation, Foshan University. He is currently a Professor and a Ph.D. Advisor with the School of Mechanical and Electrical Engineering, where he has been the Dean. He is the author of one book and more than 60 articles. He holds 26 patents. His research interests include precision and ultra-precision machining technology and precision detection technology based on digital twins.

Dr. Zhang has been honored with the First Prize of Jilin Science and Technology Award and the First Prize of Jilin Science and Technology Progress Award many times.

YIWEN ZHANG was born in Changchun, Jilin, China, in 1983. He received the B.S. degree in mechanical design, manufacturing, and automation and the M.S. degree in mechanical engineering from Changchun University of Science and Technology, in 2007 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the School of Mechanical and Electrical Engineering.

From 2009 to 2018, he was the Deputy Director of the Engineering and Technical Development Department, FAW Mold Manufacturing Company Ltd. Since 2018, he has been a Lecturer and the Deputy Director of the Mechanical Engineering Experimental Teaching Center, Changchun University of Science and Technology (a National-Level Experimental Teaching Demonstration Center). He is the author of two books and more than 11 articles. He holds 11 patents. His research interests include equipment fault diagnosis, predictive maintenance based on digital twins, and the study of human–machine effectiveness in digital twin systems.

Mr. Zhang has been awarded the Youth Innovation and Entrepreneurship Leading Talent Award in Changchun. He has received titles, such as the Innovative Figure of China Jilin Province Automotive Electronics Association. He has also been honored with the First Prize and Third Prize for Technological Innovation Progress by China FAW Group Company Ltd.

LINSEN SONG was born in Heilongjiang, China, in 1975. He received the Ph.D. degree in mechanical manufacturing and automation from Changchun University of Science and Technology, in 2010.

In 2014, he was a Professor and a Ph.D. Advisor with Changchun University of Science and Technology. In 2021, he was the Dean of the School of Mechanical and Electrical Engineering, Changchun University of Science and Technology. He is the author of more than 30 high-level articles, holds more than ten patents, and two software copyrights. His research interests include high-end equipment and intelligent manufacturing technology, complex system performance testing and simulation technology, and testing equipment in the national defense and military industry.

Dr. Song is as a Senior Member of the Chinese Mechanical Engineering Society, a Senior Member of the Chinese Optical Society, the Deputy Director of the Mechanical Engineering Committee of the Jilin Ordnance Industry Society, and the Director of China Mechanical Industry Outstanding Engineer Education Alliance. He has won one second-class National Teaching Achievement Award, one first-class Jilin Provincial Teaching Achievement Award, and one first-class and three second-class Jilin Provincial Scientific and Technological Progress Award.
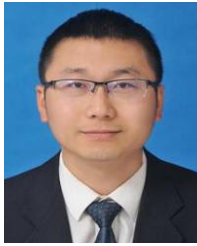
ZHENGLEI YU was born in Changchun, Jilin, in 1984. He received the B.S. degree in optical information science and technology from Changchun University of Science and Technology, in 2007, and the M.S. and Ph.D. degrees in solid mechanics from Jilin University, in 2014.

Since 2014, he has been a Professor and a Ph.D. Supervisor with the Key Laboratory of Engineering Bionics of the Ministry of Education, Jilin University. He has published 52 high-level articles, including 36 SCI papers, and has authorized ten invention patents. His research interests include bionic structure design and additive manufacturing technology research, research on the impact cooling characteristics of bionic systems, and auto parts lightweight design and crashworthiness research.

Dr. Yu is also a member of the International Society for Bionic Engineering (ISBE) and the Director and the Deputy Secretary-General of Jilin Additive Manufacturing Society. He has been honored with the Jilin Provincial Science and Technology Award (Second Prize) and the Jilin Provincial Natural Science Academic Achievement Award (Second Prize).

YITIAN WANG was born in Changchun, Jilin, China, in 1993. He received the bachelor's degree in mechanical design and manufacturing automation from Changchun University of Science and Technology, in 2016, and the Ph.D. degree in metrology and instrumentation from Jilin University, in 2023.

Since 2023, he has been a Lecturer with the School of Mechanical and Electrical Engineering, Changchun University of Science and Technology. He has published five papers in influential international journals (JCR indexed). His research interests include motion state parameter estimation for vehicles and intelligent control algorithms for autonomous vehicles.

YAN GAO was born in Baicheng, Jilin, China, in 2000. He received the B.S. degree in mechanical design, manufacturing, and automation from Changchun University of Science and Technology, in 2022, where he is currently pursuing the M.S. degree with the School of Mechanical and Electrical Engineering. His research interests include industrial equipment condition monitoring, fault diagnosis, and predictive maintenance.

LEI WANG was born in Weifang, Shandong, China, in 1998. He received the B.S. degree from the School of Mechanical Engineering, Jinan University, in 2016, and the M.S. degree in mechanical engineering from the School of Mechanical and Electrical Engineering, Changchun University of Science and Technology, in 2023. He is currently pursuing the Ph.D. degree with Changchun University of Science and Technology. He is the author of two articles and holds five patents. His research interests include ergonomics, digital twins, and predictive maintenance technologies.

. . .