

# Plunger Pump Fault Diagnosis Method Based on Wavelet Convolution and Multi-Head Self-Attention Mechanism for Multi-Channel Feature Fusion

Pu Zhou, Zhichun Qian, Dongliang Fu, Yi Zhang

Shanghai Marine Equipment Research Institute, Shanghai, China

Email: 3194801390@qq.com

**How to cite this paper:** Zhou, P., Qian, Z.C., Fu, D.L. and Zhang, Y. (2025) Plunger Pump Fault Diagnosis Method Based on Wavelet Convolution and Multi-Head Self-Attention Mechanism for Multi-Channel Feature Fusion. *Journal of Computer and Communications*, 13, 345-356.  
<https://doi.org/10.4236/jcc.2025.137018>

**Received:** June 18, 2025

**Accepted:** July 27, 2025

**Published:** July 30, 2025

---

## Abstract

To address the issues of low accuracy, high dependence on prior knowledge, and poor adaptability in fusing multi-channel features in existing plunger pump fault diagnosis methods, a new method based on single-channel wavelet convolution and multi-head self-attention mechanism is proposed. This method first applies wavelet decomposition and 2D convolution to extract local features of each channel signal individually, and then utilizes the multi-head self-attention mechanism to enhance inter-channel mutual information perception, enabling intelligent diagnosis of plunger pump conditions. Experimental results show that this method can effectively diagnose plunger pump faults with an accuracy of 99.54%, outperforming other deep learning models in terms of training efficiency and diagnostic accuracy.

## Keywords

Plunger Pump Fault Diagnosis, Wavelet Decomposition, Multi-Head Self-Attention Mechanism, Multi-Channel Fusion

---

## 1. Introduction

As the “power heart” of hydraulic transmission systems, plunger pumps play a crucial role in marine propulsion systems [1]. However, due to the harsh operating environments, internal components of the plunger pump are prone to damage [2]. Coupled with the complexity of the equipment structure, maintenance personnel often struggle to accurately and promptly locate fault positions and analyze their causes. Therefore, during operation, plunger pumps are at risk of efficiency

degradation, shutdown, or even safety accidents due to faults, which seriously threatens the system's stability and personnel safety [3]. To enhance operational reliability and performance, and to reduce economic losses caused by excessive maintenance or faults, research on fault diagnosis methods for plunger pumps has significant practical and application value.

Traditional fault diagnosis methods for plunger pumps are mainly based on signal processing techniques, which help extract features from vibration, acoustic, and other signals while filtering irrelevant noise [4]. Common methods include wavelet-based techniques such as Empirical Wavelet Transform (EWT) [5] and Wavelet Packet Decomposition (WPD) [6], mode decomposition techniques such as Empirical Mode Decomposition (EMD) [7] and Ensemble Empirical Mode Decomposition (EEMD) [8], as well as frequency-domain analysis methods like Short-Time Fourier Transform (STFT) [9] and Hilbert Transform (DCT) [10]. These methods have shown strong capabilities in feature extraction and classification.

With the rise of deep learning, which offers good adaptability and generalization, researchers have introduced convolutional neural networks (CNNs) such as LeNet-5 [11] and AlexNet [12], as well as recurrent neural networks (RNNs) like LSTM [13], into feature extraction tasks. Furthermore, since fault data of plunger pumps are often derived from multiple sensors, multi-source fusion methods like Bayesian networks [14] and Dempster-Shafer (D-S) theory [15] have been applied to integrate features from different signals, thereby improving fault characterization and diagnosis accuracy.

However, existing studies still face three main limitations:

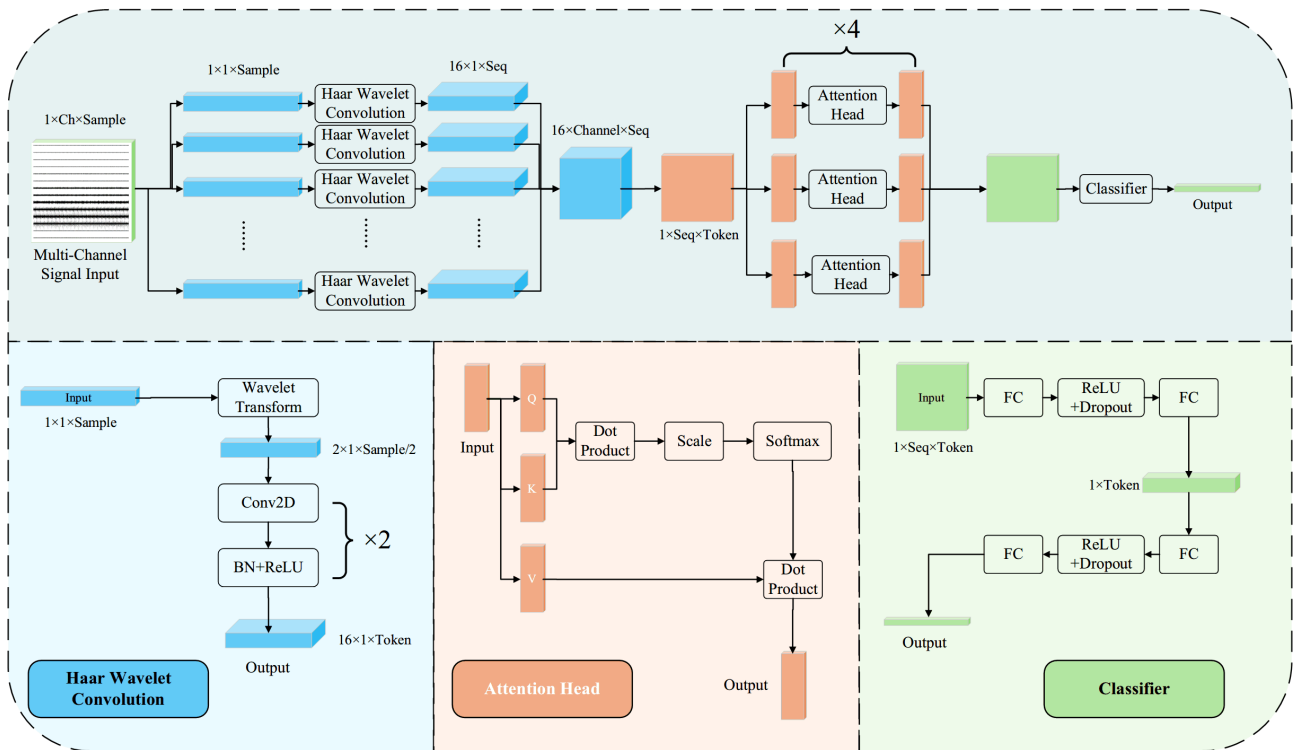
- Traditional signal processing methods struggle to extract deep features and rely heavily on expert knowledge, which limits the generalization ability of diagnosis models;
- Classical neural network architectures have inherent shortcomings: CNNs are less effective in capturing global patterns, while RNNs tend to lose temporal details and local information during iteration;
- Existing multi-channel fusion methods often rely on manually set rules or prior knowledge, lacking the ability to adaptively adjust fusion strategies based on data, thus failing to achieve true adaptive information fusion.

To address the above issues, this paper proposes a fault diagnosis method based on wavelet convolution and a multi-head self-attention mechanism for multi-channel feature fusion. First, wavelet convolution operators are employed to extract local features from each signal channel across different frequency domains, adapting to the non-stationary nature of fault signals. Then, multi-head self-attention is used to adaptively fuse features across channels and enhance the expression of key fault-related information. Experimental results demonstrate that the proposed method can capture critical features within and across channels, improving feature representation and fault classification accuracy.

## 2. Method

### 2.1. Overview

As illustrated in **Figure 1**, the proposed model mainly consists of three parts:



**Figure 1.** Structure of the proposed model.

First, the multi-sensor input signals are fed into a multi-channel Haar wavelet and 2D convolution module for initial feature extraction from the vibration signals.

Second, the extracted features are passed into a multi-head self-attention module to establish global dependencies in the time dimension.

Finally, a fully connected layer serves as the classifier to output the plunger pump fault diagnosis results.

### 2.2. Network Structure

#### 2.2.1. Haar Wavelet Convolution Module

Inspired by the Haar wavelet downsampling method proposed by Xu G. *et al.* [16], this module introduces the Haar wavelet basis for each channel, performing a first-order Haar wavelet transform to downsample the signal while preserving local features of the original signal. The result of the Haar wavelet transform can be expressed as follows:

$$a_j[k] = \frac{1}{\sqrt{2}}(a_{j-1}[2k] + a_{j-1}[2k+1]) \quad (1)$$

$$d_j[k] = \frac{1}{\sqrt{2}}(a_{j-1}[2k] - a_{j-1}[2k+1]) \quad (2)$$

In Equations (1) and (2),  $a_j[k]$  represents the low-frequency wavelet coefficients, while  $d_j[k]$  represents the high-frequency wavelet coefficients.

The obtained wavelet coefficients are concatenated along the convolution channel dimension and then sequentially passed through two 2D convolution modules. The first convolution module uses 16 kernels along the time dimension to extract features from both low- and high-frequency wavelet coefficients, expanding the signal into 16 channels. Batch normalization and ReLU activation are then applied to ensure training stability and enhance the model's nonlinear representation capacity. The second convolution module is similar to the first, except that the number of convolution channels remains unchanged. The parameters of the two convolutional modules are summarized in **Table 1**.

After passing through these two modules, the wavelet coefficients are transformed into the final output of the single-channel wavelet convolution module.

**Table 1.** Configuration of the convolution modules.

Layer	Layer parameters			
	Input Channels	Output Channels	Subhead	Stride
Conv2D	2	16	(1,16)	8
BatchNorm2D	16	16	—	—
ReLU	—	—	—	—
Conv2D	16	16	(1,16)	8
BatchNorm2D	16	16	—	—
ReLU	—	—	—	—

Finally, the wavelet convolution outputs from all channels are concatenated along the signal channel dimension and then merged with the convolutional channels.

The merged dimensions are subsequently transposed with the time dimension, resulting in a matrix of shape  $Seq \times Token$ , where  $Seq$  represents the length of the time sequence and  $Token$  represents the feature vector at each time step.

This feature representation is then fed into the self-attention module to enable more expressive and high-level representation learning via the subsequent self-attention structure.

### 2.2.2. Multi-Head Self-Attention Module

Due to the limited receptive field of convolution kernels, the aforementioned modules can only extract features within local regions of the signal. To capture the global features of plunger pump fault signals, a multi-head self-attention mechanism is introduced in this study.

First, the  $Token$  dimension in the  $Seq \times Token$  matrix is evenly divided into  $h$  parts, resulting in vectors  $x_i \in \mathbb{R}^{Seq \times Token/h}$ . Then, each  $x_i$  is linearly projected using learned weights  $W_{q_i}, W_{k_i}, W_{v_i}$ , producing the query, key, and value matrices:  $Q_i, K_i, V_i$ , respectively.

For each head, the attention output is computed as:

$$Attention_i = softmax(Q_i K_i^T / \sqrt{Token / h}) V_i \quad (3)$$

Finally, the outputs from all attention heads are concatenated to obtain the final output of the multi-head self-attention module:

$$MHA = [Attention_1, Attention_2, \dots, Attention_h] \quad (4)$$

### 2.2.3. Classifier

The output of the multi-head self-attention module retains the shape  $Seq \times Token$ . To further compress the information along the time dimension and extract more discriminative features, a classifier primarily based on fully connected layers is designed.

This classifier first reduces the temporal dimension and then compresses the feature dimension to match the number of target classes  $num\_classes$ . The specific architecture is shown in **Table 2** below:

**Table 2.** Structure of the classifier.

Layer	Input Size	Output Size
Fully Connected	$Seq * Token$	$16 * Token$
ReLU + Dropout	$16 * Token$	$16 * Token$
Fully Connected	$16 * Token$	$1 * Token$
Squeeze	$1 * Token$	$Token$
Fully Connected	$Token$	128
ReLU + Dropout	128	128
Fully Connected	128	$num\_classes$

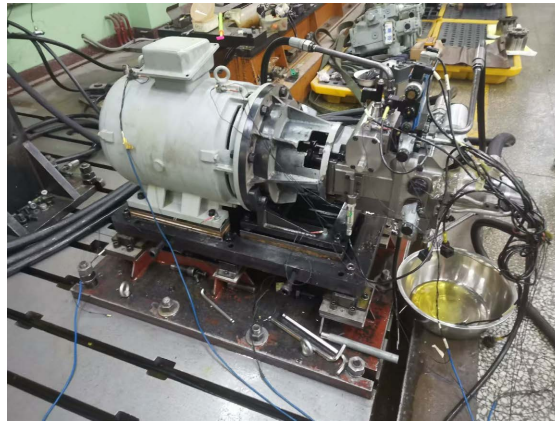
## 3. Experiments and Analysis

### 3.1. Data Acquisition and Preprocessing

**Figure 2** shows the plunger pump fault simulation test bench used in this study. Fourteen vibration measurement channels were arranged as listed in **Table 3**, using Brüel & Kjær 4514-B accelerometers, with a sampling frequency of 25,600 Hz for all channels.

To obtain fault data of the plunger pump, 12 typical fault types were injected into the test bench as listed in **Table 4**. For each fault type, data were collected before and after fault injection to obtain both normal and fault-state samples.

To increase the dataset size and improve classification accuracy, each continuous data segment was first split into training and testing sets at a 7:3 ratio over time, and then sliced into 1-second segments with a 0.5-second overlap. Each data slice has a shape of 14 channels  $\times$  25,600 samples. The number of samples for each category is summarized in **Table 5**.



**Figure 2.** Plunger pump fault simulation test bench.

**Table 3.** Sensor placement information.

Channel No.	Measurement Location	Type of Measurement	Sensor Axis Direction
1	Plunger pump	Triaxial acceleration	X-axis
2			Y-axis
3			Z-axis
4	Gear pump	Triaxial acceleration	X-axis
5			Y-axis
6			Z-axis
7	Motor front end (near coupling)	Triaxial acceleration	X-axis
8			Y-axis
9			Z-axis
10	Motor front end (far coupling)	Triaxial acceleration	X-axis
11			Y-axis
12			Z-axis
13	Motor base	Uniaxial acceleration	Z-axis
14	Pump base	Uniaxial acceleration	Z-axis

**Table 4.** Injected fault types and labels for plunger pump.

Label	Fault Type	Label	Fault Type
1	Loose fasteners	7	Poor suction in gear pump
2	Deformed fasteners	8	Gear pump wear
3	Damaged elastomer	9	Plunger pair wear
4	Overflow regulation failure	10	Inner ring bearing fault
5	Loose slipper	11	Rolling element bearing fault
6	Valve plate wear	12	Outer ring bearing fault

**Table 5.** Number of samples.

Type	Training Set	Test Set	Total
Normal	84	36	120
Faulty (12 types)	$12 \times 35$	$12 \times 15$	$12 \times 50 = 600$
<b>Total</b>	504	216	720

### 3.2. Model Training

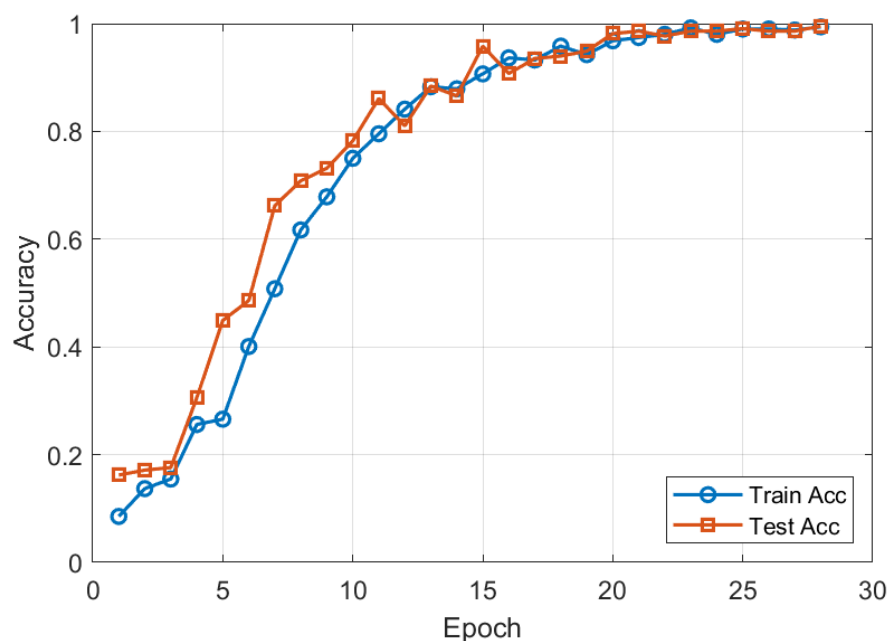
All neural network models in this section were implemented using Python 3.10.11 and the deep learning framework PyTorch 2.12, and executed on a GeForce RTX 4070 GPU using the CUDA 12.1.66 parallel computing architecture.

After experimental tuning, the training configuration was finalized as follows:

- Learning rate:  $1e-4$
- Batch size: 64
- Loss function: Cross-entropy loss
- Optimizer: Adam optimizer (with  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ )
- Dropout rate: 0.1

During training, early stopping was introduced to prevent overfitting. Specifically, training was terminated early if the validation loss did not decrease within 30 consecutive epochs, and the model was restored to the state with the best validation performance.

As shown in **Figure 3** and **Figure 4**, the model achieved a 99.54% accuracy on the test set at the 28<sup>th</sup> epoch, with a total training time of 24.07 seconds. At this point, the confusion matrix of the test set is shown in **Figure 5**, where only one normal sample was misclassified as an overflow regulation failure.

**Figure 3.** Accuracy curve of the proposed model.

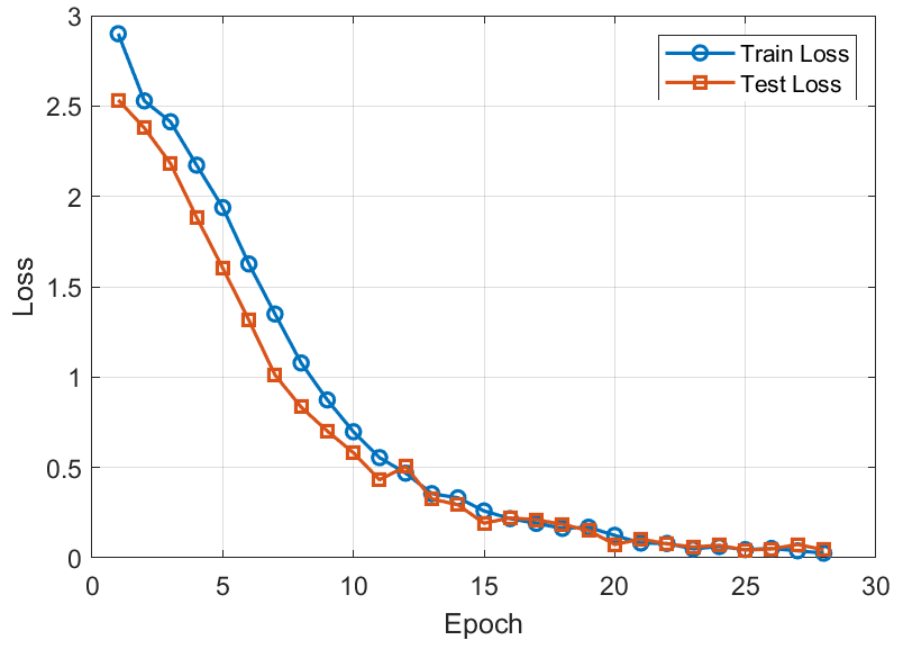


Figure 4. Loss curve of the proposed model.

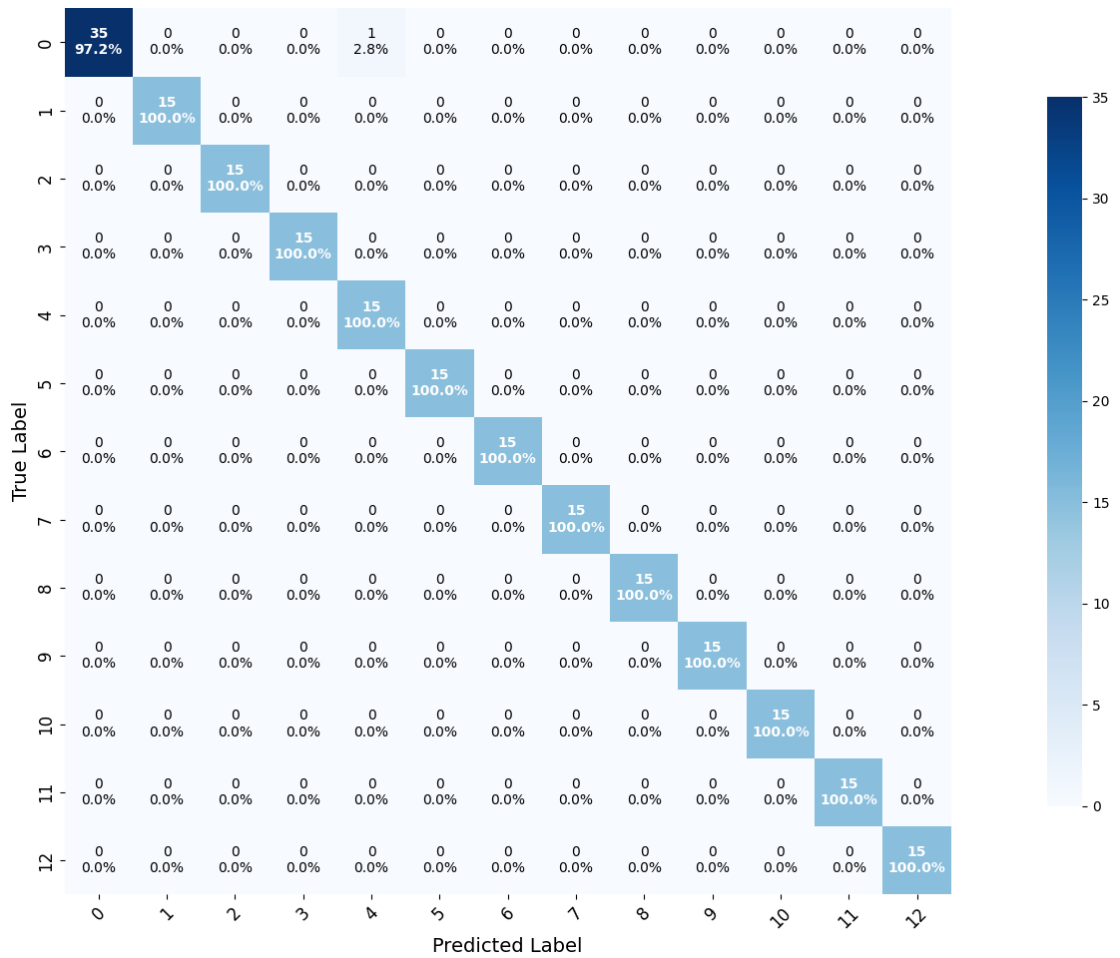


Figure 5. Confusion matrix of the proposed model on the test set.

### 3.3. Ablation Study

To verify the effectiveness of the key structural components in the proposed model, two comparative models were designed for ablation analysis against the original architecture:

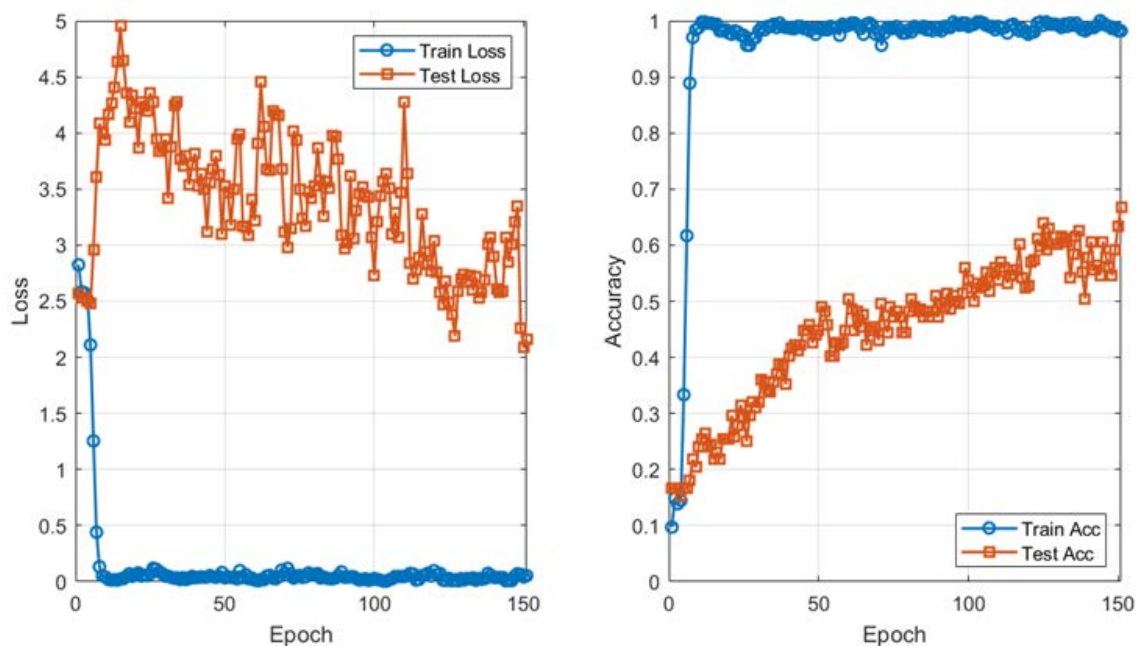
- Model A: The Haar wavelet downsampling method is replaced with a simple interpolation method, which only performs shape transformation on the raw input signal.
- Model B: The multi-head self-attention module in the proposed model is removed. Instead, features extracted by wavelet convolution from each channel are directly pooled and classified, ignoring inter-channel correlations.

All other parameters remain unchanged. The training results are shown in **Table 6**.

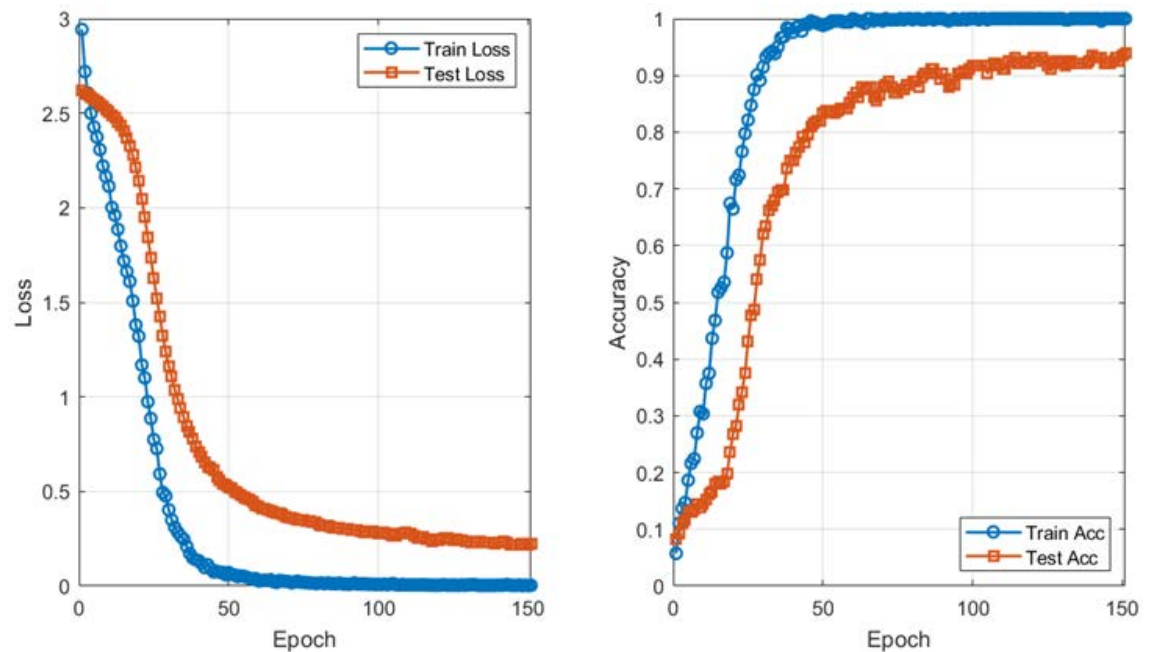
As shown in **Figure 6** and **Figure 7**, Model A fails to converge effectively, while Model B demonstrates a certain level of classification ability but still falls significantly short of the complete model. These results indicate that simple waveform reshaping through interpolation is insufficient for capturing the temporal and spatial features of the signal. Moreover, removing the attention mechanism and ignoring the relationships among channels leads to performance degradation.

**Table 6.** Ablation study results.

Model	Training Steps	Training Time (s)	Accuracy
Model A	151	121.64	66.67%
Model B	151	74.42	93.98%
Proposed Model	28	24.07	99.54%



**Figure 6.** Loss and accuracy curves of Model A.



**Figure 7.** Loss and accuracy curves of Model B.

Therefore, the proposed combination of wavelet convolution and multi-head self-attention not only achieves superior accuracy but also improves training efficiency, demonstrating the effectiveness and rationality of the model's structural design.

### 3.4. Comparative Experiments

To further validate the effectiveness of the proposed model in processing plunger pump fault signals, this section compares it with several classical convolutional neural network (CNN) models. Since these baseline models require inputs of specific dimensions, the original input signals are first expanded into a three-channel format via 2D convolution. The last two dimensions are then adjusted using reshaping and interpolation techniques to meet the input size requirements.

All models are trained using the same parameter settings and dataset partitions to ensure a fair comparison in terms of classification accuracy and training efficiency. The experimental results are presented in **Table 7**.

**Table 7.** Comparative experiment results.

Model	Training Steps	Training Time (s)	Accuracy
AlexNet	71	53.25	99.54%
GoogLeNet	75	86.54	90.74%
LeNet-5	123	56.64	44.91%
ResNet34	<b>10</b>	<b>15.38</b>	92.13%
Proposed Model	28	24.07	<b>99.54%</b>

The results show that although classical CNN models perform well in image recognition tasks, they face challenges such as low training efficiency and sub-optimal accuracy when applied to the fault diagnosis of plunger pumps. In contrast, the proposed model achieves comparable or even higher classification accuracy while maintaining a compact architecture and superior training efficiency. These findings demonstrate the proposed model's effectiveness and practicality in multi-channel temporal signal classification tasks.

## 4. Conclusions

This paper proposes a fault diagnosis method for plunger pumps based on wavelet convolution and multi-head self-attention mechanisms. The main conclusions are as follows:

- The proposed model effectively extracts instantaneous feature variations across different frequency bands within individual channels through the wavelet convolution module. Meanwhile, the multi-head self-attention mechanism enhances the model's ability to adaptively capture key inter-channel features, which helps suppress redundant information and contributes to improved detection accuracy.
- Experiments conducted on real-world datasets show that the proposed model achieves a classification accuracy of 99.54% on multi-channel vibration signals, with a training time of 24.07 seconds on an RTX 4070 GPU. Compared with classical convolutional models, the proposed approach significantly improves both training efficiency and diagnostic accuracy, demonstrating clear advantages.

This study focuses on multi-channel vibration signals of plunger pumps and achieves effective fault type classification based on the signal characteristics. Future work will explore the integration of real-time signal processing and edge computing to facilitate on-site deployment and real-time application of the proposed method in industrial environments.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Zhu, Y., Li, G., Wang, R., Tang, S., Su, H. and Cao, K. (2023) Intelligent Fault Diagnosis Methods for Hydraulic Piston Pumps: A Review. *Journal of Marine Science and Engineering*, **11**, 1609. <https://doi.org/10.3390/jmse11081609>
- [2] Kumar, S., Bergada, J.M. and Watton, J. (2009) Axial Piston Pump Grooved Slipper Analysis by CFD Simulation of Three-Dimensional NVS Equation in Cylindrical Coordinates. *Computers & Fluids*, **38**, 648-663. <https://doi.org/10.1016/j.compfluid.2008.06.007>
- [3] Khan, K., Sohaib, M., Rashid, A., Ali, S., Akbar, H., Basit, A. and Ahmad, T. (2021) Recent Trends and Challenges in Predictive Maintenance of Aircraft's Engine and Hydraulic System. *Journal of the Brazilian Society of Mechanical Sciences and Engi-*

- neering*, **43**, 403. <https://doi.org/10.1007/s40430-021-03121-2>
- [4] Yang, Y., Ding, L., Xiao, J., Fang, G. and Li, J. (2022) Current Status and Applications for Hydraulic Pump Fault Diagnosis: A Review. *Sensors*, **22**, 9714. <https://doi.org/10.3390/s22249714>
- [5] Yu, H., Li, H., Li, Y. and Li, Y. (2019) A Novel Improved Full Vector Spectrum Algorithm and Its Application in Multi-Sensor Data Fusion for Hydraulic Pumps. *Measurement*, **133**, 145-161. <https://doi.org/10.1016/j.measurement.2018.10.011>
- [6] Hou, W., Lu, C., Liu, H. and Lu, C. (2009) Fault Diagnosis Based on Wavelet Package for Hydraulic Pump: *Proceedings of the 8th International Conference on Reliability, Maintainability and Safety*, Chengdu, China, 20-24 July 2009, 831-835. <https://doi.org/10.1109/ICRMS.2009.5269950>
- [7] Meng, L., Dong, C., Zhou, J., Lu, X., Wang, Y., Deng, M., Huang, C. and Yuan, B. (2021) Typical Fault Simulation and On-Line Monitoring for Aviation Hydraulic Pump. *Machinery Tool and Hydraulics*, **49**, 170-174.
- [8] Jiang, W.L., Zhang, P.Y., Li, M. and Zhang, S.Q. (2021) Axial Piston Pump Fault Diagnosis Method Based on Symmetrical Polar Coordinate Image and Fuzzy C-Means Clustering Algorithm. *Shock and Vibration*, **2021**, 6681751. <https://doi.org/10.1155/2021/6681751>
- [9] Przystupa, K., Ambrozkiwicz, B. and Litak, G. (2020) Diagnostics of Transient States in Hydraulic Pump System with Short Time Fourier Transform. *Advances in Science and Technology Research Journal*, 178-183. <https://doi.org/10.12913/22998624/116971>
- [10] Yu, H., Li, H. and Li, Y. (2020) Vibration Signal Fusion Using Improved Empirical Wavelet Transform and Variance Contribution Rate for Weak Fault Detection of Hydraulic Pumps. *ISA Transactions*, **107**, 385-401. <https://doi.org/10.1016/j.isatra.2020.07.025>
- [11] Zhu, Y., Li, G., Wang, R., Tang, S., Su, H. and Cao, K. (2021) Intelligent Fault Diagnosis of Hydraulic Piston Pump Combining Improved LeNet-5 and PSO Hyperparameter Optimization. *Applied Acoustics*, **183**, 108336. <https://doi.org/10.1016/j.apacoust.2021.108336>
- [12] Zhu, Y., Li, G., Wang, R., Tang, S., Su, H. and Cao, K. (2021) Intelligent Fault Diagnosis of Hydraulic Piston Pump Based on Wavelet Analysis and Improved AlexNet. *Sensors*, **21**, 549. <https://doi.org/10.3390/s21020549>
- [13] Zhu, Y., Su, H., Tang, S., Zhang, S. and Wang, J. (2023) A Novel Fault Diagnosis Method Based on SWT and VGG-LSTM Model for Hydraulic Axial Piston Pump. *Journal of Marine Science and Engineering*, **11**, 594. <https://doi.org/10.3390/jmse11030594>
- [14] Liu, Z., Liu, Y., Shan, H., Cai, B. and Huang, Q. (2015) A Fault Diagnosis Methodology for Gear Pump Based on EEMD and Bayesian Network. *PLoS ONE*, **10**, e0125703. <https://doi.org/10.1371/journal.pone.0125703>
- [15] Lu, C., Wang, S. and Wang, X. (2017) A Multi-Source Information Fusion Fault Diagnosis for Aviation Hydraulic Pump Based on the New Evidence Similarity Distance. *Aerospace Science and Technology*, **71**, 392-401. <https://doi.org/10.1016/j.ast.2017.09.040>
- [16] Xu, G., Liao, W., Zhang, X., Li, C., He, X. and Wu, X. (2023) Haar Wavelet Downsampling: A Simple but Effective Downsampling Module for Semantic Segmentation. *Pattern Recognition*, **143**, 109819. <https://doi.org/10.1016/j.patcog.2023.109819>