E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.202602124134

# A Review of Industrial Fault Diagnosis Technologies Based on Machine Learning

**Siying Chen***

*College of Artificial Intelligence and Low-Altitude Technology, South China Agricultural University, Guangzhou 510000, China*

**Corresponding author: Siying Chen.*

## Abstract

The stable operation of industrial equipment is a fundamental guarantee for intelligent manufacturing, and fault diagnosis technologies have evolved from experience-driven and model-driven approaches to data-driven paradigms based on machine learning. In response to challenges such as nonlinearity, strong coupling, and the scarcity of fault samples in industrial systems, machine learning has become a core approach for fault diagnosis due to its advantages in feature learning and pattern recognition. This paper systematically reviews the research progress of machine learning in industrial fault diagnosis. From the perspective of technical frameworks, it elaborates on the application principles and engineering performance of traditional machine learning, deep learning, and improved algorithms. Furthermore, it analyzes algorithm adaptability in typical industrial scenarios, identifies key challenges such as small-sample diagnosis and model interpretability, and finally discusses future development trends from the perspectives of algorithm integration and data augmentation. This study aims to provide references for technological optimization and engineering applications in this field.

## Keywords

machine learning, industrial fault diagnosis, feature extraction, deep learning, small-sample learning, intelligent diagnosis

## 1. Introduction

Fault diagnosis is a core technology for ensuring the reliability of industrial equipment. It refers to the complete process of fault detection, fault isolation, and fault localization through the acquisition, analysis, and identification of equipment operating state signals. Its fundamental principle is to establish a mapping relationship between operational characteristics and fault types, thereby enabling the transition from “post-failure maintenance” to “predictive maintenance” and providing critical decision support for Prognostics and Health Management (PHM) systems. In industrial production, fault diagnosis technology can accurately identify early weak faults and locate their root causes, making it an essential means of ensuring the continuous and safe operation of industrial systems.

Compared with traditional experience-driven and model-driven fault diagnosis methods, machine learning-based approaches, grounded in a data-driven paradigm, exhibit significant technical advantages. They do not rely on precise physical mechanism models of equipment and can autonomously learn fault features from massive multi-source operational data. Moreover, they demonstrate strong adaptability to nonlinear and strongly coupled complex industrial systems and can effectively handle varying operating conditions such as fluctuating loads and rotational speeds. With robust generalization capabilities, these models can be rapidly adapted to different types of industrial equipment through fine-tuning. In addition, their diagnostic accuracy and efficiency significantly outperform traditional methods, making them the mainstream approach in current industrial fault diagnosis research.

Despite the widespread application and notable progress of machine learning in industrial fault diagnosis, several critical challenges remain due to the constraints of real industrial environments and data characteristics. First, fault samples in industrial settings are often scarce and highly imbalanced, while labeling costs are expensive, leading to significantly degraded feature learning capability and diagnostic accuracy in small-sample scenarios. Second, deep learning models suffer from a “black-box” nature, resulting in limited interpretability of diagnostic outcomes, which hinders trust from engineering practitioners and restricts their application in safety-critical domains. Third, complex deep learning models typically involve high computational complexity and large numbers of parameters, making it difficult to achieve millisecond-level real-time diagnosis on resource-constrained edge devices. Finally, in the presence of multi-fault coupling and cross-condition operating scenarios, single algorithms often exhibit insufficient feature extraction capability, resulting in limited diagnostic accuracy and generalization performance.

To address these issues, extensive research efforts have been conducted worldwide, yielding a series of advancements. Some studies mitigate small-sample and data imbalance problems through techniques such as data augmentation and self-supervised learning [1, 2]. Others improve model interpretability by incorporating attention mechanisms and visualization techniques [3]. In addition, model lightweighting methods have been proposed to facilitate the deployment of fault diagnosis models on edge devices [4], while multi-algorithm fusion strategies have been explored to enhance adaptability in multi-fault coupling scenarios [5].

Although existing studies have addressed certain issues in machine learning-based fault diagnosis from individual perspectives, a systematic solution framework has yet to be established. Moreover, there is insufficient research on the engineering adaptability of algorithm improvements and the selection principles of algorithms under different application scenarios. Against this background, this paper focuses on the technological applications and systematic improvements of machine learning in industrial fault diagnosis. It reviews the development trajectory of relevant technologies, analyzes improvement strategies and engineering performance of different algorithms, and aims to provide systematic insights for addressing key challenges in this field while offering guidance for future research directions.

## 2. Technical Framework and Existing Challenges

Machine learning-based industrial fault diagnosis constitutes a data-driven closed-loop process, which mainly includes three core stages: data acquisition and preprocessing, feature extraction and selection, and diagnostic model construction and evaluation. The performance of each stage directly affects the overall diagnostic effectiveness [6].

Data acquisition relies on sensors to collect four categories of key signals, including vibration, electrical, temperature, and image data, with sampling frequencies required to match the operational characteristics of the equipment [7]. In industrial environments, collected data often suffer from issues such as noise, missing values, and class imbalance, which require targeted preprocessing strategies. Specifically, Min–Max normalization is used to unify data scales [8]; wavelet transform is applied for signal denoising [9]; linear interpolation is adopted to fill missing values [10]; and SMOTE oversampling is employed to balance sample distributions.

It is noteworthy that existing preprocessing methods exhibit inherent limitations when dealing with weak early-stage fault features under strong noise conditions. In particular, threshold-based denoising using wavelet transform may remove weak fault features along with noise, making it difficult to achieve effective signal-to-noise separation. This limitation cannot be resolved solely at the preprocessing stage and requires coordinated optimization with subsequent feature extraction processes.

Feature extraction can be broadly categorized into manual feature engineering and automatic feature learning. The former relies on domain expertise to extract features from the time domain (e.g., mean, kurtosis), frequency domain (e.g., spectral peaks), and time–frequency domain (e.g., wavelet entropy), and is typically applied in traditional machine learning methods [11]. The latter leverages deep learning models to perform end-to-end feature mining without manual intervention [12]. Feature selection aims to eliminate redundant information, where filter-based methods (e.g., mutual information) are computationally efficient [13], while embedded methods (e.g., L1 regularization) provide higher selection accuracy [14].

However, certain limitations persist. Existing feature extraction methods perform inadequately in handling fault frequency drift under varying operating conditions. Manual features, constrained by fixed formulations, lack adaptability to changing conditions. Although standard convolutional neural networks (CNNs) can automatically learn features, their fixed receptive fields limit the ability to capture multi-scale fault characteristics. Meanwhile, feature selection methods are prone to overfitting in small-sample scenarios, and the contradiction between high-dimensional features and limited samples remains unresolved.

Diagnostic models mainly include traditional machine learning and deep learning approaches. The former, such as Support Vector Machines (SVM), Random Forests (RF), and Extreme Learning Machines (ELM), are relatively simple and interpretable, making them suitable for small-sample scenarios [15]. The latter, including Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Transformers, exhibit strong capability for automatic feature learning and are well suited for large-sample, end-to-end diagnosis [16]. Evaluation metrics should be selected according to data characteristics: accuracy is appropriate for balanced datasets, while F1-score or AUC is preferred for imbalanced datasets; for multi-fault diagnosis, confusion matrices are used to analyze the causes of false positives and false negatives.

Nevertheless, existing models remain insufficient when simultaneously addressing the three major challenges of small samples, strong noise, and cross-condition variability. Traditional machine learning methods rely on manually engineered features and exhibit limited robustness to noise. Although deep learning methods possess strong feature learning capabilities, they tend to overfit in small-sample scenarios, are susceptible to feature suppression under strong noise, and show degraded generalization performance across varying operating conditions. A single improvement strategy—such as enhancing denoising, augmenting samples, or modifying network architecture alone—is insufficient to comprehensively address these challenges.

In summary, although the technical framework of machine learning-based industrial fault diagnosis has matured, significant limitations remain in addressing three key challenges: preserving weak features under strong noise, capturing multi-scale features under varying operating conditions, and ensuring model generalization under small-sample scenarios. These challenges are strongly coupled and cannot be effectively resolved through isolated optimization of individual stages or single strategies. To address these issues, this paper proposes a Multi-Strategy Adaptive Fusion Deep Residual Shrinkage Network (MA-DRSN). The core idea is to collaboratively optimize noise suppression, multi-scale feature extraction, and lightweight attention mechanisms within an end-to-end deep learning framework, aiming to tackle the combined challenges of small samples, strong noise, and cross-condition variability in a unified manner.

## 3. Applications and Improvements of Traditional Machine Learning Algorithms in Industrial Fault Diagnosis

Traditional machine learning algorithms, including Support Vector Machines (SVM), Extreme Learning Machines (ELM), and Random Forests (RF), constitute the foundational methodological framework for industrial fault diagnosis. However, when applied to complex industrial scenarios, single algorithms face three major challenges: insufficient multi-class classification capability, difficulty in adaptive hyperparameter optimization, and degraded generalization performance under noisy environments. To address these issues, researchers have developed three primary technical routes: algorithm improvement, parameter optimization, and model ensemble.

## 3.1 Support Vector Machine (SVM) and Its Improvements

SVM is based on the principle of structural risk minimization and achieves high-dimensional mapping of low-dimensional data through kernel functions. It performs well in small-sample and high-dimensional handcrafted feature scenarios [17]. However, in industrial applications, SVM suffers from limitations such as weak multi-class classification capability, reliance on empirical tuning of kernel parameters, and sensitivity to noise, which constrain further improvements in diagnostic performance.

To address the multi-class classification problem, the binary tree SVM extends the multi-class recognition capability of SVM through a hierarchical decision structure. Li *et al.* applied this approach to rolling bearing fault diagnosis, achieving a 12% improvement in accuracy compared with traditional SVM [18]. This approach indicates that when the inherent capability of an algorithm does not match task requirements, structural modification to expand algorithm boundaries is an effective way to enhance diagnostic performance. For parameter optimization, swarm intelligence algorithms have demonstrated significant advantages. Ruan Wanying *et al.* employed Particle Swarm Optimization (PSO) to jointly optimize the kernel parameters and penalty factors of SVM, achieving a diagnostic accuracy of 98.5% on a rolling bearing dataset [8]. This suggests that transitioning from static parameter settings to dynamic adaptive optimization is a key pathway for improving SVM performance. To address noise and data imbalance, Fuzzy SVM introduces membership functions to reduce the influence of outlier samples. Liu *et al.* applied this method to fault detection in high-speed train braking systems, improving the fault detection rate by 15% [19]. The core idea of this approach is to embed data quality awareness into the algorithm, enabling the model to actively identify and suppress noise while highlighting critical information, which is essential for improving robustness.

## 3.2 Extreme Learning Machine (ELM) and Its Improvements

ELM overcomes the tendency of backpropagation (BP) neural networks to fall into local optima due to its fast training speed and strong generalization capability, making it particularly suitable for real-time fault diagnosis scenarios. However, its random generation of hidden layer weights leads to poor model stability and susceptibility to overfitting.

To address overfitting, regularization techniques have been introduced into ELM. Zhang Chaolin *et al.* proposed an L2-regularized ELM, which effectively suppresses overfitting in rolling bearing fault diagnosis [14]. This approach has inspired subsequent attempts to incorporate structural risk minimization into fast-training models. For weight optimization, Fu *et al.* applied PSO to optimize the hidden layer weights of ELM and combined it with Ensemble Empirical Mode Decomposition (EEMD) for signal decomposition. This method achieved rolling bearing fault diagnosis under strong noise conditions, improving accuracy by 10% [20]. This demonstrates that the co-design of signal preprocessing and model optimization is an effective strategy for enhancing diagnostic accuracy in noisy environments.

ELM and its improved variants demonstrate that fast training and generalization capability can be achieved simultaneously. However, their inherent sensitivity to input noise cannot be fundamentally resolved through algorithm-level optimization alone—this limitation foreshadows the introduction of soft-threshold denoising techniques in subsequent deep learning approaches.

## 3.3 Ensemble Learning and Algorithm Fusion

Single algorithms have limited feature learning capability, whereas ensemble learning achieves complementary advantages by combining multiple base models. As a representative Bagging method, RF effectively mitigates overfitting and improves accuracy by 12% compared with the traditional three-ratio method in transformer fault diagnosis [13]. Wen Jiangtao *et al.* combined fuzzy granular autoencoders with RF, achieving an 8% improvement in diagnostic accuracy for rotating machinery while reducing training time by 30% [11]. As a Boosting-based fusion model, ELM-AdaBoost improved diagnostic accuracy by 13% compared with standalone ELM in railway axle box bearing fault diagnosis under strong noise conditions [20]. The success of ensemble learning demonstrates that multi-model collaboration can effectively enhance diagnostic performance. However, its essence remains a combination of multiple weak learners, which does not fundamentally resolve the coupling problem between feature extraction and noise suppression. This limitation highlights the potential advantages of end-to-end architectures in deep learning.

Overall, existing studies indicate that improvements in traditional machine learning primarily focus on structural optimization and parameter adaptation. The core paradigm of innovation in the traditional machine learning era can be summarized as “active human design with passive algorithm execution.” Researchers incorporate domain knowledge into algorithms through structural modification, parameter optimization, sample weighting, and model integration, as summarized in Table 1.

*Table 1: Summary of Innovations in Traditional Machine Learning Algorithm Improvements*

| Algorithm | Improvement Target | Key Innovation | Reference |
|---|---|---|---|
| SVM | Algorithm structure | Enables multi-fault classification, accuracy improved by 12% | [18] |
| | Parameter configuration | Achieves adaptive parameter optimization, accuracy reaches 98.5% | [8] |
| | Data quality awareness | Enhances robustness under noise and imbalanced data, fault detection rate improved by 15% | [19] |
| ELM | Model complexity | Suppresses overfitting while maintaining fast training and generalization | [14] |
| | Parameter space | Improves diagnostic accuracy under strong noise, accuracy increased by 10% | [20] |
| Ensemble Learning | Model diversity | Reduces overfitting risk, accuracy improved by 12% over traditional methods | [13] |
| | Feature learning–classification decoupling | Simplifies feature engineering, improves accuracy by 8% and reduces training time by 30% | [11] |
| | Hard sample focusing | Enhances classification performance under strong noise, accuracy improved by 13% | [20] |

## 4. Applications and Improvements of Deep Learning Algorithms in Industrial Fault Diagnosis

Deep learning, with its end-to-end feature self-learning capability, overcomes the dependence of traditional machine learning on manual feature engineering and has been widely applied in fault diagnosis. However, existing deep learning models still exhibit notable limitations when addressing three major practical industrial challenges: small-sample scenarios, strong noise environments, and cross-condition variability. This section systematically reviews the improvement strategies of models such as CNN, LSTM/GRU, and Transformer, identifies current research bottlenecks, and introduces the core innovation of this paper—the Multi-Strategy Adaptive Fusion Deep Residual Shrinkage Network (MA-DRSN).

Bottleneck 1: Feature suppression under noisy environments. To address this issue, Li Weihua *et al.* constructed a Deep Residual Shrinkage Network (DRSN) by embedding soft-threshold denoising into residual blocks, achieving a 15% improvement in diagnostic accuracy under strong noise conditions with a signal-to-noise ratio of 5 dB [4]. This approach provides an important paradigm for the integration of deep learning and signal processing. However, the soft threshold in this method is fixed and lacks adaptability under varying noise intensities—this constitutes the first improvement of the proposed MA-DRSN: an adaptive threshold learning mechanism.

Bottleneck 2: Feature mismatch caused by fault frequency drift under varying operating conditions. To address this problem, Chen Renxiang *et al.* introduced multi-scale convolution and channel attention mechanisms to construct a multi-scale CNN-BiGRU-Attention model, improving diagnostic accuracy by 3.7% in hydropower unit fault diagnosis [5]. This demonstrates that multi-scale feature fusion is an effective means of enhancing model adaptability to varying operating conditions—this forms the second improvement of MA-DRSN: a parallel multi-scale convolutional kernel design.

Bottleneck 3: Conflict between edge deployment requirements and model complexity. Li Weihua *et al.* achieved CNN lightweighting through model pruning and quantization, reducing the number of parameters by 70% and achieving inference time below 50 ms [4]. This indicates that lightweight design is essential for deploying deep learning models in industrial applications—this corresponds to the third improvement of MA-DRSN: a channel–spatial collaborative attention mechanism that enhances feature discriminability while strictly controlling parameter size.

## 4.1 Recurrent Neural Networks (LSTM/GRU) and Their Improvements

LSTM and GRU capture temporal dependencies through gated structures and perform well in early fault diagnosis of time-series signals such as current and voltage. He Zhengyou *et al.* applied LSTM to learn temporal features of traction motor currents, achieving an accuracy of 97.5% for early bearing fault detection [12]. Xiong Rui *et al.* developed a GRU-based model combined with dynamic thresholds to provide a 30-minute early warning of thermal runaway in power batteries, achieving 98% accuracy with a false alarm rate below 1% [16]. Hybrid CNN-LSTM models combine spatial and temporal feature learning capabilities, achieving a 10% improvement in accuracy over standalone LSTM in fault diagnosis of motors and fans under varying operating conditions [12].

These results indicate that temporal models are particularly sensitive to early weak faults. However, their reliance on long sequence inputs makes them prone to overfitting in small-sample scenarios. How to effectively integrate temporal modeling capabilities with CNN-based spatial feature extraction while addressing small-sample learning remains an open research direction.

## 4.2 Emerging Methods: Transformer and Self-Supervised Learning

Transformer models capture global dependencies through self-attention mechanisms and are well suited for long-sequence and multi-source data fusion in fault diagnosis. Liu Ruochen *et al.* proposed a temporal Transformer model, achieving a diagnostic accuracy of 98.5% for wind turbine systems under long time-series signals [15]. Self-supervised learning has proven effective in addressing the small-sample problem in industrial settings. Zhang Chaolin *et al.* pre-trained models through signal reconstruction, achieving a mechanical fault diagnosis accuracy of 96.62% using only 1% labeled data, representing a 45.79% improvement over traditional methods [1]. Chen Tao *et al.* combined contrastive learning with self-supervised learning, achieving a 92.7% accuracy in rolling bearing fault diagnosis with only five samples per class [2].

The rise of self-supervised learning marks a promising breakthrough for small-sample scenarios. However, its two-stage paradigm (pre-training and fine-tuning) remains complex and sensitive to data augmentation strategies. How to simultaneously address small samples, strong noise, and cross-condition variability within an end-to-end framework constitutes the core motivation of this study.

## 4.3 Multi-Strategy Adaptive Fusion Deep Residual Shrinkage Network (MA-DRSN)

In summary, although deep learning models have achieved significant progress in industrial fault diagnosis, there is still a lack of a unified solution for complex industrial scenarios where small samples, strong noise, and cross-condition variability coexist. Residual shrinkage networks introduce soft-threshold denoising, but their fixed thresholds limit adaptability to varying noise intensities. Multi-scale networks [5] improve adaptability to varying conditions but fail to account for noise interference. Self-supervised methods address small-sample issues but involve complex two-stage training procedures.

To overcome these limitations, this paper proposes the Multi-Strategy Adaptive Fusion Deep Residual Shrinkage Network (MA-DRSN), as summarized in Table 2. Its core innovations include:

a) Adaptive threshold learning module: By improving the soft-threshold function in traditional residual shrinkage networks and introducing a signal-to-noise ratio estimation subnetwork, the denoising threshold can be dynamically adjusted according to the noise intensity of the input signal. This enables the preservation of weak fault features under strong noise while avoiding information loss under low-noise conditions.

b) Parallel multi-scale convolution module: Multi-branch convolutional kernels (with sizes of 3, 5, and 7) are designed to extract fault features at different scales in parallel, enhancing adaptability to fault frequency drift under varying operating conditions. Channel attention mechanisms are further employed to adaptively fuse multi-scale features.

c) Lightweight channel–spatial collaborative attention: A collaborative attention mechanism is embedded within residual blocks to simultaneously enhance fault-related features in both channel and spatial dimensions while suppressing noise interference. Group convolution is adopted to control parameter growth, ensuring suitability for edge deployment.

*Table 2: Improvements of MA-DRSN*

| Existing Bottleneck | Proposed Improvement |
|---|---|
| Feature suppression under noisy environments | Adaptive threshold learning mechanism |
| Feature mismatch due to fault frequency drift under varying conditions | Parallel multi-scale convolutional kernel design |
| Conflict between edge deployment requirements and model complexity | Channel–spatial collaborative attention mechanism with strict parameter control while enhancing feature discriminability |

## 5. Applications in Typical Industrial Scenarios

The application of machine learning algorithms in industrial fault diagnosis must be adapted to specific scenario characteristics. This section selects three representative scenarios—rotating machinery, power equipment, and new energy equipment—to analyze their differentiated algorithmic requirements and to identify the application value of the proposed MA-DRSN model.

### 5.1 Fault Diagnosis of Rotating Machinery

Rotating machinery (e.g., bearings, gearboxes, and motors) primarily relies on vibration signals as key fault features, with common fault types including wear, cracks, and eccentricity. Algorithms such as CNN, SVM, and ELM are widely applicable in this domain [3, 9]. Zhe Hongshan *et al.* applied a one-dimensional CNN to raw vibration signals of rolling bearings, achieving a diagnostic accuracy of 99.2% [3]. Wang Fengtao *et al.* combined wavelet packet decomposition features with SVM, achieving accuracy above 98% [9]. Fu *et al.* implemented gearbox fault diagnosis using EEMD combined with optimized ELM, achieving 95% accuracy under strong noise conditions [20]. Yu Fei *et al.* applied LSTM to motor bearing current signals, enabling early detection of weak faults [12].

Fault diagnosis of rotating machinery faces three main challenges: rotational speed fluctuations (cross-condition variability), environmental noise, and the extraction of weak early-stage features. The MA-DRSN model demonstrates strong application potential in this scenario. Its adaptive threshold learning module can dynamically suppress noise interference, while the multi-scale convolution module can adapt to frequency drift caused by speed variations.

### 5.2 Fault Diagnosis of Power Equipment

Power equipment (e.g., transformers and switchgear) is characterized by fault features such as dissolved gases in oil, partial discharge, and temperature signals, making algorithms such as RF and SVM particularly suitable. Liao Ruijin *et al.* applied RF to transformer dissolved gas analysis (DGA) data, achieving a 12% improvement in accuracy compared with the traditional three-ratio method [13]. Liu *et al.* proposed a fuzzy SVM to address data imbalance in fault detection of high-speed train braking systems [19]. The primary challenges in power equipment fault diagnosis are the scarcity of fault samples (small-sample problem) and the fusion of multi-source heterogeneous data. Although MA-DRSN is primarily designed for vibration signals, its lightweight attention mechanism can be extended to multi-sensor feature fusion, providing a foundation for future research in this area.

### 5.3 Fault Diagnosis of New Energy Equipment

New energy equipment (e.g., power batteries and wind turbine units) involves fault features such as time-series voltage/temperature signals and vibration signals. These scenarios require both real-time performance and early warning capability, making LSTM/GRU and self-supervised learning methods particularly suitable. Xiong Rui *et al.* used a GRU model to analyze voltage and temperature time-series signals of power batteries, enabling thermal runaway warnings 30 minutes in advance [16]. Zhao Hongshan *et al.* applied an improved K-means algorithm for unsupervised anomaly detection in wind turbine gearbox vibration signals, achieving a warning delay of less than 5 minutes [10]. Chen Tao *et al.* utilized self-supervised contrastive learning to achieve small-sample fault diagnosis for wind turbines, effectively addressing the scarcity of fault samples

[2]. Fault diagnosis in new energy equipment faces challenges such as varying operating conditions, data distribution drift, and high real-time requirements. The lightweight design of MA-DRSN (reduced parameter size and inference time below 50 ms) provides strong potential for edge deployment and can meet real-time early warning requirements.

## 6. Core Challenges of Machine Learning in Industrial Fault Diagnosis

Despite significant progress, machine learning in industrial fault diagnosis still faces four major challenges due to constraints related to industrial data characteristics, operating conditions, and engineering requirements. These challenges are interrelated and collectively define the key research focus in this field.

### 6.1 Small-Sample and Data Imbalance Issues

In industrial settings, fault samples are far fewer than normal samples, and the cost of data labeling is high. As a result, both traditional machine learning and deep learning models suffer from significantly reduced feature learning capability in small-sample scenarios. Specifically, in large-scale equipment such as wind turbines and nuclear power systems, the low occurrence rate of faults makes it difficult to accumulate sufficient high-quality fault data, thereby limiting model generalization. In equipment such as rolling bearings and motors, imbalanced distributions among different fault types lead to minority-class faults being easily overlooked, resulting in higher false negative rates [8]. This issue directly constrains the practical applicability of models in real industrial environments.

### 6.2 Lack of Model Interpretability (“Black-Box” Problem)

Although deep learning models possess strong feature learning capabilities, their “black-box” nature leads to a lack of physical interpretability in diagnostic results, making it difficult for engineers to establish trust in the models. For example, models such as CNN and LSTM can achieve high diagnostic accuracy but fail to explain the key features and reasoning processes underlying fault identification. In safety-critical domains such as power systems and nuclear energy, insufficient interpretability has become a major barrier to practical deployment.

### 6.3 Engineering Adaptation for Real-Time Edge Deployment

Industrial fault diagnosis often requires millisecond-level response times. However, complex deep learning models (e.g., Transformer and deep CNN architectures) involve high computational complexity and large parameter sizes, making them difficult to deploy directly on resource-constrained edge devices (e.g., sensors and embedded systems). For instance, real-time diagnosis of rotating machinery and power batteries typically requires inference times below 100 ms, which conventional deep learning models struggle to achieve [16]. The trade-off between model performance and deployment efficiency is becoming increasingly prominent.

### 6.4 Multi-Fault Coupling and Cross-Condition Diagnosis

Industrial equipment often operates under varying conditions (e.g., fluctuating loads and speeds) and is prone to multiple concurrent faults, leading to interference among fault features and making accurate diagnosis challenging for single algorithms. For example, when a gearbox simultaneously experiences gear wear and bearing cracks, the superposition of fault features significantly reduces the diagnostic accuracy of traditional SVM and CNN models. Similarly, under variable load conditions, shifts in vibration and current signal characteristics pose severe challenges to model generalization across operating conditions [5].

In summary, the four major challenges correspond to different dimensions: data-level (small samples and imbalance), model-level (interpretability), engineering-level (real-time deployment), and scenario-level (multi-fault coupling and varying conditions). These challenges are not independent but highly interconnected. For instance, small-sample issues exacerbate the difficulty of identifying coupled faults, while lack of interpretability reduces engineers’ trust in diagnostic results. Therefore, future research should aim for coordinated breakthroughs across multiple dimensions.

## 7. Future Development Trends

In response to the aforementioned core challenges, and considering the evolution of machine learning technologies alongside industrial engineering demands, future research will focus on four dimensions: data, model, engineering, and application. These directions are mutually reinforcing and will collectively drive fault diagnosis technologies toward greater intelligence, reliability, and practicality.

a) Data level: Small-sample learning and data augmentation techniques. To address small-sample and data imbalance issues, research will focus on self-supervised learning and meta-learning. Self-supervised learning leverages unlabeled data for pretraining, effectively uncovering intrinsic data structures. Zhang Chaolin *et al.* demonstrated that a diagnostic accuracy of 96.62% can still be achieved with only 1% labeled data [1]. Meta-learning, on the other hand, aims to enable rapid adaptation with few samples. Chen Tao *et al.* applied meta-learning to rolling bearing fault diagnosis, achieving an accuracy of 92.7% with only five samples per class [2]. Meanwhile, data augmentation techniques such as Generative Adversarial Networks (GANs) and time stretching will be further optimized to efficiently expand fault samples [8]. The synergistic development of these techniques is expected to fundamentally alleviate data scarcity in industrial scenarios.

b) Model level: Algorithm fusion and interpretability design. To address limited model capability and poor interpretability, multi-algorithm fusion and explainable artificial intelligence (XAI) have become dominant research directions. In terms of algorithm fusion, hybrid models such as CNN-LSTM and Transformer with attention mechanisms integrate spatial and temporal features, as well as local and global information, significantly improving diagnostic performance [5, 15]. Regarding interpretability, XAI techniques are increasingly integrated with fault diagnosis. Lei Yaguo *et al.* utilized Grad-CAM to visualize the key feature regions focused on by CNN during diagnosis [3], while Zhang Chaolin *et al.* employed attention mechanisms to guide models toward critical fault frequency bands, effectively enhancing interpretability and engineering trust [1]. This direction will gradually mitigate the trust barriers associated with “black-box” models.

c) Engineering level: Lightweight design and edge intelligence deployment. To meet real-time diagnostic requirements at the edge, lightweight model design and edge intelligence architectures have become research hotspots. In terms of lightweighting, techniques such as model pruning, quantization, and knowledge distillation continue to evolve. Li Weihua *et al.* reduced CNN parameters by 70% through pruning and quantization, achieving inference times below 50 ms and meeting edge deployment requirements [4]. At the architectural level, edge intelligence is driving the migration of fault diagnosis models from cloud to edge devices, enabling local data processing and real-time diagnosis, thereby reducing transmission latency and bandwidth demands [16]. In addition, the integration of digital twins with machine learning is fostering hybrid architectures that combine physical models with data-driven models, further enhancing early fault prediction capabilities [6].

d) Application level: Multimodal fusion and full lifecycle diagnosis. To address complex industrial scenarios and lifecycle management requirements, multimodal fusion and integrated diagnosis are emerging trends. The fusion of multi-source heterogeneous data (e.g., vibration, electrical, image, and temperature signals) enables the complementary use of different modalities, improving diagnostic accuracy under complex conditions. Meanwhile, machine learning is being deeply integrated with full lifecycle management of equipment, forming a closed-loop system from remaining useful life prediction and online fault diagnosis to maintenance decision optimization. The integrated model of life prediction and fault diagnosis proposed by Guo Yifan *et al.* provides an important foundation for this direction [7]. This trend indicates a shift from isolated “post-event alarms” toward full lifecycle “predictive maintenance.”

## 8. Conclusion

In summary, machine learning has become a core technological approach in industrial fault diagnosis. The field has evolved from traditional methods such as SVM and ELM, which rely on handcrafted features, to deep learning models such as CNN and LSTM with end-to-end feature learning capabilities, and further to advanced approaches such as Transformer and self-supervised learning for small-sample and multimodal scenarios. The technological framework has been continuously refined and has achieved practical applications in typical scenarios such as rotating machinery, power equipment, and new energy systems.

However, current technologies still face several critical challenges, including small-sample and data imbalance issues, lack of model interpretability, difficulty in real-time edge deployment, and insufficient capability in handling multi-fault coupling. Future research should focus on coordinated optimization across the data, model, and engineering levels, with particular emphasis on self-supervised learning, algorithm fusion, lightweight deployment, and multimodal integration.

Research on machine learning in industrial fault diagnosis will increasingly emphasize engineering adaptability, promoting the transition of technologies from laboratory settings to real industrial environments. This will enable a shift from “accurate diagnosis” to “predictive maintenance” and “intelligent operation and maintenance,” providing essential support for intelligent manufacturing and improved reliability of industrial equipment. Future work should further integrate the physical mechanisms of industrial systems with data-driven approaches, enabling deeper fusion between mechanism-driven and data-driven models to better meet practical industrial requirements. At the same time, cross-disciplinary collaboration should be strengthened to promote continuous advancement toward more intelligent and practical fault diagnosis technologies.

## References

[1] Lei, Y. G., Jia, F., et al. (2018). Opportunities and challenges of intelligent fault diagnosis for machinery in the era of big data. *Journal of Mechanical Engineering*, 54(1), 94–104.

[2] Guo, Y. F., & Tang, J. Y. (2019). A review of life prediction and fault diagnosis technologies based on machine learning algorithms. *Computer Measurement & Control*, 27(3), 1–5, 10.

[3] Zhao, H. S., Zhang, J. P., & Li, G. (2012). Fault early warning method for wind turbine gearbox based on cluster analysis. *Automation of Electric Power Systems*, 36(15), 102–106.

[4] Song, Y. Q., Zhu, Z. J., & Ji, Y. F. (2014). Intelligent fault diagnosis based on multi-sensor information fusion. *Journal of Xi’an Polytechnic University*, (5), 568–573.

[5] Bian, W. B., Deng, A. D., Liu, D. C., et al. (2023). Fault diagnosis method for rolling bearings of wind turbines based on improved deep residual shrinkage network. *Journal of Mechanical Engineering*, 59(12), 160–171.

[6] Xiao, Y., Xiang, J. W., Tang, H. S., et al. (2024). Fault diagnosis method for hydraulic waterproof valves based on multi-sensor information fusion and CNN-BiGRU-Attention model. *Mechanical & Electrical Engineering*, 41(9), 1531–1540.

[7] Xu, K., Chen, Z. H., Zhang, C. B., et al. (2019). Rolling bearing fault diagnosis based on empirical mode decomposition and support vector machine. *Control Theory & Applications*, 36(6), 915–922.

[8] Ruan, W. Y., Ma, Z. Q., & Li, Y. C. (2020). Rolling bearing fault diagnosis based on variational mode decomposition, permutation entropy, and PSO-SVM. *Journal of University of Jinan (Natural Science Edition)*, 34(2), 131–137.

[9] Zhang, Y. F., Yao, Z. Y., Tang, S. L., et al. (2016). Engine fault diagnosis method based on principal component analysis and support vector machine. *China Mechanical Engineering*, (24), 3307–3311.

[10] Wen, J. T., & Zhou, X. N. (2018). Fault diagnosis of rotating machinery based on fuzzy granular unsupervised learning combined with random forest. *Mechanical Science and Technology for Aerospace Engineering*, 37(11), 1722–1730.

[11] Wu, C. Z., Jiang, P. C., Feng, F. Z., et al. (2018). Gearbox fault diagnosis based on one-dimensional convolutional neural network. *Journal of Vibration and Shock*, (22), 51–56.

[12] Yu, F., Fan, Q. C., & Xuan, M. (2022). Motor bearing fault prediction based on grasshopper optimization Bi-LSTM network. *Electric Machines and Control*, 26(6), 9–17.

[13] Tao, D. Q., Bo, C. M., & Yi, H. (2015). Transformer fault detection method based on random forest. *Electronic Components and Materials*, 38(4), 840–844.

[14] Wang, T. T., Wang, Y., & Ji, Z. C. (2018). Rolling bearing fault diagnosis based on improved extreme learning machine. *Journal of System Simulation*, 30(11), 4413–4420.

[15] Gao, D. X., Wang, Y., Zheng, X. Y., et al. (2023). Charging state monitoring and safety early warning method for electric vehicles based on deep learning. *Electric Machines and Control*, 27(7), 122–132.

[16] Li, Y. B., Yang, Y. T., Wang, X. Z., et al. (2018). Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. *Journal of Sound and Vibration*, 428, 72–86.

[17] Fu, Q., Jing, B., He, P. J., et al. (2018). Fault feature selection and diagnosis of rolling bearings based on EEMD and optimized Elman-AdaBoost algorithm. *IEEE Sensors Journal*, 18(12), 5024–5034.

[18] Zhang, Z. J., Li, H., Huang, Y. S., et al. (2024). Application of deep residual shrinkage network in rolling bearing fault diagnosis. *Journal of Northeastern University (Natural Science)*, 45(11), 1587–1594.

[19] Chen, T., & Lei, Y. G. (2021). Cross-equipment fault diagnosis based on transfer component analysis. *Journal of Mechanical Engineering*, 57(12), 112–120.

[20] Zhou, Q., Chen, W. G., et al. (2014). Transformer multiple fault diagnosis and short-term prediction method based on cloud theory. *High Voltage Engineering*, 40(5), 1450–1456.

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.