

A convolutional neural network based on MSCAM for intelligent diagnosis of ball screws

Qin WU^{1,2*}, Jianxiong LI¹ , and Xinglian WANG³

¹ School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou, 730050, China

² CEPE, Centre for Mechanical Efficiency and Performance Engineering, University of Huddersfield, HD1 3DH, Huddersfield, UK

³ Mechanical and Electrical Operation and Maintenance Center, Lanzhou Petrochemical Company, Lanzhou 730060, China

Abstract. In response to the challenge of identifying fault types in ball screws of CNC machine tools, particularly under complex operating conditions with often low classification accuracy, we propose a convolutional neural network fault diagnosis model that incorporates multi-scale convolution and an attention mechanism (MSCAM). First, we collect fault data corresponding to various fault types of the ball screw and establish a comprehensive fault dataset. Next, we apply the S transform to the original data to generate time-frequency diagrams, which serve as input for the two-dimensional neural network. In this paper, we present a multi-scale convolutional layer integrated with an attention mechanism, designed to highlight key features in fault information and extract more comprehensive characteristics. Ultimately, the superior recognition and classification capabilities of the model are validated through experimental datasets, and its robustness is thoroughly analyzed.

Keywords: complex working conditions; ball screw; S transform; convolutional neural network; attention mechanism.

1. INTRODUCTION

With the advent of the Industry 4.0 era, the growth of sectors such as aviation, automotive, and shipbuilding has led to an increasing demand for high-speed and high-precision CNC machine tools [1]. As a critical transmission component of CNC machine tools, the ball screw pair plays a vital role in maintaining accuracy and ensuring performance reliability [2]. To enhance accuracy retention, it is essential to implement intelligent detection and diagnosis of ball screw faults [3]. Currently, the rapid advancement of machine learning algorithms has significantly accelerated progress across various fields. Today, nearly all modern fault diagnosis technologies depend on machine learning algorithms [4]. In the realm of fault diagnosis, the application of machine learning models alleviates the maintenance burden on researchers and enhances equipment reliability [5, 6].

The working environment of CNC machine tool ball screws is complex, and the vibration signals they produce are often contaminated by significant amounts of noise. These signals exhibit nonlinear and nonstationary characteristics [7], complicating the extraction of feature information and increasing the difficulty of fault diagnosis. Therefore, to extract fault features from the vibration signals more comprehensively, it is essential to analyze the time-frequency domain of the fault signals. Time domain analysis is suitable for intuitive feature extraction. Frequency domain analysis captures frequency information more comprehensively, but time-frequency analysis can

distinguish dynamic characteristics and is more suitable for processing complex signals. At present, the commonly used time-frequency analysis methods include short-time Fourier transform (STFT) [8], wavelet transform [9] and S transform [10]. Because the short-time Fourier transform [11] uses a fixed short-time window function, it is essentially a single-resolution signal analysis method, and it is difficult to maintain good resolution in the time domain and frequency domain of nonstationary signals. Wavelet transform [12] has certain difficulty in selecting wavelet bases, the data redundancy is serious, and the analysis results of different wavelet bases are also different. The S transform [13] is a new time-frequency analysis tool with adjustable time-frequency resolution, which can meet the time-frequency analysis requirements of different frequency signals. Because of its superior anti-noise ability, the S transform is particularly suitable for the analysis and processing of vibration signals.

Guo *et al.* [14] employed the orthogonal matching pursuit (OMP) algorithm to eliminate harmonic signals while preserving impulse signals and noise. Wavelet analysis was utilized to perform a time-frequency transformation on the signal, and a deformable convolutional neural network was implemented for feature extraction and classification. The experimental results indicate that the accuracy of this method can reach 99.9% across various fault modes, enabling precise identification of rolling bearing faults. To address the challenge of end-to-end fault diagnosis, Wu *et al.* [15] developed a convolutional neural network that learns features directly from the original vibration signal before conducting fault diagnosis. The effectiveness of this proposed method was validated using the PHM (Prediction and Health Management) 2009 gearbox challenge data and a

*e-mail: llgqz@lut.edu.cn

Manuscript submitted 2025-03-04, revised 2025-05-11, initially accepted for publication 2025-06-01, published in September 2025.

planetary gearbox test bench. Liu *et al.* [16] identified the bearing fault frequency band based on the physical parameters of the bearing, constructed a sparse wavelet decomposition structure, and integrated it with a one-dimensional convolutional neural network for fault diagnosis. In [17], an adaptive convolutional neural network fault diagnosis model based on end-to-end recognition was employed to diagnose faults in cylindrical roller bearing cages, addressing the issues of instability and the lack of impact characteristics in rolling bearing cage fault signals. However, due to the limited receptive field of the 1D-CNN, insufficient network depth, and the propensity for overfitting during the training process, the diagnostic accuracy remains low. Consequently, the performance of the 1D-CNN in processing time-frequency signals is inferior to that of the 2D-CNN.

Wang *et al.* [18] proposed a fault classification method based on multi-sensor information fusion. In this approach, time-domain vibration signals from multiple sensors positioned at different locations are organized into a two-dimensional rectangular matrix. An improved two-dimensional convolutional neural network (CNN) is then employed to perform signal classification. Zhang *et al.* [19] transformed the original one-dimensional signal into a two-dimensional image, thereby eliminating the influence of expert experience on the feature extraction process. This method facilitates automatic feature extraction and fault diagnosis through a two-dimensional CNN. Wang *et al.* [20] introduced a general bearing fault diagnosis model that converts the original acceleration signal into a time-frequency image of the same dimensions. Furthermore, the standardized images generated by eight different time-frequency analysis methods are utilized to validate the effectiveness of the proposed method in two distinct cases. Additionally, Xie *et al.* [21] combined continuous wavelet transforms with a two-dimensional CNN for the fault diagnosis of ball screws. The diagnostic results for various types of faults indicate that this method can reduce the uncertainty associated with manual feature extraction. Although the two-dimensional CNN possesses robust image processing capabilities and offers significant advantages in extracting fault features and enhancing diagnostic accuracy, it may not adequately prioritize certain critical features.

Although the aforementioned research demonstrates strong performance in fault identification under a single working condition, challenges remain in accurately identifying fault types and achieving high classification accuracy for ball screws operating under complex conditions. To address these issues, this paper introduces a two-dimensional convolutional neural network (CNN) model integrated with an attention mechanism. The model transforms the collected one-dimensional vibration signals into two-dimensional images using the S transform as input. This time-frequency representation conveys more comprehensive fault information. Additionally, the traditional convolutional layer has been enhanced, and a multi-scale convolutional layer has been designed to extract more subtle and significant features in the horizontal direction. The integration of the attention mechanism with the two-dimensional CNN facilitates improved recognition of ball screw faults in CNC machine tools. The attention mechanism enhances focus on critical features, thereby increasing the accuracy of fault diagnosis.

2. RELATED THEORIES

2.1. S transform

The S transform is a reversible time-frequency analysis technique that integrates the features of both the short-time Fourier transform and the wavelet transform. It overcomes the limitation of the short-time Fourier transform, which cannot modify the frequency of the analysis window, and incorporates the multi-resolution analysis of the wavelet transform while preserving a direct relationship with the Fourier spectrum [22]. The S transform is defined as follows

$$S(\tau f) = \int_{-\infty}^{\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-i2\pi f t} dt, \quad (1)$$

where τ is time, control the position of the window function on the time axis, $h(t)$ is the analysis signal, f is the frequency, and $S(\tau, f)$ is the time-frequency spectrum matrix obtained by transformation.

2.2. Convolutional neural network

Convolutional neural networks (CNNs) are specialized neural networks designed for processing image data. They possess the capability to learn features and utilize a hierarchical structure to classify input information in a translation-invariant manner. The structure of a CNN is illustrated in Fig. 1.

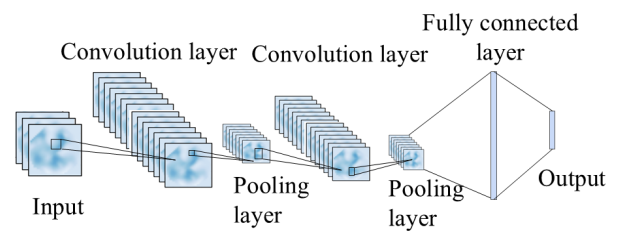


Fig. 1. The structure of CNN

2.2.1. Convolution layer

The convolutional neural network (CNN) extracts local features from the input data through convolution operations. Each convolutional kernel in the convolutional layer is designed to extract a specific feature, and multiple kernels can operate in parallel to capture various types of features. The mathematical model can be expressed as follows

$$X_j^l = f \left(\sum_{i \in M_j} X_i^{l-1} \cdot \omega_{ij}^l + b_j^l \right), \quad (2)$$

where X is the input of the convolution layer, M_j is the set of output feature maps of layer $l-1$, ω is the weight matrix of the corresponding convolution kernel, b is the bias term, l is the number of convolution layers, i and j are two connected neurons, and f is the activation function, which can improve the nonlinear expression ability of the network. The commonly used activation function of CNN is ReLU, which can be expressed as

$$f(x) = \max \{0, \log [1 + \exp(x)]\}. \quad (3)$$

2.2.2. Pooling layer

The pooling layer, also referred to as the downsampling layer, is primarily utilized to downsample the output of the convolutional layer, facilitating dimension reduction. Its main purpose is to decrease the number of parameters and enhance computational efficiency. The most common pooling methods are max pooling and average pooling, which are defined as follows

$$y = p \left[X_{\text{down}}(x_i^{l-1}) + b_j^l \right], \quad (4)$$

where y is the output of the pooling layer, X_{down} is the down-sampling function, x is the input, and b_j^l is the bias term.

2.2.3. Full connected layer

Before the output layer of the network is the fully connected layer. This layer operates as a fully connected neural network, integrating the features extracted from the preceding layer for classification or regression tasks. Each neuron in the fully connected layer is connected to all neurons in the previous layer. Additionally, each neuron in the network is interconnected with other neurons at various levels, thereby maximizing the number of parameters throughout the entire network. Its mathematical model can be expressed as follows

$$y = f(\omega \cdot x + b), \quad (5)$$

where x is the input matrix, y is the output matrix, f is the activation function, and b is the bias of the full connection layer.

2.2.4. Multi-scale convolutional layer

In this paper, we propose an enhanced version of the convolutional neural network (CNN), referred to as the multi-scale convolutional neural network, for the fault diagnosis of ball screws in CNC machine tools. This method employs convolutional kernels of three different sizes to extract features from images processed using the short-time Fourier transform, as shown in Fig. 2. Unlike the traditional vertical deepening approach (which involves convolution, pooling, and re-convolution), our method comprehensively extracts subtle and significant features in the horizontal direction.

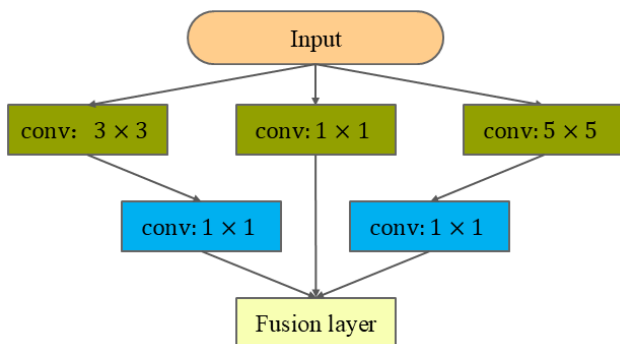


Fig. 2. Multi-scale convolution layer diagram

Through multi-layer convolution, the network progressively learns and extracts more abstract and semantically rich features

from the original image, thereby achieving more accurate and effective feature extraction and processing of the input data.

2.3. Attention mechanism

The ball screw of CNC machine tools typically operates in environments subject to significant external loads, resulting in time-varying and nonlinear characteristics of its vibration signals. Consequently, under identical conditions, the signal characteristics obtained at different times may vary. Some features effectively convey fault information, while others may introduce interference, thereby impacting the model generalization ability. The attention mechanism adaptively assigns weights to the features of different signal segments, filters information, emphasizes critical fault features, and suppresses irrelevant features. Its structure is illustrated in Fig. 3. When provided with an intermediate feature map, the CBAM module infers the attention map along two independent dimensions: the channel attention mechanism and the spatial attention mechanism. It then multiplies the attention map with the input feature map for adaptive feature optimization.

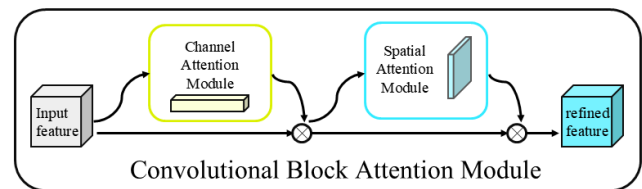


Fig. 3. CBAM diagram

The attention mechanism [23] was originally developed for machine translation, typically employing self-encoding to facilitate sequence conversion. This mechanism is inspired by the study of human vision and has been extensively applied in natural language processing and various other fields. The convolutional block attention module (CBAM) comprises two sub-modules: the channel attention module and the spatial attention module, which focus on channel and spatial information, respectively. The input image F (with high channel dimensions) is processed through both a maximum pooling layer and a global average pooling layer to generate a feature map with a height and width of 1. Subsequently, these feature maps are fed into two shared perceptron networks, which output results by summing them sequentially. The channel attention features are then produced after applying the activation function

$$F' = M_C(F) \otimes F, \quad (6)$$

$$F'' = M_S(F) \otimes F', \quad (7)$$

where $F \in R^{C \times W \times H}$, $M_C \in R^{C \times 1 \times 1}$, $M_S \in R^{1 \times H \times W}$, F is the output image after fusion, $\otimes$ is the multiplication of corresponding elements, C , W , H represents the number of channels, width and height, respectively; F' is the channel 'attention' module picture, F'' is the spatial 'attention' module picture, M_S is a one-dimensional channel attention map, and M_C is a two-dimensional spatial attention map.

2.4. Fault extraction based on MSCAM-CNN

The single-scale convolution kernel often faces challenges in fully extracting fault features. To address this issue, we propose the MSCAM-CNN model, which facilitates the extraction and classification of multi-scale feature information (Fig. 4). The feature extraction block consists of a convolutional layer, an activation layer, a batch normalization layer, and a pooling layer. The classification block includes a fully connected layer followed by a Softmax layer. Unlike single-scale convolution kernels, convolution kernels of varying scales can capture frequency features at different resolutions. Consequently, utilizing multi-scale convolution kernels for fault feature extraction provides richer feature information, thereby enhancing the accuracy and robustness of fault recognition.

In the process of collecting the original fault signal, noise interference is inevitable. Compared to smaller convolution kernels, larger convolution kernels are more effective at suppressing high-frequency noise. Therefore, larger convolution kernels,

specifically (3×3) and (5×5) are employed in the feature extraction block to mitigate high-frequency noise across multi-scale information. By utilizing multiple parallel convolution kernels of varying sizes within the feature extraction block, fault information features of different scales are extracted and further abstracted. Ultimately, the fault features are identified and classified by classification blocks, where the fully connected layer consists of 200 neurons, and the softmax layer contains 10 neurons.

The details of the MSCAM-CNN model presented in this paper are shown in Table 1. After extracting the features, it is essential to apply an activation function to enhance the model's nonlinear representation. The rectified linear unit (ReLU) function was chosen as the activation function. The expression is as follows

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (8)$$

where x is the output value of the convolution operation.

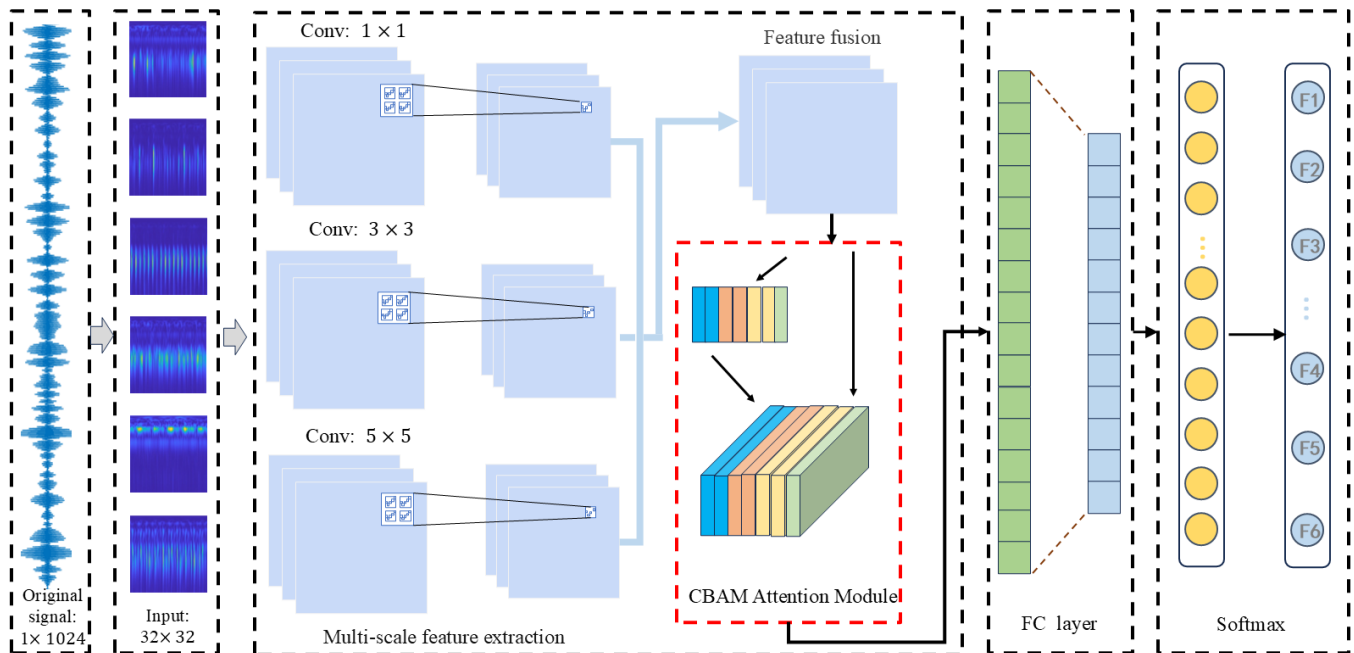


Fig. 4. The structure of MSCAM-CNN

Table 1
The structure of MSCAM-CNN

Type of layer	Parameters			
Input	[batch, -1, 32×32, 1]			[batch, 1@32*32]
Convolution layer	filter = [3×3, 1, 8]	strides = [1, 1, 1, 1]	padding = "SAME"	[batch, 8@32*32]
Pooling layer	ksize = [1, 2×2, 1]	strides = [1, 2, 2, 1]	padding = "SAME"	[batch, 8@16*16]
Convolution layer	filter = [3×3, 8, 16]	strides = [1, 1, 1, 1]	padding = "SAME"	[batch, 16@16*16]
Pooling layer	ksize = [1, 2×2, 1]	strides = [1, 2, 2, 1]	padding = "SAME"	[batch, 16@8*8]
Fully connected layer				[batch, 1*6]
Classification layer				[batch, 1*6]

3. BALL SCREW FAULT DIAGNOSIS PROCESS BASED ON MULTI-SCALE CONVOLUTION AND ATTENTION MECHANISM

The structure of the MSCAM-CNN fault diagnosis model proposed in this paper is illustrated in Fig. 5. As shown in Fig. 5, the method presented here consists of three main stages: dataset construction, feature extraction, and fault identification and classification. Vibration signals from six different fault types are collected. Following sliding sampling, the time-frequency diagrams are generated, and the dataset is divided accordingly. In the feature extraction phase, fault information is obtained through a multi-scale convolution layer and an attention mechanism, ultimately leading to the identification and classification of fault types. A multi-scale convolution kernel is utilized to integrate features from the data across various time scales, facilitating the extraction of deep features. The attention mechanism effectively highlights significant information within the

horizontal features, thereby amplifying the influential factors associated with this subset of features. This approach enhances the model accuracy while reducing the risk of overfitting. The detailed steps are outlined below.

1. Data preprocessing and division of the fault dataset. The S transform is employed to convert one-dimensional vibration signals into two-dimensional images. Concurrently, classification labels are assigned, and the data is divided into a training set and a test set.
2. Fault feature extraction. Convolutional kernels of various sizes (1×1 , 3×3 , and 5×5) are employed to design a multi-scale convolutional layer, enabling the extraction of features from images processed using the S transform.
3. Initialize the network architecture and configure the hyper-parameters.
4. The training dataset for the model is input into the proposed MSCAM-CNN, after which the model is trained.

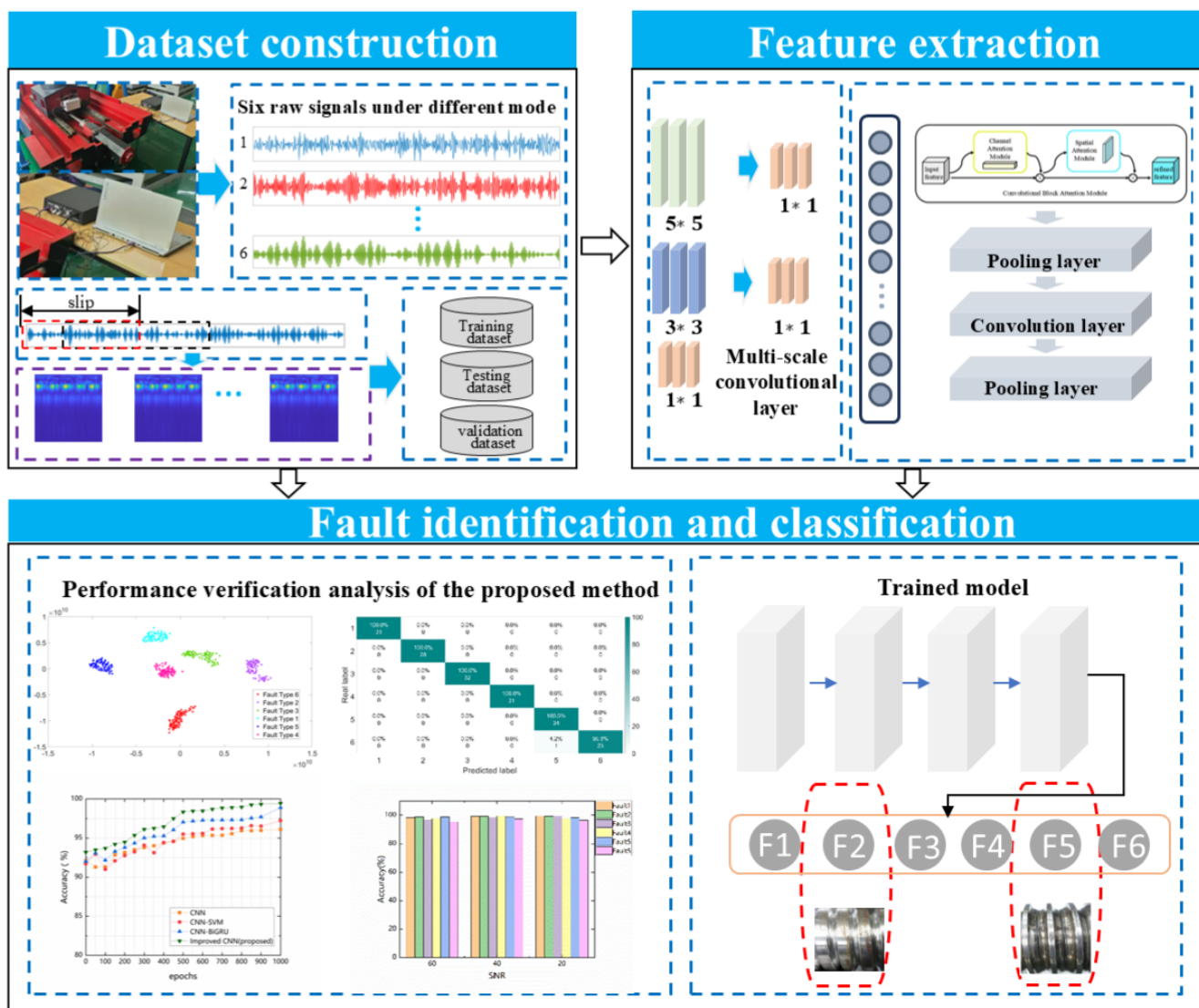


Fig. 5. Fault diagnosis model based on multi-scale convolution and attention mechanism

5. The trained MSCAM-CNN model is employed for the fault diagnosis of mechanical equipment to evaluate its effectiveness and robustness in identifying faults.

4. EXPERIMENTAL VERIFICATION ANALYSIS

4.1. Experimental data acquisition

The ball screw of the CNC machine tool is used to collect fault data. The GD4010 screw is employed in the experiment, and its specific parameters are presented in Table 2.

The experimental device primarily consists of a YMC121A100 unidirectional IEPE acceleration sensor, a YMC9216 signal collector, and YMC9800 signal analysis software, along with a measured screw, motor, coupling, and power amplifier (Fig. 6).

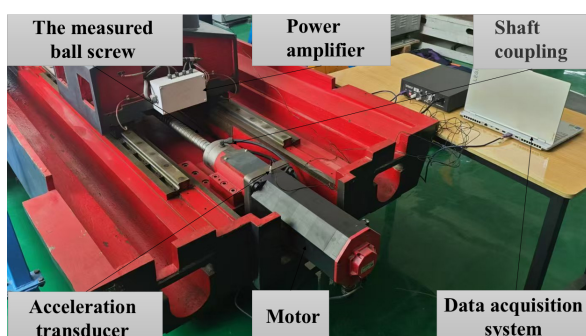


Fig. 6. Fault information acquisition test bench

The speed is 1772 revolutions per minute (r/min), and vibration signals from various fault types are collected. Additionally, a 16-channel data recorder is employed to gather these vibration signals. The sample length is 1024, and the sampling interval is 128. The specific method for dividing the training set and test set is presented in Table 3.

Table 2

Process parameters of GD4010 ball screw

Name	Screw diameter	Ball diameter	Contact angle	Screw lead angle
Unit	d_0 [mm]	d_b [mm]	α [°]	λ [°]
Value	40	5.953	45	4.55

Table 3

Fault types and labels

Label	Ball screw state	Number of trainings	Number of tests
1	Normal	700	300
2	Screw raceway wear fault	700	300
3	Rolling element wear fault	700	300
4	Misalignment fault of screw	700	300
5	Screw bending fault	700	300
6	Screw pitting fault	700	300

A sliding window technique was employed to perform non-overlapping slicing operations on the original vibration signal, resulting in a time series sample for every 1200 sample points. Each fault category contains 1000 samples, leading to a total dataset of 6000 samples. The training set and test set were divided in a ratio of 7:3.

One-dimensional vibration signals were collected, and the time-domain waveforms of each fault signal are presented in Fig. 7. As illustrated in Fig. 7, the time-domain waveform can only capture the fault characteristics within the time-domain, and the extracted features do not fully represent the fault characteristics of the screw. Consequently, we employ the S transform to process the time-domain signal, resulting in the time-frequency diagram shown in Fig. 8. The horizontal axis rep-

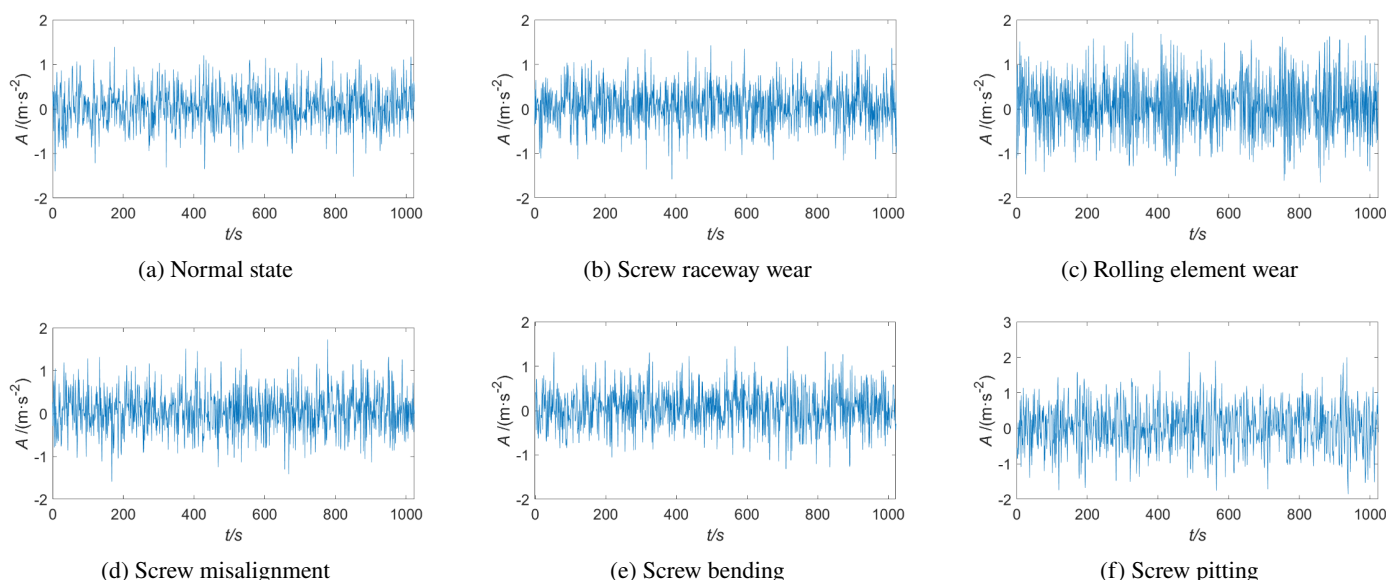


Fig. 7. Time-domain waveform of fault signal

A convolutional neural network based on MSCAM for intelligent diagnosis of ball screws

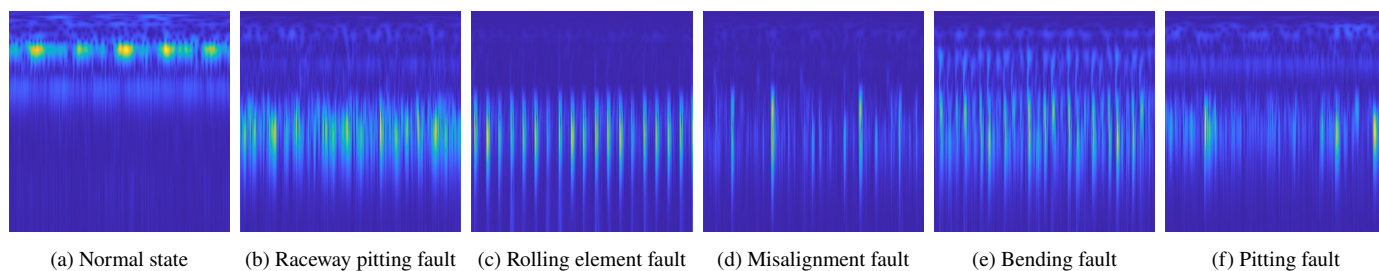


Fig. 8. S transform diagram of fault signal

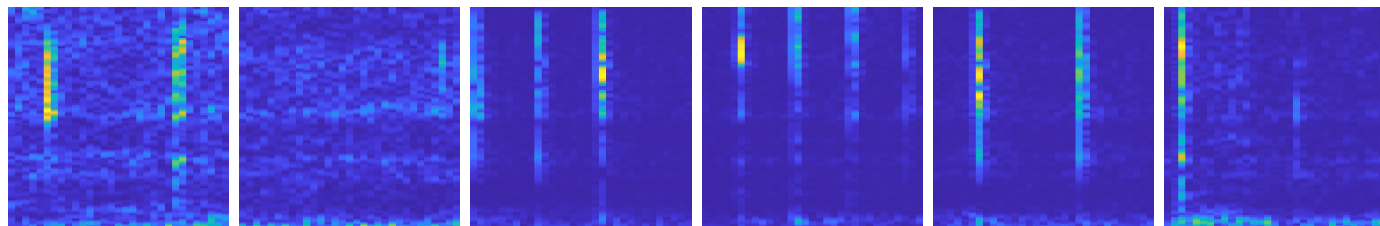


Fig. 9. STFT diagram of fault signal

resents time, while the vertical axis denotes frequency. It is evident that the two-dimensional feature matrix generated by the vibration signal processed through the S transform contains a richer array of fault information, providing a solid foundation for subsequent fault classification.

In order to compare the time-frequency conversion effects of the S transform, the short-time Fourier transform (STFT) is employed to perform time-frequency conversion on one-dimensional time series signals. The results are illustrated in Fig. 9.

4.2. Model performance verification

The hyperparameters of the model cannot be adjusted during training. They are typically established prior to the training process. The optimization and adjustment of hyperparameters play a crucial role in fault diagnosis research. The batch size for input samples is set to 32, and the learning rate is 0.01. The widely used Adam optimizer was selected, and both training and validation samples were fed into the model for parameter initialization and training. As training progresses, the model performance gradually improves. The accuracy of the model during training is illustrated in Fig. 10.

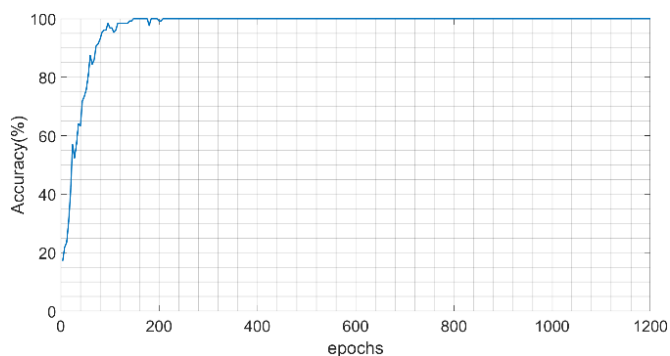


Fig. 10. Change curve of model training accuracy

It can be observed from Fig. 9 that the network achieves an accuracy of 90% on the training set after approximately 75 iterations, which improves to 99% after 125 iterations. Concurrently, as the number of iterations increases, the network loss value continues to decrease, as illustrated in Fig. 11, indicating that the network was not overfitted.

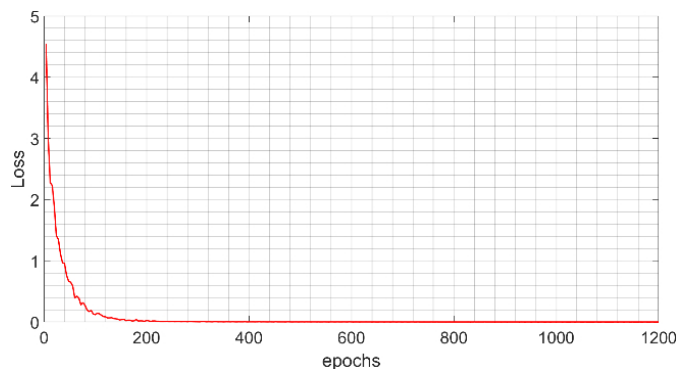


Fig. 11. Loss variation curve

To demonstrate the effectiveness of fault diagnosis, a confusion matrix is employed to visualize the model results. The recognition outcomes for each fault sample are presented in the form of a confusion matrix, as shown in Fig. 12. The horizontal axis of the confusion matrix represents the predicted labels of the screws, while the vertical axis represents the actual labels of the screws.

From Fig. 12, it is evident that the recognition of screw pitting is frequently misclassified as other conditions, while the recognition rates for the other five states are exceptionally high. The fault recognition rate reaches 100%, and the overall accuracy rate is 99.44%. This demonstrates that utilizing multi-feature extraction in conjunction with a fusion convolutional neural network is highly effective for fault recognition.

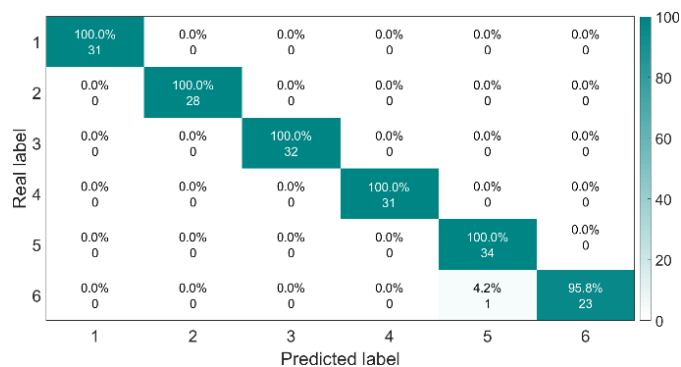
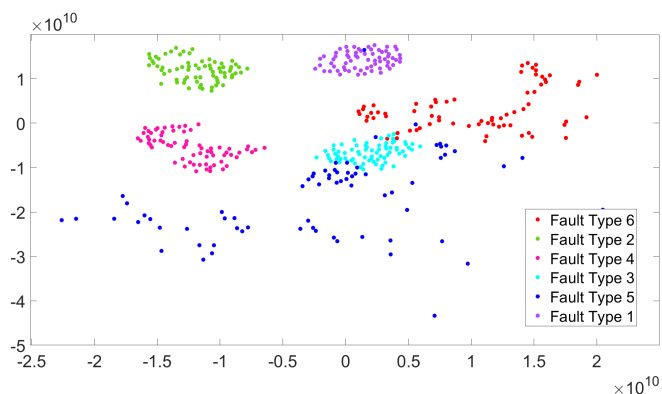
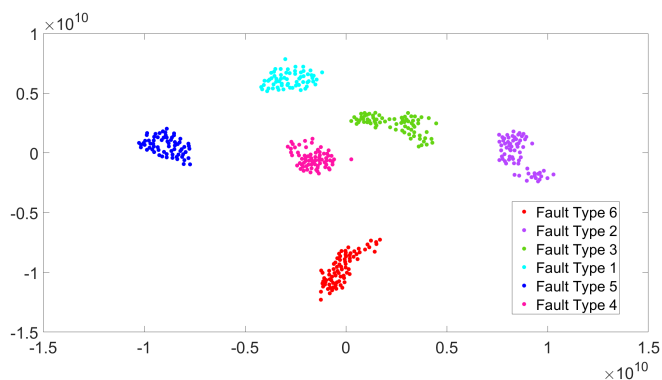


Fig. 12. Confusion matrix

t-SNE (t-distributed stochastic neighbor embedding) [24] is a nonlinear manifold learning algorithm that represents the similarity between high-dimensional spatial data points in the form of probabilities. This technique utilizes t-SNE diagrams to evaluate the fault identification and classification capabilities of the model. As illustrated in Fig. 13, panel 13a depicts the degree of aggregation of the input layer feature vectors, while panel 13b illustrates the degree of aggregation of the classification layer feature vectors. Prior to identification, the fault features are dispersed, exhibiting a high degree of confusion, which complicates the effective differentiation between various faults. However, following identification, the aggregation of fault features is



(a) Input layer feature vector t-SNE diagram



(b) Classification layer feature vector t-SNE diagram

Fig. 13. t-SNE diagram before and after recognition

enhanced, and the degree of confusion is significantly reduced. This improvement indicates that the identification and classification performance of the model presented in this paper was enhanced.

In order to verify the effectiveness of the proposed model, the S transform-CNN model is compared with the S transform-CNN-SVM model and the S transform-CNN-BiGRU model. The original one-dimensional data is utilized, and the S transform is applied to generate a time-frequency diagram. The total number of data iterations is set to 1000.

It can be observed from Fig. 14 that the accuracy of network fault recognition using multi-scale feature extraction and a spatial attention mechanism is 3.33% higher than that of traditional convolutional neural networks. Compared to the CNN-SVM and CNN-BiGRU models, the recognition accuracy of the model presented in this paper is significantly greater. This indicates that the proposed model effectively extracts fault features and demonstrates superior fault diagnosis capabilities.

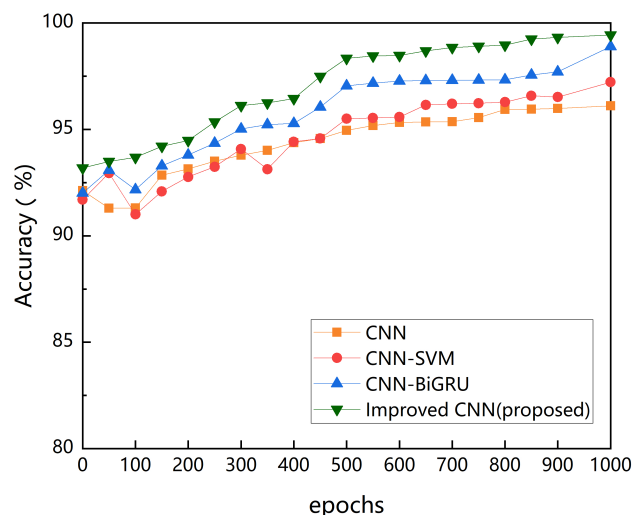


Fig. 14. Comparison of accuracy of different models

To compare the diagnostic accuracy of various time-frequency images using the model proposed in this paper, we employed the short-time Fourier transform (STFT) and the S transform for time-frequency image conversion. The results of the diagnostic accuracy are presented in Table 4. It is evident that the average diagnostic accuracy of the method proposed in this paper reaches 98.48%, while the average diagnostic accuracy of the STFT method is 96.79%. This demonstrates that the method outlined in this paper exhibits superior performance in terms of model accuracy.

Table 4

Diagnostic accuracy results

Type of time-frequency diagram	Average diagnostic accuracy
S transform	98.48%
STFT	96.79%

4.3. Robustness analysis

Due to the potential disturbances caused by noise during the operation of machine tools under real working conditions, it is essential to consider the impact of noise on fault diagnosis, particularly concerning the ball screw of CNC machine tools. Various noises with differing signal-to-noise ratios were added to the original signal, and the diagnostic results are illustrated in Fig. 15. The data indicates that when the signal-to-noise ratio ranges from 40 to 60 dB, the addition of noise to the original signal results in a slight decrease in accuracy. However, the accuracy consistently remains above 95%. The model presented in this paper demonstrates high diagnostic accuracy, indicating that its structure possesses significant robustness.

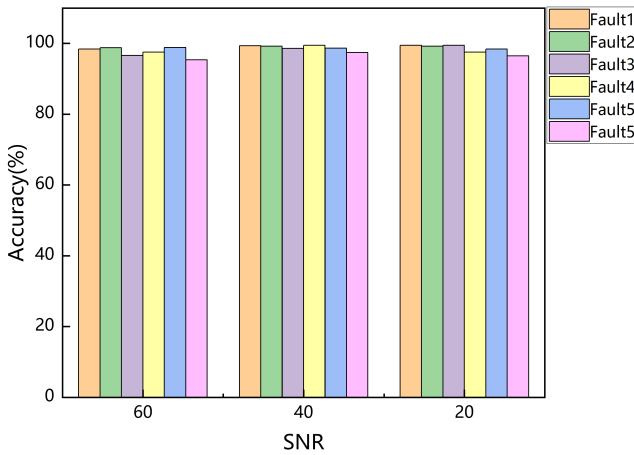


Fig. 15. Robustness comparison

This robustness is attributed to the S transform, which converts the vibration signal from the time domain to the time-frequency domain. This transformation allows for flexible selection of the noise suppression region across various frequency ranges, effectively mitigating the influence of noise while preserving the essential information of the signal. Additionally, the convolutional and pooling layers enhance the filtering effects.

4.4. Analysis of generalization ability

Precision, Recall, and $F1_score$ are introduced as the primary evaluation metrics for assessing the generalization ability of the model. These values characterize the adaptability of diagnostic algorithms to various fault modes and evaluate the overall performance of the model. The specific definitions are as follows

$$\begin{cases} \text{Precision} = \frac{TP}{TP + FP}, \\ \text{Recall} = \frac{TP}{TP + FN}, \\ F1_score = \frac{2 \times P \times R}{P + R}. \end{cases} \quad (9)$$

Among these terms, TP stands for true positive (normal samples correctly identified as normal), FN denotes false negative (nor-

mal samples incorrectly identified as faulty), and FP signifies false positive (faulty samples incorrectly identified as normal).

By comparing the evaluation metrics of CNN-BiGRU, CNN-SVM, CNN, and the model presented in this paper, the results are illustrated in Fig. 16. It is evident that the accuracy, recall rate, and $F1_score$ of the model proposed in this paper surpass those of the other models by approximately 2% or more. This indicates that the model also demonstrates strong performance in terms of generalization ability.

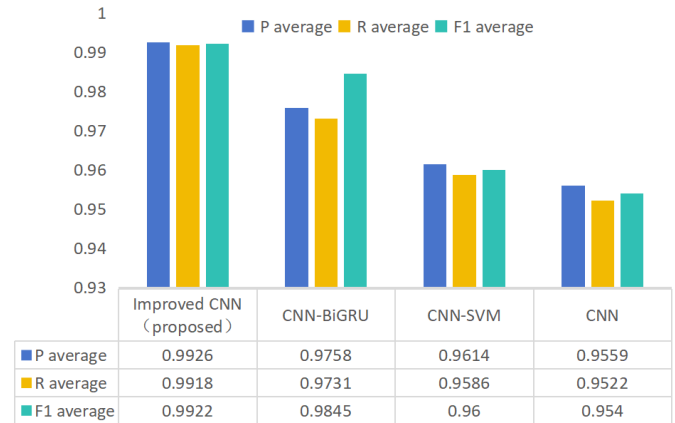


Fig. 16. Precision, recall, and $F1_score$ of different models

5. CONCLUSIONS

In this paper, the S transform is employed to convert a one-dimensional vibration signal into a two-dimensional time-frequency image. Multi-scale feature extraction and an attention mechanism are utilized to extract fault information. Ultimately, fault recognition and classification are achieved using a two-dimensional convolutional neural network. The main conclusions are as follows:

1. The time-frequency analysis of nonlinear and unstable vibration signals is performed using the S transform. This method capitalizes on the advantages of the S transform in both the time and frequency domains, thereby offering more comprehensive information for input into two-dimensional neural networks.
2. The designed multi-scale feature extraction module effectively captures fault information across an extended time scale while achieving a larger receptive field, thereby enhancing the model feature extraction capabilities. The attention mechanism prioritizes the critical features present in the fault information, which improves the accuracy of fault diagnosis.
3. The experimental results demonstrate that the proposed method outperforms traditional machine learning fault diagnosis techniques in terms of accuracy and robustness in fault identification. While the model exhibits exceptional diagnostic capabilities in the presence of significant noise, it does not sufficiently evaluate the effectiveness of fault identification under varying operational conditions. Future research will focus on fault diagnosis across diverse working environments.

REFERENCES

- [1] J. Chen *et al.*, "Toward intelligent machine tool," *Engineering*, vol. 5, no. 4, pp. 679–690, 2019, doi: [10.1016/j.eng.2019.07.018](https://doi.org/10.1016/j.eng.2019.07.018).
- [2] X. Wei and J. Mao, "Research advance on geometric error recognition algorithm for CNC machine tools," *2020 3rd World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM)*, 2020, pp. 170–174, doi: [10.1109/WCMEIM52463.2020.00042](https://doi.org/10.1109/WCMEIM52463.2020.00042).
- [3] Z. Wang *et al.*, "A high-accuracy intelligent fault diagnosis method for aero-engine bearings with limited samples," *Comput. Ind.*, vol. 159, p. 104099, 2024, doi: [10.1016/j.compind.2024.104099](https://doi.org/10.1016/j.compind.2024.104099).
- [4] C. Li *et al.*, "A review of static and dynamic analysis of ball screw feed drives, recirculating linear guideway, and ball screw," *Int. J. Mach. Tools Manuf.*, vol. 188, no. 1, p. 104021, 2023, doi: [10.1016/j.ijmachtools.2023.104021](https://doi.org/10.1016/j.ijmachtools.2023.104021).
- [5] J. Li and L. Hu, "Review of machine learning for predictive maintenance," *Comput. Eng. Appl.*, vol. 56, no. 21, pp. 11–19, 2020.
- [6] P. Kumar and A. S. Hati, "Review on machine learning algorithm based fault detection in induction motors," *Arch. Comput. Meth. Eng.*, vol. 28, no. 3, pp. 1929–1940, 2021, doi: [10.1007/s11831-020-09446-w](https://doi.org/10.1007/s11831-020-09446-w).
- [7] M. Saman Azari, F. Flammini, S. Santini, and M. Caporuscio, "A systematic literature review on transfer learning for predictive maintenance in industry 4.0," *IEEE Access*, vol. 11, pp. 12887–12910, 2023, doi: [10.1109/ACCESS.2023.3239784](https://doi.org/10.1109/ACCESS.2023.3239784).
- [8] L. Wang, Y. Xie, and Z. Zhou, "Asynchronous motor fault diagnosis based on convolutional neural network," *J. Vib. Meas. Diag.*, vol. 37, no. 6, pp. 1208–1215, 2017, doi: [10.16450/j.cnki.issn.1004-6801.2017.06.021](https://doi.org/10.16450/j.cnki.issn.1004-6801.2017.06.021).
- [9] K. Hadad, M. Pourahmadi, and H. Majidi-Maraghi, "Fault diagnosis and classification based on wavelet transform and neural network," *Prog. Nucl. Energy*, vol. 53, no. 1, pp. 41–47, 2011, doi: [10.1016/j.pnucene.2010.09.006](https://doi.org/10.1016/j.pnucene.2010.09.006).
- [10] G. Li *et al.*, "Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform," *Sensors*, vol. 19, no. 12, p. 2750, 2019, doi: [10.3390/s19122750](https://doi.org/10.3390/s19122750).
- [11] H. Geraei, E. A. V. Rodriguez, E. Majma, S. Habibi and D. Al-Ani, "A noise invariant method for bearing fault detection and diagnosis using adapted local binary pattern (ALBP) and short-time Fourier transform (STFT)," *IEEE Access*, vol. 12, pp. 107247–107260, 2024, doi: [10.1109/ACCESS.2024.3438106](https://doi.org/10.1109/ACCESS.2024.3438106).
- [12] B. Li *et al.*, "Self-iterated extracting wavelet transform and its application to fault diagnosis of rotating machinery," *IEEE Trans. Instr. Meas.*, vol. 73, pp. 1–17, 2024, doi: [10.1109/TIM.2024.3370784](https://doi.org/10.1109/TIM.2024.3370784).
- [13] W. Liu *et al.*, "Time-Reassigned Multisynchrosqueezing S-Transform for Bearing Fault Diagnosis," *IEEE Sens. J.*, vol. 23, no. 19, pp. 22813–22822, 2023, doi: [10.1109/JSEN.2023.3303879](https://doi.org/10.1109/JSEN.2023.3303879).
- [14] J. Guo *et al.*, "Bearing intelligent fault diagnosis based on wavelet transform and convolutional neural network," *Shock Vib.*, vol. 2020, no. 1, pp. 1–14, 2020, doi: [10.1155/2020/6380486](https://doi.org/10.1155/2020/6380486).
- [15] C. Wu *et al.*, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network," *Comput. Ind.*, vol. 108, pp. 53–61, 2019, doi: [10.1016/j.compind.2018.12.001](https://doi.org/10.1016/j.compind.2018.12.001).
- [16] X. Liu *et al.*, "One dimensional convolutional neural networks using sparse wavelet decomposition for bearing fault diagnosis," *IEEE Access*, vol. 10, pp. 86998–87007, 2022, doi: [10.1109/ACCESS.2022.3199381](https://doi.org/10.1109/ACCESS.2022.3199381).
- [17] Y. Zhen *et al.*, "Fault diagnosis of cylindrical roller bearing cage based on 1D convolution neural network," *J. Vib. Shock*, vol. 40, no. 19, pp. 230–238, 2021, doi: [10.13465/j.cnki.jvs.2021.19.029](https://doi.org/10.13465/j.cnki.jvs.2021.19.029).
- [18] J. Wang *et al.*, "Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network," *IEEE Access*, vol. 9, pp. 23717–23725, 2021, doi: [10.1109/ACCESS.2021.3056767](https://doi.org/10.1109/ACCESS.2021.3056767).
- [19] J. Zhang *et al.*, "A new bearing fault diagnosis method based on modified convolutional neural networks," *Chin. J. Aeronaut.*, vol. 33, no. 2, pp. 439–447, 2020, doi: [10.1016/j.cja.2019.07.011](https://doi.org/10.1016/j.cja.2019.07.011).
- [20] J. Wang *et al.*, "A deep learning method for bearing fault diagnosis based on time-frequency image," *IEEE Access*, vol. 7, pp. 42373–42383, 2019, doi: [10.1109/ACCESS.2019.2907131](https://doi.org/10.1109/ACCESS.2019.2907131).
- [21] Z. Xie *et al.*, "Ball screw fault diagnosis based on continuous wavelet transform and two-dimensional convolution neural network," *Meas. Control*, vol. 56, no. 3–4, pp. 518–528, 2023, doi: [10.1177/00202940221107620](https://doi.org/10.1177/00202940221107620).
- [22] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: the S transform," *IEEE Trans. Signal Process.*, vol. 44, pp. 998–1001, 1996, doi: [10.1109/78.492555](https://doi.org/10.1109/78.492555).
- [23] H. Wang *et al.*, "Data-Augmentation Based CBAM-ResNet-GCN Method for Unbalance Fault Diagnosis of Rotating Machinery," *IEEE Access*, vol. 12, pp. 34785–34799, 2024, doi: [10.1109/ACCESS.2024.3368755](https://doi.org/10.1109/ACCESS.2024.3368755).
- [24] J. Soni, N. Prabakar, and H. Upadhyay, "Visualizing high-dimensional data using t-distributed stochastic neighbor embedding algorithm," *Princ. Data Sci.*, pp. 189–206, 2020, doi: [10.1007/978-3-030-43981-1_9](https://doi.org/10.1007/978-3-030-43981-1_9).