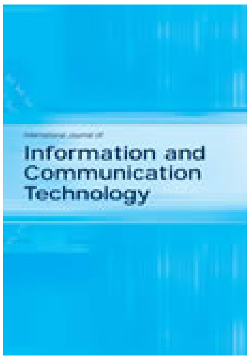


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Intelligent fault diagnosis of mechanical equipment based on industrial big data

Jieqi Zhang, Shuai Huang

Article History:

Received:	30 December 2024
Last revised:	13 January 2025
Accepted:	14 January 2025
Published online:	21 March 2025

Intelligent fault diagnosis of mechanical equipment based on industrial big data

Jieqi Zhang* and Shuai Huang

Mechanical Engineering Department,
Jiangxi Polytechnic University,
JiuJiang 332007, China
Email: zhangjieqi666@163.com
Email: cjxjj2020@163.com
*Corresponding author

Abstract: Effective fault diagnosis will greatly improve the operational efficiency of industrial machinery and equipment. In this paper, for the issues of multi-fault coupling and low diagnostic accuracy that exist in the current research. Firstly, the mechanical equipment signals are pre-processed. The empirical modal decomposition is introduced to construct the fault eigenvectors of industrial mechanical equipment. Then the improved principal component analysis is used to map the high-dimensional features to the low-dimensional space, the dual attention mechanism (DAM) is introduced to improve the transformer model (ODAT), an ODAT model is trained for each fault for diagnosis, and a fault set is generated based on the diagnosis results of all ODAT models. Comparative experiments were conducted on the PHM15 dataset, and the results show that the fault diagnosis accuracy of the proposed model is 93.71%.

Keywords: mechanical equipment fault diagnosis; empirical modal decomposition; principal component analysis; PCA; dual attention mechanism; DAM; transformer model.

Reference to this paper should be made as follows: Zhang, J. and Huang, S. (2025) 'Intelligent fault diagnosis of mechanical equipment based on industrial big data', *Int. J. Information and Communication Technology*, Vol. 26, No. 5, pp.84–99.

Biographical notes: Jieqi Zhang obtained her Master's degree from the North University of China in 2010. She is currently a Lecturer at the Jiangxi Polytechnic University. Her research interests include mechanical optimisation design and system scheduling.

Shuai Huang received his Master's degree from the Chongqing University in 2012. He is currently a Lecturer in the Jiangxi Vocational Technical University. His research interests include optimisation design of high-precision mechanical structures and the application of artificial intelligence in mechanical design.

1 Introduction

With the growth of industrial society and the continuous progress of technology, modern industry began to totally adopt mechanical equipment in lieu of manual labour. However, the operation and maintenance of machinery and equipment is crucial while the efficiency of work increases. If mechanical failure occurs in the process of industrial production, it may affect the productivity or cause safety accidents (Kareem and Jewo, 2015). Failure of mechanical equipment generally manifests itself in the inability to accomplish a specific functionally required task, and there are challenges in identifying failures due to the presence of more complex parts in the equipment (Huo et al., 2020). To ensure the reliable and stable operation of mechanical equipment, it is very essential to carry out equipment fault diagnosis. The conventional mechanical equipment fault diagnosis approach chiefly depends on the maintenance personnel's experience judgement, which has the issues of low diagnostic accuracy and low maintenance efficiency (Sun et al., 2023). Therefore, it is urgent to develop automated and intelligent mechanical equipment fault diagnosis approaches.

Traditional study is relied on statistical modelling to diagnose the faults of mechanical equipment, Jiang et al. (2013) projected the monitoring data onto a predefined operation mode to obtain the specified element, and memorised the equipment fault diagnosis by the size of the specified element's significance. Sarita et al. (2022) adopted principal component analysis (PCA) to decompose each set of fault data and project the data to be detected in the direction of known faults to diagnose the type of equipment faults, but it was difficult to diagnose faults caused by non-vibration factors. Conventional statistical model based mechanical equipment fault diagnosis methods suffer from severe pattern compounding effects, leading to poor diagnostic accuracy. As the industrial big data rapidly growing, data-driven techniques relied on machine learning (ML) directly deep mine the fault data to achieve high-precision fault diagnosis. Goyal et al. (2020) proposed a fault diagnosis method for mechanical equipment based on restricted Boltzmann machines (RBM) and support vector machines (SVM), and the faults can be well detected by extracting the features by RBM and inputting them into SVM. Wang et al. (2023) adopted a decision tree (DT) to categorise historical faults and then used Jaccard similarity to calculate the similarity between new faults and historical fault classes for case-based reasoning, but the diagnostic efficiency is not high.

Data-driven models based on deep learning (DL) overcome this shortcoming by mining valuable information from raw data to enable end-to-end diagnosis, thereby improving diagnostic efficiency. Tang et al. (2020) used CNNs to automatically learn features conducive to bearing fault detection and used them in a mechanical equipment fault diagnosis task with higher accuracy compared to DT methods. Shi et al. (2022) proposed an LSTM-based fault diagnosis method for mechanical equipment, which can well diagnose a variety of faults. Liu et al. (2021) collected the timing signals generated during the operation of the equipment, used the CRITIC weighting method to determine the key parameters that affect the occurrence of equipment faults, and used the GRU to perform a predictive analysis. Xing et al. (2020) identified multiple faults by deep belief network (DBN), but did not consider the class imbalance and multi-labelling problem that exists between faults, and was unable to learn the complex dependencies between labels.

In the latest research, the attention mechanism (AM) allows the model to automatically learn and selectively focus on important information when processing input data. By assigning different weights to different information, the model is able to focus on the most relevant information instead of treating all inputs equally. This ability allows the model to more accurately capture key features in the input data to improve the performance of the DL. Xu et al. (2022) utilised temporal convolutional neural networks (TCNs) to extract features from time series for fault prediction by assigning weights to multiple dimensions of the input features through a distributed AM. Pan et al. (2022) proposed a generative adversarial network (GAN)-based engine fault prediction method to reconstruct the data in the pre-training phase via GAN, and feature enhancement via AM in order to classify the equipment faults via softmax. Xiao et al. (2023) proposed an engine fault prediction method based on the transformer architecture, which extracts mechanical equipment features in an unsupervised manner and further enhances the features through AM for fault diagnosis, which greatly improves the diagnosis efficiency.

Intending to the issues of fault imbalance and multi-fault coupling in the above studies, an intelligent fault diagnosis method for mechanical equipment based on industrial big data is suggested. Firstly, the box plot method and wavelet packet noise reduction method are used to pre-process the mechanical equipment signals, and secondly, the empirical mode decomposition (EMD) algorithm is utilised to decompose the equipment signals into multiple intrinsic mode function (IMF) components, and multiple feature parameters are computed for the effective IMF components to construct the high-dimensional fault features. These high dimensional feature data are then mapped to a low dimensional space using improved PCA to reduce feature redundancy. Finally, the transformer model is improved by introducing the dual attention mechanism (ODAT), so that the model focuses on the more important information in the features from multiple dimensions. To reduce the difficulty of fault diagnosis, a binary correlation method is used to train an ODAT model for each fault for diagnosis, and after the diagnosis is completed, a fault set is generated relied on the diagnosis results of all ODAT models. The experimental outcome indicates that the diagnostic accuracy and F1 value of the proposed model are improved by 3.54%–16% compared with the comparison model, which can effectively improve the diagnostic accuracy and has good application prospects in the field of industrial machinery and equipment fault diagnosis.

2 Relevant theoretical foundations

2.1 Attention mechanism

AM improves the information extraction ability of the model by modelling the human AM and focusing the model's attention on the information that is more important to the current task, i.e., assigning appropriate weights to the input information according to its importance to the current task (Lv et al., 2022). The role of the AM is to find the most important and relevant information for the output from the input time series, which can be described as a mapping of a query *Query* to multiple key-value pairs *key-value* as shown in Figure 1. AM first maps the input time series into multiple key-values and the output into a query. Here, three weight matrices W^K, W^V, and W^Q are introduced and multiplied with the input time series X and output Y to obtain the corresponding K (Key), V (Value), and Q (Query) parameter matrices. In the AM, K, Q and V are computed from the input data by specific linear transformations, i.e., $K = XW^K$, $V = XW^V$, and $Q = YW^Q$, respectively.

The similarity between the output Q and multiple K of the input time series is then computed to obtain the attention weight α for each V, as shown in equation (1). It is transformed into a probability distribution with a sum of 1 by means of the softmax function, as shown in equation (2).

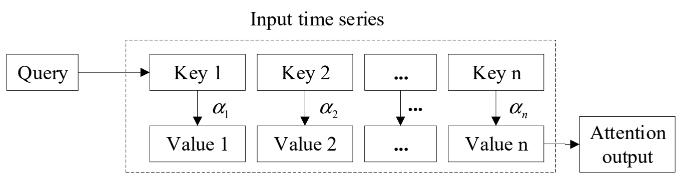
$$f\left(Q, K_i\right) = Q^T W_\alpha K_i \tag{1}$$

$$\alpha_i = softmax\left(f\left(Q, K_i\right)\right) = \frac{e^{f(Q,K_i)}}{\sum_j e^{f(Q,K_j)}} \tag{2}$$

Finally, the attention output corresponding to the input time series is obtained based on the weighted summation of α to V, as shown in equation (3).

$$Attention(Q, K, V) = \sum_i \alpha_i V_i \tag{3}$$

Figure 1 Structure of the AM



2.2 *Transformer model*

Traditional models tend to lose a lot of original information when dealing with long time series and are unable to learn the effect of different time steps in the input time series on the output. Transformer is a model of multiple encoder-decoder stacking relied on the AM (Guerra and Mota, 2006), as shown in Figure 2. The transformer model was introduced mainly to address the shortcomings of traditional models in dealing with sequential data, especially the limitations of recurrent neural networks (RNNs) and their variants, (e.g., LSTMs and GRUs) in dealing with long-range dependencies and parallel computation. The model encodes the input time series into multiple intermediate state vectors that change with the output, and can fully learn the importance of different time steps in the input time series to the output, which can not only solve the long time series dependency problem well, but also can be computed in parallel.

1 Encoder (E). E consists of a position coding (PC) layer, a multi-head self-attention (MAM) layer, a feed-forward neural network (FNN) layer and a residual normalisation (RN) layer. First, the sequence location information is added to the input feature sequences through the PC level; then, the attention weights are assigned to the input feature sequences through the MAM level; finally, the outputs of the MAM level are transformed to the nonlinear space through the FNN level to enhance the nonlinear representation of the model. In addition, the RN level is added after the MAM and FNN levels to speed up the model convergence speed to enhance the model generalisation ability.

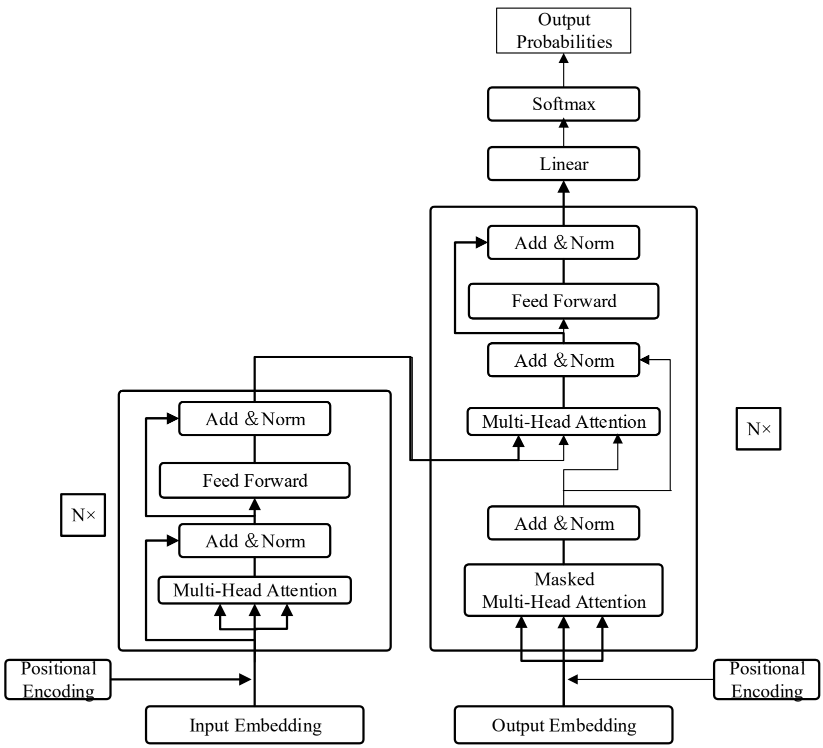
2 Decoder (D). D is composed of PC level, Masking MAM level, MAM level, FNN level and RN level. First, the sequence position information is added to the input label sequence through the PC level; then, the model is made to focus more on the important information in the input label sequence by masking the MAM level, as shown in equation (4). Then, the dependencies between the intermediate state vectors

of the encoder output and the input label sequence are learned through the MAM level; finally, the output of the MAM level is transformed to a nonlinear space through the FNN layer. The role of the RN level is the same as that of the FN level in the encoder.

$$Masked(Q, K, V) = softmax\left(\frac{QK^T M}{\sqrt{d_{model}}}\right) \tag{4}$$

where M is the lower triangular unit matrix.

Figure 2 Transformer model structure



3 Data acquisition and pre-processing of industrial machinery and equipment

Industrial machinery and equipment generate a variety of signals during operation, which represent their voltage and current changes or vibration amplitude, etc. To capture the signals generated during the operation of machinery and equipment, the study installs different sensors on the machinery and equipment to capture the industrial big data of the operation. However, due to the presence of more types of equipment and many types of industrial data, there are also many data with missing values, redundant values, outliers, etc. (Fernandes et al., 2022). If these data are used directly as inputs to the diagnostic model, the complexity of the input variables will be greatly enhanced, thus reducing the diagnostic accuracy. For this reason, pre-processing of signal data is required.

There are many outliers in the collected data, which affect the accuracy of fault diagnosis, and the outliers need to be eliminated, for this purpose, the study utilises the box plot method to screen the outlier anomalous data. It was found that if these anomalies are simply eliminated, the continuity of the time series is destroyed, for this reason; the screened outliers are uniformly treated as missing values as shown below.

$$x_i = \frac{x_{i-1} + x_{i+1}}{2} \tag{5}$$

where $x_i \notin [D, U]$, $[D, U]$ are the upper and lower value ranges of the box plot, and x_i is the observed value at time point i.

In signal acquisition, a lot of noise is generated, and the data needs to be denoised in order to obtain more effective diagnostic results. Currently the commonly used denoising methods are smoothing denoising and signal transformation. The signal transform noise reduction method can better preserve the original characteristics of the data while denoising. The wavelet transform has the property of multiresolution analysis, which means that it is able to decompose the signal at different scales. This property allows the wavelet transform to portray the non-smooth features of the signal well, such as edges, spikes, and breakpoints. When dealing with signals containing noise, the wavelet transform is able to separate the signal from the noise at different scales, thus removing the noise more effectively. Therefore, the study used wavelet transform (Almounajjed et al., 2022) in the signal transform noise reduction method for noise reduction of the data. First the signal is decomposed and after decomposition the wavelet coefficients of each node are obtained. The wavelet coefficients obtained from the data with too much noise will exceed the threshold, so some of the data can be filtered according to the calculated wavelet coefficients. After filtering the data signal is reconstructed according to equation (6).

$$d_l^{j+1,n} = \sum_{k \in z} \left[h_{l-2k} d_k^{j,2n} + g_{l-2k} d_k^{j,2n+1} \right] \tag{6}$$

where k is the number of layers to continue decomposition, z is the set of all integers, j is the frequency index, $d_l^{j+1,n}$ is the orthogonal wavelet packet decomposition coefficients of the signal at a resolution of $j + 1$, n is the number of discrete data points, h_{l-2k} and g_{l-2k} are sequences consisting of filter coefficients. In determining the threshold value, the study chooses the wavelet threshold estimation method for threshold estimation, as shown in equation (7).

$$T(j, n) = \frac{\sigma^2(j, n)}{\sigma_x(j, n)} \tag{7}$$

where σ_x is the original signal variance, σ^2 is the noise variance, and T is the resulting threshold. The wavelet packet coefficients are filtered by a soft threshold function as shown in equation (8).

$$d_l^{j,n} = \begin{cases} \operatorname{sgn}\left(d_l^{j,n}\right)\left(\left|d_l^{j,n} - T(j,n)\right|\right), \left|d_l^{j,n}\right| \geqslant T(j,n) \\ 0, \left|d_l^{j,n}\right| < T(j,n) \end{cases} \tag{8}$$

Due to the existence of diversity in mechanical equipment, the collected data has different magnitudes, and the diagnostic model needs to be trained with a uniform magnitude. For this reason, the study operates on the data to eliminate the magnitude, and the standardised calculation method is shown below.

$$\hat{x} = \frac{x - E(x)}{\sqrt{\operatorname{var}(x)}} \tag{9}$$

where x and $\hat{x}$ are the values before and after treatment, respectively, and $E(x)$ is the sample mean.

4 Intelligent fault diagnosis of mechanical equipment based on industrial big data

4.1 Constructing fault eigenvectors of industrial machinery and equipment based on EMD

Intending to the fault class imbalance and multi-fault coupling issues existing in the fault diagnosis of industrial machinery and equipment, a fault diagnosis method based on the DAM improvement transformer (ODAT) model under the class imbalance of industrial big data is proposed, as shown in Figure 3. Firstly, the mechanical equipment signals are decomposed into intrinsic modal function components (IMFs) by EMD and the IMFs containing feature information are obtained, and then multiple feature parameters are computed for these effective components to form the high-dimensional features, and then these high-dimensional data are mapped to the low-dimensional space by using the improved PCA. Finally, the ODAT model is used to extract important features from multiple dimensions, and the binary correlation method is used to reduce the difficulty of training the multi-label model, and after the diagnosis is completed, the fault set is generated based on the diagnostic results of all ODAT models.

After the industrial machinery and equipment signal acquisition and pre-processing, the principle of EMD method (Lei et al., 2013) is utilised to make a multi-dimensional decomposition of the machinery and equipment signals, and then construct the fault feature vector. The EMD method is a big data-driven decomposition method that does not require prior assumptions about the distribution or structure of the signal and is capable of adaptively generating the IMF.

Since any industrial machinery signal consists of independent IMF components, the components are interrelated and coupled. Firstly, EMD is used to decompose the complex signal into a series of IMF components and a sum of residual components to extract the local features and dynamic information in the signal, the formula is as follows.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r(t) \tag{10}$$

where $x(t)$ is the mechanical equipment signal, $c_i(t)$ is the intrinsic mode component of the vibration signal, and $r(t)$ is the residual component of the vibration signal. Next, all local extremes are identified from $x(t)$. Find the mean value of the upper and lower envelopes, denote it as m_1, and derive the difference between the original vibration signal and the mean value of the envelope as follows.

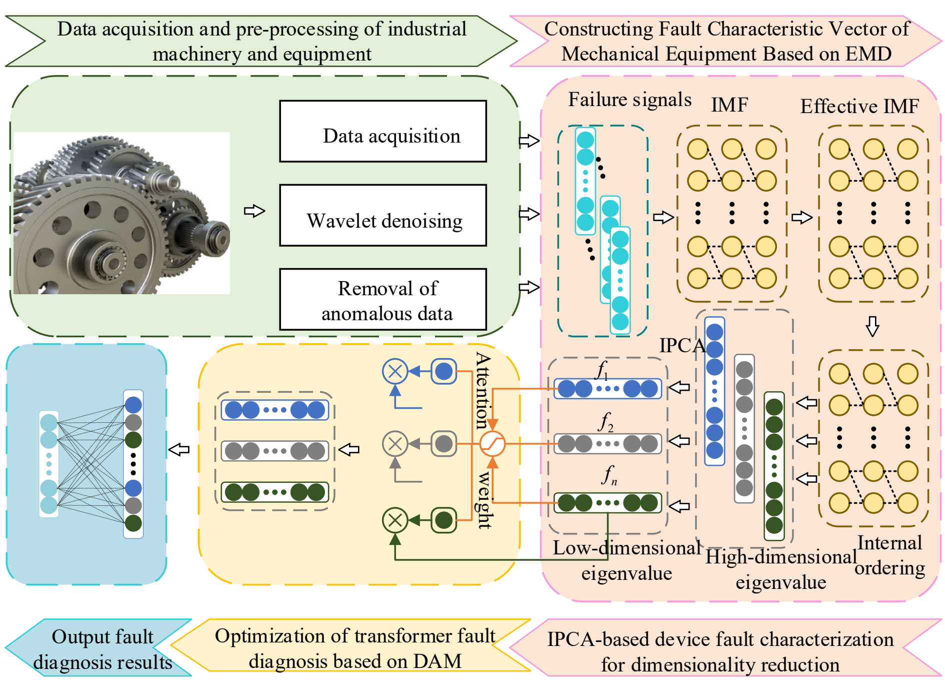
$$h_1 = x(t) - m_1 \tag{11}$$

If h_1 meets the two necessary conditions of the IMF component of the mechanical equipment signal, it is considered to be the first IMF component of the equipment signal; if either condition is not met, it is treated as the original signal, and the above decomposition steps of the equipment signal are repeated. Calculate the correlation coefficient between the IMF component and the original signal as follows.

$$Q_{xy} = \frac{\sum_{k=1}^{n} x(k)y(k)}{\sqrt{\sum_{k=0}^{n} x^2(k) \sum_{k=0}^{n} y^2(k)}} \tag{12}$$

where $x(k)$ and $y(k)$ are IMF components. Calculate the correlation coefficient between the two, rank the correlation coefficients of each IMF component, take the IMF component with the highest correlation coefficient, and then extract the features in the time and frequency domains of the valid IMF components to get the high-dimensional eigenvalue $f_1, f_2, \ldots, f_n$ related to the failure of the mechanical equipment.

Figure 3 Intelligent fault diagnosis model of mechanical equipment based on industrial big data (see online version for colours)



4.2 Dimensionality reduction of industrial machinery and equipment fault characteristics based on improved PCA

After obtaining the eigenvalues associated with the mechanical equipment faults, PCA is utilised to perform data dimensionality reduction on the above high dimensional feature matrix. Due to the high computational complexity of traditional PCA (Gewers et al., 2021), this paper uses kernel function to optimise PCA, and carries out dimensionality reduction processing for high-dimensional data features of decomposed useful components, making the data structure more concise while retaining the signal feature information to the maximum extent. Thus, the correct rate of mechanical equipment fault diagnosis can be significantly improved.

Suppose that a nonlinear mapping function $\varphi(x)$ transforms the set of high-voltage features $F = [f_1, f_2, \ldots, f_n]^T$ into a set of kernel features $F' = [\varphi(f_1), \varphi(f_2), \ldots, \varphi(f_n)]^T$, where n is the number of samples, the covariance matrix F of the kernel feature set is as follows.

$$\begin{cases} F = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \psi(f_i)\psi(f_i)^T \\ \psi(f_i) = \varphi(f_i) - \dfrac{1}{n}\displaystyle\sum_{j=1}^{n} \varphi(f_i) \end{cases} \tag{13}$$

The eigenvalues λ_i and eigenvectors ξ_i of the matrix F are then computed as follows.

$$\begin{cases} \xi_i = \displaystyle\sum_{i=1}^{n} b_i \psi(f_i) \\ b_i = \dfrac{1}{n\lambda_i}\displaystyle\sum_{i=1}^{n} \left(\psi(f_i)^T \xi_i\right) \end{cases} \tag{14}$$

A Gaussian radial basis function is chosen as the kernel function for the inner product operation of PCA in high dimensional space.

$$\hat{E}(f_i, f_j) = \varphi(f_i)^T \varphi(f_j) \tag{15}$$

Then the e eigenvectors corresponding to the e largest eigenvalues are selected as the column vectors of the transform matrix γ. The result of the improved PCA after dimensionality reduction is as follows.

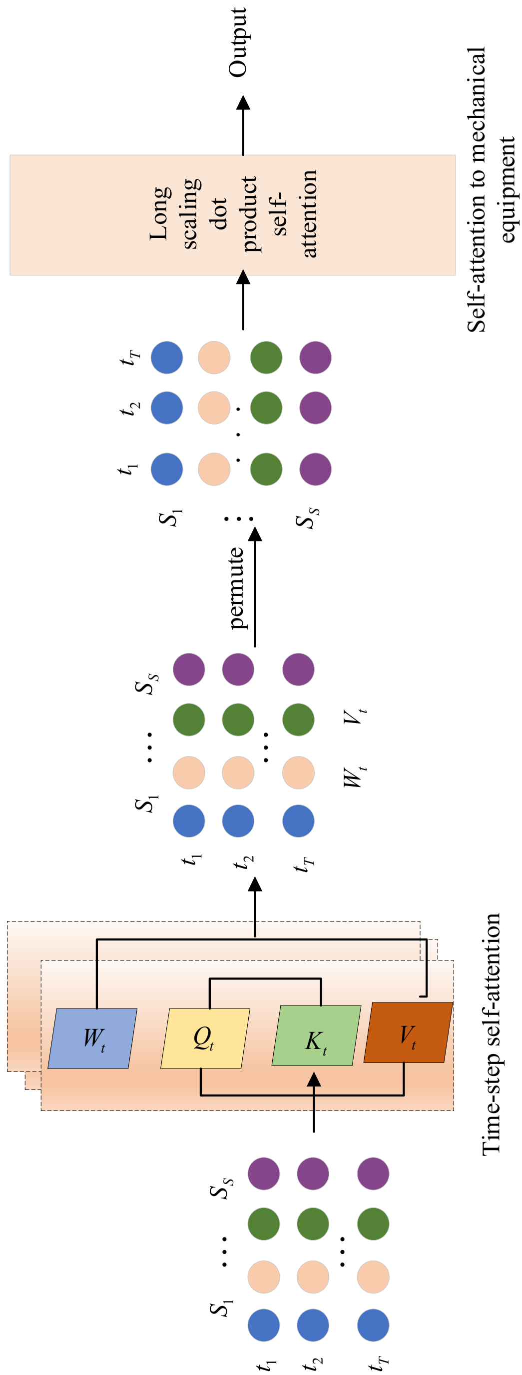
$$R = \gamma^{\mathrm{T}} \cdot F' \tag{16}$$

To illustrate the degree of correlation between the extracted feature vectors and the equipment fault sample category vectors, the expression is defined as follows.

$$\tau = \frac{\operatorname{cov}(\xi_i, \xi_i)}{\sqrt{\operatorname{cov}(\xi_i, \xi_i)\operatorname{cov}(\xi_n, \xi_n)}} \tag{17}$$

where τ is the correlation degree coefficient, ξ_i is the extracted feature vector, and ξ_n is the sample category vector of the original feature set.

Figure 4 Dual AM (see online version for colours)



4.3 Improved transformer based on dual AM for fault diagnosis of industrial machinery and equipment

The traditional transformer model only focuses on the attention of different time steps in the input feature sequence, but ignores the attention of different mechanical devices in the input feature sequence. For this reason, the industrial machinery and equipment troubleshooting model (ODAT) proposed in this chapter improve the transformer model by replacing the MAM with the DAM. The DAM mechanism is added to the encoder to extract the important information from the signals of different mechanical devices. An embedding layer is added to the input side of the decoder to enhance the dimension of the input tag sequence. The ODAT model can dig deep into the complex mapping relationship between multi-dimensional mechanical equipment and multiple fault labels.

DAM is shown in Figure 4, firstly, the IPCA dimensionality reduction features are passed through time step AM to get the weighted time step feature sequence; then it is transposed to get the mechanical device feature sequence; finally, the weighted signal feature sequence is passed through device AM. So DAM can extract important features not only in different time steps but also in different mechanical devices as shown in equation (18) to equation (20). In addition, both time-step attention and mechanical device attention in DAM use a multi-head mechanism to improve the learning ability of the model.

$$Timestep(Q_t, K_t, V_t) = Softmax\left(\frac{Q_t K_t^T}{\sqrt{d_{model}}}\right)V_t = W_t V_t \tag{18}$$

$$(W_t V_t)^T = Q_s = K_s = V_s \tag{19}$$

$$ME(Q_s, K_s, V_s) = Softmax\left(\frac{Q_s K_s^T}{\sqrt{d_{model}}}\right)V_s = W_s V_s \tag{20}$$

where Q_t, K_t, and V_t are the query, key, value, and weight matrices for the time-step feature, respectively; Q_s, K_s, and V_s are the query, key, value, and weight matrices for the mechanical equipment feature, respectively, and d_{model} is the dimension of the model.

To reduce the complexity of industrial equipment fault diagnosis under multiple class imbalance, this paper adopts the binary association method to convert the class-imbalanced multi-label multiclassification issue into an independent *m* class-imbalanced single-label binary classification issue. The complex industrial equipment fault dataset D is split into m class-unbalanced binary datasets D_g (g = 1, 2, …, m) based on fault labels as shown in equation (21) to equation (24).

$$D_g = \{(x_i, \varnothing(G_i, g)) | 1 \le i \le n\} \tag{21}$$

$$\varnothing(G_i, g) = \begin{cases} 1, & g \in G_i \\ 0, & otherwise \end{cases} \tag{22}$$

$$D_g^+ = \{(x_i, 1) | g \in G_i, 1 \le i \le n\} \tag{23}$$

$$D_g^- = \{(x_i, 0) | g \notin G_i, 1 \le i \le n\} \tag{24}$$

where $\varnothing(F_i, g)$ is the labelling category of sample x_i in dichotomous dataset D_g corresponding to fault g, D_g^+ and D_g^- are the positive and negative sample sets, respectively.

The class imbalance rates of D_g^+ and D_g^- with fault g in D_g are shown in equation (25). To address the class imbalance between D_g^+ and D_g^- after decomposition, this chapter increases the sample size of D_g^+, the minority fault class, through SMOTE oversampling to ensure a relatively balanced sample distribution in the dataset. For the transformed binary learning problem, this chapter utilises D_g to train an ODAT model for each fault g for fault diagnosis, uses $ODAT_g$ to generate the corresponding set of fault labels according to equation (26).

$$IR_g = Max\left(\left|D_g^+\right|, \left|D_g^-\right|\right) \Big/ Min\left(\left|D_g^+\right|, \left|D_g^-\right|\right) \tag{25}$$

$$G_i = \left\{g \left| ODAT_g\left(x_i\right) > 0, 1 \le g \le m \right.\right\} \tag{26}$$

5 Experimental results and analyses

This chapter evaluates the proposed model using the dataset provided by the PHM Society Association in the PHM 2015 Data Challenge. The dataset is real historical monitoring data of complex industrial equipment after desensitisation of the plant, and each piece of equipment consists of normal events and six types of failure events. There are about 33 plants, and the data are collected over a period of about 3 to 4 years, with sampling intervals of about 15 minutes. This paper mainly uses data from Plant 1, which consists of six components, each with four sensors (s1–s4), four operating parameters (r1–r4), and six fault labels (f1–f6). The number of samples and percentage of each type of faults after processing by SMOTE oversampling are shown in Table 1.

Table 1 Sample size and percentage of each type of failure in Plant 1

Fault type	*f1*	*f2*	*f3*	*f4*	*f5*	*f6*
Quantity	6,015	3,796	5,194	3,687	4,123	8,319
Ratio (%)	19.3	12.2	16.7	11.8	13.2	26.8

The model training and evaluation in this chapter are based on the Python programming language and the TensorFlow, Keras DL framework, and Scikit-learn ML framework. The experimental environment is Intel Xeon Silver 4210R CPU 64G RAM, NVIDIA Tesla T4 GPU 16G RAM, and Windows Server 2016 Standard system. When training the model, the time step is 24, the batch size is 64, the maximum number of iterations for training is set to 200, the loss function is set to binary cross-entropy; the optimiser is set to Adam, and the initial learning rate is set to 0.001.

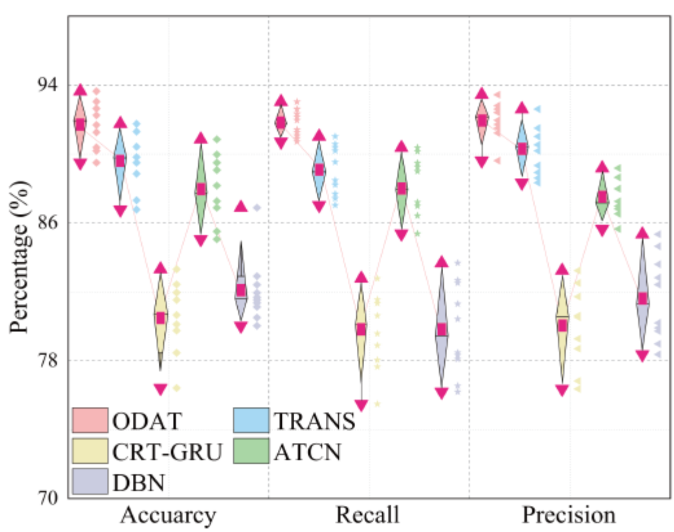
To validate the diagnostic effectiveness of the ODAT model, it will be compared with the CRT-GRU (Liu et al., 2021), DBN, ATCN (Xu et al., 2022), and TRANS (Xiao et al., 2023) models, and the evaluation metrics will be used as the accuracy, precision, recall, F1 and G-mean (Jian and Ao, 2023). Comparison of diagnostic accuracies of different fault types for each model is shown in Table 2. The diagnostic accuracies of ODAT are higher than the other five models for all six fault types. When the fault types are f2 and f5, the diagnostic accuracy of the five models is lower than the diagnostic accuracy of the other fault types, this is because f2 and f5 are the tooth breakage and tooth wear of the mechanical equipment, respectively, which are difficult to recognise, resulting in low accuracy. When the fault types are f1, f3, f4, and f6, the diagnostic accuracy of ODAT is improved by 3.31%–17.84% compared to other models.

Table 2 Diagnostic accuracy of different fault types for various models (%)

Model	*f1*	*f2*	*f3*	*f4*	*f5*	*f6*
CRT-GRU	76.89	71.12	77.15	76.93	70.68	75.14
DBN	79.63	75.31	80.09	79.84	76.04	78.52
ATCN	83.46	80.01	82.19	84.94	80.56	82.41
TRANS	89.81	85.64	90.25	88.16	84.12	89.53
ODAT	93.62	89.25	93.56	93.05	88.15	92.98

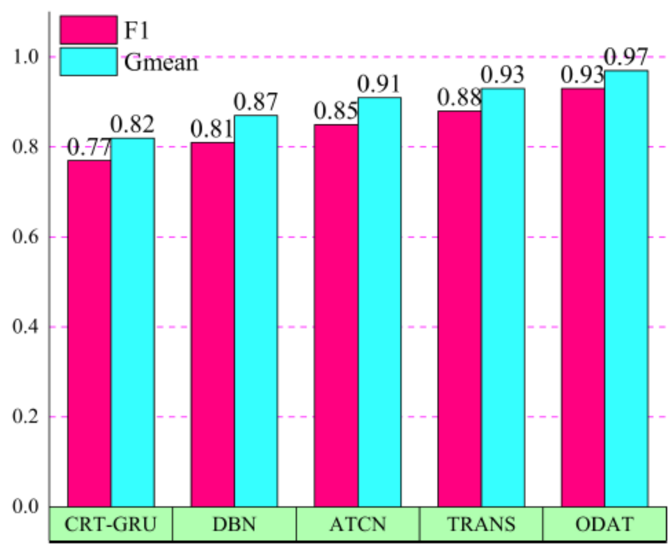
Comparison of the average accuracy, recall and precision of different fault diagnosis models is shown in Figure 5. The accuracy of ODAT is 93.71%, which is 15.73%, 13.15%, 8.54%, and 3.54% higher than CRT-GRU, DBN, ATCN, and TRANS, respectively. CRT-GRU did not denoise the mechanical equipment signals and determined the key parameters only by the CRITIC weighting method, resulting in poor diagnostic accuracy. DBN model identifies faults through DBN but does not consider the class imbalance and multi-labelling that exists between faults. ATCN does not consider the long term dependency of the time series although it highlights the key features of the device through AM. TRANS performs mechanical equipment fault diagnosis through transformer, but transformer is not optimised, so the diagnostic accuracy is not as good as that of ODAT; therefore, the ODAT model is more comprehensive for complex industrial equipment fault diagnosis, and is able to diagnose multiple faults occurring at the same time in a timely and accurate manner.

Figure 5 Comparison of the average accuracy, recall and precision of different fault diagnosis model (see online version for colours)



In addition to accuracy, recall and precision, F1 and Gmean are also important indicators for evaluating the diagnostic effectiveness. F1 is the reconciled average of precision and recall, which comprehensively reflects the diagnostic efficiency of faults. Gmean is the geometric mean of the true case rate (TPR) and the true negative case rate (TNR), the larger the Gmean value, the more efficient the diagnosis. As implied in Figure 6, the F1 and Gmean values of ODAT are 0.93 and 0.97 respectively, which are 16% and 15% higher compared to CRT-GRU, 12% and 10% higher compared to DBN, 8% and 6% higher compared to ATCN, and 5% and 4% higher compared to TRANS, suggesting that ODAT exhibits excellent diagnostic performance. ODAT not only denoises the mechanical equipment signals and uses IPCA to downscale the high-dimensional equipment feature vector constructed by EMD, but also improves the Transformer model by DAM, which effectively solves the fault class imbalance and multi-fault coupling problems, thus improving the diagnostic efficiency.

Figure 6 Comparison of F1 and Gmean for different models (see online version for colours)



6 Conclusions

Fault diagnosis of industrial machinery and equipment is essentially a problem of fault pattern recognition, and the selection of appropriate diagnostic methods is crucial to the accuracy of the diagnostic results. In this paper, for the current mechanical equipment fault diagnosis method diagnostic results of the problem of low accuracy, first of all, the industrial machinery and equipment big data pre-processing, the use of EMD on the pre-processed machinery and equipment signals to make a multi-dimensional decomposition, and then build the fault feature vector. The kernel function is then introduced to improve the PCA algorithm (IPCA), which is utilised to map the high-dimensional features of the mechanical device to the low-dimensional space and reduce feature redundancy. Finally, the transformer model is improved (ODAT) by DAM so that the model focuses on the key features from multiple dimensions, and in order to alleviate the difficulty of fault diagnosis, the binary association method is used to train one ODAT model for each fault for diagnosis, and after the diagnosis is completed, a fault set is generated based on the diagnostic results of all ODAT models. The experimental outcome implies that the model has high fault diagnosis accuracy and can accurately diagnose multiple faults occurring at the same time for targeted repairs. This paper focuses on industrial big data-driven fault diagnosis methods for mechanical equipment, but there are still some shortcomings, and subsequent consideration is given to the physical knowledge of industrial equipment and related domain knowledge as constraints to be incorporated into the proposed model, in order to improve the robustness and generalisability of the model.

Declarations

All authors declare that they have no conflicts of interest.

References

Almounajjed, A., Sahoo, A.K. and Kumar, M.K. (2022) 'Condition monitoring and fault detection of induction motor based on wavelet denoising with ensemble learning', *Electrical Engineering*, Vol. 104, No. 5, pp.2859–2877.

Fernandes, M., Corchado, J.M. and Marreiros, G. (2022) 'Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review', *Applied Intelligence*, Vol. 52, No. 12, pp.14246–14280.

Gewers, F.L., Ferreira, G.R., Arruda, H.F.D. et al. (2021) 'Principal component analysis: a natural approach to data exploration', *ACM Computing Surveys (CSUR)*, Vol. 54, No. 4, pp.1–34.

Goyal, D., Choudhary, A., Pabla, B. et al. (2020) 'Support vector machines based non-contact fault diagnosis system for bearings', *Journal of Intelligent Manufacturing*, Vol. 31, pp.1275–1289.

Guerra, F.D.C.F. and Mota, W.S. (2006) 'Current transformer model', *IEEE Transactions on Power Delivery*, Vol. 22, No. 1, pp.187–194.

Huo, J., Lin, D. and Qi, W. (2020) 'Intelligent fault diagnosis method of mechanical equipment based on fuzzy pattern recognition', *Journal of Intelligent & Fuzzy Systems*, Vol. 38, No. 4, pp.3657–3664.

Jian, C. and Ao, Y. (2023) 'Imbalanced fault diagnosis based on semi-supervised ensemble learning', *Journal of Intelligent Manufacturing*, Vol. 34 No. 7, pp.3143–3158.

Jiang, F., Zhu, Z., Li, W. et al. (2013) 'Robust condition monitoring and fault diagnosis of rolling element bearings using improved EEMD and statistical features', *Measurement Science and Technology*, Vol. 25, No. 2, p.025003.

Kareem, B. and Jewo, A. (2015) 'Development of a model for failure prediction on critical equipment in the petrochemical industry', *Engineering Failure Analysis*, Vol. 56, pp.338–347.

Lei, Y., Lin, J., He, Z. et al. (2013) 'A review on empirical mode decomposition in fault diagnosis of rotating machinery', *Mechanical Systems and Signal Processing*, Vol. 35, Nos. 1–2, pp.108–126.

Liu, H., Ma, R., Li, D. et al. (2021) 'Machinery fault diagnosis based on deep learning for time series analysis and knowledge graphs', *Journal of Signal Processing Systems*, Vol. 93, pp.1433–1455.

Lv, H., Chen, J., Pan, T. et al. (2022) 'Attention mechanism in intelligent fault diagnosis of machinery: a review of technique and application', *Measurement*, Vol. 199, p.111594.

Pan, T., Chen, J., Zhang, T. et al. (2022) 'Generative adversarial network in mechanical fault diagnosis under small sample: a systematic review on applications and future perspectives', *ISA Transactions*, Vol. 128, pp.1–10.

Sarita, K., Devarapalli, R., Kumar, S. et al. (2022) 'Principal component analysis technique for early fault detection', *Journal of Intelligent & Fuzzy Systems*, Vol. 42, No. 2, pp.861–872.

Shi, J., Peng, D., Peng, Z. et al. (2022) 'Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks', *Mechanical Systems and Signal Processing*, Vol. 162, p.107996.

Sun, Y., Wang, J. and Wang, X. (2023) 'Fault diagnosis of mechanical equipment in high energy consumption industries in China: a review', *Mechanical Systems and Signal Processing*, Vol. 186, p.109833.

Tang, S., Yuan, S. and Zhu, Y. (2020) 'Convolutional neural network in intelligent fault diagnosis toward rotatory machinery', *IEEE Access*, Vol. 8, pp.86510–86519.

Wang, H., Zheng, J. and Xiang, J. (2023) 'Online bearing fault diagnosis using numerical simulation models and machine learning classifications', *Reliability Engineering & System Safety*, Vol. 234, p.109142.

Xiao, Y., Shao, H., Feng, M. et al. (2023) 'Towards trustworthy rotating machinery fault diagnosis via attention uncertainty in transformer', *Journal of Manufacturing Systems*, Vol. 70, pp.186–201.

Xing, S., Lei, Y., Wang, S. et al. (2020) 'Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions', *IEEE Transactions on Industrial Electronics*, Vol. 68, No. 3, pp.2617–2625.

Xu, Y., Yan, X., Feng, K. et al. (2022) 'Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery', *Reliability Engineering & System Safety*, Vol. 226, p.108714.