Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 9, No. 2, 2026

Overview of Rolling Bearing Fault Diagnosis Technology Based on Variational Mode Decomposition

Tao Zhang, Lianhai Wang, Yang Yu *

School of Shipping, Shandong Jiaotong University, Weihai 264209, China

Abstract: As a core component of rotating machinery, rolling bearings are directly related to the safety and reliability of the equipment. Variational Mode Decomposition (VMD) has become a research hotspot in the field of rolling bearing fault diagnosis due to its solid mathematical theoretical foundation, good noise immunity and adaptive decomposition capabilities. Based on 9 representative research papers published in recent years, this paper systematically reviews the application status and technical evolution of VMD in rolling bearing fault diagnosis. Key technical aspects such as adaptive optimization of VMD parameters, intelligent screening of intrinsic mode functions (IMF), multi-domain feature extraction, and deep learning fusion classification are analyzed. This paper compares in detail the performance differences of intelligent algorithms such as Gray Wolf Optimization (GWO), Improved Mustang Algorithm (IWHO), and Beluga Whale Optimization (BWO) in VMD parameter optimization; deeply discusses the application scenarios of IMF screening strategies such as kurtosis criterion, effective weighted correlation sparsity (EWCS), and multi-entropy fusion; comprehensively evaluates the diagnostic effects of deep learning models such as CNN, GRU, and Transformer with VMD fusion. The accuracy, robustness and computing efficiency of various methods are verified through experimental comparisons with multiple datasets such as CWRU, XJTU-SY, and DDS test benches. Finally, based on various research results, the current technical bottlenecks are pointed out, and future development directions such as lightweight deployment, interpretable AI, and cross-operating condition generalization are prospected, providing systematic technical reference for intelligent fault diagnosis of rolling bearings.

Keywords: Rolling bearing, fault diagnosis, variational modal decomposition, parameter optimization, feature extraction, deep learning.

1. Introduction

Rolling bearings are the most widely used key basic components in rotating machinery. They assume the core functions of supporting rotating bodies and transmitting power in aerospace, rail transportation, wind power generation, intelligent manufacturing and other fields. According to statistics, about 30% of rotating machinery failures are caused by bearing failures, which may lead to equipment shutdown, major property losses and even casualties. With the advancement of Industry 4.0 and intelligent manufacturing, early fault warning and accurate diagnosis of rolling bearings has great engineering value for ensuring production safety and reducing operation and maintenance costs. The operating environment of rolling bearings is complex and harsh, and they endure alternating loads, high-speed rotation and strong impact vibration for a long time. Their vibration signals show obvious nonlinear and non-stationary characteristics. Early fault features are weak and easily submerged by strong background noise, making it difficult for traditional signal processing methods to effectively extract fault features. How to accurately extract fault characteristics and achieve intelligent classification under complex working conditions has become a core challenge in the field of fault diagnosis.

Limitations of traditional signal processing methods. Early research mainly relied on traditional methods such as empirical mode decomposition (EMD) and wavelet transform. However, these methods have significant limitations: Although empirical mode decomposition (EMD) can adaptively decompose non-stationary signals, it has inherent problems such as modal mixing, endpoint effects and lack of mathematical theoretical support. Xiang Mingsheng and others pointed out that the decomposition effect of EMD is seriously distorted in strong noise environments, and some intrinsic mode functions (IMF) contain multiple time-scale components, resulting in the inability to extract effective fault features. Integrated empirical mode decomposition (EEMD) alleviates modal aliasing by adding auxiliary noise, but introduces new problems such as reconstruction errors and large computational complexity. Research by Li Cuisheng and others showed that the computational complexity of EEMD increased significantly, and the residual noise was difficult to completely eliminate. Although wavelet transform can extract weak features, the selection of basis functions relies on expert experience and is poor in adaptability, making it difficult to adapt to complex working conditions such as variable speeds and variable loads. The inherent flaws of these methods limit their reliable application in engineering sites, and more advanced signal decomposition technology is urgently needed.

Technical advantages of Variational Mode Decomposition (VMD). In 2014, the Variational Mode Decomposition (VMD) proposed by Dragomiretskiy et al. provided a new paradigm for fault diagnosis. VMD transforms signal decomposition into a constrained variational problem solution. By constraining the frequency bandwidth of each mode, the original signal is decomposed into K intrinsic mode functions (IMFs) with limited bandwidth. Compared with EMD-based methods, VMD has the following outstanding advantages: 1. Rigorous mathematical theory: Based on variational theory and Wiener filtering, EMD's recursive mode mixing problem is overcome. Each IMF has a clear center frequency and bandwidth, and its physical meaning is clear; 2. Superior anti-noise performance: The modal bandwidth is controlled by the penalty factor α, which is highly robust to noise, especially

suitable for strong background noise environments; 3. Mode-less mixing: Each mode is separated in the frequency domain, effectively avoiding modal mixing in EMD; 4. High computational efficiency: the alternating direction multiplier method (ADMM) is used to solve the problem iteratively, and the convergence speed is faster than that of EMD-based methods. He Xiaoliang [1] et al. pointed out that VMD performs well in fault feature extraction, but its decomposition effect relies heavily on two key parameters: the modal number K and the penalty factor α. The value K determines the number of decomposed modes. Too small will lead to underdecomposition, and too large will produce false modes; α controls the modal bandwidth, and too small will cause modal mixing, and too large will lose local information. Traditional trial and error methods or empirical setting are inefficient and highly subjective, and adaptive optimization of parameters has become a core issue in VMD engineering applications. Trend of integration of intelligent optimization and deep learning In recent years, swarm intelligent optimization algorithms (such as Gray Wolf Optimization GWO, Improved Mustang Algorithm IWHO, Beluga Whale Optimization BWO, etc.) have been introduced into VMD parameter optimization to achieve adaptive parameter selection. At the same time, deep learning models (CNN, GRU, Transformer, etc.) rely on their powerful feature learning capabilities to avoid the shortcoming of traditional methods relying on expert experience. Combining optimized VMD with deep learning to build an end-to-end diagnostic framework of "signal decomposition-feature screening-intelligent classification" has become the mainstream research direction.

2. Intelligent Optimization Methods for VMD Parameters

2.1. The Necessity and Challenge of Parameter Optimization

VMD performance is highly sensitive to parameters. Fixed parameters are difficult to adapt to different working conditions and fault types. He Xiaoliang and others pointed out through research that in different application scenarios such as CNC machine tools, MEMS sensors, and gearboxes, the common optimization range of the penalty factor α is 100-10,000. However, considering the high-frequency noise interference in bearing fault signals, it is necessary to shrink α to [500, 4000] in practical applications to balance calculation efficiency and decomposition quality. The value of the modal number K depends on the number of potential fault characteristics in the signal. Early faults of rolling bearings usually contain 1-4 main impact components. The effective range of K is mostly [2, 10], and composite faults may need to be expanded to [1, 15]. Traditional grid search requires a huge amount of computation. Intelligent optimization algorithms can efficiently find optimizations in high-dimensional nonlinear solution spaces by designing fitness functions. Current research mainly focuses on two core issues: fitness function design and optimization algorithm selection.

2.2. Application of Gray Wolf Optimization Algorithm (GWO)

GWO was proposed by Mirjalili et al. It simulates the social hierarchy and hunting behavior of wolf packs. It is divided into four levels: α, β, δ, and ω. It is optimized by tracking, enclosing, and attacking prey. It has strong global search capabilities, few parameters, and fast convergence speed. Jiang Liying [2] et al. research: GWO was used to optimize VMD parameters, using permutation entropy as the fitness function. The permutation entropy reflects the complexity of the signal. The more noise components, the greater the entropy value. The optimization goal is to minimize the permutation entropy to obtain the best denoising effect. The experimental termination condition was that the fitness difference between the front and back was less than 0.0001, and the minimum permutation entropy of 0.356115 was obtained in the second generation, corresponding to the optimal parameter combination (K=6, α=2308.64). Compared with traditional EMD, VMD and GA-VMD, GWO-VMD achieves a higher signal-to-noise ratio (28.89dB) and a lower mean-square error (0.0075), verifying the superiority of the GWO algorithm. Jiang Yonghua [3] et al. research: Taking the minimum average envelope entropy (MAEE) as the fitness function, MAEE is defined as the average value of envelope entropies of each modal component. The envelope entropy characterizes the sparse nature of the signal. The smaller the entropy value, the more regular the fault pulse. After GWO optimization, the optimal parameter combination (K=11, α=2958) is obtained for the bearing inner ring fault signal. Comparative experiments show that the parametric optimization VMD decomposition effect is much better than that of fixed parameters (K=5, α=1000), and the fault characteristic frequency extraction is clearer. Analysis of advantages of GWO algorithm: Coefficient matrices A and C change dynamically with iteration, |A| Expand the search range at 1:00, |A| When ≤1, local fine search takes into account exploration and development; in VMD parameter optimization, convergence usually takes 20-30 iterations, and the calculation efficiency is high; the standard deviation of 20 independent experiments is less than 0.5%, which is highly robust.

2.3. Improved Mustang Algorithm (IWHO) Innovation

The Wild Horse Optimization Algorithm (WHO) was proposed by Naruei et al. to simulate the behaviors of horse herds such as grazing, mating, and territory division. Aiming at the problem of a single population initialization by WHO, He Xiaoliang et al. introduced Tent chaotic mapping to improve population diversity and proposed the IWHO algorithm. Improvement strategy: Tent chaotic map is used to generate the initial population to avoid local concentration problems caused by random initialization and ensure population diversity; through the "dynamic territory division" and "individual differentiated search" mechanisms, the first horse information transmission is combined with the update of the group position, global and efficient exploration of parameter space; a search strategy based on group collaboration quickly converges to a global optimal solution through the first horse guidance. Experimental results: On the rolling bearing data set, IWHO converges within 40 iterations, which is much lower than the maximum number of iterations of 100. After optimization, the optimal parameters for rolling element faults are K=10, α=3000, and the inner ring faults are K=7, α=1000. The diagnostic accuracy rate of IWHO-VMD-SVM reaches 99.17%, which is 5.31% higher than that of unoptimized VMD, and is better than PSO and WOA algorithms. In a simulated noise interference environment, the diagnostic accuracy of 95.8% is still maintained at SNR=5dB.

Performance comparison: IWHO has advantages over GWO in balancing search accuracy and efficiency, and is especially suitable for solving complex, multi-constraint nonlinear problems such as VMD parameter optimization.

2.4. Characteristics of Beluga Whale Optimization Algorithm (BWO)

The Beluga Whale Optimization Algorithm (BWO) is a new meta-heuristic algorithm proposed in 2025. It simulates the migration, predation and whale landing behaviors of beluga whales, and has an exploration-development balance mechanism. Li Chaoyang [4] et al research: Using BWO to optimize VMD parameters, using the minimum envelope entropy as the fitness function. Set the number of optimization variables to 2, the penalty parameter $\alpha \in [100, 2400]$, the number of modes $K \in [3, 10]$, the maximum iteration is 50, and the population size is 10. For the 0.007-inch inner ring fault signal, BWO converges to 7.3563 in the 9th iteration, corresponding to the optimal parameter (K=6, α=1461). Algorithm features: Levy flight strategy is introduced to improve local search capabilities and avoid premature convergence; whale landing mechanism simulates the natural selection process to enhance global optimal capture capabilities; balance factor $B=2\times rand\times(1-t/T)$ dynamically adjusts exploration and development stage. Application effect: BWO-VMD combined with the CNN-Transformer dual-channel model has an accuracy rate of 99.1% on the CWRU dataset, which is significantly better than CNN (92.2%), Transformer (94.1%) and CNN-BiLSTM (96.2%), proving that the BWO optimization effect is outstanding.

2.5. Comprehensive Comparison of Algorithm Performance

Table 1. Performance comparison of intelligent optimization algorithms

Algorithm	Convergence speed	Global search capability	Parameter Settings	Typical Accuracy	Robustness
GWO	Fast (ages 20-30)	Strong	Simple (population size, number of iterations)	99.1%	High (standard deviation <0.5%)
IWHO	Relatively fast (<40s age group)	Relatively strong	Relatively simple (PS, G/N)	99.17%	High (95.8% at SNR=5 dB)
BWO	Fast (<10s of generations)	Strong	Simple (population size, number of iterations)	99.1%	High (significant improvement with hybrid model)
PSO	Medium (50-100 generation)	Weaker	Simple (inertia weight, learning factor)	95.83%	Medium
GA	Slow (>100 generations)	Medium	Complex (crossover rate, mutation rate)	94.17%	Lower

Conclusion: GWO, IWHO and BWO perform well in VMD parameter optimization, with fast convergence speed and strong global search ability. They are the mainstream choices in current research. Among them, IWHO and BWO, as emerging algorithms, are slightly better than GWO in convergence speed and accuracy, and have simple parameter settings and stronger engineering applicability.

3. Feature Extraction and IMF Screening Strategies

3.1. The Necessity and Challenges of IMF Screening

The VMD is decomposed to produce K IMFs, but not all components contain valid fault information. Some IMFs are dominated by false modes or noise, and direct input into the classifier will increase the computational burden and reduce accuracy. Therefore, screening sensitive IMFs is a key step in feature extraction. The main challenge facing current research is how to objectively and adaptively evaluate the fault information content of each IMF and avoid subjective threshold setting.

3.2. Kurtosis Criterion Screening Method

Kurtosis is a dimensionless parameter that is highly sensitive to impact signals and can effectively identify surface damage faults. He Xiaoliang et al. used kurtosis criterion to select the top 5 IMFs with the largest kurtosis values, and calculated time-frequency domain features to build a 13-dimensional feature vector. Among the four categories of tasks of rolling element, outer ring, inner ring failure and normal state, the diagnosis accuracy rate reaches 99.17%. Under simulated noise interference, the accuracy rate is 98.30% when SNR=20dB, and still remains 95.8% when SNR= 5 dB, reflecting strong robustness. Advantages: Simple calculation, sensitive to early impact faults, and good real-time performance; Disadvantages: It mainly reflects impact strength, and may ignore wear-like weak periodic fault characteristics; it is sensitive to noise interference.

3.3. Effective Weighted Correlation Sparsity Indicator (EWCS)

In view of the limitations of a single indicator, Jiang Yonghua and others proposed that EWCS comprehensively evaluate IMF components and integrate the two dimensions of sparsity and correlation coefficient. This indicator first calculates the sparsity of each modal component (reflecting the degree of signal energy concentration) and the correlation coefficient with the original signal, then weights and combines it to obtain a comprehensive evaluation value, and finally subtracts the average value to obtain the effective indicator. When the effective indicator is greater than 0, it is determined to be sensitive IMF. Experimental verification: On the DDS test bench, GWO optimized the VMD, calculated the EWCS values of each IMF, and selected IMF1-IMF4 as effective components for reconstruction. Envelope spectrum analysis clearly shows that the fault characteristic frequency is 162Hz, but fixed parameter VMD or single correlation coefficient-kurtosis method cannot effectively extract this frequency. The final CNN classification accuracy rate reaches 98.61%, which is significantly better than traditional methods. Advantages: Fusion sparsity and correlation, strong ability to identify false modes, and high robustness; Calculation cost: slightly higher than a single kurtosis indicator, but still acceptable.

3.4. Envelope Spectrum Kurtosis Screening Method

For strong noise environments, Shi Yifei [5] and others proposed using the square envelope spectral kurtosis (KSES) criterion to select the main modal components. This indicator can characterize the cyclostationarity of bearing faults by calculating high-order statistics of the envelope spectrum. The larger the value, the more regular the fault impact. In the -14dB strong noise simulation signal, the envelope spectrum of the maximum IMF of KSES can clearly display the fault frequency and its 2-5 times, while the smaller mode of KSES cannot extract effective features. Application effect: Shi Yifei and others adopted POFMD (Parameter Optimization Characteristic Mode Decomposition, the optimization idea is similar to that of VMD) and selected the main modal component based on the principle of maximizing KSES. In simulation signals, CWRU public data sets and high-speed train wheel set test bench data, they are significantly superior to VMD, MCKD and spectral kurtosis (SK) methods. The KSES value is increased by more than 30%, and the fault feature extraction ability is the strongest.

3.5. Multi-entropy Fusion Feature Extraction Method

Aiming at the problem of weak fault signals and difficult feature extraction of wind turbine bearing fault signals, Zhang Tianrui [6] and others used CEEMDAN decomposition to calculate the approximate entropy, permutation entropy and fuzzy entropy of each IMF, and constructed a multi-entropy fusion matrix. Complementarity of the three entropies: Approximate entropy: quantitatively describes the complexity of the time series, is insensitive to the amount of data, and the results are stable; Permutation entropy: Introduce the concept of permutation, which has high computing efficiency, strong anti-noise ability, and sensitivity to signal mutations; Fuzzy entropy: The fuzzy membership function is integrated into the sample entropy to more intuitively reflect the internal characteristics of the signal. Experimental results: In the 10 operating states of the CWRU data set, a single entropy value has characteristic aliasing problems (for example, the approximate entropy of an inner ring fault of 0.1778mm is similar to that of a rolling element fault). After multi-entropy fusion, the diagnosis accuracy of MSCNN reaches 99.8%, which is 2.6% higher than the single fuzzy entropy (97.2%), 5.8% higher than the traditional CNN (94.0%), and 11.3% higher than the SVM (88.5%), which fully proves the effectiveness of multi-entropy fusion.

4. Deep Learning Fusion Diagnosis Model

4.1. Support Vector Machine (SVM) Applications

SVM is based on statistical learning theory and has high accuracy under limited samples. It maps features to high-dimensional space through a kernel function to construct an optimal classification surface. Research by He Xiaoliang and others: Build a 13-dimensional feature vector from the time-frequency domain features of the optimized IMF, and input them into SVM to achieve four-classification. The diagnostic accuracy rate reaches 99.17% on the CWRU-style dataset, and still maintains 95.8% accuracy at SNR=5dB noise, demonstrating strong robustness. Advantages: Strong generalization ability under small samples, fast training speed, and simple and interpretable models; Limitations: The calculation complexity is high under large-scale data, and the ability to identify multiple faults is limited; feature engineering relies on manual design and is difficult to mine deep nonlinear relationships.

4.2. Convolutional Neural Network (CNN) and Its Improvements

CNN automatically extracts local features through convolutional layers to avoid manual design and achieve end-to-end diagnosis. In recent years, research has mainly focused on network structure optimization and multi-scale feature extraction.

4.2.1. Standard CNN Structure

Jiang Liying and others built a three-layer CNN: Conv1(16×3×3)→Conv2 (32×3×3) →MaxPool→ FC (750) → FC(10). The input is a 2048×1 feature vector, and the accuracy rate of identifying four types of faults on the DDS test bench reaches 100%. Key parameters: Small size convolution kernels extract details, large size convolution kernels suppress high-frequency noise, maximum pooling retains significant features, and ReLU activation function accelerates convergence.

4.2.2. Multi-scale CNN (MSCNN)

Bearing vibration signals have multi-scale characteristics, and it is difficult for a single-scale convolution kernel to fully capture the characteristics. Current research mainly achieves multi-scale extraction through parallel multi-branch structures. Research by Chen Chao [7] and others: designed the MSCNN-GRU model, which includes three branches (64×1, 32×1, and 16×1 convolution kernels). Each branch is followed by a GRU to extract timing dependencies, and then expanded the receptive field through hole convolution (DR= 1, 2, 7). On the CWRU dataset, the noise-free accuracy rate is 100%, and still reaches 86.75% at -6dB, which is significantly better than WDCNN (44.7%) and ResNet18 (47.88%). Zhang Xiong [8] et al research: Build a 2D-CNN-GRU model, use three-scale convolution layers (20×20×32, 9×9×64, 4×4×128) to extract features in parallel, reduce dimensions through global average pooling, and GRU captures timing dependence. The accuracy rate on the XJTU-SY dataset reaches 98%, with only 191,149 parameters, which is much lower than AlexNet's 21.6 million, making it suitable for edge deployment.

4.2.3. CNN-Transformer Mixed Model

Transformer's self-attention mechanism can capture the global dependence of sequences and make up for the limitations of CNN's local receptive fields, making it a recent research hotspot. Li Chaoyang et al research: Constructing a CNN-Transformer dual-channel model. CNN channel extracts local time-frequency features; Transformer channel captures long-range dependence using multi-head self-attention encoder; outputs of two channels are spliced and sent to classifier. The accuracy of 99.1% on CWRU dataset is significantly better than that of CNN (92.2%), Transformer (94.1%) and CNN-BiLSTM (96.2%), demonstrating the effectiveness of two-channel fusion. 4.2.4 GRU Timing Modeling Applications

As a lightweight variant of LSTM, Gated Cyclic Unit (GRU) reduces training parameters and speeds up convergence, while maintaining modeling capability for long sequences. Chen Chao et al: Add GRU after each branch of MSCNN to fully extract time series features of different

scales. Experiments show that after adding GRU, the accuracy of the model is improved by more than 15% under-6dB noise. Zhang Xiong et al: Input the spatial features extracted by 2D-CNN into the GRU layer, integrate the spatial and time series information, and the composite fault classification accuracy rate reaches more than 98% under various working conditions.

Table 2. Comprehensive comparison of classifier performance

model	Core mechanism	Accuracy	Noise resistance	Computational efficiency	Applicable Scenarios
SVM	Statistical learning theory, kernel function mapping	99.17%	Strong	Tall	Small sample, binary/quadruple classification
CNN	Convolutional feature extraction, end-to-end	100%	Medium	Tall	Large-scale data, single operating condition
MSCNN-GRU	Multi-scale Temporal Modeling	100%	Extremely strong	Medium	Strong noise, changing working conditions
CNN-Transformer	Local and global feature fusion	99.1%	Strong	Medium	Complex faults, long sequences

5. Performance Evaluation and Future Development

5.1. Experimental Verification and Performance Evaluatiom

Various studies have verified the effectiveness of methods using multiple datasets such as CWRU, XJTU-SY, and DDS test benches.

Noise resistance: Tests by He Xiaoliang et al. show that IWHO-VMD-SVM still achieves an accuracy of 95.8% at SNR=5dB. Comparisons by Chen Chao et al. indicate that MSCNN-GRU maintains 86.75% accuracy at -6dB, with performance degradation of only 13.25%, while models like WDCNN and ResNet18 degrade by over 50%, demonstrating the strong noise resistance of the multi-scale GRU structure. Key strategy: Dropout preprocessing at the input layer allows the model to train with a small number of features, improving noise resistance by more than 20%.

Cross-condition adaptability: The MICNN model proposed by Li Qing [9] et al. achieves an average accuracy of 98.74% on the PU dataset, 96.59% on the JNU dataset, and 93.81% on a self-built dataset in single-source domain generalization tasks, significantly outperforming traditional WDCNN (80.17%). The core lies in the collaborative normalization (CNM) and multi-dimensional spatiotemporal feature fusion (MSTFF) modules, enhancing the extraction of domain-invariant features. Variable-speed tests by Chen Chao et al. show that training with a small number of cross-condition samples can increase accuracy from 80% to over 90%.

Few-shot learning: MSCNN-GRU achieves over 80% accuracy with only 20 samples per fault type (10% proportion); when the sample size increases to 60 (30%), accuracy exceeds 90%. In contrast, ResNet18 achieves less than 40% accuracy with 20 samples, indicating poor few-shot learning ability. Key strategies: overlapping sampling data augmentation, dropout regularization, and global average pooling reduce the number of parameters and effectively prevent overfitting.

Computational efficiency: The 2D-CNN-GRU model by Zhang Xiong et al. has only 191,149 parameters and a training time of 609 seconds, much less than AlexNet. Chen Chao et al.'s MSCNN-GRU has only 71,550 parameters and faster inference speed. Parameters can be reduced after VMD optimization, and GWO optimization by Jiang Yonghua et al. shortens decomposition time by 30%.

5.2. Technical Challenges and Limitations

Parameter optimization: Swarm intelligence algorithms have the risk of local optima, high computational costs, and lack a unified standard for the selection of fitness functions. Shi Yifei et al. pointed out that although WGISES performs well under strong noise, it is insufficiently sensitive to weak faults.

IMF selection: The kurtosis criterion may overlook weak periodic features; threshold settings in EWCS and KSES are subjective; interaction information between modes is not utilized. He Xiaoliang et al. clearly indicated that kurtosis is sensitive to impact features but may overlook weak periodic wear-related fault information.

Deep learning models: There is severe overfitting with small samples, poor interpretability, and insufficient generalization across devices. Chen Chao et al.'s experiments showed that ResNet18 achieves less than 40% accuracy with small samples and has poor generalization.

Complex working conditions: Diagnosis is difficult under variable speed and compound faults, and performance drops sharply under strong noise interference (SNR<-10dB). Li Qing et al. mentioned that current methods mainly target constant speed conditions, and performance declines significantly under variable speed.

Engineering applications: Online real-time performance is insufficient, data labeling is costly, and there is a lack of unified performance evaluation standards. Jiang Liying et al. noted that a single diagnosis takes several seconds, which cannot meet real-time warning requirements.

5.3. Future Development Trends

Parameter Optimization Innovation: Hybrid optimization strategies (GWO+Simulated Annealing), meta-learning for rapid adaptation, adaptive fitness function design. Yonghua Jiang et al. suggest exploring GWO+SA hybrid algorithms to enhance global search and local refinement capabilities.

IMF Screening Innovation: Multi-indicator fusion decision system, attention mechanism for automatic weight learning, graph neural network modeling of modal correlations. Xiaoliang He et al. suggest studying adaptive weight allocation mechanisms to avoid bias from single indicators.

Deep Learning Evolution: Lightweight designs (knowledge distillation, pruning, quantization), self-supervised pretraining, physics-informed integration, federated learning architectures. Xiong Zhang et al. suggest exploring GCNs to mine higher-order interactions between modalities.

Cross-Condition Diagnosis: Domain generalization techniques (MixStyle, CNM), decoupled representation learning, multi-task joint learning. Qing Li et al. note that MICNN performs well in single-source domain generalization, and in the future could be extended to multi-source domains.

Digital Twin Integration: Construct bearing digital twin models, generate virtual samples to augment data, perform online parameter corrections, manage predictive maintenance in a closed loop, achieving intelligent full-lifecycle management.

6. Conclusion

This paper systematically reviews VMD-based rolling bearing fault diagnosis technology, with the main conclusions as follows:

(1) VMD has obvious advantages: Compared with EMD-type methods, it performs outstandingly in mathematical rigor, noise resistance, and computational efficiency, but parameter sensitivity is the main obstacle.

(2) Intelligent optimization is highly effective: Algorithms such as GWO, IWHO, and BWO achieve adaptive parameter selection, generally improving diagnostic accuracy by more than 5%. IWHO and BWO converge faster and are suitable for real-time applications.

(3) IMF selection requires comprehensive evaluation: The kurtosis criterion is efficient for impact faults, EWCS and KSES are more robust under strong noise, and multi-entropy fusion can comprehensively characterize signal features.

(4) Deep learning integration is highly effective: VMD CNN/Transformer frameworks generally achieve over 98% accuracy on benchmark datasets, and MSCNN-GRU still maintains 86.75% under -6dB, showing excellent noise resistance.

(5) Engineering bottlenecks need to be overcome: Real-time performance, small sample size, and cross-device generalization are the main challenges, and lightweight models and self-supervised learning are important directions.

In the future, VMD-based intelligent fault diagnosis will evolve towards adaptiveness, interpretability, lightweight design, and full lifecycle management, providing reliable technical support for smart manufacturing and Industry 4.0.

References

[1] He, X., Su, C., & Zhang, Y. (2025). Fault diagnosis of rolling bearings based on improved VMD and SVM methods. Journal of Southeast University (Natural Science Edition), 55(9), 1–13.

[2] Jiang, L., Wang, T., & Cui, J. (2023). Fault diagnosis of rolling bearings based on parameter optimization VMD and CNN. Journal of Shenyang University of Aeronautics and Astronautics, 40(3), 37–43.

[3] Jiang, Y., Xu, C., & Chen, H. (2023). Fault diagnosis method of rolling bearings based on improved VMD and CNN. Journal of Zhejiang Normal University (Natural Science Edition), 46(3), 265–273.

[4] Li, C., Li, X., & Sun, Y. (2025). Research on fault diagnosis of rolling bearings based on beluga optimization VMD algorithm and CNN-Transformer. Machinery Management and Development, 40(12), 79–82.

[5] Shi, Y., Huang, Y., & Wang, F. (2024). Fault diagnosis of rolling bearings under strong background noise based on parameter optimization feature mode decomposition. Vibration & Shock, 43(21), 108–116.

[6] Zhang, T., & Zhou, L. (2025). Fault diagnosis method of wind turbine bearings based on multi-entropy fusion and multi-scale convolutional neural network. Journal of Solar Energy, 46(7), 429–438.

[7] Chen, C., Guo, J., & Qu, H. (2025). Fault diagnosis of rolling bearings based on multiscale convolutional neural network and gated circulation unit. Bearings, (9), 109–119.

[8] Zhang, X., Qu, W., & Wang, W. (2025). Multi-fault diagnosis model of rolling bearings based on improved CNN-GRU model. Mechanical and Electrical Engineering, 42(5), 1–12.

[9] Li, Q., Zhou, G., & Zhou, P. (2025). Fault diagnosis of bearings cross-condition based on multi-dimensional information collaborative neural network. Vibration and Shock, 44(13), 29–38.