



OPEN

## Hydraulic system fault diagnosis decoupling method based on 2D time-series modeling and self-attention fusion

Haicheng Wang<sup>1</sup>, Juan Zhou<sup>1✉</sup>, Hu Chen<sup>2</sup>, Bo Xu<sup>2</sup> & Zhengxiang Shen<sup>2</sup>

Hydraulic systems play a pivotal and extensive role in mechanics and energy. However, the performance of intelligent fault diagnosis models for multiple components is often hindered by the complexity, variability, strong hermeticity, intricate structures, and fault concealment in real-world conditions. This study proposes a new approach for hydraulic fault diagnosis that leverages 2D temporal modeling and attention mechanisms for decoupling compound faults and extracting features from multisample rate sensor data. Initially, to address the issue of oversampling in some high-frequency sensors within the dataset, variable frequency data sampling is employed during the data preprocessing stage to resample redundant data. Subsequently, two-dimensional convolution simultaneously captures both the instantaneous and long-term features of the sensor signals for the coupling signals of hydraulic system sensors. Lastly, to address the challenge of feature fusion with multisample rate sensor data, where direct merging of features through maximum or average pooling might dilute crucial information, a feature fusion and decoupling method based on a probabilistic sparse self-attention mechanism is designed, avoiding the issue of long-tail distribution in multisample rate sensor data. Experimental validation showed that the proposed model can effectively utilize samples to achieve accurate fault decoupling and classification for different components, achieving a diagnostic accuracy exceeding 97% and demonstrating robust performance in hydraulic system fault diagnosis under noise conditions.

Hydraulic systems play a pivotal role in modern mechanical design, boasting advantages such as rapid response characteristics, high load stiffness, remarkable power density, and commendable stability performance, leading to widespread adoption in various applications. Typically forming the core of engineering equipment, especially in control and power transmission, they are often operated in field environments. However, hydraulic systems can be vulnerable to damage due to handling high loads, experiencing severe shocks, or facing unstable working conditions. Notably, the failure of internal components within a hydraulic system can initiate a chain reaction, resulting in the cascading failure of multiple parts<sup>1,2</sup>. Consequently, timely fault diagnosis and troubleshooting of hydraulic systems to ensure their smooth operation and the efficiency of engineering equipment are of paramount importance.

Nevertheless, hydraulic systems exhibit a coupling characteristic encompassing mechanical, electrical, and hydraulic elements, rendering accurate fault diagnosis a formidable challenge. Compared to the more transparent failures in common electromechanical structures, malfunctions within hydraulic systems in engineering equipment are notably more obscure and ambiguous, characterized by the following aspects:

- Compared to conventional electrical and mechanical systems, hydraulic systems boast enhanced hermeticity, intricate structures, and a propensity for concealed malfunctions, rendering them susceptible to disturbances. For instance, the reciprocating motion of cylinders generates a significant amount of excitation sources<sup>3</sup>.
- The mapping between the characteristics of malfunctions and their underlying causes is intricate. A single fault may stem from numerous sources, making the precise identification of the origin challenging. The failure of internal components can initiate a domino effect, leading to the cascading failure of multiple parts.

<sup>1</sup>College of Energy Environment and Safety Engineering and College of Carbon Metrology, China Jiliang University, Hangzhou 310018, China. <sup>2</sup>Ningbo Special Equipment Inspection and Research Institute, Ningbo 315000, China. ✉email: zhoujuan@cjl.edu.cn

- Failures within the hydraulic system can occur either concurrently or independently. When composite malfunctions arise, the signals transmitted through the intricate channels of the hydraulic system may become distorted, thereby impacting the accuracy of fault detection outcomes.

The quintessence of diagnostic methodologies for hydraulic system malfunctions lies in extracting signal characteristics and decoupling fault diagnoses, thus establishing a correlation between fault manifestations and their underlying causes. To date, many scholars have delved into the realm of fault diagnosis within hydraulic systems, employing fundamental techniques such as signal processing, statistical methodologies, machine learning, and deep learning to forge intelligent fault diagnosis methodologies.

In the initial stages, experts and technicians directly observed and analyzed the components' failure mechanisms for subjective diagnostics or engaged in modeling and simulation<sup>1</sup>. For instance, they constructed simulation models using AMESim software to analyze the fault modes and mechanisms of critical components<sup>4</sup>; Subsequently, scholars introduced sensor technology and signal processing techniques into hydraulic systems for feature extraction. They described the relationship between measured signals and fault modes through mathematical models, which typically encompassed traditional statistical methods and machine learning approaches<sup>5</sup>. This includes, but is not limited to, various signal decomposition techniques such as Continuous Wavelet Transform (CWT), Filter Bank-based Empirical Wavelet Transform (FBEWT), Ensemble Empirical Mode Decomposition (EEMD), and Wavelet Packet Decomposition Reconstruction (WPDR). Fault decoupling algorithms mainly consist of physical model correction methods based on cosine similarity<sup>6</sup>, Linear Discriminant Analysis (LDA)<sup>7</sup>, and Extreme Gradient Boosting (XGBoost)<sup>8</sup>.

As the discipline of fault diagnosis increasingly transitions into the vast and intricate landscape of big data, the plethora of collected signals is both immense in volume and multifaceted. Conventional methodologies for identification are finding themselves poorly equipped to cater to the exigencies presented by the tremendous influx of data. Therein lies the importance of harnessing deep learning techniques to forge intelligent diagnostic algorithms within this domain. The prowess of Convolutional Neural Networks (CNNs) in deep learning, with their inherent capability for autonomous feature learning and the advantages presented by their local receptive fields in extracting features, has been recognized with accolades in computer vision. Presently, these methodologies are being zealously adopted in fault diagnosis. For example, Tang SN, along with associates, made notable advancements by applying the continuous wavelet transformation to morph raw vibration signals from hydraulic axial piston pumps into bidimensional time-frequency representations, subsequently employing profound CNN networks for the extraction of features and the diagnosis of faults.

Similarly, Zhu and his team<sup>9</sup>, after transforming signals into the time-frequency domain via wavelet changes, improved and stacked the classic CNN network, AlexNet, for fault diagnosis tasks. Jiang et al.<sup>10</sup>, replaced ordinary convolutional kernels with stochastic ones for feature extraction from the vibration signals of axial piston pumps. Another approach involves the direct use of one-dimensional convolutional methods, the advantage of which lies in their suitability for handling sequential data. This allows for capturing local patterns and long-range dependencies within sequences while preserving the sequential order information. Such methods have extensively applied in numerous natural language processing and time series analysis tasks<sup>11</sup>. Huang et al.<sup>12</sup>, have harnessed the capabilities of multi-layer, multi-channel 1D CNNs to devise a diagnostic model for hydraulic system components, leveraging multispeed data samples collected from various sensors. In pursuit of enhancing the reliability of deep neural networks for industrial fault diagnosis and condition monitoring, Keleko AT, among others<sup>13</sup>, has integrated the game-theoretic approach of Shapley Additive exPlanations into neural networks. This integration aims to furnish credible outcomes by elucidating each sensor's significant weights and contributions towards the DNN's decision-making process, a methodology whose efficacy has been corroborated through tests on hydraulic system data.

The aforementioned deep learning methodologies are adept at adaptively extracting fault characteristics, exhibiting robust fault classification capabilities. However, applying deep learning algorithms for fault diagnosis requires a substantial volume of samples, requiring single-label samples of various fault modes. In hydraulic systems, faults tend to be covert and intricate, often manifesting as composite and coupled. Therefore, the utilization of multi-sensor feature information fusion technology is advisable.

Integrating multi-sensor feature information technology is an emerging trend in fault diagnosis. It can generate more comprehensive information than individual sensors, thereby improving the robustness of fault diagnosis and addressing the issue of complex fault decoupling<sup>12</sup>. The focus of current research lies in how to extract features from time series data of different sensors and fuse the features from sensors with different sampling rates. The fundamental approach to feature information fusion is to utilize multiple channels for extracting feature information, where each channel independently extracts features based on dimensions such as time and frequency. Afterward, the information from these channels is integrated.

Thus, extracting multi-channel features from sensor arrays operating at various sampling rates is a pivotal challenge in diagnosing hydraulic system malfunctions.

The employment of temporal features derived from sensor data timestamps stands as a critical solution, with time series analysis extensively applied in detecting anomalies and predicting life expectancy in industrial monitoring data<sup>14</sup>. Methods utilizing Recurrent Neural Networks (RNN)<sup>15,16</sup>, predicated on the Markov assumption for modeling continuous time, and those employing Convolutional Neural Networks (CNN) on the temporal dimension to capture changes<sup>17,18</sup>, underscore the fundamental task of time series feature extraction: to distill both long-term characteristics and short-term features, mirroring the temporal data of industrial sensors in capturing the immediate variability and enduring trends of sensor data<sup>19,20</sup>.

Another critical aspect of diagnosing hydraulic system failures is the fusion of features from sensors operating at multiple sampling rates. Despite the formidable feature extraction capabilities of CNNs, critical information may be diminished when features within the model are directly amalgamated using maximum or average pooling.

Attention mechanisms have effectively solved this problem. For example, Song et al.<sup>21</sup> constructed a new deep transfer learning network model for planetary gearbox fault diagnosis by integrating one-dimensional convolutional neural networks, attention mechanisms, and domain adaptation methods; Zhao et al.<sup>22</sup> utilized attention mechanisms to share information between channels, and their constructed rotor system fault diagnosis model demonstrated excellent load adaptability and noise resistance; Cheng et al.<sup>23</sup> designed a lightweight and efficient channel attention mechanism (LECA) combined with one-dimensional convolution, achieving good gearbox fault diagnosis results; Zhang et al.<sup>24</sup> applied attention mechanisms in unsupervised fault diagnosis, developing a semi-supervised attention mechanism to solve the problems of insufficient fault diagnosis data and false alarms in online fault diagnosis.

This study proposes a hydraulic system fault diagnosis method based on temporal two-dimensional variation modeling and the fusion of multi-rate data samples. This method addresses the issues above, facilitating the decoupled diagnosis of fault-coupled hydraulic systems in critical components. The primary contributions of this paper are as follows:

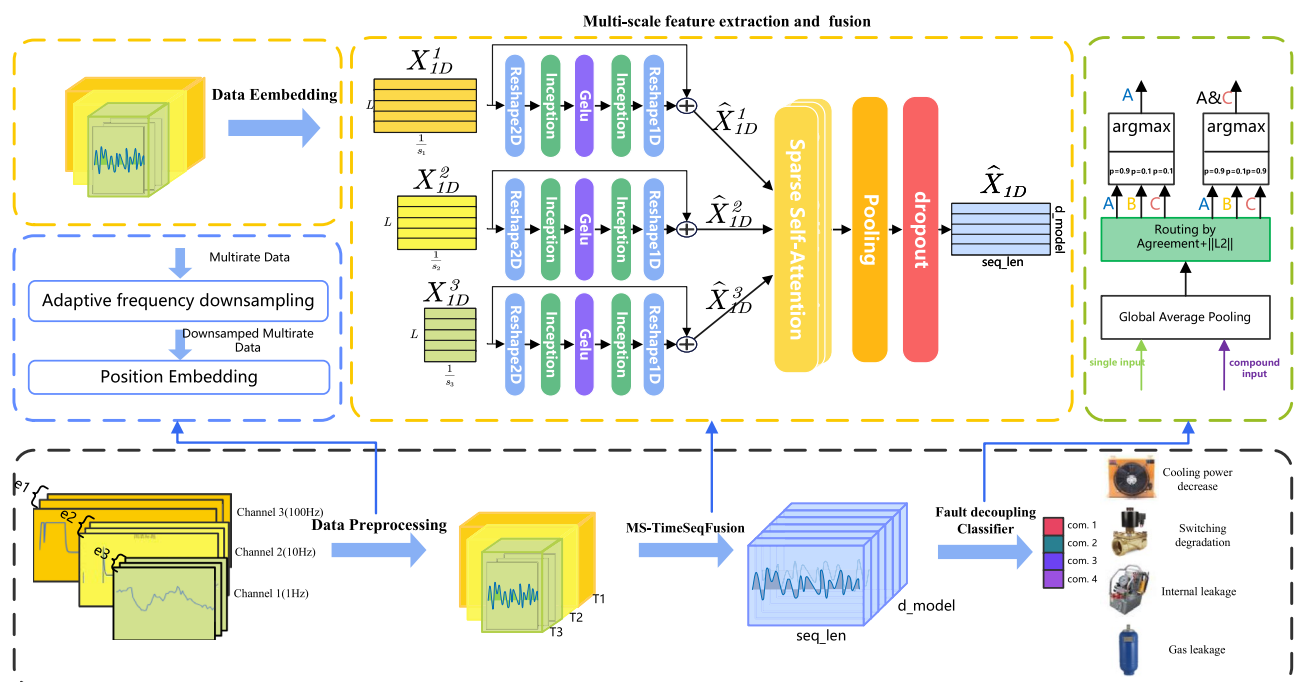
1. To mitigate the issue of oversampling in high-frequency sensors, a variable frequency data sampling method is introduced to streamline redundant high-frequency data.
2. In response to the strong concealment and high coupling of faults in hydraulic systems, a two-dimensional change modeling of sensor temporal data is employed, utilizing a two-dimensional convolution method to extract both the instantaneous variability and long-term trends of sensor data, overcoming the limitations of one-dimensional feature extraction methods.
3. The proposed feature fusion framework, based on the attention mechanism, automatically integrates features from multi-rate data samples, superseding the traditional approach of merging features within CNN networks through direct maximum or average pooling, thereby ameliorating the issue of crucial information loss.

## Methods

The process of diagnosing faults in multi-scale time series involves three stages, as illustrated in Fig. 1: Multi-rate data preprocessing, multi-scale time series feature extraction and fusion, and faulty decoupling and classification. During data preprocessing, high-frequency signals are resampled, and all signals are encoded with their sequence positions. For multi-scale feature extraction and fusion, sensor data with different sampling rates are transformed into 2D representations per channel. Convolutions are then employed to extract short-term and long-term period features from a 2D perspective. Feature fusion uses self-attention to focus on deep features and feature decoupling at various scales.

### Multi-rate data preprocessing

The extraction and fusion of data features from multiple sensors have always been topics of intense research. Different sensors measure various physical quantities, and their data features vary in scale. These data feature



**Figure 1.** Fault diagnosis process of multi-scale time series decoupling.

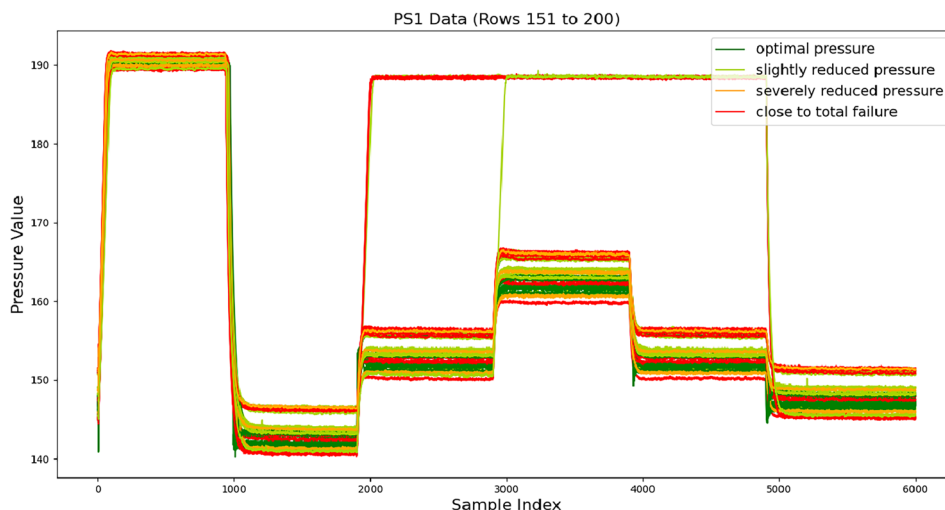
mappings provide insights into the conditions of various components. By utilizing data fusion methods, the accuracy of fault diagnosis can be enhanced. However, due to the multi-rate characteristics of measurement data and the different sampling frequencies of multiple sensors, it is impossible to process the input data in the same dimension. Therefore, it is necessary to preprocess the data. Upsampling and downsampling methods are not suitable for the current application scenario. For example, downsampling involves sampling in high-frequency data, and the detection performance of accumulator component faults relies on instantaneous variability.

Figure 2 demonstrates the data of pressure sensor PS1. It has been observed that high-frequency sensors, including pressure sensors, exhibit significant changes in data frequency and amplitude when malfunctioning in the hydraulic system. This type of data characteristic is common in mechanical systems, and if high-frequency data is downsampled, the information on these data changes may be lost. Therefore, analyzing time series data aims to capture the instantaneous changes during signal transients and the long-term trends of signal trends. It is generally difficult and ineffective to extract two features on a one-dimensional time series simultaneously. The sampling frequency of the pressure sensor is 100 Hz, which results in oversampling due to the relatively lower frequency of pipeline pressure changes. Additionally, the hydraulic platform simulation in this dataset encompasses testing variable work cycles with pseudo-random load variations and fixed work cycles with predetermined load levels. These scenarios represent typical cyclic operations and repetitive load characteristics found in industrial applications, as well as load fluctuations in mobile machinery that undergo local abrupt changes during operation. High-frequency sensors within hydraulic systems will exhibit significant alterations in both frequency and amplitude upon failure. This type of data feature is prevalent in mechanical systems; direct downsampling of high-frequency data may lead to a loss of crucial information regarding these data changes. Therefore, it is necessary to conduct variable-frequency sampling on high-frequency data and analyze time series data to capture instantaneous signal changes during transient conditions and long-term trends.

Industrial sensors include those with high sampling frequencies, such as pressure sensors for hydraulic systems, which have a sampling frequency of 100 Hz. However, the pressure variation in the hydraulic system does not reach this frequency, resulting in oversampled data. Traditional uniform downsampling methods have issues. Setting the sampling frequency too high can capture complete data information, but if there are many sensors and components, it generates a large amount of data, reducing the efficiency of fault diagnosis models. On the other hand, setting the sampling frequency too low requires more data to guarantee the accuracy of data collection, which may lead to the loss of valuable data, hindering the diagnosis of faults in hydraulic coupling systems under complex conditions.

The problem can be solved using variable-frequency data sampling. Variable-frequency data sampling has been used in data collection, where sensors analyze the variations in previously collected data and automatically adjust the time interval for data sampling. In the data processing stage, this method is used to downsample the collected data. In simple terms, variable-frequency data sampling adjusts the sampling frequency based on data changes.

This paper utilizes an integrated variable-frequency downsampling method based on the jitter ratio and swing door algorithm<sup>25</sup>. Firstly, the basic principle of the jitter ratio method involves fitting  $n$  sampling points into a straight line using least squares, and the distance between the data points and the line represents the fluctuation level of the data. The jitter ratio for  $n$  sampling points is defined as the ratio of the average jitter size between the closest  $m$  sampling points within a data window to the remaining  $m - n$  data points, denoted as  $\delta$ .



**Figure 2.** The first 50 sets of data from pressure sensor PS1.

$$\delta = \frac{\frac{1}{m} \sum_{k=i-m+1}^{i-m} |\beta_0 + \beta_1 t_k - y_k| \sin \theta}{\frac{1}{n-m} \sum_{k=i-m+1}^{i-m} |\beta_0 + \beta_1 t_k - y_k| \sin \theta} = \frac{(n-m) \sum_{k=i-m+1}^{i-m} |\beta_0 + \beta_1 t_k - y_k|}{m \sum_{k=i-m+1}^{i-m} |\beta_0 + \beta_1 t_k - y_k|} \quad (1)$$

The rotating gate algorithm constructs an expanding quadrilateral to store data outside the quadrilateral. It determines whether the data is inside the quadrilateral by comparing the slopes of the upper and lower rotating gates. If the upper slope is less than the lower slope, it means that the data is within the rotating gate, indicating a smooth data change. The frequency acquisition method based on the jitter ratio is more suitable for situations where the data change is relatively smooth, while the acquisition method based on the rotating gate algorithm is suitable for data acquisition processes with large data fluctuation.

To judge the severity of parameter changes using the jitter ratio, first collect historical data of the process at a relatively high frequency. Calculate the jitter ratio for each sampling point within an interval. Within a sequence of  $k$  consecutive jitter ratios, if the number of sampling points where  $\delta$  exceeds the threshold is  $h$ , and if  $h$  is greater than  $k/2$ , then the parameter is considered to have undergone significant changes during the production process; otherwise, the parameter is considered to have changed smoothly. For parameters with smooth changes, data collection is conducted using a method based on the jitter ratio, with slightly longer initial data collection intervals. For parameters with significant changes, data collection is conducted using a method based on the rotating gate algorithm, with slightly shorter initial data collection intervals. The formula is expressed as follows:

$$T_{next} = \begin{cases} \min(T + T_{inc}), \delta < \delta_1 \\ \min(T - T_{dec}), \delta > \delta_2 \\ T, \delta_1 \leq \delta \leq \delta_2 \\ \min(T + T_{inc}, T_{max}) \\ \max(T/2, T_{min}) \end{cases}, \text{ if } h > k/2 \quad (2)$$

In the formula,  $T$  is the initial sampling interval,  $T_{inc}$  and  $T_{dec}$  are the time interval step sizes for increase and decrease in jitter-based variable frequency sampling method,  $T_{max}$  and  $T_{min}$  are the upper and lower limits of sampling interval for the gate-based variable frequency sampling method.

When variable time sampling is used, it can cause a problem where the lengths of each time series are not consistent. In order to establish multi-scale information fusion, it is important to pad the time series lengths so that they match the original length of the sensor data.

A sine and cosine positional encoding technique is used to capture the temporal information in time series data. The values of sine and cosine are derived from the time intervals recorded with the variable frequency sampling mentioned previously. For an input observation sequence of a sensor  $\vec{o}_s = (o_1, o_2, \dots, o_T)$ ,  $o_i \in \mathbb{R}^f$ , it is mapped to a high-dimensional space as  $\vec{V} = (v_1, v_2, \dots, v_T)$ ,  $v_i \in \mathbb{R}^d$ . For the input sequence position vector  $p = (0, \dots, i, \dots, T)$ , where  $i$  is the position index of each sample in the sequence, it is mapped to the same high-dimensional space as  $\vec{P} = (p_1, p_2, \dots, p_T)$ ,  $p_i \in \mathbb{R}^d$ . The final output of an embedding layer for a scale is.

$$X_e = V_e + P_e \quad (3)$$

### Multi-scale time series feature extraction and fusion methods

Because a one-dimensional time series structure can only present changes between adjacent time points, frequency features of time series data are extracted through signal processing methods, and the time series is transformed into a two-dimensional space based on these frequency features. This explicit representation can reveal the long-term characteristics of trends and capture the short-term characteristics during transient periods, which is beneficial for subsequent representation learning. For each frequency, the 1D time series is transformed into a 2D space, and the Inception Block module with two-dimensional convolution is used to extract features for both types of changes simultaneously. As shown in Fig. 3, the top- $k$  periods obtained from the fast Fourier transform are the data's characteristic periods, and the two-dimensional convolution can extract both the short-term characteristics within each period and the long-term characteristics between periods. In the frequency domain, the Fast Fourier Transform (FFT) decomposes the frequencies of superimposed waves. In the resulting spectrum, frequencies with higher amplitudes are more significant.

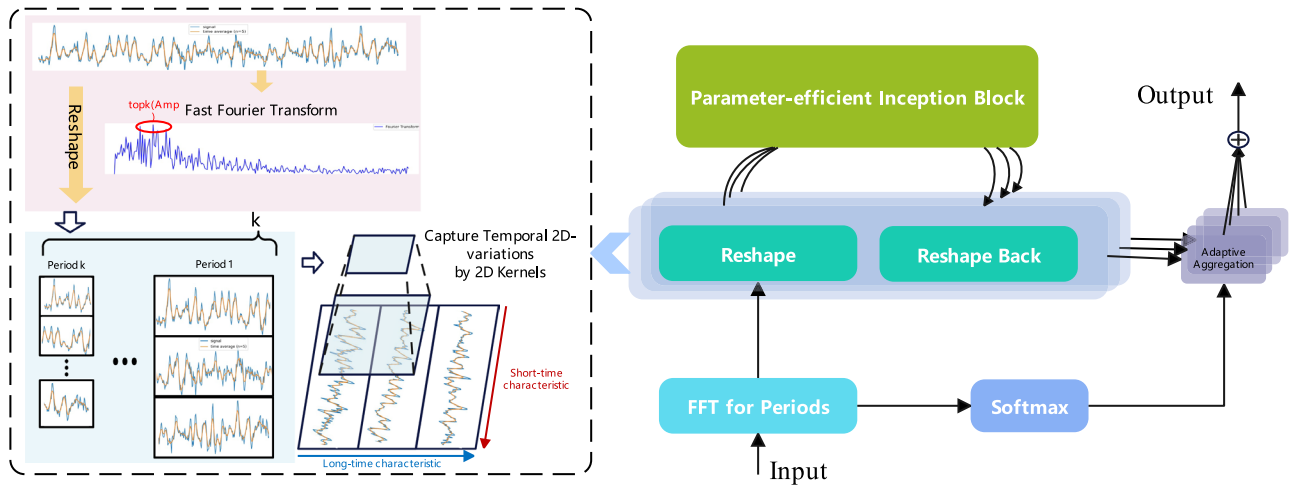
The Pan, X's research<sup>26</sup> has shown that convolution is advantageous for extracting shallow features, while attention mechanisms are more effective for deep features. Based on this, this paper performs convolution to extract features at various scales and then uses a self-attention mechanism for feature fusion.

The comprehensive methodology for multi-scale time series feature extraction and fusion is depicted in Fig. 1. The specific process of time series feature extraction is presented in Fig. 3. Detailed outputs and data transformations at each stage are illustrated in the Supplementary Material, the specific methods include the following steps.

#### Method to convert a one-dimensional time series to two-dimensional

Time series data encompasses both short-term fluctuations and long-term trends. For instance, in a hydraulic system, there can be randomly variable working loads as well as fixed working loads. During normal operations, the hydraulic system exhibits stable long-term load changes, whereas different operational environments or fault conditions introduce random short-term load variations. Modeling these two types of features using one-dimensional time series data is challenging because the original one-dimensional structure only captures changes between adjacent time points. To address this, the time series is first transformed from the time domain to the frequency domain to capture its periodicity. This frequency domain information is then used to create a 2D representation.





**Figure 3.** 2D feature extraction methods.

For a given time series with a length of  $T$  and  $C$  dimensions, the initial one-dimensional representation is  $X_{1D} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_C)$ . Assuming the input data consists of  $C$  channels, each with a sampling rate of  $s$ , and each channel contains  $T$  sensors, with the top- $k$  parameter set to  $k$ .

For the  $j$ -th channel, there are  $T_j$  sensors. The input time series data for this channel is represented as  $(X_{1D})_j = (X_{1D}^1, X_{1D}^2, \dots, X_{1D}^T, \dots, X_{1D}^{T_j})$ . To represent the multi-periodic changes, it is necessary to first capture the different frequencies of the time series. This can be achieved by analyzing the time series in the frequency domain using the Fast Fourier Transform (FFT), as shown below:

$$\begin{aligned} X_{1D}^l &= \text{Embedding}(X_{1D}^t) \\ A &= \text{Mean}\left(\text{Amp}\left(\text{FFT}\left(X_{1D}^l\right)\right)\right), \\ \{f_1, \dots, f_k\} &= \text{Topk}(A), p_i = \left\lfloor \frac{T}{f_i} \right\rfloor, i \in \{1, \dots, k\} \end{aligned} \quad (4)$$

In the equation,  $X_{1D}^t$  represents the raw data from the sensor, mapped to the input one-dimensional time series data. Embedding(.) denotes a linear layer mapping, Amp(.) represents the calculation of amplitude.  $A$  represents the amplitude for each frequency, which is averaged over the first dimension using Mean(.). Top- $k$  represents selecting the  $k$  frequencies with the largest amplitudes from  $A$ , corresponding to a period of  $p$ . Considering the sparsity in the frequency domain and the common occurrence of noise in industrial systems, only the first  $k$  amplitude values are chosen, and the highest effective frequencies are obtained, denoted as  $\{A_{f_1}, A_{f_2}, \dots, A_{f_k}\}$  and  $\{f_1, f_2, \dots, f_k\}$ , respectively, where  $k$  is a hyper-parameter.

Based on the selected frequencies  $\{f_1, f_2, \dots, f_k\}$  and their corresponding periods  $\{p_1, p_2, \dots, p_k\}$ , the 1D time series  $X_{1D}^l$  can be transformed into multiple 2D tensors  $X_{2D}^{l,i}$  using the following equation:

$$X_{2D}^{l,i} = \text{Reshape}_{p_i, f_i}(\text{Padding}(X_{1D}^l)), i \in \{1, \dots, k\} \quad (5)$$

Padding is to extend the time series along the time dimension with zeros in order to match its period  $p_i$  with the reshaped 2D tensor, where  $p_i$  and  $f_i$  represent the number of rows and columns of the transformed 2D tensor, respectively.  $X_{2D}^{l,i}$  represents the 2D features of the  $i$ -th time series based on frequency  $f_i$ , with its columns and rows representing changes within the period of length  $p_i$  and changes between periods, respectively. Ultimately, a set of 2D tensors  $\{X_{2D}^{l,1}, \dots, X_{2D}^{l,k}\}$  are obtained based on the selected frequency and estimated periods, representing  $k$  different temporal changes derived from different periods  $p_i$ .

The reshaping process introduces two categories of data to the transformed 2D tensor: the interplay of instantaneous variations (columns, changes occurring within a specific timeframe) and long-term features (rows, changes spanning across timeframes). Consequently, the spatial fluctuations of time-dependent modifications can be effectively handled by 2D kernels.

#### Parameter-efficient inception block

Sequence information representation can be easily obtained by using parameter-efficient Inception blocks, this method reduces the number of parameters required for 2D convolution in each channel, defined as

$$\hat{X}_{2D}^{l,i} = \text{Inception}\left(X_{2D}^{l,i}\right) \quad (6)$$

This approach comes from the visual task mechanism<sup>27</sup>, as the time series data has been transformed into 2D, making it possible to utilize feature extraction methods from the visual domain. Here,  $\text{Trunc}(\cdot)$  is used to restore the sequence  $\hat{X}_{2D}^{l,i}$  to its original 1D dimension:

$$\hat{X}_{1D}^l = \text{Trunc}\left(\text{Reshape}_{1, (p_i \times f_i)}\left(\hat{X}_{2D}^{l,i}\right)\right), i \in \{1, \dots, k\} \quad (7)$$

#### Multi-scale feature fusion and decoupling

The above method achieves feature extraction for time-series data with varying rates and sampling frequencies, addressing the issue of input data that cannot be processed in the same dimension. A modified Time Block is employed to develop a multi-scale feature extraction method, with hyperparameters specifically designed for high-frequency and low-frequency data, respectively.

As shown in Fig. 1, in order to effectively utilize temporal information, the data is input into the multi-scale network after being encoded with sine and cosine position encoding. After feature extraction, each channel will have an output sequence denoted as  $X_{1D}^l(e)$ .

Please refer to Tables A1–A3 in the Appendix. Through correlation analysis, it is discovered that hydraulic data contains interrelated redundant information. By leveraging the advantages of the self-attention mechanism, the diagnostic model automatically integrates the features from different channels. The traditional self-attention mechanism is characterized by high time and space complexity. Studies have found that the attention weight distributions for different query-value pairs are not all highly focused, with some approaching a uniform distribution, which is referred to as lazy distribution, while the parts with higher focus are called active distribution. To improve computational efficiency, a probabilistic sparse self-attention mechanism is adopted<sup>28</sup>.

The standard self-attention mechanism is defined based on tuple inputs (i.e., query, key, and value). It performs scaled dot-product operation  $A(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$ , where  $Q \in \mathbb{R}^{L_Q \times d}$ ,  $K \in \mathbb{R}^{L_K \times d}$ ,  $V \in \mathbb{R}^{L_V \times d}$ , and  $d$  refers to the input dimension. To further discuss the self-attention mechanism, let  $q_i, k_i, v_i$  respectively represent the  $i$ -th row of  $Q, K, V$ . The attention for the  $i$ -th query is defined in the form of probability:

$$A(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_j k(q_i, k_j)} v_j = E_{p(k_j|q_i)}[v_j] \quad (8)$$

From Eq. 8, the attention value of the  $i$ -th query to all keys is defined as the probability  $p(k_j|q_i)$ , and the output is the dot product with its corresponding value  $v$ . The dominant dot product shifts the attention probability distribution of the corresponding query away from a uniform distribution. The “similarity” between distributions  $p$  and  $q$  can be used to distinguish important or unimportant queries. Through the Kullback-Leibler divergence  $KL(q||p) = \ln \sum_{l=1}^{L_K} e^{q_i k_l^T / \sqrt{d}} - \frac{1}{L_K} \sum_{l=1}^{L_K} q_i k_l^T / \sqrt{d} - \ln L_K$ . Removing constants, the sparsity measure of the  $i$ -th query is defined as

$$M(q_i, K) = \ln \sum_{l=1}^{L_K} e^{\frac{q_i k_l^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{l=1}^{L_K} \frac{q_i k_l^T}{\sqrt{d}} \quad (9)$$

The first term is the Log Sum Exponential (LSE) of all keys  $q_i$ , and the second term is their arithmetic mean. If the  $M(q_i, K)$  of the  $i$ -th query is greater, the probability  $p$  of its attention is more “diverse”, and there is a high chance of including dominant dot products in the head fields of the long-tailed self-attention distribution.

Based on the proposed measurement, probability-sparse self-attention can be achieved by allowing each key to only attend to  $u$  dominant queries.

$$A(Q, K, V) = \text{Softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (10)$$

where  $\bar{Q}$  is a sparse matrix with the same size as  $q$ , and it only contains the top  $u$  scores within the sparsity measure  $M(q, K)$ . We control the value of  $u$  by a constant sampling factor  $c$ , setting  $u = c \cdot \ln L_Q$ . This ensures that ProbSparse self-attention only requires computing dot products between  $\bar{Q}$  and  $K$  with a complexity of  $\mathcal{O}(\ln L_Q)$ .

From a multi-head perspective, attention generates different sparse query-key pairs for each head, focusing on different features and their representations in different subspaces. In other words, the output of multi-head attention is the decoupled characteristics of multi-component composite faults.

#### Fault decoupling classification method

To decouple the multi-head output features of self-attention into different single-fault classifications, a multi-classification head is needed.

First, for the matrix  $y = [y_1, y_2, \dots, y_{K_i}]^T \in \mathbb{R}^{K_i \times R_i}$ , where  $y_i \in \mathbb{R}^{R_i \times 1}$  represents the feature (vector) at time point  $i$  outputted by self-attention, and  $K_i$  is the total length of the feature sequence.

Then, Eq. 11 utilizes global average pooling (*gap*) and *reshape* to transform the output matrix  $y$  into a one-dimensional vector. This vector is then mapped to the output layer using a fully connected layer (*projection*), resulting in the final prediction vector  $y_{pred} = [y^1, y^2, \dots, y^H]^T \in \mathbb{R}^{H \times 1}$ . The magnitude of each component  $y^i$  in  $y_{pred}$  represents the probability of the input sample belonging to the  $i$ th class. When the magnitude of  $y^i$  is close to 1, the probability of the predicted class is higher.

$$y_{pred} = \text{projection}(\text{reshape}(\text{gap}(y))) \quad (11)$$

Finally, as shown in Eq. 12, a threshold value  $\varphi$  is chosen to limit the number of predicted labels for classification. If  $y^i$  is greater than the selected threshold, the classifier outputs a label of 1 for the  $i$ th class. For reliable classification results, good predictions should have  $y_i$  approaching 1 for the corresponding actual class and approaching 0 for other classes.

$$y_{output}^i = \begin{cases} 1, & y_{pred}^i \geq \varphi \\ 0, & y_{pred}^i < \varphi \end{cases} \quad (12)$$

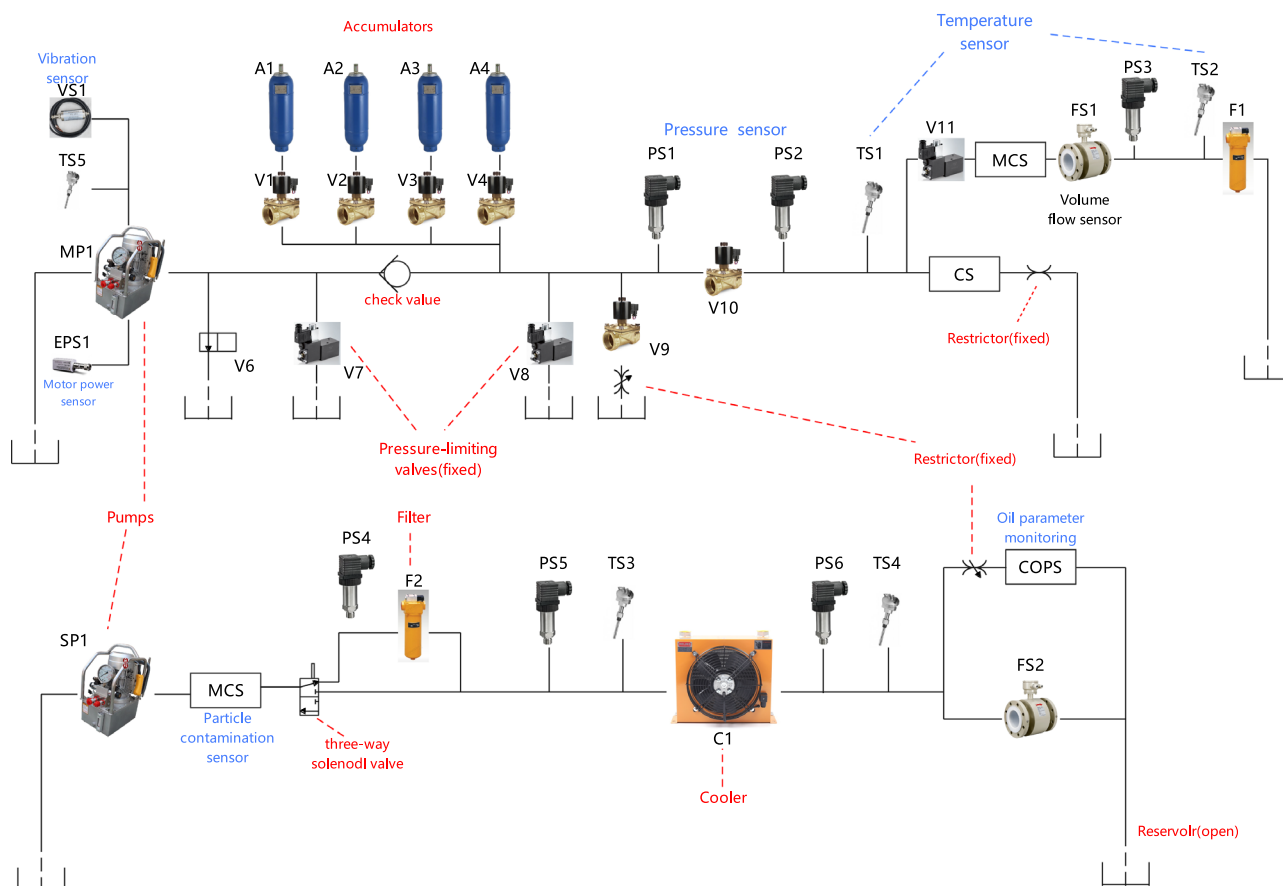
## Experimental verification

### Dataset introduction

To validate the effectiveness of the designed model, several experiments will be conducted in this section. The fault diagnosis experimental data is sourced from an open dataset<sup>7</sup>.

The hydraulic engineering test bench, as shown in Fig. 4, consists of two hydraulic circuits connected through an oil tank: the primary working circuit and the secondary cooling and filtering circuit. The red-labeled components in the figure represent the critical components of the hydraulic system, while the blue labels represent sensors. The primary working circuit is the upper part of the oil circuit in Fig. 4, mainly composed of the leading pump MP1, switchable accumulators A1–A4, and filter F1. The proportional pressure relief valve (V11) generates different load levels in the main pump working circuit to simulate different conditions. The auxiliary cooling and filtering circuit is responsible for cooling and filtering the hydraulic oil, composed of a hydraulic pump SP1, three-way solenoid valve, filter F2, cooler C1, and various sensors. The micro-particle metal contamination sensor (MCS) detects the degree of hydraulic oil pollution. It determines the position detection results of the solenoid valve spool, selecting the appropriate detection method. In the figure, the solenoid valve spool is connected to the first interface, and the hydraulic oil flows through the filter to reach the condenser. If the solenoid valve slide valve is connected to another interface, the hydraulic oil will flow to the condenser without filtration.

Several types of faults can occur in the hydraulic system, with different fault types, severity, and duration. In cooler C1, the degradation level of cooling power (CP) is simulated by changing the cooling fan duty cycle. In valve V10, different switching degradation states are simulated by changing the valve current setpoint. The internal leakage level of the main pump MP1 is caused by three cascaded 0.2 mm and 0.25 mm orifices. Different leakage levels are simulated by changing the pre-charge pressure step A1–A4. The breakdown performance of



**Figure 4.** Structure of hydraulic system: primary working circuit and secondary cooling-filtration circuit.



different components varies. Fault characteristics such as CP reduction and valve switching degradation are obvious, making the diagnosis of these components relatively easy. However, the fault characteristics of the hydraulic pump and accumulator are more covert and easily influenced by environmental noise. Therefore, it is necessary to adopt advanced diagnostic methods to identify the faults of these components. In addition, the samples are collected under different conditions. Some data samples are taken under stable conditions (1449 samples), while others are recorded when static conditions may not yet have been reached (756 samples). The experiment uses data from these two situations, which may result in class imbalance and increase the difficulty of fault detection.

Multiple sensors are used to monitor the status of the hydraulic system. As summarized in Table 1, the physical sensors include pressure sensors (PS1–PS6), electric power sensor (EPS1), flow sensors (FS1, FS2), temperature sensors (TS1–TS5), and vibration sensor (VS1). In addition, multiple virtual sensors are utilized to provide references for the system, such as cooling efficiency (CE), CP, and system efficiency factor (SE). The data is collected on a programmable logic controller (PLC) and then transmitted to the computer through EtherCAT. The sampling rate of these sensors varies according to the dynamics of the physical quantities.

To validate the effectiveness of the model, several experiments will be conducted. Firstly, fault diagnosis of four components will be performed: cooler C1, hydraulic valve V10, main hydraulic pump MP1, and accumulators A1–A4. This will test the model’s ability to decouple single-component fault patterns from complex nested data. An exemplary fault diagnosis algorithm should be able to identify whether a component has a fault and determine the category and severity of the fault. Secondly, as the dataset is collected in a laboratory setting, it represents relatively ideal data without the complex interference of actual industrial conditions. Therefore, experiments in a low signal-to-noise ratio environment will be designed, incorporating Gaussian white noise of different decibels to test the model’s noise resistance.

In the single-component fault diagnosis experiments, X represents the data samples with multiple sampling rates, and Y represents the labels. One-hot encoding is used to diagnose a single fault in different components. Specifically,

For the faults in cooler C1:

$$Y \in \{\text{Full efficiency, Efficiency reduction, Close to complete failure}\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For hydraulic valve V10:

$$Y \in \{\text{Optimal switching behavior, Slight lag, Severe lag, Close to complete failure}\} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For main hydraulic pump MP1:

$$Y \in \{\text{No leakage, Weak leakage, Severe leakage}\} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For accumulators A1–A4:

$$Y \in \{\text{Optimal pressure, Mild pressure reduction, Severe pressure reduction, Close to complete failure}\} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

All labels of the training dataset for single-component fault diagnosis are single labels, which can be used as supervised indicators to calculate the marginal loss and supervise the model’s learning of distinguishing features. In the given hydraulic system dataset, there are 566 samples with only one type of component fault occurring, 688 samples with two types of component faults occurring simultaneously, 612 samples with three types, and 360 samples with all four components having faults.

	Sensor	Physical quantity	Unit	Sampling rate
Physical sensors	PS1–PS6	Pressure	Bar	100 Hz
	FS1–FS2	Volume flow	l/min	10 Hz
	TS1–TS4	Temperature	°C	1 Hz
	EPS1	Motor power	W	100 Hz
	VS1	Vibration	mm/s	1 Hz
Virtual sensors	CE	Cooling efficiency	%	1 Hz
	CP	Cooling power	kW	1 Hz
	SE	Efficiency factor	%	1 Hz

Table 1. Monitored parameters of the hydraulic system.

Table 2 summarizes the hyper-parameters used in the proposed method for this experiment, including batch\_size, top\_k frequencies selected post FFT application, learning\_rate, train\_epochs, loss function, multi-scale channels “Multi-scale”, seq\_len\_list parameter indicating various time series data lengths in the dataset, e\_layers and d\_layers specifying encoder and decoder layers within the attention mechanism respectively. Additionally, it includes parameters such as d\_model representing input dimension to the attention mechanism from multi-scale feature extraction output feature dimension.

Single fault diagnosis of different components

According to the three sampling rates of the experimental dataset, set the number of multi-scale channels of the time series feature extractor to 3. Table 2 summarizes the hyper-parameters of the proposed method in this experiment.

To verify the effectiveness of the proposed method, several methods are introduced for comparison. First, compare with classical fault diagnosis methods, namely Linear Discriminant Analysis (LDA)<sup>7</sup>, 1D Convolutional Neural Network (1D-CNN), and Multi-channel 1D Convolutional Network<sup>29</sup>; Secondly, compare with the Time Convolutional Network (TCN) in time series methods<sup>30</sup>; Then, compare with the cutting-edge Transformer model constructed time series method. The experimental results are recorded in Table 3, and the changes in training loss and validation accuracy during the training process are shown in Figs. 5 and 6, respectively. The model’s convergence speed during the training process is significantly higher than the other models.

In order to display the results more intuitively and provide a misclassification analysis, the confusion matrix for the detection of four types of individual components is shown in Fig. 7. The vertical axis represents the true labels of the samples, while the horizontal axis represents the predicted labels of the samples. In the fault diagnosis of the cooler and hydraulic valve, all samples are classified into their respective categories. Although in the fault detection of the main pump and accumulator, only a few samples are misclassified into adjacent categories. The described misclassifications reveal that the similarity of sensor data between continuously degrading states may lead to confusion in fault recognition<sup>31</sup>.

Hyper-parameters	Setting values
batch_size	32
top_k	3
learning_rate	0.001
train_epochs	30
loss	MSE
Multi-scale	3
seq_len_list	60/600/6000
e_layers	3
d_layers	1
d_model	64
d_ff	64
d_pooling	8
d_output	3/4

Table 2. List of model parameter settings.

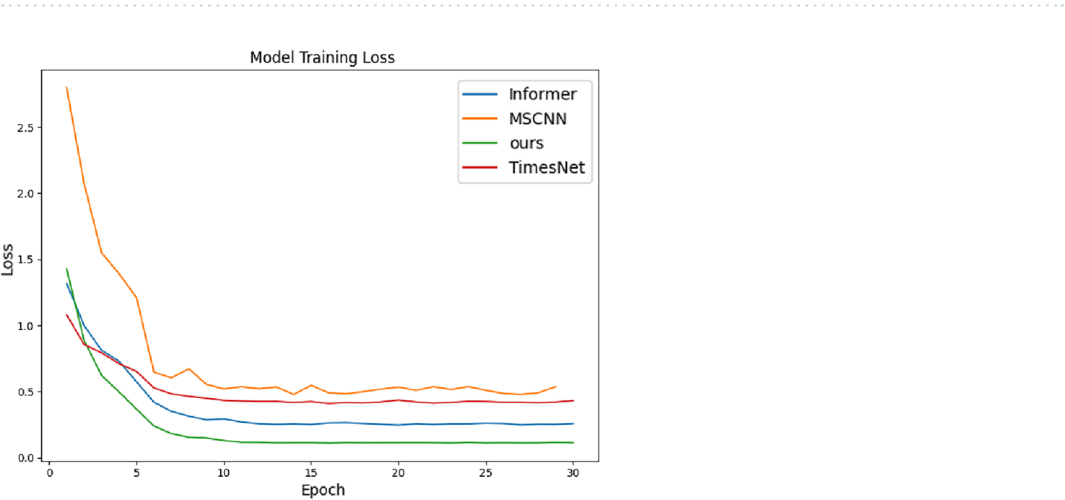
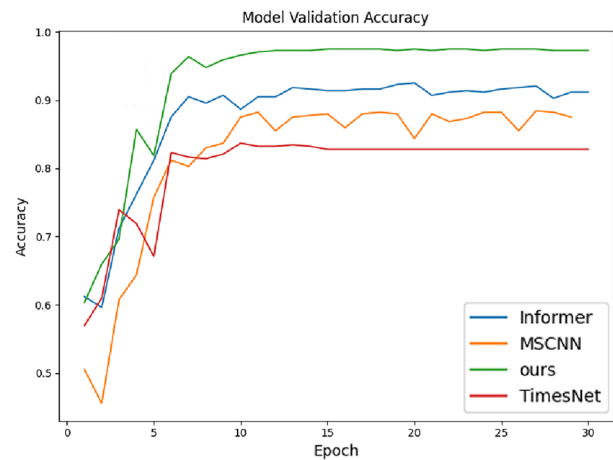


Figure 5. Loss changes in 30 iterations of four fault diagnosis methods in accumulator component diagnosis.



**Figure 6.** Accuracy changes of 30 iterations of four fault diagnosis methods in accumulator component diagnosis.

Component		Classical methods		TCN	Transformers		Ours
		LDA	CNN	–	Informer	TimesNet	
Cooler	Accuracy	100%	100%	100%	100%	100%	<b>100%</b>
	Precision	100%	100%	100%	100%	100%	<b>100%</b>
	Recall	100%	100%	100%	100%	100%	<b>100%</b>
	F1 (weighted)	100%	100%	100%	100%	100%	<b>100%</b>
Valve	Accuracy	93.88%	94.78%	46.71%	95.92%	95.69%	<b>100%</b>
	Precision	93.86%	95.05%	21.82%	95.94%	95.77%	<b>100%</b>
	Recall	93.88%	94.78%	46.71%	95.92%	95.69%	<b>100%</b>
	F1 (weighted)	93.70%	94.55%	29.75%	95.83%	95.56%	<b>100%</b>
Pump	Accuracy	98.64%	97.27%	57.60%	99.77%	<b>100%</b>	99.77%
	Precision	98.72%	97.39%	56.47%	99.78%	<b>100%</b>	98.42%
	Recall	98.64%	97.27%	57.60%	99.77%	<b>100%</b>	98.42%
	F1 (weighted)	98.64%	97.27%	47.42%	99.77%	<b>100%</b>	98.41%
Acc.	Accuracy	68.71%	72.78%	55.78%	91.38%	83.67%	<b>97.27%</b>
	Precision	68.30%	71.27%	47.18%	91.47%	84.23%	<b>97.06%</b>
	Recall	68.71%	72.79%	55.78%	91.38%	83.67%	<b>97.05%</b>
	F1 (weighted)	68.60%	71.40%	48.81%	91.39%	83.58%	<b>97.03%</b>

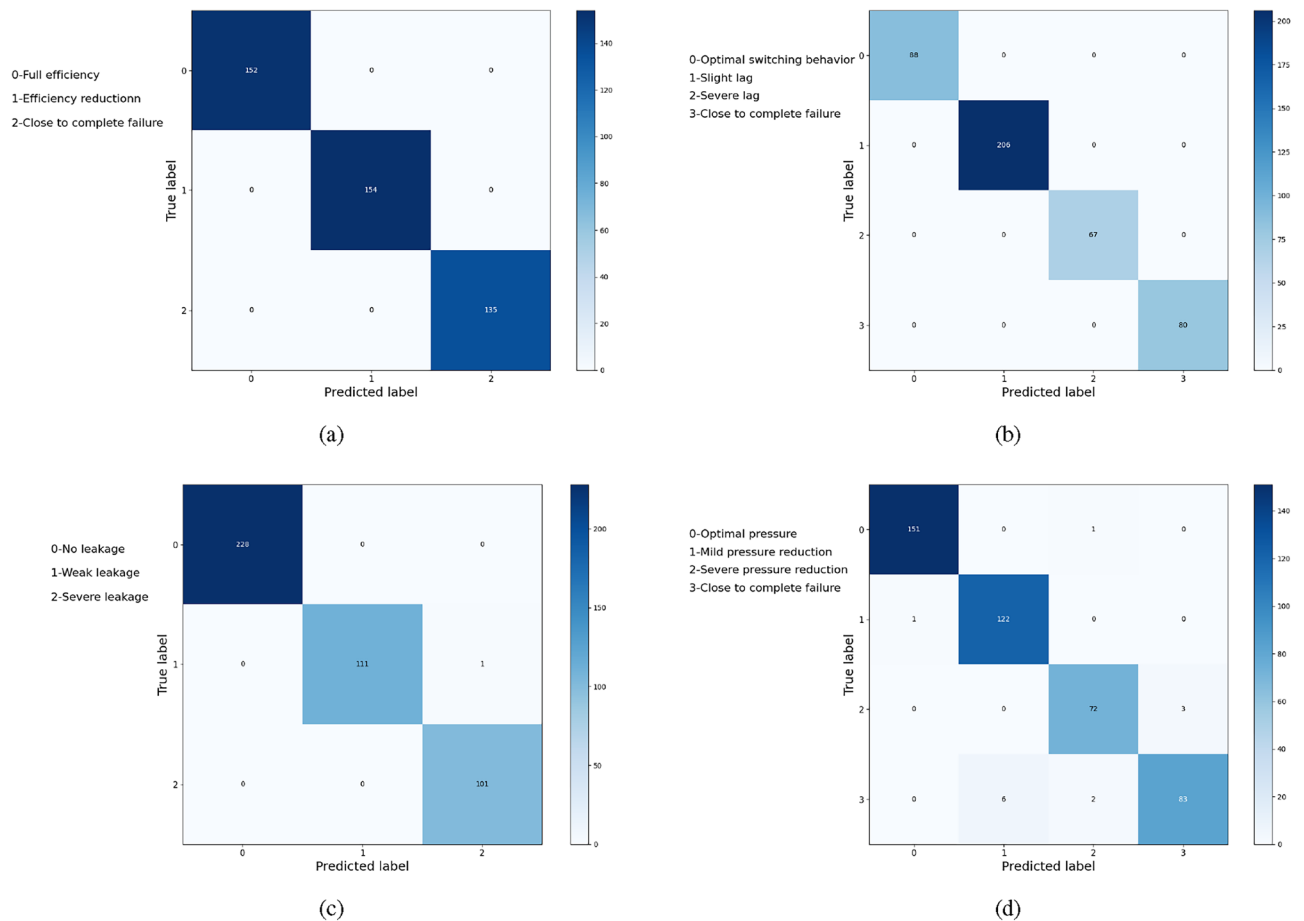
**Table 3.** Classification accuracy of fault diagnosis in different components. Bold indicates the best value for the indicator in that row.

Performance under noise environment

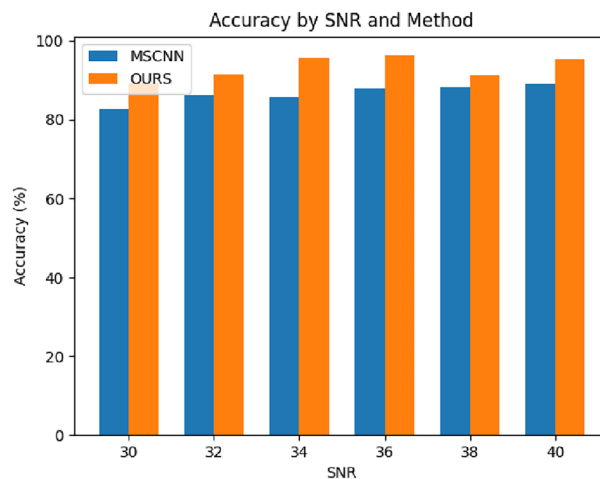
To better match real industrial control scenarios and test the model’s resistance to noise, and to verify the proposed method’s ability to suppress noise, Gaussian white noise with different signal-to-noise ratios is added to the selected fault signals. The signal-to-noise ratio is calculated as follows:

$$\text{SNR} = 10 \cdot \lg \left( \frac{P_s}{P_n} \right) \tag{13}$$

In the equation,  $P_s$  represents the power of the signal, and  $P_n$  represents the power of the added noise. As shown in Fig. 8, in the experiment, the SNR is set to increase from 30 to 40, which means the power of the noise is set to increase from 0.1 to 1% of the signal power. Generally speaking, the accuracy of fault detection increases with the increase of signal-to-noise ratio, and the noise environment has different effects on the diagnosis results of different components. In the fault detection of the cooler and valve, the influence of noise on the results is very small, and the diagnostic accuracy is close to 100.0%. However, in the fault diagnosis of the pump and accumulator with more complex fault scenarios, the noise environment has a significant impact on the diagnostic accuracy. With the increase of signal-to-noise ratio, the accuracy of this method approaches to the accuracy in a noise-free environment. In hydraulic system fault diagnosis, the accuracy of this method only decreases slightly, indicating its robustness.



**Figure 7.** Confusion matrix under different components: (a) cooler, (b) valve, (c) pump, (d) accumulator.



**Figure 8.** Accuracy of hydraulic accumulator fault diagnosis under different SNR.

### Visualization analysis with t-SNE

To gain an intuitive understanding of the effectiveness of the proposed method, t-distributed stochastic neighbor embedding (t-SNE) was used in experiments to visualize the representations of test samples in a two-dimensional space. t-SNE is a non-linear dimensionality reduction technique, easy to optimize, and particularly suitable for visualizing high-dimensional datasets.

The accumulator is a critical energy storage component of the hydraulic system, and its failure may lead to the collapse of the entire system, with diagnostic difficulties being the most complex. In the experiment,

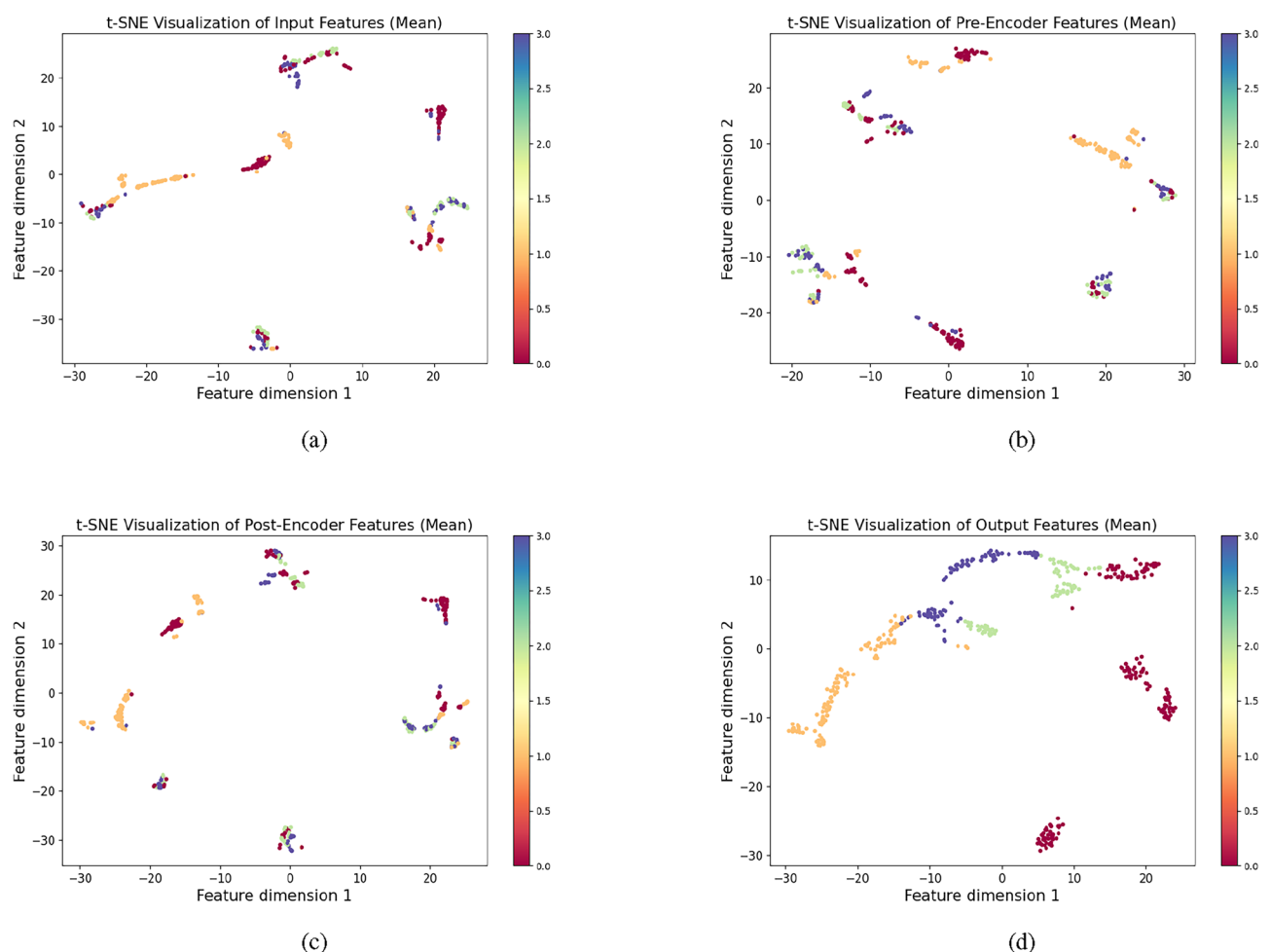
visualization analysis of accumulator fault diagnosis was conducted. Firstly, test samples were fed into a trained model, and then features were extracted from the feature extractor and classifier. Subsequently, t-SNE was used to transform high-dimensional features into two-dimensional coordinates. Based on these coordinates, scatter plots were drawn and displayed in Fig. 9.

As shown in Fig. 9a, points from different categories overlap with each other, indicating they cannot be simply separated by linear divisions. In contrast, features learned by the feature extractor exhibit better separability. Separation of data samples close to complete failure from samples of other categories indicates that the representations learned by the feature extractor facilitate preliminary distinction of fault categories. In the final Fig. 9d, data points from different categories are separated from each other, indicating that the method possesses strong non-linear mapping capabilities and can effectively identify faults.

## Results and discussion

This article proposes a new method for gear fault diagnosis based on a lightweight channel attention mechanism and transfer learning. The method addresses the problems of feature extraction, feature fusion, and fault decoupling for multi-sample rate sensor time series data and solves the issues of hidden and compound faults in hydraulic systems. The hydraulic system dataset is used to validate the classification and generalization ability of the proposed model. The conclusions are as follows:

1. This method has been proven to achieve accuracies of 100%, 100%, 100%, 99.77%, and 97.27% on the cooler C1, hydraulic valve V10, main hydraulic pump MP1, and accumulator A1–A4, respectively. Compared to classical methods and TCN, this method has absolute advantages in diagnosing the main hydraulic pump MP1 and accumulator A1–A4 datasets. It can extract more detailed features and effectively perform fault diagnosis. Compared to the latest Transformers time series models, it still has an advantage of around 5–10%.
2. Noise experiments show that this method has optimal diagnostic performance and generalization ability for different hydraulic components. It can achieve fast and accurate classification of multi-component faults in hydraulic systems. This is of great significance for solving compound fault problems in practical engineering applications.



**Figure 9.** Use t-SNE to visualize the distribution changes of sample features: (a) Original signal, (b) Multi-scale feature extract signal, (c) Feature fusion signal, (d) Fault decoupled signal.



In industrial environments, the differences in operating conditions and the issue of sample imbalance are more prominent. Future work will further explore the following aspects: (1) Expand the application scope, enhance the model's generalization ability, and attempt to achieve condition adaptation through transfer learning methods. (2) Further research is needed to improve models in unlabeled industrial datasets, realizing semi-supervised or unsupervised learning while maintaining high accuracy.

## Data availability

The fault diagnosis experimental data is sourced from an open dataset. The data that support the findings of this study are available from [doi: 10.1109/I2MTC.2015.7151267]. The datasets analyzed during the current study are available in the Kaggle repository, [<https://www.kaggle.com/datasets/jjacostupa/condition-monitoring-of-hydraulic-systems/data>].

Received: 25 March 2024; Accepted: 2 July 2024

Published online: 07 July 2024

## References

- Dai, J., Tang, J., Huang, S. & Wang, Y. Signal-based intelligent hydraulic fault diagnosis methods: Review and prospects. *Chin. J. Mech. Eng.* <https://doi.org/10.1186/s10033-019-0388-9> (2019).
- Yang, Y., Ding, L., Xiao, J., Fang, G. & Li, J. Current status and applications for hydraulic pump fault diagnosis: A review. *Sensors* <https://doi.org/10.3390/s22249714> (2022).
- Zhao, Z., Wang, F.-L., Jia, M.-X. & Wang, S. Intermittent-chaos-and-cepstrum-analysis-based early fault detection on shuttle valve of hydraulic tube tester. *IEEE Trans. Ind. Electron.* **56**, 2764–2770. <https://doi.org/10.1109/TIE.2009.2020710> (2009).
- Ma, Q., Song, D. & Liu, B. Fault mode analysis and simulation verification of hydraulic system based on amesim. *J. Phys. Conf. Ser.* **2006**, 012013. <https://doi.org/10.1088/1742-6596/2006/1/012013> (2021).
- Qiaolian, D. & Kehua, Z. Condition monitoring and fault diagnosis of hydraulic pump based on inherent vibration signals. *J. Agric. Eng.* (2007) (in Chinese).
- Qiu, Z., Yuan, X., Wang, D., Fan, S. & Wang, Q. Physical model driven fault diagnosis method for shield machine hydraulic system. *Measurement* <https://doi.org/10.1016/j.measurement.2023.113436> (2023).
- Helwig, N., Pignatelli, E. & Schuetze, A. Condition monitoring of a complex hydraulic system using multivariate statistics. In *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE Instrumentation and Measurement Technology Conference. 32nd Annual IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Pisa, Italy, May 11–14*, 210–215 (IEEE, 2015).
- Lei, Y. et al. Fault diagnosis method for hydraulic directional valves integrating PCA and xgboost. *Processes* <https://doi.org/10.3390/pr7090589> (2019).
- Zhu, Y. et al. Intelligent fault diagnosis of hydraulic piston pump based on wavelet analysis and improved alexnet. *Sensors* <https://doi.org/10.3390/s21020549> (2021).
- Jiang, W., Ma, L., Zhang, P., Zheng, Y. & Zhang, S. Anomaly detection of axial piston pump based on the dtw-rck-if composite method using pressure signals. *Appl. Sci.-Basel* <https://doi.org/10.3390/app132413133> (2023).
- Wang, K. et al. A hybrid deep learning model with 1dcnn-lstm-attention networks for short-term traffic flow prediction. *Phys. A Stat. Mech. Appl.* <https://doi.org/10.1016/j.physa.2021.126293> (2021).
- Huang, K., Wu, S., Li, F., Yang, C. & Gui, W. Fault diagnosis of hydraulic systems based on deep learning model with multirate data samples. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 6789–6801. <https://doi.org/10.1109/TNNLS.2021.3083401> (2022).
- Keleko, A. T., Kamsu-Foguem, B., Ngouna, R. H. & Tongne, A. Health condition monitoring of a complex hydraulic system using deep neural network and deepshap explainable xai. *Adv. Eng. Softw.* <https://doi.org/10.1016/j.advengsoft.2022.103339> (2023).
- Fu, S. et al. A time series anomaly detection method based on series-parallel transformers with spatial and temporal association discrepancies. *Inf. Sci.* <https://doi.org/10.1016/j.ins.2023.119978> (2024).
- Li, J., Liu, Y. & Li, Q. Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method. *Measurement* <https://doi.org/10.1016/j.measurement.2021.110500> (2022).
- Wang, H.-K., Zhang, Y. & Huang, M. A conditional random field based feature learning framework for battery capacity prediction. *Sci. Rep.* <https://doi.org/10.1038/s41598-022-17455-x> (2022).
- Franceschi, J.-Y., Dieuleveut, A. & Jaggi, M. Unsupervised scalable representation learning for multivariate time series. In Walach, H. et al. (eds.) *Advances in neural information processing systems 32 (NIPS 2019)*, vol. 32 of *Advances in Neural Information Processing Systems. 33rd Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada, Dec 08–14, 2019* (2019).
- Park, J., Park, Y. & Kim, C.-I. Tcae: Temporal convolutional autoencoders for time series anomaly detection. In *2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN)*, International Conference on Ubiquitous and Future Networks, 421–426, (IEEE; IEEE Commun Soc; IEICE Commun Soc; Korean Inst Commun & Informat Sci; Elect & Telecommunicat Res Inst; Samsung Elect; Lg Elect; SK Telecom; KT; LGU+, 2022). 13th International Conference on Ubiquitous and Future Networks (ICUFN), Electr Network, Jul 05–08, 2022
- Lai, G., Chang, W.-C., Yang, Y. & Liu, H. Modeling long- and short-term temporal patterns with deep neural networks. In *ACM/SIGIR Proceedings 2018*, 95–104, (Assoc Comp Machinery; Special Interest Grp Informat Retrieval, 2018). 41st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Univ Michigan, Ann Arbor, MI, Jul 08–12, 2018. <https://doi.org/10.1145/3209978.3210006>
- Huang, Y., Zhou, Z., Wang, Z., Zhi, X. & Liu, X. Timesnet-pm2.5: Interpretable timesnet for disentangling intraperiod and interperiod variations in pm2.5 prediction. *Atmosphere* <https://doi.org/10.3390/atmos14111604> (2023).
- Song, M.-M., Xiong, Z.-C., Zhong, J.-H., Xiao, S.-G. & Tang, Y.-H. Research on fault diagnosis method of planetary gearbox based on dynamic simulation and deep transfer learning. *Sci. Rep.* <https://doi.org/10.1038/s41598-022-21339-5> (2022).
- Zhao, Z., Jiao, Y. & Zhang, X. A fault diagnosis method of rotor system based on parallel convolutional neural network architecture with attention mechanism. *J. Signal Process. Syst. Signal Image Video Technol.* **95**, 965–977. <https://doi.org/10.1007/s11265-023-01846-y> (2023).
- Cheng, X., Dou, S., Du, Y. & Wang, Z. Gearbox fault diagnosis method based on lightweight channel attention mechanism and transfer learning. *Sci. Rep.* <https://doi.org/10.1038/s41598-023-50826-6> (2024).
- Zhang, L. & Shi, Y. Sparse and semi-attention guided faults diagnosis approach for distributed online services. *Appl. Soft Comput.* <https://doi.org/10.1016/j.asoc.2023.110911> (2023).
- Liu, X. & Si, R. Research on data compression algorithm of heating data center based on adaptive swinging door trending. In *2019 4TH International Conference on Mechanical, Control and Computer Engineering (ICMCCE 2019)*, 880–884 (2019). 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Hohhot, PEOPLES R CHINA, OCT 25–27, 2019. <https://doi.org/10.1109/ICMCCE48743.2019.00200>.

26. Pan, X. *et al.* On the integration of self-attention and convolution (2022). [arXiv:2111.14556](https://arxiv.org/abs/2111.14556).
27. Szegedy, C. *et al.* Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1–9, <https://doi.org/10.1109/CVPR.2015.7298594> (2015).
28. Zhou, H. *et al.* Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Thirty-fifth AAAI Conference on Artificial Intelligence, Thirty-third Conference on Innovative Applications of Artificial Intelligence and the Eleventh Symposium on Educational Advances in Artificial Intelligence*, vol. 35 of *AAAI Conference on Artificial Intelligence*, 11106–11115 (Assoc Advancement Artificial Intelligence, 2021). 35th AAAI Conference on Artificial Intelligence / 33rd Conference on Innovative Applications of Artificial Intelligence / 11th Symposium on Educational Advances in Artificial Intelligence, Electr Network, Feb 02–09, 2021.
29. Chen, J. *et al.* Multiscale convolutional neural network with feature alignment for bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* <https://doi.org/10.1109/TIM.2021.3077673> (2021).
30. Guo, Q., Zhang, X., Li, J. & Li, G. Fault diagnosis of modular multilevel converter based on adaptive chirp mode decomposition and temporal convolutional network. *Eng. Appl. Artif. Intell.* <https://doi.org/10.1016/j.engappai.2021.104544> (2022).
31. König, C. & Helmi, A. M. Sensitivity analysis of sensors in a hydraulic condition monitoring system using CNN models. *Sensor* <https://doi.org/10.3390/s20113307> (2020).

## Acknowledgements

Thesis project supported by: Zhejiang Provincial Administration for Market Regulation Young Eagle Program Cultivation Project (Project Number: CY2022349), and State Administration for Market Regulation Science and Technology Program Project (Project Number: 2023MK227).

## Author contributions

H.C.W. and H.C. jointly collected data and analyzed data; H.C.W., B.X., and Z.X.S. prepared the draft; H.C.W. and J.Z. conceived and designed the experiment; H.C.W. and J.Z. completed the writing of the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-66541-9>.

**Correspondence** and requests for materials should be addressed to J.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024