



Fault diagnosis of water injection pump via wavelet-enhanced attention guided Inception-LSTM networks

Xiao Wu¹, Zelin Wu¹, Feng Luo¹, Jiawei Wang¹, Tangbin Xia^{1,2,3}, Lifeng Xi^{1,2}

Keywords:

Water injection pump, unbalanced data, convolutional neural network, long short-term memory network, attention mechanism

Citation:

Wu, X.; Wu, Z.; Luo, F.; Wang, J.; Xia, T.; Xi, L. Fault diagnosis of water injection pump via wavelet-enhanced attention guided Inception-LSTM networks. *Complex Eng. Syst.* 2026, 6, 11. <https://dx.doi.org/10.20517/ces.2026.13>

Received: 16 Mar 2026

First Decision: 29 Apr 2026

Revised: 12 May 2026

Accepted: 11 Jun 2026

Published: 22 Jun 2026

Academic Editor:

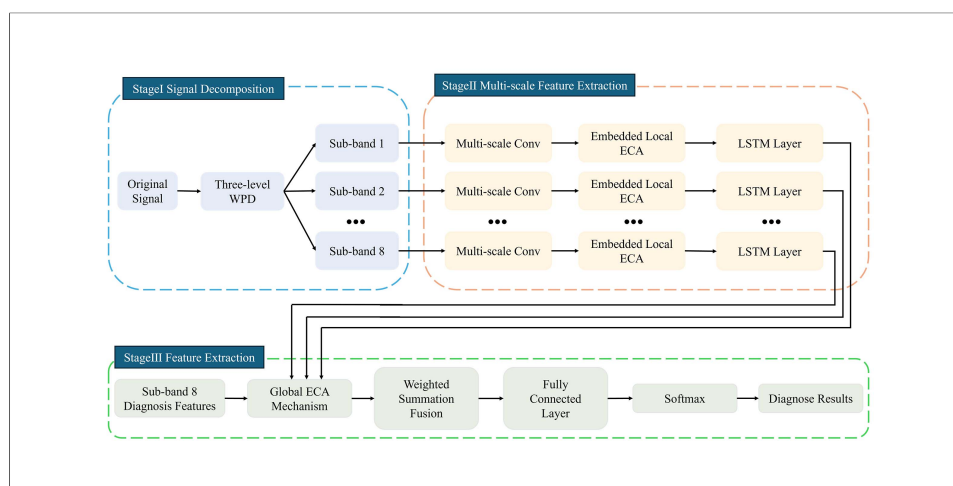
Zhiqiang Ge

Copy Editor:

Fangling Lan

Production Editor:

Fangling Lan



Abstract

Accurate fault diagnosis of water injection pump is essential for ensuring operational safety and efficiency in oil and gas exploitation. However, traditional diagnostic methods often struggle with non-stationary vibration signals and severe category imbalance in complex industrial environments. To address these challenges, this paper proposes a multi-level Inception-long short-term memory (Inception-LSTM) network integrated with wavelet packet decomposition (WPD) and efficient channel attention (ECA), termed the multi-level Inception-LSTM network with WPD and ECA (MILN-WE). The proposed framework first employs WPD to decompose complex vibration signals into fine-grained frequency sub-bands, capturing subtle fault characteristics. Subsequently, a multi-scale Inception module is utilized to extract diverse spatial features, while an LSTM layer captures the long-term temporal dependencies of the signals. The integration of the ECA mechanism further enhances the model's ability to focus on critical diagnostic information. The effectiveness of MILN-WE is validated using a private oilfield water injection pump dataset and a public rotating machinery dataset. Experimental results demonstrate that the proposed model achieves higher diagnostic accuracy and robustness compared to



¹School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

²Special Environment Digital Manufacturing Equipment Technology Innovation Center, Mianyang 621900, Sichuan, China.

³Shanghai Changxing Ocean Laboratory, Shanghai 201913, China.

Correspondence to: Dr. Tangbin Xia, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: xtbxtb@sjtu.edu.cn

state-of-the-art methods, particularly under conditions of strong noise interference and data imbalance. Specifically, on the private oilfield water injection pump dataset, the model achieved an accuracy of 99.38%, improving upon traditional convolutional neural network (CNN) and class-balanced-CNN models by 6.05% and 3.24%, respectively. This study provides a high-precision and robust solution for the intelligent predictive maintenance of critical energy equipment, offering significant theoretical and practical value for industrial health monitoring systems.

1. INTRODUCTION

Hydraulic water injection pumps play a vital role in modern oil and gas exploitation, where they are responsible for injecting high-pressure fracturing fluids into underground formations to enhance hydrocarbon production. The reliability of water injection pump systems directly affects operational safety and production efficiency. Water injection pumps are critical flow-control elements that operate under severe working conditions, including high pressure, intense vibration, and rapidly changing loads. Long-term operation under such harsh environments may lead to valve wear, fatigue damage, and sealing failures, which can eventually cause equipment malfunction or even catastrophic accidents. Therefore, the development of reliable and accurate fault diagnosis techniques for water injection pump has become an important research topic in industrial condition monitoring and predictive maintenance.

Traditional fault diagnosis approaches for rotating machinery mainly rely on signal processing techniques combined with handcrafted feature extraction. Time-frequency analysis methods such as wavelet transform, empirical mode decomposition, and variational mode decomposition have been widely applied to extract representative features from vibration signals. These approaches aim to capture the nonstationary characteristics of mechanical vibration signals and improve fault identification performance. For instance, He *et al.*^[1] proposed a fault diagnosis framework integrating wavelet packet transform and convolutional neural networks (CNNs), where time-frequency representations were used to enhance feature discrimination. Guo *et al.*^[2] further improved the Morlet wavelet transform to enhance time-frequency resolution and combined it with a shallow residual neural network for bearing fault classification. Zhai *et al.*^[3] developed a diagnostic method based on synchro squeezing wavelet transform and a transfer residual convolutional neural network to address feature extraction challenges caused by complex vibration signals.

Although these signal processing methods can effectively reveal certain fault characteristics, their performance strongly depends on expert knowledge and manual feature engineering. In complex industrial environments, vibration signals are often contaminated by strong noise and nonlinear disturbances, which makes it difficult for handcrafted features to maintain stable diagnostic performance. Consequently, traditional diagnostic approaches often suffer from limited adaptability and generalization ability when applied to practical industrial scenarios.

With the rapid development of artificial intelligence technologies, deep learning has emerged as a powerful tool for intelligent fault diagnosis. Compared with traditional methods, deep neural networks can automatically learn hierarchical feature representations directly from raw sensor signals, thereby significantly reducing the reliance on manual feature extraction. CNNs, in particular, have demonstrated strong capability in feature learning and pattern recognition and have been widely adopted in rotating machinery fault diagnosis.

Recent studies have reported significant progress in CNN-based diagnostic frameworks. Tan *et al.*^[4] demonstrated the efficacy of coupling long short-term memory (LSTM) with CNNs for diagnosing mixed-flow pumps under complex cavitation conditions, highlighting the potential of recurrent architectures in capturing fluid-induced vibrations. Gao *et al.*^[5] introduced a diagnostic approach combining continuous

wavelet transform and deep convolutional generative adversarial networks to address the issue of imbalanced datasets in machinery fault classification tasks. Deng *et al.*^[6] further developed an attention-based CNN that enhances feature representation capability by adaptively focusing on important signal components.

In addition to convolutional architectures, more advanced deep learning models have also been explored to improve diagnostic performance. Transformer-based architectures have recently attracted considerable attention due to their ability to capture long-range dependencies and global contextual information in signals. Lai *et al.*^[7] proposed a residual attention vision transformer network for rolling bearing fault diagnosis, which integrates convolutional feature extraction with self-attention mechanisms to capture both local and global signal features. Liu *et al.*^[8] further proposed a Transformer transfer learning framework capable of improving fault diagnosis performance under cross-condition scenarios.

Another important research direction focuses on improving the interpretability and physical consistency of deep learning models. Several studies have attempted to integrate signal processing knowledge into neural network structures. Hassannejad *et al.*^[9] proposed a physics-informed CNN in which wavelet-based feature extraction was embedded into the network architecture to improve both interpretability and diagnostic accuracy. Similarly, Deng *et al.*^[10] proposed a multi-sensor fusion framework for axial piston pumps, while Kim *et al.*^[11] validated the superiority of Transformer-based models in identifying pressure signal anomalies.

Despite the promising results achieved by deep learning methods, most existing studies assume that training and testing data follow the same distribution. However, in real industrial applications, operating conditions such as load, rotational speed, and pressure frequently vary over time. These variations often lead to distribution discrepancies between training and testing data, which can notably degrade the performance of deep learning models. Therefore, improving the generalization capability of intelligent diagnostic models under varying operating conditions has become a critical challenge in machinery fault diagnosis.

To address this issue, transfer learning has been widely investigated in recent years. Transfer learning aims to transfer knowledge learned from a source domain to a target domain with limited labeled data, thereby improving diagnostic performance under varying operating conditions. Zhao *et al.*^[12] proposed a wavelet convolution-based transfer learning framework for cross-machine fault diagnosis, demonstrating improved robustness under different working conditions. Yu *et al.*^[13] introduced a domain adaptation neural network based on maximum mean discrepancy to align feature distributions between source and target domains. Sun *et al.*^[14] further proposed an adversarial domain adaptation framework that employs a domain discriminator to reduce distribution discrepancies across domains.

More recently, contrastive learning and self-supervised learning strategies have also been applied to improve feature representation and transferability. Li *et al.*^[15] proposed a contrastive learning-based diagnostic framework capable of learning more discriminative feature representations for rotating machinery faults. Zhu *et al.*^[16] further developed a supervised contrastive transfer learning method that combines domain adaptation with contrastive loss to enhance feature discrimination across different domains. Moreover, to handle the distribution discrepancies in decentralized data, Zhou *et al.*^[17] proposed a modular federated learning framework using dynamic routing to collaboratively optimize local models under multiple working conditions.

In addition to domain adaptation strategies, several studies have explored other deep learning frameworks for machinery health monitoring and fault diagnosis. Shao *et al.*^[18] proposed a deep autoencoder-based feature learning method for rotating machinery diagnosis. Wang *et al.*^[19] developed a recurrent neural network-based health indicator for equipment health monitoring and remaining useful life prediction.

Zhang *et al.*^[20] optimized CNN architectures to improve diagnostic accuracy under complex working conditions. Zhang *et al.*^[21] and Zhu *et al.*^[22] further demonstrated the effectiveness of deep convolutional networks in machinery fault diagnosis tasks.

Beyond pure fault identification, the ultimate goal of machinery health monitoring is to facilitate systemic and intelligent predictive maintenance. Recently, advanced prognosis and dynamic maintenance strategies have attracted substantial attention in the industrial engineering field. For example, researchers have developed prognosis-centered intelligent maintenance optimization frameworks that systematically account for uncertain failure thresholds^[23], as well as customized multi-agent reinforcement learning approaches for systemic condition-based maintenance under inspection uncertainties^[24]. Furthermore, for large-scale industrial systems, adaptive health prediction has been integrated with global dynamic maintenance decision-making to optimize group machinery operations^[25]. Since robust and dynamic maintenance policies heavily rely on accurate condition sensing, developing a highly reliable fault diagnosis model under complex industrial constraints becomes a critical prerequisite for successfully implementing these downstream predictive maintenance systems.

Despite the promising results achieved by advanced deep learning methods, real-world industrial applications still face critical challenges. A systematic comparison reveals the distinct limitations of existing methods in handling non-stationary signals and severe data imbalance. First, traditional signal processing methods struggle with the highly non-stationary nature of vibration signals, as they rely heavily on static, handcrafted features that fail to dynamically capture transient fault shocks submerged in strong background noise. Second, while standard deep learning models can automatically extract features, they fundamentally assume a balanced data distribution. In practical scenarios, severe category imbalance is inevitable; critical faults force immediate equipment shutdown, resulting in a severe scarcity of fault samples compared to normal operating data. Consequently, traditional deep models tend to overfit the majority class while ignoring minority fault states. Although some recent studies have employed generative models or data resampling to supplement minority classes, these data-level approaches often introduce artificial noise and struggle to extract truly discriminative features under extreme imbalance. Therefore, a critical research gap exists: there is a lack of an integrated diagnostic framework capable of simultaneously isolating non-stationary fault features under strong noise and achieving precise minority-class identification without relying on potentially unreliable data resampling techniques. This distinct gap directly motivates the development of our proposed robust diagnostic architecture.

To address the aforementioned challenges, this paper proposes a novel diagnostic framework termed the multi-level Inception-LSTM network with WPD and ECA (MILN-WE), which employs wavelet packet decomposition (WPD) to adaptively decompose non-stationary vibration signals into multiple frequency sub-bands, thereby isolating subtle fault-related transients from background noise. To capture the complex spatial-temporal patterns, a multi-scale Inception module is integrated to extract features across different receptive fields, while a LSTM layer is utilized to model the long-term temporal dependencies inherent in the water injection pump cycles. Importantly, instead of conventional feature concatenation, we uniquely embed the efficient channel attention (ECA) mechanism to perform adaptive weighted fusion of the multi-scale spatial-temporal features extracted independently from each WPD sub-band. By utilizing a non-dimensionality-reduction local cross-channel interaction strategy, the model autonomously assigns higher fusion weights to the specific frequency sub-bands that contain critical minority-class fault signatures, while actively suppressing noise-dominated sub-bands. This integrated architecture ensures high diagnostic precision and robust performance under complex industrial operating conditions. The main contributions of this paper are as follows:

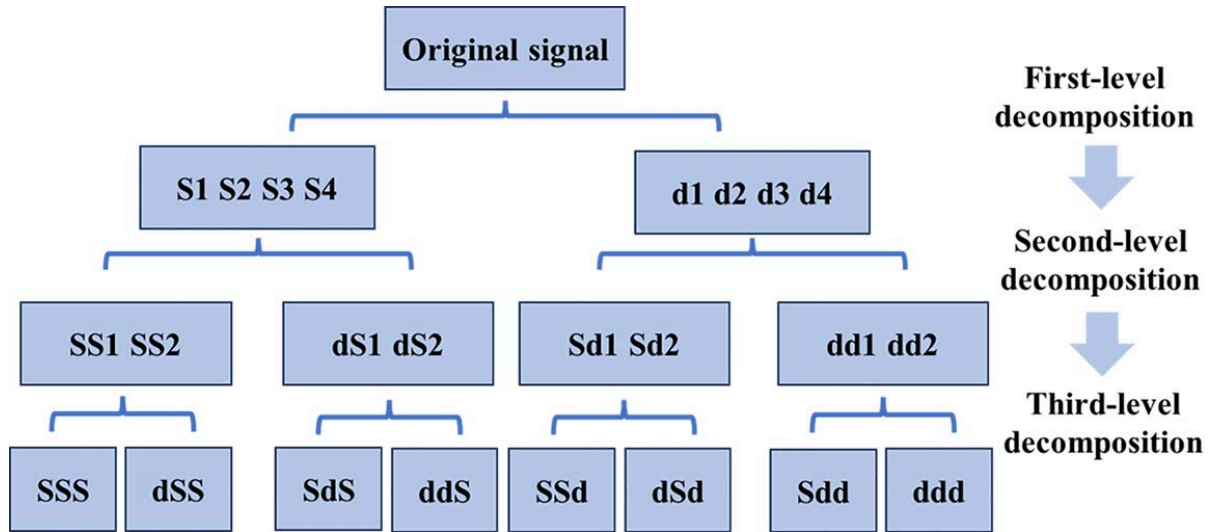


Figure 1. Three-level wavelet packet decomposition process.

(1) A novel hybrid diagnostic framework, MILN-WE, is proposed, which effectively integrates WPD, multi-scale Inception-LSTM, and an ECA mechanism to isolate subtle fault-related transients from background noise;

(2) The proposed model provides a robust solution for multi-scale feature extraction and precise minority-class identification in scenarios characterized by severe category imbalance, without relying on traditional data resampling techniques;

(3) Extensive empirical validation on both a private oilfield water injection pump dataset and a public rotating machinery dataset demonstrates the model's superior diagnostic accuracy, generalization capability, and robustness against strong noise interference compared to existing state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 and 3 introduce the details of the proposed framework and sub-modules. Section 4 presents the experimental study, and Section 5 concludes the paper with a summary of findings and suggestions for future work.

2. PRELIMINARY MATERIALS

2.1 Wavelet packet decomposition

WPD, as an extension of the Discrete Wavelet Transform (DWT), provides a more refined time-frequency analysis capability. Through recursive filtering and down-sampling processes, WPD not only iteratively decomposes low-frequency parts but also further segments high-frequency parts, thereby achieving multi-level decomposition. For an n -level decomposition, WPD can generate multiple different sets of coefficients or nodes, rather than just $(n + 1)$ sets as in DWT. Although the total number of coefficients remains unchanged due to down-sampling, this provides greater flexibility in adapting to different signal characteristics. The basic process of WPD is shown in Figure 1.

The formula for WPD is as follows:

$$P_{(t)} = \sum_{j=-\infty}^{\infty} \alpha_{jk} \varphi_{jk}(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \beta_{jk} \varphi_{jk}(t) \quad (1)$$

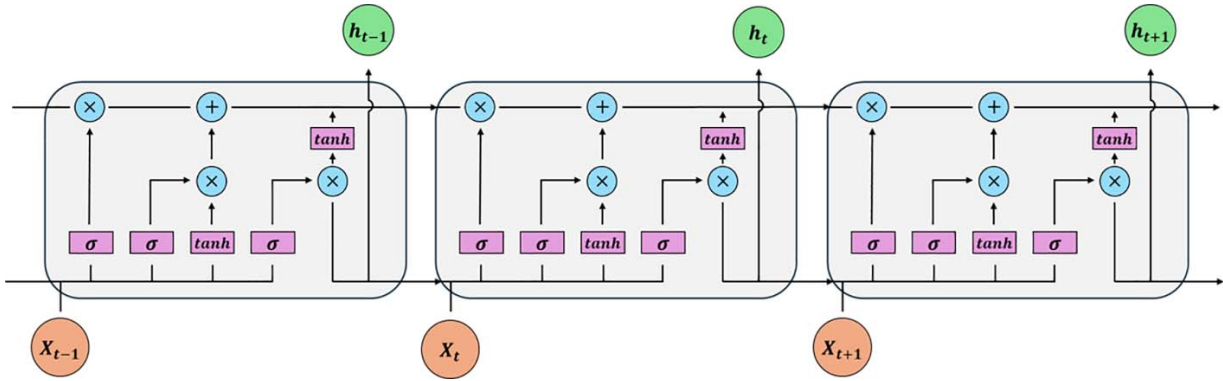


Figure 2. The structure of LSTM module.

2.2 LSTM module

LSTM is a variant of the Recurrent Neural Network that can process sequential data while addressing the problems of gradient vanishing and gradient exploding that occur when training on long sequences. It allows the model to retain information from previous data, thereby further enhancing the model's capability to capture the features of each sample, as shown in Figure 2.

An LSTM unit consists of an input gate, a forget gate, an output gate, and a cell state. The transition formulas at time t are as follows:

$$\text{Input gate : } \begin{cases} i_t = \sigma(W_i \cdot [h_{t-1} \quad x_t] + b_i) \\ \bar{C}_t = \tanh(W_c \cdot [h_{t-1} \quad x_t] + b_c) \end{cases} \quad (2)$$

$$\text{Forget gate : } f_t = \sigma(W_f \cdot [h_{t-1} \quad x_t] + b_f) \quad (3)$$

$$\text{Output gate : } \begin{cases} o_t = \sigma(W_o \cdot [h_{t-1} \quad x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \quad (4)$$

$$\text{Cell state : } C_t = f_t * C_{t-1} + i_t * \bar{C}_t \quad (5)$$

where x_t represents the sequence of the input unit, h_t denotes the hidden state of the unit, and C_t signifies the cell state. W_p , W_c , b_p , and b_c are the weight matrices and bias terms for the input gate; W_f and b_f are those for the forget gate; W_o and b_o are those for the output gate; and σ represents the sigmoid function.

2.3 ECA mechanism

The ECA mechanism is a lightweight attention module specifically designed for deep CNNs. It achieves performance gains with minimal increases in complexity by utilizing an appropriate cross-channel interaction strategy. This strategy is implemented via one-dimensional convolutions, which notably reduce model complexity while maintaining performance, as shown in Figure 3. The kernel size (k) for the convolution operation is adaptively selected based on the following formula to ensure adequate coverage of local cross-channel interactions:

$$k = \varphi(D) = \left\lfloor \frac{\log_2 D}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (6)$$

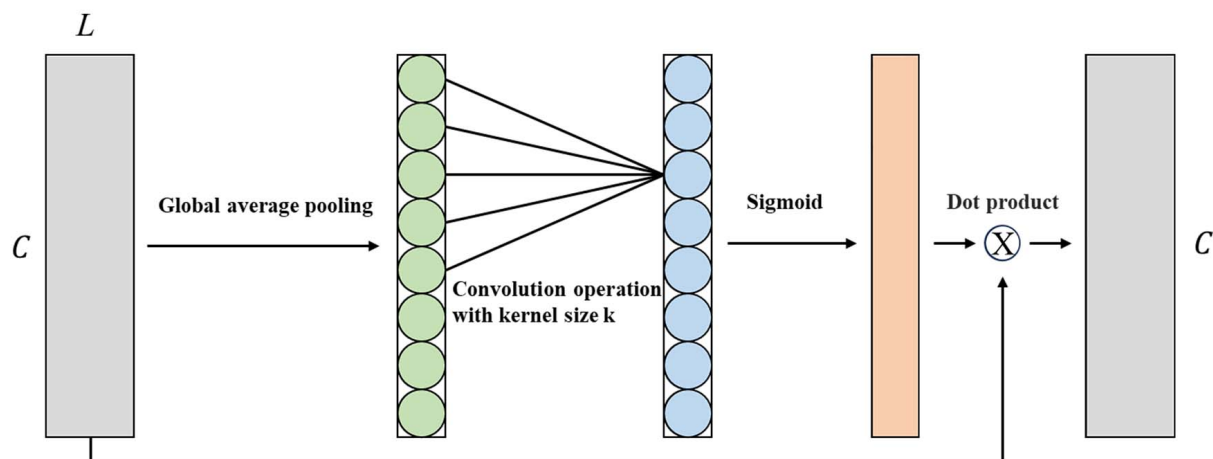


Figure 3. Schematic diagram of ECA mechanism.

Where k represents the size of the convolutional kernel. D represents the dimension of the input sequence. $|n|_{odd}$ indicates the nearest odd number to n . Additionally, the mapping parameters γ and b are fixed at 2.0 and 1.0^[26], which enables the network to adaptively determine the optimal kernel size k based on the channel dimension D , ensuring efficient cross-channel interaction without manual tuning.

3. THE METHODOLOGY

3.1 Overall framework of MILN-WE

To achieve high-precision fault diagnosis of water injection pumps under complex industrial conditions, this paper proposes a multi-level Inception-LSTM network integrated with Wavelet-enhanced attention mechanism, termed MILN-WE. The overall architecture is designed to handle the non-stationary nature of vibration signals and the challenges of feature extraction from imbalanced data.

The proposed MILN-WE framework consists of three main stages: signal decomposition, multi-scale feature extraction, and feature fusion classification. The systematic flowchart of the MILN-WE is illustrated in Figure 4.

Specifically, the raw vibration signals collected from the water injection pump are first processed using a three-level WPD. A three-level decomposition was selected because it provides a sufficient frequency resolution to isolate subtle fault characteristics without introducing excessive computational complexity or over-segmenting the signal^[27,28]. This process decomposes the original complex signal into eight distinct frequency sub-bands. By transforming the 1D time-series signal into multiple frequency components, the model can capture subtle fault characteristics that are often submerged in noise in the original domain.

Subsequently, the features extracted from the eight frequency sub-bands are individually processed through the multi-scale Inception-LSTM branches. Within each branch, an ECA mechanism is integrated to adaptively recalibrate the importance of various feature channels. Unlike conventional hybrid models that typically rely on simple feature concatenation or late-stage attention pooling, this design leverages ECA to perform a non-dimensionality-reduction local cross-channel interaction specifically tailored for the independent WPD sub-bands. This advanced weighted fusion strategy allows the model to dynamically evaluate and effectively highlight fault-related transients while simultaneously suppressing background noise. To integrate information from different frequency domains, the enhanced latent features F_i from all branches are fused into a global representation F_s via a weighted summation strategy, expressed as $F_s = \sum_i W_{C_i} F_i$ (where

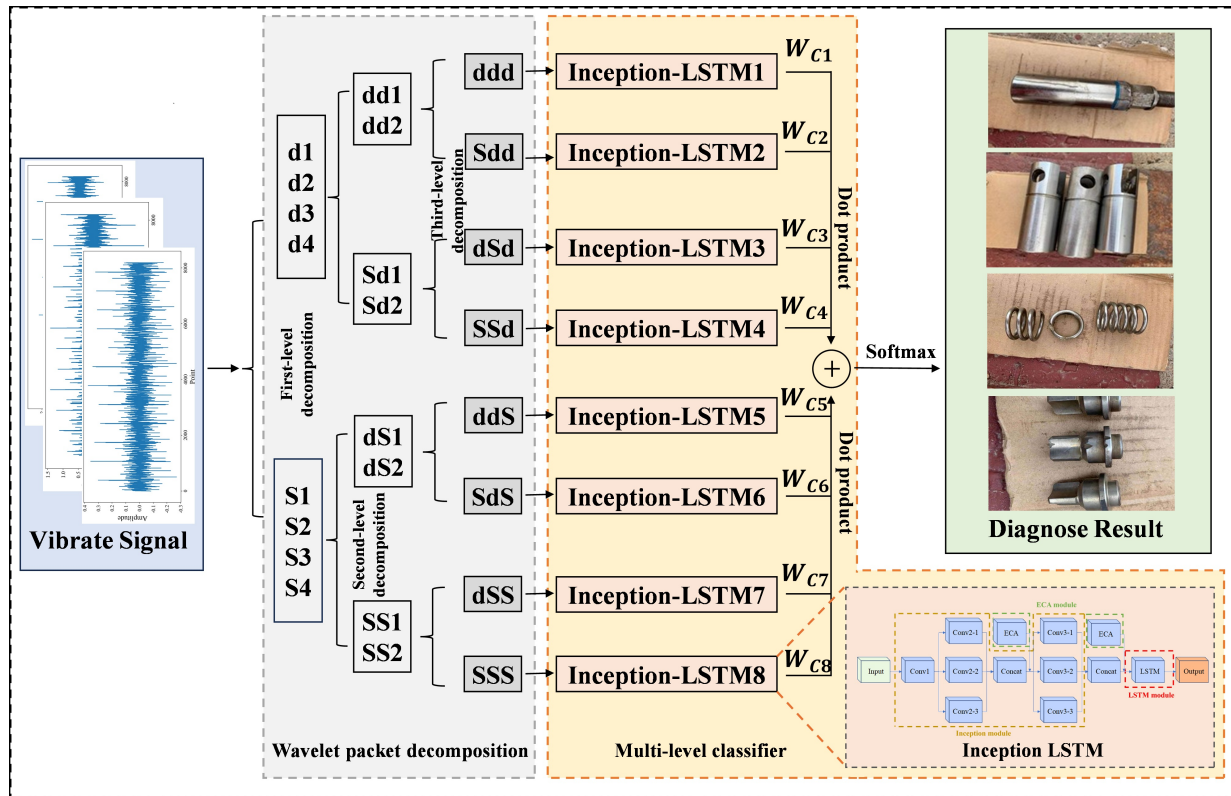


Figure 4. Overall architecture of the proposed MILN-WE framework.

$i = 1, 2, \dots, 8$), where W_{C_i} denotes the learned contribution weight of the i -th sub-band.

Finally, this fused feature map is mapped into the label space through a fully connected layer. A Softmax activation function is then employed to calculate the probability distribution across various fault types, where the category with the highest probability determines the final diagnostic result of the water injection pump.

3.2 Inception-LSTM module

In the fault diagnosis of water injection pumps, the complexity of vibration signals and the severe category imbalance in industrial environments present significant challenges. Traditional CNNs often employ a single-size convolutional kernel, which results in a fixed receptive field that may fail to capture diverse and subtle fault features, especially from minority class samples. To address this limitation, the proposed MILN-WE model utilizes a multi-scale Inception module instead of the standard CNN architecture. By incorporating parallel convolutional layers with different kernel sizes, the module can perceive local spatial features across multiple scales simultaneously. This design notably enhances the model's ability to extract discriminative information from non-stationary signals and improves the diagnostic sensitivity for rare fault types, as shown in Figure 5. Specifically, to further enhance the discriminative power of the spatial features, local ECA modules are embedded within the Inception block following the Conv2-1 and Conv3-1 layers, prior to the feature concatenation.

To complement these spatial features, the architecture further integrates LSTM layers to model the temporal dependencies within the signal. The multi-scale feature maps generated by the Inception module are fed into the LSTM unit, which leverages its unique gating mechanisms—the input, forget, and output gates—to capture long-term correlations in the vibration sequences. By synergizing multi-scale spatial perception with temporal sequential modeling, the MILN-WE framework can derive highly robust and representative

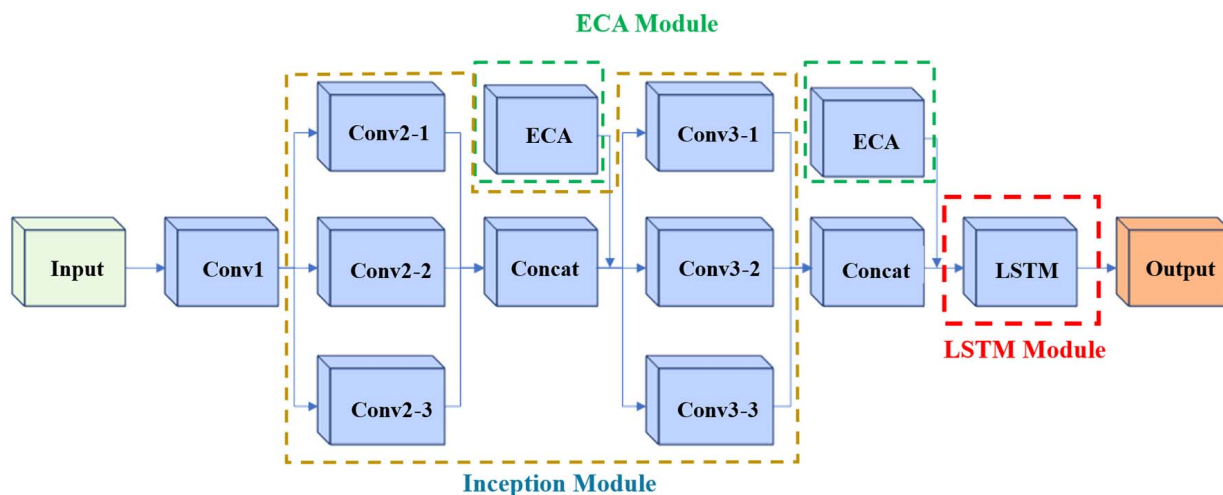


Figure 5. Hybrid structure of the Inception-LSTM network.

features from the sub-bands decomposed by WPD, providing a solid foundation for accurate fault classification under complex operational conditions.

3.3 Adaptive feature weighting via ECA

Traditional feature concatenation or simple averaging fusion methods often overlook the varying contributions of different feature branches to the diagnosis of the current operating state, which can easily introduce redundant information and increase subsequent computational overhead. Therefore, during the feature enhancement and fusion stages, this paper introduces the ECA mechanism to perform adaptive weighted fusion of multi-scale features. It achieves cross-channel local interactions with an extremely low number of parameters, dynamically focusing on the most discriminative key features while avoiding information loss from feature dimensionality reduction. Specifically, complementing the local ECA modules inside the Inception blocks, an ECA-based global weighting mechanism is employed after the parallel Inception-LSTM branches and before the final classification layer to evaluate the eight frequency sub-bands. Through this design, the model captures cross-channel interactions via a non-dimension-reduction local cross-channel interaction strategy. This enables the network to adaptively assign higher weights to key feature channels that significantly contribute to fault diagnosis, while suppressing irrelevant background noise, thereby further enhancing the architecture's overall performance under complex interference, as shown in Figure 6.

The ECA-based feature fusion process comprises three strategic phases. First, multi-scale features are compressed into channel descriptors via Global Average Pooling (GAP) to capture a global receptive field. Subsequently, instead of using computationally expensive fully connected layers, the module employs a one-dimensional convolution (1D Conv) with an adaptive kernel size $k = \varphi(D)$ to facilitate Local Cross-Channel Interaction. This approach effectively captures local dependencies and avoids information loss from dimensionality reduction. Finally, the generated attention weights W_{c_1} to W_{c_8} are applied to the original features through element-wise multiplication and summation. By amplifying fault-sensitive features and suppressing redundant noise, this adaptive strategy notably enhances the discriminative power of the fused feature, ensuring precise identification of the 16 complex operating states.

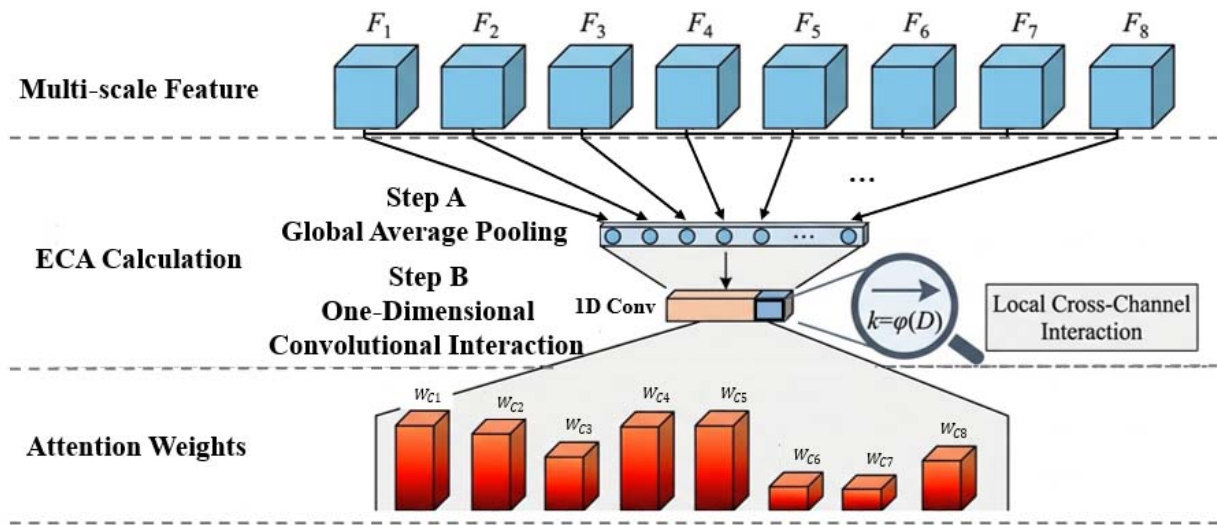


Figure 6. Schematic diagram of adaptive feature weighting via ECA mechanism.

4. EXPERIMENTAL VERIFICATION

4.1 Dataset processing

The private water injection pump dataset is derived from the field operation data of an oil field in China. The water injection pump model used in this study is 3H-8/450II, as shown in Figure 7. Data collection was performed using 15 vibration sensors installed at different locations, the collected data were treated as independent single-channel input sequences. Additionally, all vibration signals were globally normalized prior to model training to eliminate amplitude variations caused by differences in distance and mounting positions between the sensors and the vibration sources, thereby ensuring that the diagnostic process is based on inherent time-frequency fault patterns. The specific installation positions of these sensors are listed in Table 1. The sampling frequency of the sensors is 8,192 Hz with a sampling duration of 1 s, meaning each sample contains 8,192 sampling points.

During the operation of the water injection pump, the motor at the power end drives the crankshaft to rotate, which in turn moves the plunger through the connecting rod, causing it to reciprocate within a high-sealing cylinder. The resulting ultra-high pressure pumps the sand-bearing fluid out of the hydraulic end. During this process, the pump head of the plunger pump is subjected to continuous high-intensity impacts. Under the constant erosion of the sand-bearing fluid, sealing components such as the plunger and valve seats are highly susceptible to wear and tear.

The dataset contains a total of 5,190 samples across 16 different states (including the normal operating state), covering various common failures of plunger pumps, such as plunger wear, pump head spring wear, and bearing bush wear. Figures 8-10 display the raw data sequence plots for some of these states. It can be observed that when the equipment is in a normal state, the cycles are clear, noise is relatively low, and the signal components are relatively simple. Conversely, when a fault occurs, the noise in the data increases notably, the data cycles change, and the underlying components become much more complex.

The sample sizes collected for each state are shown in Figure 11. In actual operations, field engineers strive to keep the machinery in a normal state as much as possible. When certain severe faults occur—such as plunger looseness, bearing bracket damage, or motor bolt looseness—the machine may be shut down immediately, forcing data collection to stop. This further increases the difficulty of diagnosing such faults. Consequently,

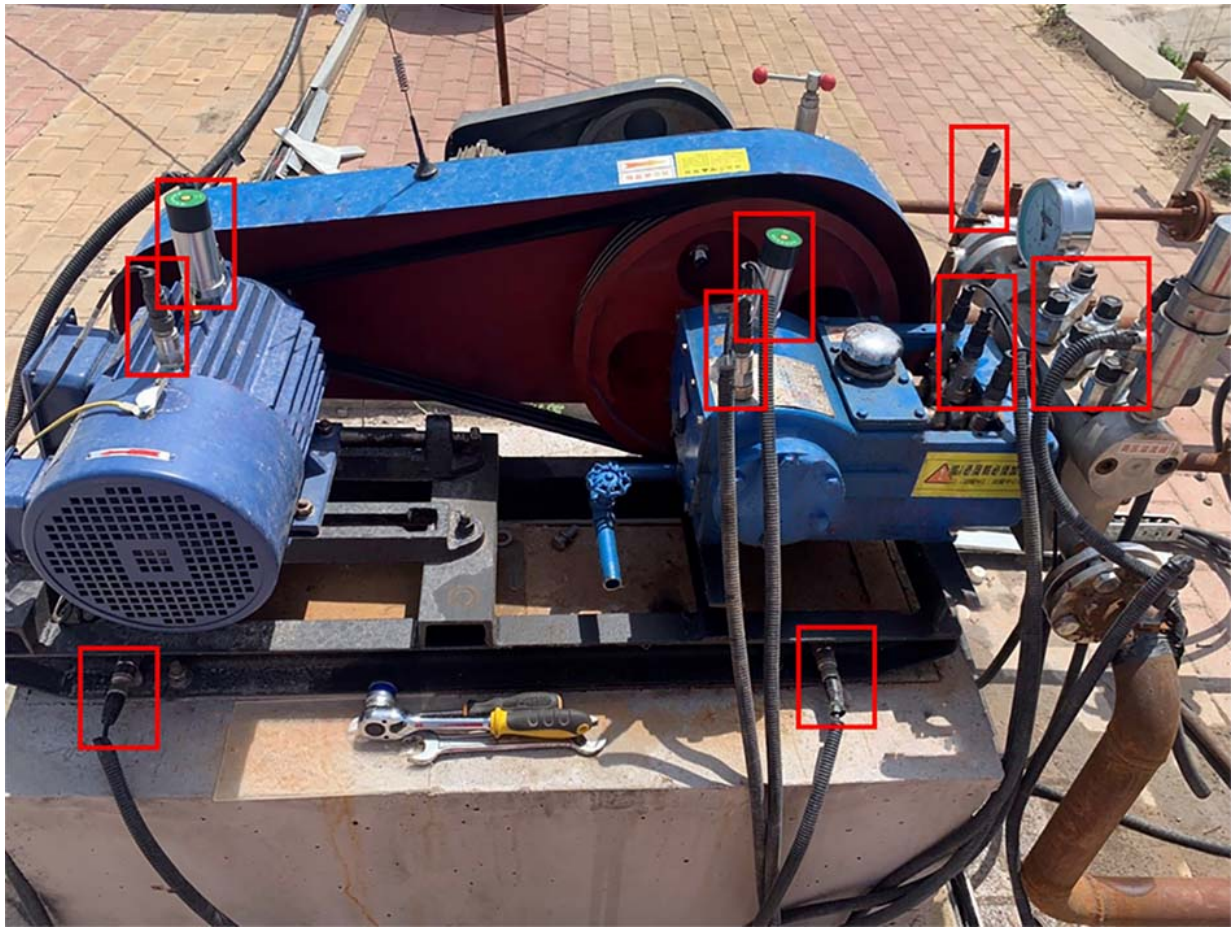


Figure 7. Schematic diagram of the 3H-8/450II water injection pump structure and sensor layout (Photographed by the authors).

Table 1. Vibration sensor mounting areas

Sensor layout	Sensor layout
Base Southeast	West plunger stuffing box
Base Northeast	Center plunger stuffing box
Base Northwest	East plunger stuffing box
Base Southwest	Directly above pump head
Crankshaft bearing	Front of pump head
Motor East	Pump inlet pipeline
Motor West	Pump outlet pipeline
Crankshaft East	/

there is a severe imbalance in the amount of data across different operating states, with the maximum gap between sample sizes exceeding 20 times. Therefore, it is difficult to use data reconstruction methods like resampling to supplement the number of minority class samples, and the quality of such supplemented data is hard to guarantee.

To strictly prevent data leakage between the training and testing phases, the continuous raw data sequences of each state are first chronologically divided into training, validation, and test sets. Specifically, based on the temporal order of data collection, the first 80% of the continuous time period for each state is designated as

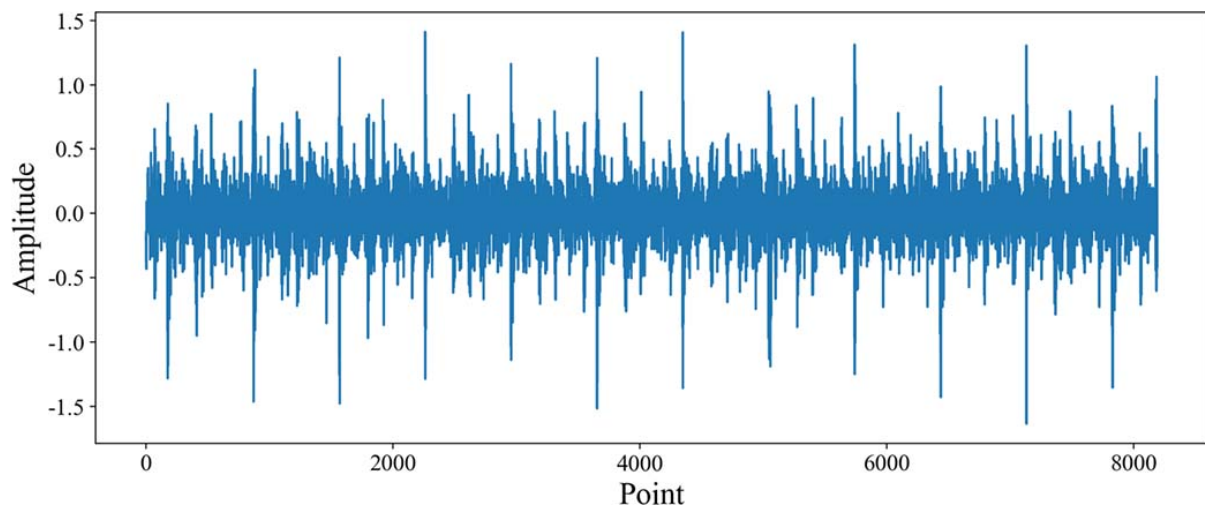


Figure 8. Vibration data of normal state.

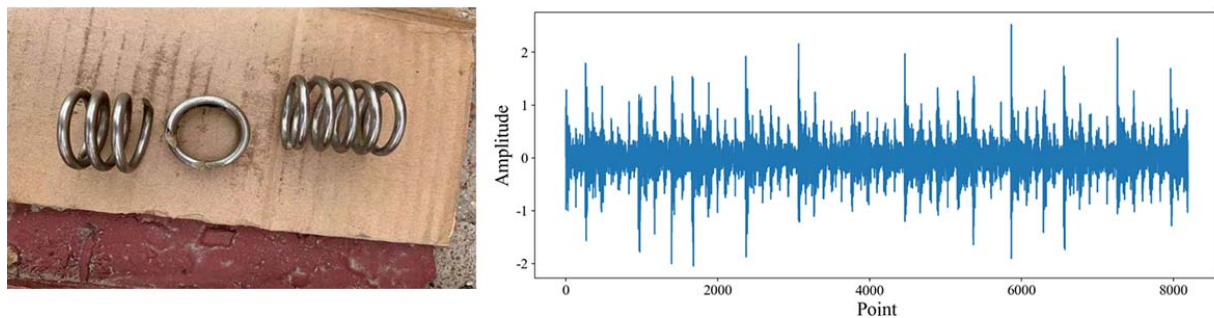


Figure 9. Vibration data with worn pump head spring.

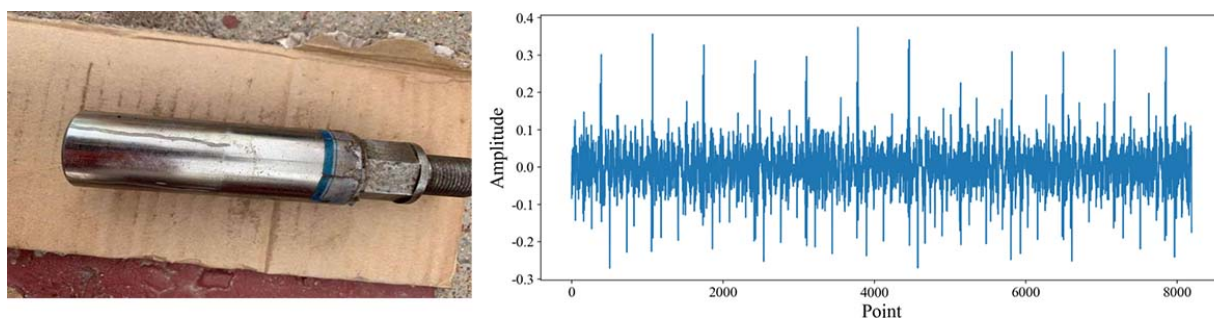


Figure 10. Vibration data with worn plunger.

the training period. The subsequent 10% of the time period is allocated for validation, and the final 10% of the time period is strictly reserved for testing. Following this strict chronological division, a sliding window with overlap is independently employed to segment the data within each respective set to increase the data volume and ensure better training performance, as shown in [Figure 12](#). Specifically, the width of the sliding window is 2,048 and the step size is 1,024. After independent segmentation, the total number of samples effectively generated reach 36,330.

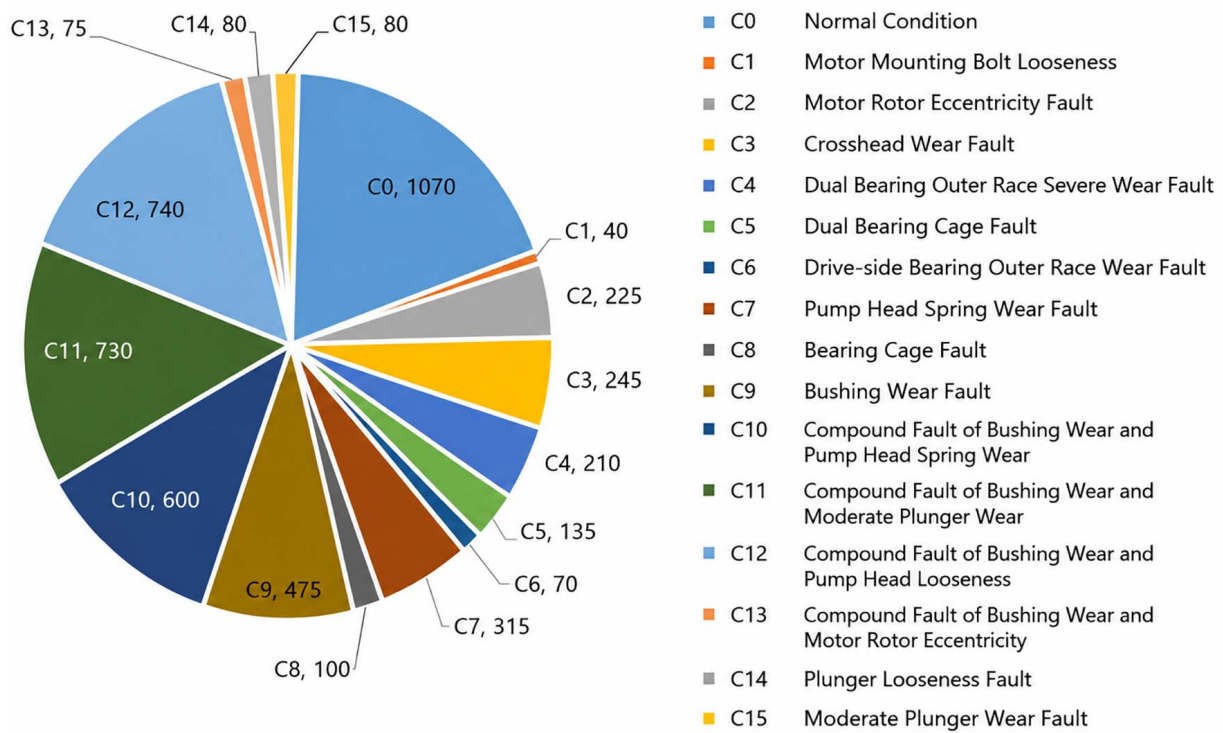


Figure 11. Sample size statistics by state.

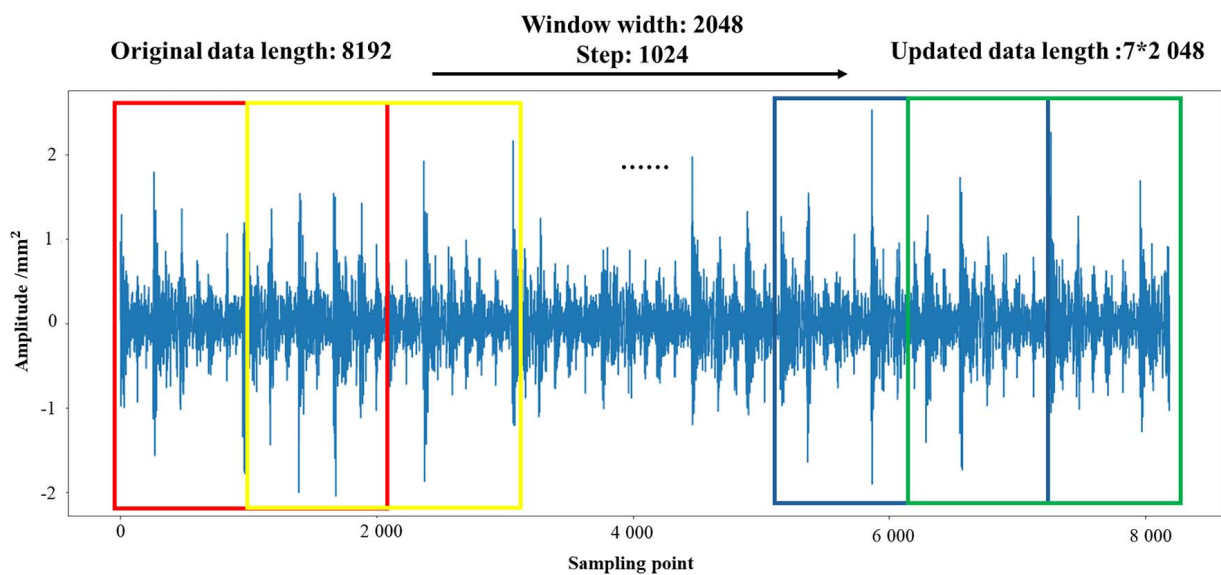


Figure 12. Data splitting with sliding window.

4.2 Experimental setup

In this experimental section, Accuracy, Precision, Recall, and F1-Score are employed as evaluation metrics to assess the diagnostic performance of the proposed models. Class-balanced CNN (CB-CNN)^[29] and SMOTE-CNN^[30] are selected as baseline methods, representing loss function-based improvement and oversampling-based approaches for handling class imbalance in water injection pump fault diagnosis, respectively. To further validate the contribution of individual modules in addressing sample imbalance, a comprehensive ablation study is conducted with traditional CNN as the reference model. Several variant

architectures are constructed for comparison, including CNN-LSTM, Inception-LSTM, Inception-LSTM-ECA, CNN-LSTM-WE, Inception-LSTM-WPD, supervised contrastive learning^[31] (SCL), the deep-stable CNN^[32] (DSCNN), Inception-WE, and the proposed MILN-WE.

Each variant systematically excludes specific components from the complete architecture: CNN-LSTM removes WPD, Inception, and ECA modules; Inception-LSTM excludes the WPD and ECA modules; Inception-LSTM-ECA removes WPD module; CNN-LSTM-WE and Inception-WE remove the Inception and LSTM modules, respectively; and Inception-LSTM-WPD excludes the ECA module. This systematic comparison enables quantitative assessment of each module's contribution to the overall diagnostic performance under imbalanced data conditions.

Based on the PyTorch deep learning framework, the proposed Multi-level Inception-LSTM network was constructed, integrating WPD and ECA mechanism. The hardware configuration included an Intel i5-13500HX CPU, an NVIDIA 4060 GPU, and 16GB of RAM. Based on the aforementioned independent data splitting strategy, the segmented samples from each state strictly maintain the 8:1:1 ratio for the training, validation, and test sets, respectively. This rigorous procedure ensures zero information crossover, resulting in 29,064 training samples, 3,633 validation samples and 3,633 testing samples.

Cross-Entropy Loss was employed as the loss function, and the Adam optimizer was used for model training. The learning rate was set to 0.002. To ensure statistical reliability and mitigate the influence of random initialization, all models evaluated in this study were independently trained and tested 5 times under the same hardware and software configurations. Within the Inception module, the kernel sizes for each convolutional layer are specified in Table 2, with a Batch Normalization layer added after each convolution. The LSTM module was configured with 2 layers. Additionally, as depicted in Figure 5, local ECA modules were integrated specifically after the Conv2-1 and Conv3-1 layers to enhance the cross-channel interaction capability of the Inception module.

4.3 Experimental results

The confusion matrix for the diagnostic results of the MILN-WE model is shown in Figure 13, where the horizontal axis represents the true labels and the vertical axis represents the predicted labels. The numbers within each colored block correspond to the number of samples classified into that category. As can be observed from the figure, except for fault samples of label 7, the number of misclassified samples for all other faults is below 10.

To verify the effectiveness of the improved method and evaluate its performance against common fault diagnosis algorithms for imbalanced data, other baseline models were trained using the same parameters for verification. Figures 14 and 15, and Table 3 illustrate the diagnostic performance comparison between the proposed model and eleven benchmark models: CNN, CB-CNN, SMOTE-CNN, CNN-LSTM, Inception-LSTM, Inception-LSTM-ECA, CNN-LSTM-WE, Inception-WE, and Inception-LSTM-WPD, SCL, DSCNN.

During the model training process, the proposed model achieved faster convergence speed in terms of both training and testing losses compared to other models. Regarding the testing set, the SMOTE-CNN, which incorporates SMOTE resampling, showed a performance similar to that of the traditional CNN model. This indicates that when facing severe data imbalance, resampling methods struggle to generate appropriate new data and fail to effectively distinguish minority class faults.

Table 2. Network parameter settings

Network layer	Parameter settings	Data dimensions
Vibration signal	/	[256,1,2048]
Wavelet decomposition	/	[256,1,256]
Conv 1-1	Kernel size: 5 Output channel: 16	[256,16,252]
Maximum pooling layer	Pooling length: 3 Pooling stride: 2	[256,16,125]
Conv2-1 to 2-3	Kernel size: 5/7/9 Output channel: 48	[256,48,125]
ECA module	Output channel: 48	[256,48,125]
Maximum pooling layer	Pooling length: 3 Pooling stride: 2	[256,48,62]
Conv3-1 to 3-3	Kernel size: 5/7/9 Output channel: 144	[256,144,62]
ECA module	Output channel: 144	[256,144,62]
Maximum pooling layer	Pooling length: 3 Pooling stride: 2	[256,144,30]
Dropout layer	/	[256,144,30]
LSTM layer	Output dimensions: 256 Number of layers: 2 Dropout probability: 0.2	[256,256,30]
Fully connected layer	Output dimensions: 960	[256,960]
Fully connected layer	Output dimensions:16	[256,16]

ECA: Efficient channel attention; LSTM: long short-term memory.

The diagnostic Accuracy, Precision, Recall and F1-Score of the proposed model on the test set reached 99.38%, 99.42%, 99.27%, and 99.34%, respectively, all of which showed improvements over the other benchmark models. Compared to the traditional CNN, the diagnostic accuracy improved by 6.05%. Compared with CB-CNN, a representative imbalance-aware fault diagnosis model based on improved loss function, the proposed MILN-WE model achieves a 3.24% improvement in diagnostic accuracy. Furthermore, when compared with SMOTE-CNN, which employs oversampling techniques for handling class imbalance, the diagnostic accuracy is enhanced by 5.28%. These results demonstrate that the proposed MILN-WE model exhibits better performance in addressing the imbalanced data diagnosis problem of plunger-type water injection pumps when compared with other advanced fault diagnosis methods.

Besides, during the process of network optimization, the diagnostic performance enhanced progressively. By comparing the diagnostic results of the proposed model with the Inception-LSTM-ECA, CNN-LSTM-WE, Inception-WE, and Inception-LSTM-WPD, SCL, DSCNN models, it is evident that each module contributes to the improvement of the final diagnostic accuracy. In summary, these results demonstrate the effectiveness and superiority of the proposed model for fault diagnosis under class-imbalanced conditions.

4.4 Public dataset verification

To verify the cross-platform robustness of the MILN-WE architecture, we extended our evaluation to a widely recognized public repository: the centrifugal multistage impeller blower dataset^[33].

In this validation phase, five operational states are considered: normal baseline (C0), along with four localized defects—outer-race (C1), inner-race (C2), rolling-element (C3), and gear (C4) failures. To mirror the sparsity of fault samples in actual industrial production, we intentionally constructed a non-uniform

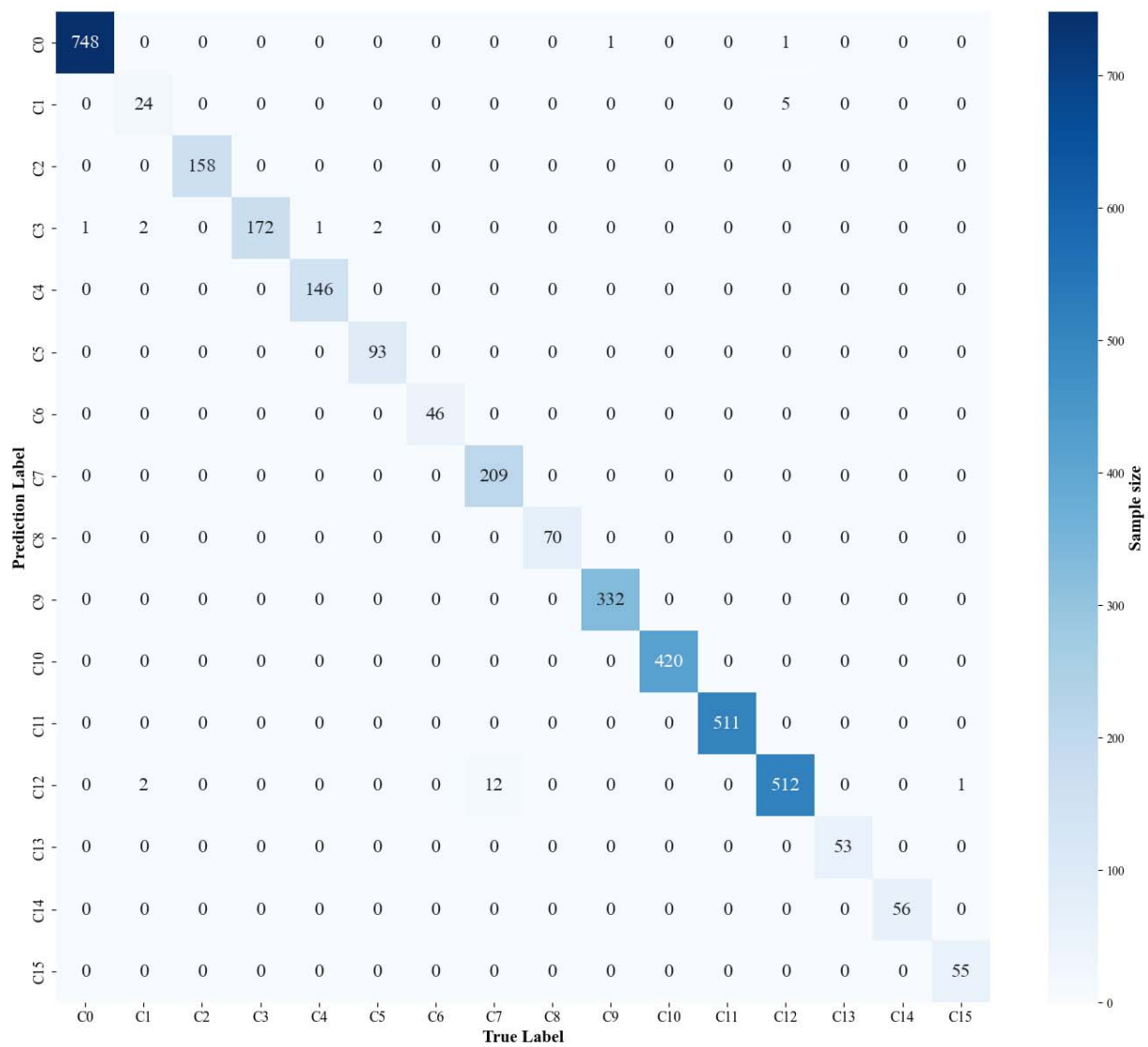


Figure 13. Confusion matrix of diagnosis results.

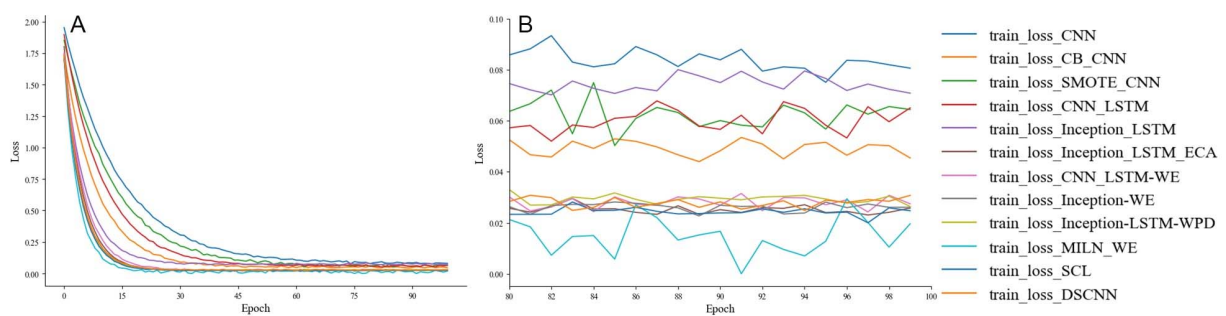


Figure 14. Training loss function. (A) Overall training loss curve, (B) Local training loss curve (80-100).

distribution. As detailed in Table 4, for classes C0-C2, 600 training and 180 testing samples per class are used; for minority classes C3 and C4, 200 training and 60 testing samples per class are used. Additionally, 10% of the training samples from each class are partitioned as a validation set during the training process. By testing the model on this skewed dataset, we can more effectively demonstrate MILN-WE's proficiency in

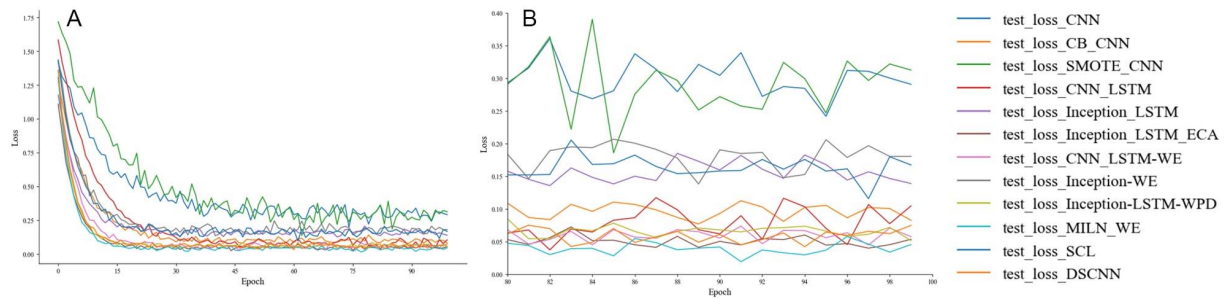


Figure 15. Testing loss function. (A) Overall test loss curve, (B) Local test loss curve(80-100).

Table 3. Comparison of diagnostic results

Model	Accuracy	Precision	Recall	F1-score
CNN	93.33	90.26	93.50	91.85
CB-CNN	96.14	95.83	96.68	96.25
SMOTE-CNN	94.10	92.37	94.19	93.27
CNN-LSTM	95.79	94.01	95.49	94.74
Inception-LSTM	96.31	95.14	97.17	96.14
Inception -LSTM-ECA	98.82	98.49	98.81	98.65
CNN-LSTM-WE	98.39	96.94	99.03	97.97
Inception-WE	97.84	98.42	96.35	97.37
SCL	98.03	97.86	97.21	97.53
Inception-LSTM-WPD	97.77	97.25	98.42	97.83
DSCNN	97.71	97.34	98.02	97.68
MILN-WE	99.38	99.42	99.27	99.34

*Bold values indicate the optimal results under different evaluation metrics. CNN: Convolutional neural network; CB: class-balanced; LSTM: long short-term memory; ECA: efficient channel attention; SCL: supervised contrastive learning; WPD: wavelet packet decomposition; DSCNN: the deep-stable CNN; MILN-WE: the multi-level Inception-LSTM network with WPD and ECA.

Table 4. Details of the public dataset

Label	Fault type	Training samples	Testing samples
C0	No fault	600	180
C1	Bearing outer race fault	600	180
C2	Bearing inner race fault	600	180
C3	Bearing rolling element fault	200	60
C4	Gear fault	200	60

identifying underrepresented fault signatures amidst dominant healthy signals.

In the training process, the Adam optimizer was deployed for 50 epochs with a starting learning rate of 0.001. To prevent the model from overfitting, a decay strategy was integrated: the learning rate was reduced by 50% whenever the loss metric failed to decrease over 15 successive epochs. Additionally, we introduced Gaussian white noise with a signal-to-noise ratio of 1 into the raw vibration data to replicate the complex noise interference typically encountered in actual industrial operations.

Table 5. Results of comparison experiments on public dataset

Model	Accuracy	Precision	Recall	F1-score
CNN	81.24	80.15	82.30	81.21
CB-CNN	84.50	83.92	84.88	84.40
SMOTE-CNN	82.10	81.05	82.95	81.99
CNN-LSTM	85.65	84.33	86.12	85.22
INCEPTION-LSTM	88.42	87.90	88.75	88.32
INCEPTION-LSTM-ECA	91.15	90.85	91.50	91.17
CNN-LSTM-WE	92.30	91.75	92.65	92.20
INCEPTION-WE	92.85	92.10	93.40	92.75
SCL	93.08	92.76	93.35	93.05
INCEPTION-LSTM-WPD	93.65	93.20	94.05	93.62
DSCNN	94.57	94.21	94.88	94.54
MILN-WE	95.82	95.60	96.15	95.87

*Bold values indicate the optimal results under different evaluation metrics. CNN: Convolutional neural network; CB: class-balanced; LSTM: long short-term memory; ECA: efficient channel attention; SCL: supervised contrastive learning; WPD: wavelet packet decomposition; DSCNN: the deep-stable CNN; MILN-WE: the multi-level Inception-LSTM network with WPD and ECA.

As detailed in Table 5, our MILN-WE framework attained a 95.82% diagnostic accuracy. This performance shows higher diagnostic accuracy compared to the benchmark models, exhibiting improvements over CNN, SMOTE-CNN, CNN-LSTM, INCEPTION-LSTM-WPD, SCL, DSCNN by margins of 14.58%, 13.72%, 10.17%, 2.17%, 2.74%, and 1.25%, respectively. Beyond accuracy, our approach consistently yielded competitive results across the remaining evaluation metrics. These outcomes illustrate the model's capability to maintain effective fault identification on the public dataset, even under severe background noise conditions.

5. DISCUSSION

The study proposes a MILN-WE network for fault diagnosis of oilfield water injection pump and public rotating machinery datasets. Experimental results demonstrate that MILN-WE is effective in processing complex, non-stationary vibration signals. Dual validation on both private and public datasets confirms its superior robustness and generalization across diverse mechanical structures and operating conditions. This research contributes a high-precision feature extraction and classification scheme to address industrial challenges such as class imbalance and strong noise, providing a theoretical foundation for intelligent predictive maintenance.

6. CONCLUSION

This study addresses the critical challenges of non-stationary vibration signals and severe category imbalance in the fault diagnosis of water injection pumps operating in complex industrial environments. Looking forward, a prospective summary of our future research will focus on exploring cross-condition transfer learning and few-shot learning to improve model adaptability in data-scarce environments, and pursuing model lightweighting for real-time edge-side deployment.

DECLARATIONS

Authors' contributions

Software, Writing-Original draft preparation: Wu, X.

Writing-Reviewing and Editing: Wu, Z.

Data Curation: Luo, F.

Software: Wang, J.

Conceptualization, Methodology, Funding acquisition: Xia, T.

Project administration: Xi, L.

Availability of data and materials

The private dataset used in the study are available from the corresponding author upon reasonable request. The public dataset used in the study is openly available in CFD_datasets at https://github.com/THUFDD/CFD_datasets.

AI and AI-assisted tools Statement

Not applicable.

Financial support and sponsorship

This research is supported by National Natural Science Foundation of China (72571173), Natural Science Foundation of Shanghai (25ZR1401196), and National Key Research and Development Program of China (2022YFF0605700).

Conflicts of interest

Xia, T. is an Editorial Board Member of the journal *Complex Engineering Systems*. Xia, T. was not involved in any steps of editorial processing, notably including reviewers' selection, manuscript handling and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. He, F.; Ye, Q. A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm. *Sensors* **2022**, *22*, 1410. DOI PubMed PMC
2. Guo, L.; Han, B.; Huang, Q. Bearing fault diagnosis based on improved morlet wavelet transform and shallow residual neural network. *Appl. Sci.* **2024**, *14*, 4542. DOI
3. Zhai, Z.; Luo, L.; Chen, Y.; Zhang, X. Rolling bearing fault diagnosis based on a synchrosqueezing wavelet transform and a transfer residual convolutional neural network. *Sensors* **2025**, *25*, 325. DOI PubMed PMC
4. Tan, Y.; Wu, G.; Qiu, Y.; Fan, H.; Wan, J. Fault diagnosis of a mixed-flow pump under cavitation condition based on deep learning techniques. *Front. Energy. Res.* **2023**, *10*, 1109214. DOI
5. Gao, Y.; Piltan, F.; Kim, J. A novel image-based diagnosis method using improved DCGAN for rotating machinery. *Sensors* **2022**, *22*, 7534. DOI PubMed PMC
6. Deng, C.; Deng, Z.; Lu, S.; He, M.; Miao, J.; Peng, Y. Fault diagnosis method for imbalanced data based on multi-signal fusion and improved deep convolution generative adversarial network. *Sensors* **2023**, *23*, 2542. DOI PubMed PMC
7. Lai, S.; Cheung, T.; Zhao, J.; Xue, K.; Fung, K.; Lam, K. Residual attention single-head vision transformer network for rolling bearing fault diagnosis in noisy environments. In Proceedings of the 2024 6th International Conference on Video, Signal and Image Processing; Ningbo Hainan, China. New York, NY, USA: ACM; 2024. pp. 136-50. DOI
8. Liu, D.; Cui, L.; Wang, G.; Cheng, W. Interpretable domain adaptation transformer: a transfer learning method for fault diagnosis of rotating machinery. *Struct. Health. Monit.* **2024**, *24*, 1187-200. DOI
9. Hassannejad, R.; Ettefagh, M. M.; Bahrami Mossayebi, Y. Adaptive wavelet-based physics-informed CNN for bearing fault diagnosis. *Int. J. Progn. Health. Manag.* **2025**, *16*, 4234. DOI
10. Deng, R.; Chen, D.; Yao, C.; Shao, M.; Hu, D. A multi-scale sensor importance-aware attention fusion network and its applications in fault diagnosis of centrifugal pumps and axial piston pumps. *Measurement* **2026**, *258*, 119315. DOI
11. Kim, A. R.; Seon, Kim. H.; Young, Kim. S. Transformer-based fault detection using pressure signals for hydraulic pumps. *IEEE. Access.* **2024**, *12*, 145795-808. DOI

12. Zhao, L.; He, Y.; Zheng, H.; Dai, D. A novel multistep wavelet convolutional transfer diagnostic framework for cross-machine bearing fault diagnosis. *Sensors* **2025**, *25*, 3141. [DOI PubMed PMC](#)
13. Yu, S.; Song, L.; Pang, S.; Wang, M.; He, X.; Xie, P. M-net: a novel unsupervised domain adaptation framework based on multi-kernel maximum mean discrepancy for fault diagnosis of rotating machinery. *Complex. Intell. Syst.* **2024**, *10*, 3259-72. [DOI](#)
14. Sun, K.; Xu, X.; Lu, N.; Xia, H.; Han, M. Joint discriminative adversarial domain adaptation for cross-domain fault diagnosis. *IEEE. Trans. Instrum. Meas.* **2023**, *72*, 1-11. [DOI](#)
15. Li, G.; Wu, J.; Deng, C.; Wei, M.; Xu, X. Self-supervised learning for intelligent fault diagnosis of rotating machinery with limited labeled data. *Appl. Acoust.* **2022**, *191*, 108663. [DOI](#)
16. Zhu, P.; Ma, S.; Han, Q.; Chu, F. Deep contrastive transfer learning for rotating machinery fault diagnosis. *IEEE. Trans. Instrum. Meas.* **2025**, *74*, 1-10. [DOI](#)
17. Zhou, F.; Zhang, Z.; Li, S. Research on federated learning method for fault diagnosis in multiple working conditions. *Complex. Eng. Syst.* **2021**, *1*, 7. [DOI](#)
18. Shao, H.; Xia, M.; Wan, J.; De Silva, C. W. Modified stacked autoencoder using adaptive morlet wavelet for intelligent fault diagnosis of rotating machinery. *IEEE/ASME. Trans. Mechatron.* **2022**, *27*, 24-33. [DOI](#)
19. Wang, Y.; Liu, Y.; Chow, T. W. S.; Gu, J.; Zhang, M. A balanced adversarial domain adaptation method for partial transfer intelligent fault diagnosis. *IEEE. Trans. Instrum. Meas.* **2022**, *71*, 1-11. [DOI](#)
20. Zhang, Y.; Liu, Z.; Huang, Q. A contrastive learning-based fault diagnosis method for rotating machinery with limited and imbalanced labels. *IEEE. Sensors. J.* **2023**, *23*, 16402-12. [DOI](#)
21. Zhang, Y.; Ren, Z.; Zhou, S.; Feng, K.; Yu, K.; Liu, Z. Supervised contrastive learning-based domain adaptation network for intelligent unsupervised fault diagnosis of rolling bearing. *IEEE/ASME. Trans. Mechatron.* **2022**, *27*, 5371-80. [DOI](#)
22. Zhu, J.; Chen, N.; Shen, C. A new multiple source domain adaptation fault diagnosis method between different rotating machines. *IEEE. Trans. Ind. Inf.* **2021**, *17*, 4788-97. [DOI](#)
23. Yang, L.; Chen, Y.; Ma, X.; Qiu, Q.; Peng, R. A prognosis-centered intelligent maintenance optimization framework under uncertain failure threshold. *IEEE. Trans. Rel.* **2024**, *73*, 115-30. [DOI](#)
24. Tan, L.; Wei, F.; Ma, X.; Peng, R.; Xiao, H.; Yang, L. Systemic condition-based maintenance optimization under inspection uncertainties: a customized multiagent reinforcement learning approach. *IEEE. Trans. Rel.* **2025**, *74*, 5848-62. [DOI](#)
25. Yang, L.; Zhou, S.; Ma, X.; Chen, Y.; Jia, H.; Dai, W. Group machinery intelligent maintenance: Adaptive health prediction and global dynamic maintenance decision-making. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110426. [DOI](#)
26. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-net: efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13-19; Seattle, WA, USA. IEEE; 2020. pp. 11531-9. [DOI](#)
27. Aburakhia, S. A.; Myers, R.; Shami, A. A hybrid method for condition monitoring and fault diagnosis of rolling bearings with low system delay. *IEEE. Trans. Instrum. Meas.* **2022**, *71*, 1-13. [DOI](#)
28. Dubaish, A. A.; Jaber, A. A. Comparative analysis of SVM and ANN for machine condition monitoring and fault diagnosis in gearboxes. *Math. Model. Eng. Probl.* **2024**, *11*, 976-86. [DOI](#)
29. Cui, Y.; Jia, M.; Lin, T.; Song, Y.; Belongie, S. Class-balanced loss based on effective number of samples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15-20; Long Beach, CA, USA. IEEE; 2019. pp. 9260-9. [DOI](#)
30. Joloudari, J. H.; Marefat, A.; Nematollahi, M. A.; Oyelere, S. S.; Hussain, S. Effective class-imbalance learning based on SMOTE and convolutional neural networks. *Appl. Sci.* **2023**, *13*, 4006. [DOI](#)
31. Pan, C.; Shang, Z.; Tang, L.; Cheng, H.; Li, W. Open-set domain adaptive fault diagnosis based on supervised contrastive learning and a complementary weighted dual adversarial network. *Mech. Syst. Signal. Process.* **2025**, *222*, 111780. [DOI](#)
32. Xu, Z.; Lee, C. K. M.; Wong, C. A novel fault diagnosis method based on deep stable learning for bearings with imbalanced data samples. *Expert. Syst. Appl.* **2025**, *281*, 127634. [DOI](#)
33. Liu, Z.; Zhang, J.; He, X.; Zhang, Q.; Sun, G.; Zhou, D. Fault diagnosis of rotating machinery with limited expert interaction: a multicriteria active learning approach based on broad learning system. *IEEE. Trans. Contr. Syst. Technol.* **2023**, *31*, 953-60. [DOI](#)

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.



© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.