World Journal of Engineering Research
ISSN: 2959-9873
DOI: https://doi.org/10.61784/wjer3029

INTELLIGENT FAULT DIAGNOSIS OF ROLLING BEARINGS BASED ON VMD-CNN-TRANSFORMER

JinYuan Hu
School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China.
Corresponding Email: jyhu825@gmail.com

Abstract: The rapid development of deep learning has brought transformative advances to intelligent fault diagnosis, providing powerful end-to-end feature learning capabilities that enable more effective analysis of rolling bearing vibration signals. However, conventional convolutional neural network (CNN), with their fixed architectures, have difficulty capturing the dynamically changing time-frequency features of vibration signals. In addition, most existing models lack effective mechanisms to suppress noise and vibration interference during monitoring, leading to a marked drop in diagnostic accuracy under non-stationary and noisy conditions.To improve the model's ability to process non-stationary signals, this study introduces a multi-module diagnostic framework, VMD-CNN-Transformer, which integrates Variational Mode Decomposition (VMD), CNN, and Transformer architectures. The framework first applies VMD to decompose the vibration signals into representative intrinsic mode functions, enhancing the multi-scale representation of the original signals. The CNN module then extracts key local features and integrates multi-scale information. Finally, the Transformer captures long-range dependencies, allowing detailed characterization of complex fault patterns.Comparative experiments on benchmark datasets, including CWRU, XJTU, and DIRG, show that the proposed method achieves superior robustness and generalization under challenging conditions with noise and varying operating states. The framework outperforms mainstream methods and provides a novel technical solution for intelligent industrial equipment monitoring, demonstrating strong potential for practical engineering applications.
Keywords: Rolling bearing; Variational mode decomposition; Convolutional neural network; Transformer; Fault diagnosis

1 INTRODUCTION

Rolling bearings, as essential components of rotating machinery, play a key role in supporting rotational motion and minimizing frictional losses in high-end manufacturing sectors, including aero engines, wind turbines, and rail transit systems [1]. The operating condition of rolling bearings is closely tied to economic performance and has critical implications for public safety [2]. Therefore, the development of high-precision intelligent fault diagnosis systems for rolling bearings is crucial for ensuring motor stability. These systems also form a core technological foundation for intelligent maintenance of industrial equipment, improving both operational safety and economic efficiency [3].

Recent breakthroughs in artificial intelligence have revitalized the field of intelligent fault diagnosis. Deep learning, owing to its strong nonlinear feature extraction and end-to-end adaptive learning capabilities, has shown significant technical advantages in this field. Convolutional Neural Network (CNN) [4], with their hierarchical architectures, effectively capture spatial correlations in signals and are particularly suited to extracting localized fault features. Long Short-Term Memory (LSTM) networks [5], via gating mechanisms, model the dynamic evolution of temporal signals. Furthermore, the Transformer architecture [6], employing self-attention mechanisms, overcomes sequence length limitations of traditional models and provides an innovative solution for modeling long-range dependencies. The combined development of these technologies offers diverse technical approaches for fault diagnosis under complex operating conditions. Zhilin et al. [7] proposed a one-dimensional improved self-attention-enhanced CNN (1D-ISACNN) based on empirical wavelet transform, achieving 100% classification accuracy on three bearing datasets. A hybrid CNN-LSTM model was developed [8] to classify bearing faults under progressive wear conditions using vibration signals, achieving 99% accuracy in experiments. However, despite promising results, most deep learning models lack robust data preprocessing procedures [9]. Under complex operating conditions, fault features often appear as weak signals overlapped by strong noise, severely disrupting feature extraction and significantly reducing model robustness. To address these challenges, Xia et al. [10] proposed a hybrid model combining optimized Variational Mode Decomposition (VMD), Fuzzy Dispersion Entropy (FDE), and Support Vector Machines (SVM), demonstrating effective diagnosis across various fault types and severities in rolling bearings. Additionally, Chen et al. [11] presented a fault diagnosis method integrating VMD-based denoising and feature enhancement with Transformer-based classification, achieving 98.1% accuracy in experiments.

Despite recent progress, numerous challenges persist in real-world industrial environments. To begin with, vibration signals often exhibit strong non-stationary characteristics [12], with statistical properties that vary significantly over time. Traditional signal processing techniques and static models often fail to capture these time-varying features, limiting their effectiveness in representing meaningful information. Moreover, most deep learning models focus on extracting features from local windows but struggle to capture global temporal dependencies, making it difficult to recognize long-term fault evolution patterns. This limitation hinders the interpretation and classification of complex

temporal features. Therefore, achieving robust and accurate feature extraction and temporal modeling in non-stationary environments remains a key challenge in advancing intelligent fault diagnosis systems.

To address these challenges, this study proposes a novel intelligent diagnostic model that integrates VMD with a hybrid CNN–Transformer architecture. The model uses VMD for data denoising and combines the strengths of CNN and Transformer architectures, thereby significantly improving accuracy and robustness in noisy and complex operational settings.

Specifically, the proposed method employs a multimodal fusion architecture, where VMD is used in signal preprocessing to extract physically meaningful intrinsic mode functions (IMFs), thereby improving the multi-scale representation capability of the original signal. During feature extraction, the CNN module leverages its local receptive fields and weight-sharing mechanism to effectively capture transient impulses and localized fault patterns in the signal. Simultaneously, the Transformer module utilizes a multi-head self-attention mechanism to overcome the limitations of traditional convolutional networks, enabling global modeling of long-range dependencies in sequential signals. This hierarchical feature extraction strategy preserves local details and builds global contextual relationships, enabling comprehensive characterization and accurate identification of complex fault features. The main contributions of this study are summarized as follows:

(1) VMD is used in signal preprocessing to extract physically meaningful IMFs, enhancing the multi-scale representation capability of the original signal;

(2) CNN is employed to extract local fault features across multiple scales and perform feature fusion;

(3) The Transformer architecture models global dependencies in long sequences, enabling precise identification and representation of complex fault patterns.

The paper is organized as follows: Section 2 elaborates the overall structure and key module principles of the proposed model; Section 3 presents specific experimental setups and performance evaluation results, comparing them with existing methods; Section 4 concludes the paper, discussing the engineering significance and future directions of the research.

2 METHODS AND MODELS

2.1 Variational Mode Decomposition

VMD, introduced by Dragomiretskiy et al. [13], is an adaptive signal decomposition technique. Unlike Empirical Mode Decomposition (EMD) [14], VMD effectively suppresses endpoint effects and mode mixing, allowing for improved separation of complex, nonlinear, and non-stationary signals into distinct spectral components. The core concept of VMD is to decompose the original signal $f(t)$ into K IMFs, each centered at a specific frequency, with their bandwidths minimized. The variational model is formulated as follows:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ &\text{s.t.} \quad \sum_{k=1}^{K} u_k(t) = f(t) \end{aligned} \tag{1}$$

Here, K denotes the predefined number of modes, $u_k(t)$ is the $k-th$ IMF, and w_k is its center frequency. $\delta(t)$ denotes the Dirac delta function, and $\partial(t)$ indicates the time derivative. To solve the constrained optimization problem, a quadratic penalty term α and a Lagrange multiplier $\lambda(t)$ are introduced, resulting in the augmented Lagrangian formulation:

$$\begin{aligned} &\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) = \\ &\alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ &\left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \end{aligned} \tag{2}$$

VMD performance depends on choosing its key parameters: the number of IMFs K and the penalty factor α .A too small K causes mode mixing and hampers the separation of critical fault information.In contrast, too large a K introduces redundant modes, lowers computational efficiency, and adds noise.The penalty factor α determines the bandwidth compactness of each mode.A larger α yields smoother components, favoring low-frequency feature extraction. Conversely, a smaller α produces more abrupt variations, aiding detection of high-frequency impulsive faults.Therefore, optimizing VMD requires selecting the optimal combination of K and α .

2.2 Convolutional Neural Network

CNN have shown excellent performance in image recognition and sequence modeling [15]. They offer strong local perception and feature-sharing capabilities, allowing automatic extraction of deep and discriminative features from raw

signals. This approach addresses the limitations of traditional methods that depend heavily on handcrafted features and expert knowledge.

CNNs mainly consist of convolutional layers, pooling layers, and nonlinear activation functions, such as ReLU. These components together form a mechanism for local receptive fields and hierarchical feature abstraction. For a one-dimensional input sequence $x \in \mathbb{R}^n$, the convolution operation is defined as:

$$y_i = \sigma\left(\sum_{j=1}^{k} w_j \cdot x_{i+j-1} + b\right), \tag{3}$$

where $w \in \mathbb{R}^k$ denotes the convolution kernel weights, b is the bias, and k is the kernel size. The activation function $\sigma(\bullet)$, such as the Rectified Linear Unit (ReLU) [16], is defined as:

$$\sigma(x) = \max(0, x). \tag{4}$$

Pooling layers perform downsampling to reduce feature dimensionality and improve translational invariance. Mathematically, the pooling operation is defined as:

$$z_i = \max\{y_i, y_{i+1}, \ldots, y_{i+p-1}\}, \tag{5}$$

where p denotes the pooling window size and z_i is the pooled output. In this study, multiple modules combining convolution, activation, and pooling are employed to progressively extract local features at various levels.

Fault signals often exhibit a range of localized feature patterns—such as transients, modulated components, and frequency drifts—that are typically restricted to specific time intervals. CNN, leveraging local receptive fields and weight-sharing mechanisms, effectively capture these localized and non-stationary structures. This design increases the network's sensitivity to local anomalies and enhances its ability to detect incipient faults. Furthermore, the use of multi-scale convolutional kernels enhances the network's ability to extract information across various temporal scales, thereby facilitating a more comprehensive representation of complex signal characteristics.

2.3 Transformer

Transformer was originally developed for natural language processing tasks [17]. Due to its powerful sequence-modeling and parallel-computing capabilities, it has found wide applications in fields such as time-series analysis and fault diagnosis. The core of the Transformer architecture is the multi-head self-attention mechanism, which captures dependencies in different subspaces by computing multiple attention mappings in parallel.

Given an input $\boldsymbol{X} \in \mathbb{R}^{n \times d}$, the query, key, and value matrices are computed using linear projections as follows:

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{Q}}, \quad \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{K}}, \quad \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}^{\boldsymbol{V}}. \tag{6}$$

The attention scores are calculated using scaled dot-product attention:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}}\right)\boldsymbol{V}. \tag{7}$$

The outputs of multiple attention heads are concatenated and passed through a linear transformation:

$$\text{MultiHead}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)\boldsymbol{W}^{\boldsymbol{O}}. \tag{8}$$

In Equs. (6)–(8), $\boldsymbol{W}^{\boldsymbol{Q}}, \boldsymbol{W}^{\boldsymbol{K}}, \boldsymbol{W}^{\boldsymbol{V}} \in \mathbb{R}^{d \times d_k}$ are the linear projection matrices for queries, keys, and values, respectively; The matrices $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$, each of size $n \times d_k$, represent the query, key, and value vectors, respectively. d_k is the dimensionality of each attention head, and h denotes the number of heads. $\boldsymbol{W}^{\boldsymbol{O}} \in \mathbb{R}^{hd_k \times d}$ is the projection matrix applied after concatenating all attention-head outputs.The softmax() function normalizes the attention weights, and $head_i$ denotes the output of the $i-th$ attention head.

Each Transformer encoder layer comprises a multi-head attention sub-layer and a feedforward neural network (FFN) sub-layer. The FFN includes two linear transformations with a ReLU activation function applied between them, mathematically defined as:

$$\text{FFN}(x) = \max(0, x\boldsymbol{W}_1 + b_1)\boldsymbol{W}_2 + b_2 \tag{9}$$

where $x \in \mathbb{R}^{n \times d}$ represents the encoder layer input; $W_1 \in \mathbb{R}^{d \times d_{ff}}, W_2 \in \mathbb{R}^{d_{ff} \times d}$ are the FFN weight matrices, where d_{ff} indicates the hidden layer dimension. b_1, b_2 are bias terms, and $\max(\bullet)$ denotes the ReLU activation function.Each sub-layer employs residual connections followed by layer normalization, expressed as:

$$\text{Output} = \text{LayerNorm}(x + \text{SubLayer}(x)) \tag{10}$$

In this expression, $\text{SubLayer}(x)$ denotes a sub-layer transformation applied to the input x, while LayerNorm refers to the layer normalization function, which accelerates convergence and enhances model stability.

Unlike recurrent neural networks (RNNs) and long short-term memory (LSTM) networks [18], Transformers enable direct information exchange between arbitrary time steps via self-attention, effectively mitigating the gradient vanishing issue commonly seen in long-sequence training. This results in an improved capacity for modeling long-term dependencies. Furthermore, the Transformer's parallel computation mechanism greatly enhances training efficiency, making it well-suited for modeling complex long-range dependencies in non-stationary vibration signals.

2.4 Bearing Intelligent Diagnosis Model Based on VMD-CNN-Transformer

This study develops a VMD-CNN-Transformer model for intelligent rolling bearing diagnosis, comprising three key modules: VMD signal decomposition, CNN-based local feature extraction, and Transformer-based global modeling. The overall architecture is illustrated in Figure 1. First, to effectively handle the strong non-stationarity and multi-frequency modulation in rolling bearing vibration signals, the model front end applies the VMD algorithm for adaptive decomposition of the raw signals. During feature extraction, the CNN module inputs multi-scale signals reconstructed by VMD and employs multi-layer convolutional kernels and nonlinear activation functions to progressively abstract local signal features. Pooling and normalization strategies are applied to suppress overfitting and improve the robustness of local fault feature detection, including transient impacts and periodic modulations. Finally, the Transformer module receives temporal feature maps from the CNN and captures long-term dependencies across sequences using a multi-head self-attention mechanism. It also integrates positional encoding and residual connections to enhance modeling of non-stationary dynamic evolutions. With its three-level structure—signal decomposition, local feature extraction, and global modeling—this model achieves high fault identification accuracy and strong generalization in multi-condition and noisy environments. It offers an efficient and scalable intelligent solution for rolling bearing health monitoring under complex industrial conditions.

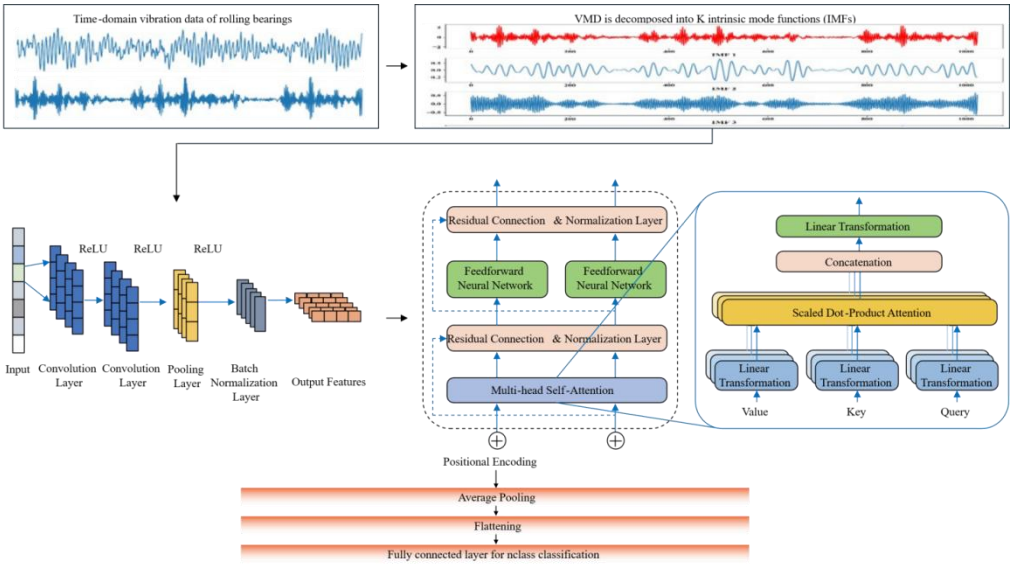


Figure 1 Bearing Intelligent Diagnosis Model Based on VMD-CNN-Transformer

3 EXPERIMENTAL DESIGN AND PERFORMANCE EVALUATION

3.1 Dataset Description

To thoroughly assess the adaptability and generalization performance of the proposed VMD-CNN-Transformer model in multi-source and multi-condition settings, three representative public rolling bearing datasets were selected. These datasets span laboratory, industrial, and high-speed aerospace application environments, as detailed below:

(1) Case Western Reserve University Bearing Dataset (CWRU Dataset) : This widely used benchmark for bearing fault diagnosis includes four fault types—Normal, Inner Race Fault (IF), Outer Race Fault (OF), and Ball Fault (BF)—all generated via electrical discharge machining. The dataset was collected under varying loads (0–3 hp) and speeds (1730–1797 rpm), using a 16-channel acquisition system at 12 kHz. A torque sensor recorded power and speed data to ensure high experimental repeatability.

(2) Xi'an Jiaotong University Bearing Dataset (XJTU Dataset) : Acquired from a bearing life-cycle test platform, this dataset includes IF, OF, BF, and Compound Fault (CF) types, with a sampling rate of 20.48 kHz. Continuous long-term monitoring enables clear degradation trends. A selected subset of the vibration signals was used to evaluate the model's robustness under noise and progressive degradation.

(3) Politecnico di Torino Aerospace Bearing Dataset (DIRG Dataset) : Designed for high-speed aerospace bearing diagnostics, this dataset was collected at 51.2 kHz under rotational speeds up to 30,000 rpm. Faults were introduced via Rockwell indentations, with severity graded from 0A (healthy) to 6A (severe). Fourteen condition signals, acquired at 200 Hz under two load scenarios, were used to assess diagnostic stability in dynamic environments.

3.2 Data Preprocessing and Experimental Setup

To ensure computational efficiency and experimental reproducibility, all experiments were conducted on a platform featuring a 13th-generation Intel® Core™ i9-13900H processor and integrated Intel® Iris® Xe Graphics. The

experimental workflow was developed using Python 3.9, with model construction and training performed via the PyTorch deep learning framework. Performance metrics were calculated using the Scikit-learn library, resulting in an end-to-end integrated pipeline for model development and evaluation.

Before training, all vibration signals were normalized using Min-Max scaling, which linearly maps feature values to the [0, 1] range. This preprocessing step minimizes the influence of feature scale differences on learning and speeds up convergence. To improve training stability and maintain evaluation independence, the dataset was divided into training (60%), validation (10%), and test (30%) sets. As shown in Figure 2, both the model's loss and classification accuracy converged rapidly within 50 epochs, demonstrating robust fitting and convergence even under complex data distributions.

Three standard classification metrics were adopted to comprehensively evaluate the model's performance. Accuracy measured overall prediction correctness, while recall evaluated the model's ability to identify positive samples. The F1-score, balancing precision and recall, was especially useful in scenarios with class imbalance. Collectively, these metrics provide a systematic evaluation of the model's generalization ability and diagnostic performance under diverse operating conditions and sample distributions.

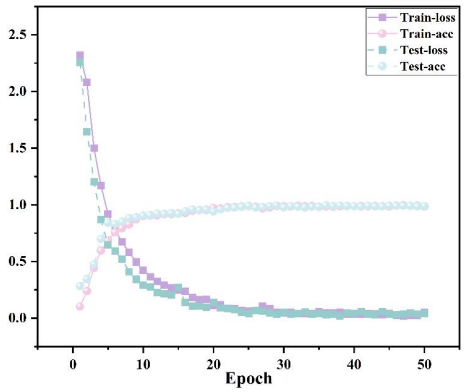


Figure 2 Convergence Curves of Loss and Accuracy During Training

3.3 Results and Analysis

3.3.1 Comparison of multi-model performance and advantage validation

To thoroughly assess the fault diagnosis performance of the proposed VMD-CNN-Transformer model, five representative baseline methods were evaluated on three publicly available bearing datasets. The baseline methods include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Multilayer Perceptron (MLP), a standard CNN, and an unoptimized CNN-Transformer model. The classification accuracies of all methods across the three datasets are summarized in Table 1.

The proposed VMD–CNN–Transformer model achieves the highest classification accuracy across all datasets, reaching 99.73%, 94.86%, and 97.96% on the CWRU, XJTU, and DIRG datasets, respectively. These results significantly outperform those of other methods, demonstrating the model's superior capability in extracting features from multi-source signals and modeling complex data distributions.

Traditional methods such as KNN and SVM consistently show lower performance across all datasets, especially on the XJTU dataset, where they achieve only 78.05% and 75.36% accuracy, respectively. These methods struggle to handle the challenges posed by complex operating conditions and variations in modal characteristics. In contrast, MLP and CNN, as representative deep neural networks, offer certain advantages in feature extraction. However, they still inadequately capture local or global features, resulting in slightly reduced performance on the DIRG dataset, with accuracies of 86.28% and 90.74%, respectively.

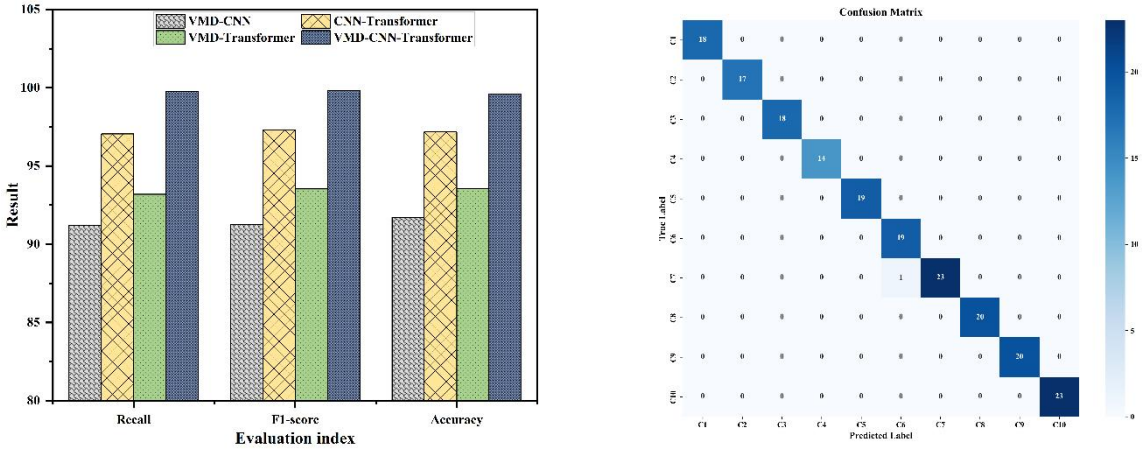
Table 1 Performance Comparison of Different Models on Three Datasets

Methods	CWRU	XJTU	DIRG
KNN	84.62	78.05	80.66
SVM	80.42	75.36	77.53
MLP	90.12	85.41	86.28
CNN	92.64	89.52	90.74
CNN-Transformer	96.52	91.93	93.46
VMD-CNN-Transformer	99.73	94.86	97.96

The CNN–Transformer model, which incorporates multi-scale convolution and attention mechanisms, performs well on all three datasets, confirming the effectiveness of the Transformer architecture in enhancing local feature awareness and modeling long-range dependencies. However, compared to the proposed VMD–CNN–Transformer model, it still shows a noticeable accuracy gap. This discrepancy is primarily due to the VMD module's ability to adaptively decompose and denoise raw signals at the input stage, thereby enhancing the network's sensitivity to critical time-frequency features and improving overall classification robustness and generalization.

3.3.2 Ablation study

To comprehensively evaluate the contribution of each component in the proposed VMD–CNN–Transformer model, ablation experiments were conducted using three simplified variants: VMD–CNN (containing only the CNN structure with VMD-decomposed signals as input), VMD–Transformer (containing only the Transformer structure with VMD-decomposed signals as input), and CNN–Transformer (which omits VMD decomposition and directly uses raw signals). All models were evaluated under identical experimental conditions and dataset configurations using three key metrics: recall, F1-score, and accuracy. The experimental results are illustrated in Figure 3(a).



(a) Results of the ablation study (b) Visualization of the confusion matrix

Figure 3 Results of the Experiment

Overall, the complete VMD–CNN–Transformer model achieved superior performance over all simplified variants, with recall, F1-score, and accuracy reaching 99.76%, 99.83%, and 99.61%, respectively. These results highlight the model's synergistic advantages in feature extraction, fault sensitivity, and global recognition capabilities. Furthermore, confusion matrices were utilized to provide a more intuitive evaluation of the model's diagnostic performance, as shown in Figure 3(b).

In the structural component analysis, the VMD–CNN model exhibited the lowest performance across all three metrics. This indicates that while VMD offers basic time-frequency decomposition, its integration with a shallow CNN is inadequate for capturing deep patterns and long-range dependencies present in complex fault signals. By contrast, the CNN–Transformer model showed notable performance improvements owing to the attention mechanism, confirming the Transformer's effectiveness in enhancing feature representation and capturing global temporal dependencies. However, the absence of a front-end decomposition process limits its capability to suppress high-frequency noise and address local ambiguities in raw signals.The VMD–Transformer model, excluding the CNN module, still achieved relatively strong performance. This result suggests that VMD plays a critical role in enhancing signal separability and mitigating feature aliasing. It also highlights the Transformer's ability to effectively integrate high-quality time-frequency features extracted through VMD processing.

In summary, the VMD module strengthens the model's capacity to extract key frequency components, the CNN module enhances local spatial feature learning, and the Transformer significantly improves modeling of temporal dependencies. The integration of these modules in the VMD–CNN–Transformer model yields optimal performance across multiple evaluation metrics, demonstrating superior robustness and generalization under complex operating conditions. These findings validate the rationality and complementarity of each module, offering theoretical support for model design and a practical architectural reference for real-world fault diagnosis systems.

4 CONCLUSIONS

To address the non-stationary and nonlinear characteristics of rolling bearing vibration signals, and to capture their global temporal dependencies and deep fault patterns, this paper proposes an intelligent diagnostic framework based on VMD–CNN–Transformer. The proposed method significantly improves diagnostic accuracy and robustness under high noise interference. The main conclusions are as follows:

(1)The model utilizes Variational Mode Decomposition (VMD) to adaptively decompose raw signals, enhancing fault-relevant components and suppressing redundant noise, thereby improving the quality of signal representation. During feature extraction, a Convolutional Neural Network (CNN) module captures local time-frequency features of the vibration signals, while a multi-scale fusion strategy further enriches hierarchical feature representations. Additionally, a Transformer module models long-range dependencies in temporal sequences, enabling deep modeling and accurate identification of complex fault patterns.

(2)The proposed model is trained and evaluated on three real-world bearing datasets. Performance is comprehensively evaluated using classification accuracy, recall, F1-score, and confusion matrices. The results confirm the model's high diagnostic accuracy and robustness under diverse conditions.

(3) Comparative experiments are conducted between the proposed VMD–CNN–Transformer and several state-of-the-art fault diagnosis methods. Results show that the proposed model surpasses others in fault identification accuracy and stability, highlighting its broad adaptability and application potential in practical engineering scenarios.

The VMD–CNN–Transformer effectively extracts key features and captures deep temporal representations of sequential data, achieving highly accurate fault identification for rolling bearings even under heavy noise interference. However, in real-world industrial applications, the lack of high-quality, accurately labeled training samples remains a major barrier to large-scale model deployment. Future research should therefore focus on leveraging operational and maintenance data from existing equipment to develop efficient and reliable diagnostic models.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Yang Y, Zhai J, Wang H, et al. An Improved Fault Diagnosis Method for Rolling Bearing Based on Relief-F and Optimized Random Forests Algorithm. Machines, 2025, 13(3): 183-183.

[2] Yan H, Shang L, Chen W, et al. An adaptive hierarchical hybrid kernel ELM optimized by aquila optimizer algorithm for bearing fault diagnosis. Scientific Reports, 2025, 15(1): 11990-11990.

[3] Liao W, Fu W,Yang K, et al. Multi-scale residual neural network with enhanced gated recurrent unit for fault diagnosis of rolling bearing. Measurement Science and Technology, 2024, 35(5).

[4] Feisa T T, Gebremedhen S H, Kibrete F, et al. One-Dimensional Deep Convolutional Neural Network-Based Intelligent Fault Diagnosis Method for Bearings Under Unbalanced Health and High-Class Health States. Structural Control and Health Monitoring, 2025, 2025(1): 6498371-6498371.

[5] Bharatheedasan K, Maity T, Kumaraswamidhas L, et al. Enhanced fault diagnosis and remaining useful life prediction of rolling bearings using a hybrid multilayer perceptron and LSTM network model. Alexandria Engineering Journal, 2025: 115355-369.

[6] Li X, Ma J,Wu J, et al. Transformer-based conditional generative transfer learning network for cross domain fault diagnosis under limited data. Scientific Reports, 2025, 15(1): 6836-6836.

[7] Zhilin D, Dezun Z, Lingli C. An intelligent bearing fault diagnosis framework: one-dimensional improved self-attention-enhanced CNN and empirical wavelet transform. Nonlinear Dynamics, 2024, 112(8): 6439-6459.

[8] Sahu D, Dewangan K R, Matharu S P S. Hybrid CNN-LSTM model for fault diagnosis of rolling element bearings with operational defects. International Journal on Interactive Design and Manufacturing (IJIDeM), 2024: 1-12.

[9] Sarunyoo B, Pradit F, Chitchai S, et al. Adaptive meta-learning extreme learning machine with golden eagle optimization and logistic map for forecasting the incomplete data of solar irradiance. Energy and AI, 2023: 13.

[10] XiaX, WangX, ChenW. A Hybrid Fault Diagnosis Model for Rolling Bearing With Optimized VMD and Fuzzy Dispersion Entropy. International Journal of Rotating Machinery, 2025, 2025(1): 7990867-7990867.

[11] Chen H,Yu Y . Acoustic Emission Diagnosis of Rolling Bearing Faults based on Optimized VMD-Transformer. Frontiers in Computing and Intelligent Systems, 2025, 11(3): 19-24.

[12] Wang Y, Zhu K, Wang X, et al. An extended iterative filtering and composite multiscale fractional-order Boltzmann-Shannon interaction entropy for rolling bearing fault diagnosis. Applied Acoustics, 2025: 236110699-110699.

[13] Dragomiretskiy K, Zosso D. Variational Mode Decomposition. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.

[14] Liu T, Diao F, Yao W, et al. Study on Motion Response Prediction of Offshore Platform Based on Multi-Sea State Samples and EMD Algorithm. Water, 2024, 16(23): 3441-3441.

[15] Chen Q, Zhang F, Wang Y, et al. Bearing fault diagnosis based on efficient cross space multiscale CNN transformer parallelism. Scientific Reports, 2025, 15(1): 12344-12344.

[16] Layton W O, Peng S, Steinmetz T S. ReLU, Sparseness, and the Encoding of Optic Flow in Neural Networks. Sensors, 2024, 24(23): 7453-7453.

[17] Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. arXiv, 2017.

[18] Zhang F, Yin J, Wu N, et al. A dual-path model merging CNN and RNN with attention mechanism for crop classification. European Journal of Agronomy, 2024: 159127273-127273.