Computers and Artificial Intelligence; Vol.3, No.6; 2026
E-ISSN: 3006-3728 P-ISSN: 3006-371X
DOI: https://doi.org/10.70267/cai.v3n6.100119
Published by: Zeus Press

# Research Progress in Machine-Learning-Based Fault Diagnosis for Mechanical Equipment

**Yanlin Chen***

*School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China*

**Corresponding author: Yanlin Chen.*

## Abstract

Fault diagnosis of mechanical equipment is essential for ensuring safe and reliable operation and reducing losses caused by unplanned downtime. With the large-scale deployment of sensors and the Industrial Internet of Things, condition-monitoring data have grown rapidly in both volume and dimensionality. Conventional diagnostic approaches that rely on human expertise and signal processing are increasingly inadequate, while machine learning, especially deep learning, has become the dominant research paradigm because of its automatic feature extraction and powerful nonlinear modeling capabilities. Following a data-feature-model-application framework, this paper systematically reviews advances in machine-learning applications for mechanical equipment fault diagnosis. It first outlines the basic diagnostic workflow, commonly used monitoring signals, and representative public datasets. It then reviews diagnostic methods based on conventional machine learning, including feature engineering, support vector machines, and ensemble learning, as well as deep-learning methods based on convolutional neural networks, recurrent neural networks, autoencoders, attention mechanisms, and Transformers, comparing their principles, applicable conditions, strengths, and limitations. For real industrial environments, the paper focuses on emerging directions such as few-shot learning, transfer learning, multisensor fusion and graph neural networks, federated learning, and generative-adversarial-network-based data augmentation, and summarizes their applications to representative systems including rotating machinery, motors, computer numerical control machine tools, and pumps. Finally, current challenges are analyzed from the perspectives of data quality, model generalization, interpretability, and real-time edge deployment, and future trends involving physics-data integration, foundation models, and intelligent operation and maintenance are discussed. The analysis shows that the field is evolving from discriminative diagnosis dependent on large amounts of labeled data toward few-shot, cross-condition, interpretable, and deployable solutions. Data governance and engineering implementation will therefore become the principal bottlenecks constraining practical adoption.

### Keywords

fault diagnosis, machine learning, deep learning, mechanical equipment, transfer learning, review



© 2026 The Author(s). Authors retain copyright in their articles.

# 1. Introduction

## 1.1 Research Background

Rotating machinery constitutes the core actuation components of electric power, metallurgy, rail transit, and advanced equipment. Sudden failures can cause substantial unplanned-downtime losses and may also trigger safety and environmental accidents [1]. Jardine et al. [2] divided condition-based maintenance (CBM) into data acquisition, data processing, and maintenance decision-making, establishing the role of diagnosis and prognosis in equipment lifecycle management. Chen et al. [3] noted that high-speed, precision, and heavy-duty equipment must rely on condition monitoring and fault diagnosis to ensure lifecycle performance. As low-cost vibration, current, and temperature sensors are deployed at scale and connected to the Industrial Internet, monitoring data exhibit large volume, multisource heterogeneity, rapid growth, and sparse value [4]. Experience-based signal analysis can no longer cope with these characteristics, leading to the rise of data-driven intelligent diagnosis [5].

## 1.2 Research Significance

From an academic perspective, machine learning has transformed diagnosis from a two-stage process of manual feature engineering plus a shallow classifier into an end-to-end process in which a model adaptively extracts fault features and directly maps them to health states, providing new tools for nonlinear and nonstationary vibration-signal modeling [4, 6]. From an engineering perspective, accurate and timely diagnosis is a prerequisite for predictive maintenance: maintenance decisions shift from scheduled replacement and corrective repair to targeted intervention based on actual degradation states, thereby shortening downtime, reducing spare-parts and labor costs, and supporting lifecycle health management of critical equipment [2, 1]. A model's ability to generalize across operating conditions also directly determines whether an algorithm can move from public laboratory datasets to industrial sites.

## 1.3 Overview of Research in China and Abroad

Intelligent fault diagnosis has advanced along three main lines. The first is conventional machine learning: domain experts manually construct features from wavelet-packet results, envelope spectra, higher-order spectra, and other processed signals, and then feed them into shallow classifiers such as support vector machines (SVMs), k-nearest neighbors, and backpropagation networks. Lei et al. [7] noted that this stage relies heavily on manual feature engineering and expert knowledge, with limited generalization under variable conditions and noise. The second is deep learning: stacked autoencoders, convolutional neural networks, and recurrent networks learn hierarchical features directly from raw vibration signals and outperform manually engineered features on public benchmarks such as CWRU [8, 9]. However, Khan and Yairi [6] pointed out that deep models' dependence on labeled samples and their limited interpretability hinder large-scale deployment. The third comprises emerging methods for complex industrial environments. Few-shot, transfer, and federated learning have developed rapidly in response to scarce fault samples, cross-condition domain shifts, and data silos [7]. Accordingly, this paper systematically reviews the methodological framework and engineering pathway of machine-learning-based fault diagnosis.

## 1.4 Scope and Organization of This Paper

This paper is organized around a data-feature-model-application framework. Section 2 introduces monitoring-signal types, sensors, and public datasets, clarifying the data inputs to diagnosis. Section 3 reviews feature engineering and shallow classifiers in conventional machine learning. Section 4 evaluates deep-learning models including convolutional networks, autoencoders, recurrent networks, and Transformers. Section 5 discusses emerging directions such as few-shot learning, transfer learning, multisensor fusion, and federated learning. Section 6 summarizes applications to representative equipment including bearings, gearboxes, motors, and machine tools. Section 7 analyzes challenges involving data quality, interpretability, edge deployment, and physics-data integration. Section 8 presents the conclusions and future trends. Compared with existing reviews, this paper connects the various branches through the evolution of methods and identifies the applicable conditions and unresolved issues of each method category.



© 2026 The Author(s). Authors retain copyright in their articles.

## 2. Fundamentals of Mechanical Equipment Fault Diagnosis

### 2.1 Basic Concepts and General Workflow of Fault Diagnosis

Mechanical equipment is considered faulty when wear, fatigue, misalignment, or material degradation causes a component to deviate from its specified function. Fault diagnosis uses measurable information to identify the operating state, locate the faulty component, and estimate fault severity without disassembling the equipment [3]. Condition monitoring continuously tracks parameters and captures signs of abnormality, while prognostics and health management (PHM) extends diagnosis to remaining-useful-life estimation and maintenance decision-making. Jardine et al. [2] summarized CBM as data acquisition, data processing, and maintenance decision-making, a division that remains central to diagnostic tasks. Following this framework, the general process comprises five stages: signal acquisition, preprocessing, feature extraction, fault identification, and maintenance decision-making [2]. This process underpins the reduction of unplanned downtime in high-availability equipment.

### 2.2 Common Monitoring Signals, Their Characteristics, and Public Datasets

Vibration signals reflect the dynamic response of rotating components and are generally acquired by industrial vibration accelerometers such as that shown in Fig. 1. Early bearing spalling generates periodic impacts, whose characteristic frequencies can be identified through envelope demodulation; vibration is therefore the preferred measurement for rotating-machinery diagnosis. Motor-current signals can be obtained from variable-frequency-drive circuits, enabling motor current signature analysis (MCSA) without additional hardware costs [10]. However, current signals respond weakly to incipient mechanical defects. Using the Paderborn University (PU) dataset, Lessmeier et al. [11] showed that current-only bearing-damage classification still underperforms vibration-based diagnosis. Acoustic emission is sensitive to crack initiation and is suitable for low-speed heavy-duty equipment and leakage detection, whereas temperature and oil analyses are mainly used to monitor lubrication and thermal conditions. Complementary multisource signals can improve coverage of incipient faults.

*Figure 1: Industrial vibration accelerometer (Source: publicly available product image)*




Public benchmarks are essential for the horizontal comparison of algorithms. Table 1 compares the objects, fault origins, operating-condition channels, and principal uses of three representative datasets: CWRU, XJTU-SY, and PU. CWRU uses electrical-discharge machining to introduce single-point defects and records vibration under multiple loads; it has long served as a de facto standard test bed [12]. To address the limited conditions and fault types in CWRU, XJTU-SY provides full-lifecycle vibration data from healthy operation to failure for 15 rolling bearings at three rotational speeds, offering continuous observations for degradation modeling and remaining-useful-life prediction [13]. PU includes both artificial notches and naturally



© 2026 The Author(s). Authors retain copyright in their articles.

developed damage and provides dual current and vibration channels [11]. The value of open-source code and unified evaluation for reproducibility has also received increasing attention [1]. Under Internet-of-Things acquisition, monitoring information already exhibits large volume, multisource heterogeneity, rapid growth, and sparse value [4].

*Table 1: Comparison of Common Public Datasets*

| Dataset | Object | Fault Origin | Operating Conditions and Channels | Primary Use |
|---|---|---|---|---|
| CWRU | Rolling bearings | Single-point defects introduced by electrical-discharge machining | Vibration under multiple loads | Classification benchmark (de facto standard test bed) |
| XJTU-SY | 15 rolling bearings | Natural run-to-failure degradation | Full-lifecycle vibration at three rotational speeds | Degradation modeling and remaining-useful-life prediction |
| PU (Paderborn) | Rolling bearings | Artificial notches and naturally developed damage | Dual current and vibration channels | Comparative multichannel diagnosis |

*Note: Compiled from Refs. [11]-[13] cited in Section 2.2.*

## 2.3 Conventional Fault-Diagnosis Methods and Their Limitations

Before machine learning became widespread, fault diagnosis developed along signal-processing and physics/knowledge-based lines. The former uses the fast Fourier transform to identify rotational frequency and its harmonics, and the short-time Fourier transform, wavelet transform, and empirical mode decomposition/ensemble empirical mode decomposition (EMD/EEMD) to handle nonstationary components caused by speed fluctuations and impacts. The EEMD method proposed by Wu and Huang [14] uses ensemble averaging with white noise to alleviate mode mixing and has been widely combined with envelope analysis to extract bearing fault frequencies [15]. However, the number of decomposition levels and white-noise amplitude must still be set empirically. The latter line builds degradation models from dynamic equations or converts expert knowledge into rules for expert systems [3].

These methods are effective under steady-state, single-fault conditions but share evident shortcomings. Filter passbands, wavelet bases, and decomposition levels are selected by trial and error, and features must be designed individually. When signals are strongly nonlinear and time-varying and operating conditions change frequently, fixed thresholds and shallow decision rules are prone to failure [6]. Their cross-condition robustness is also weak: a criterion calibrated at one load often fails at another rotational speed, requiring repeated on-site readjustment [3].

## 2.4 Technical Necessity of Introducing Machine Learning into Fault Diagnosis

The Industrial Internet of Things integrates low-cost sensing, edge acquisition, and interconnected networks, and the daily monitoring data generated by equipment now far exceed what humans can analyze [5]. The paradigm of manually selected features plus a shallow classifier is no longer sustainable: feature engineering is time-consuming and difficult to transfer, while shallow models have limited capacity to represent complex nonlinear mappings [4]. This is precisely the appeal of deep learning. Convolutional and autoencoding structures can automatically learn hierarchical fault representations from raw signals or time-frequency images, eliminating manual design and modeling the complex relationship between inputs and states [5, 8]. Khan and Yairi [6] concluded that data-driven methods are becoming a dominant paradigm alongside physics-based methods. The combination of data expansion, the feature-engineering bottleneck, and greater model expressiveness has pushed fault diagnosis into a data-centric intelligent stage.

# 3. Fault-Diagnosis Methods Based on Conventional Machine Learning

## 3.1 Feature-Engineering Methods

Conventional machine learning first compresses long time-series signals into a small set of indicators and then passes them to a shallow classifier; Lei et al. [7] described this as the manual-features-plus-shallow-classifier paradigm. Time-domain statistical features are the easiest to implement. Samanta and Al-Balushi [16] used only root mean square, variance, skewness, kurtosis, and the normalized sixth central moment as inputs



© 2026 The Author(s). Authors retain copyright in their articles.

to a backpropagation network and successfully distinguished healthy from defective rolling bearings, establishing an online-monitoring paradigm based on a few features and a simple network. Its limitation is sensitivity to rotational speed and load.

Impact faults are often modulated onto a carrier, making it difficult for time-domain indicators to locate fault frequencies; frequency-domain and time-frequency features have therefore become mainstream. Yang et al. [15] used empirical mode decomposition to separate intrinsic mode functions, extracted the envelope spectra of key components, and formed feature vectors from amplitude ratios at fault-characteristic frequencies, avoiding empirical passband selection. Saidi et al. [17] introduced the bispectrum, a third-order higher-order spectrum, to characterize non-Gaussian information. Guo et al. [18] used wavelet-packet energy spectra to characterize nonstationary distributions and then applied principal component analysis for dimensionality reduction and noise suppression. Wang et al. [19] showed that inappropriate cumulative contribution thresholds, decomposition levels, or mother-wavelet types can cause information loss or leave residual noise.

## 3.2 Support Vector Machine Methods

SVMs occupy a central position among conventional methods. Widodo and Yang [20] noted that SVMs follow the principle of structural risk minimization and control generalization by maximizing the classification margin. They are less prone than neural networks to overfitting in small-sample, high-dimensional nonlinear classification, matching the scarcity of fault samples and the high dimensionality of engineered features. Decisions are determined by only a small number of support vectors, while kernel functions map linearly inseparable problems into a high-dimensional space for solution.

Practical implementation must also address multiclass classification. Widodo and Yang [20] summarized one-versus-one, one-versus-rest, and related strategies that extend binary classifiers to multistate recognition of bearings, induction motors, and rotors. SVMs are often combined with feature engineering in a two-stage feature-extraction-plus-classification framework. Yang et al. [15] fed amplitude ratios from intrinsic-mode-function envelope spectra into an SVM to identify roller-bearing operating conditions and fault types. Saidi et al. [17] reduced the dimensionality of bispectral features using principal component analysis and then used a multiclass SVM to distinguish four levels of bearing-fault severity. Wang et al. [19] combined wavelet-packet reconstruction energy with an SVM to identify fault severity at different locations. Saidi et al. [17] also used grid search with ten-fold cross-validation to optimize the penalty and kernel parameters, an approach that has become nearly standard.

## 3.3 Tree Models and Ensemble-Learning Methods

Decision trees and their ensembles are engineering-friendly approaches. A single decision tree recursively partitions the feature space using axis-aligned splits. Its rules can be translated into human-readable criteria, and it trains quickly without sensitivity to feature scale; however, it is prone to overfitting and high variance. Random forests reduce variance by constructing multiple decorrelated trees through bootstrap sampling and random feature subspaces and then aggregating their votes. Xu et al. [21] converted time-domain signals into time-frequency grayscale images using the continuous wavelet transform, extracted multilevel features with a convolutional network, and passed them to a random-forest ensemble for decision-making. This fused low-level impact information with high-level operating-condition information and improved both accuracy and generalization under complex conditions.

Gradient-boosted trees iteratively fit the residuals from the preceding round, while XGBoost adds regularization and second-order derivatives to impose stronger constraints on overfitting. Zhang et al. [22] fed a feature matrix extracted by a deep convolutional network into grid-tuned XGBoost for compound gearbox-fault diagnosis and compared it with deep backpropagation networks, random forests, and SVMs. XGBoost provided better overfitting resistance and stability, along with superior accuracy and computation time. Neither random forests nor XGBoost depends on kernel functions or distributional assumptions, and both are intuitive to tune, making them suitable classifier back ends.



© 2026 The Author(s). Authors retain copyright in their articles.

### 3.4 Other Classical Methods Including k-Nearest Neighbors, Bayesian Methods, and Shallow Neural Networks

k-Nearest neighbors has no explicit training stage; it classifies a test sample by voting among its neighbors in feature space. The method is intuitive and makes no distributional assumptions. Wang [23] constructed redundant statistical features using a db44 wavelet-packet transform, reduced their dimensionality, and used k-nearest neighbors to identify five gear-crack levels. In comparisons with linear discriminant analysis, quadratic discriminant analysis, CART, and naive Bayes, k-nearest neighbors achieved the highest accuracy under variable speed and load. This indicates that when the features already characterize differences among operating conditions, neighbor-based decisions can be more robust than explicit modeling, at the cost of storing all training samples.

Naive Bayes makes posterior-probability decisions under the assumption of class-conditional independence. Its computational cost is extremely low, and it can be useful when features are approximately independent and the sample size is modest; performance declines markedly when the independence assumption is violated [24]. Shallow backpropagation networks are trained by error backpropagation, as in the work of Samanta and Al-Balushi [16]. They can fit nonlinear boundaries, but their hidden-layer structure and weight initialization depend on experience, and they readily overfit when samples are limited. Liu et al. [24] noted that shallow networks are constrained by their single-hidden-layer structure and have limited ability to extract discriminative information automatically under complex conditions and incipient weak faults.

### 3.5 Comparative Analysis of Conventional Machine-Learning Methods

Each of the above methods has its own range of applicability. Reviews by Liu et al. [24] and Lei et al. [7] compared them in terms of applicable conditions, data requirements, interpretability, computational cost, and noise robustness. For small samples with high-dimensional features, SVMs are highly competitive because of structural risk minimization. When feature engineering is adequate and operating conditions vary, k-nearest neighbors also performs robustly because it makes no distributional assumptions [23]. Naive Bayes and shallow backpropagation networks are worthwhile only when samples are sufficient and features are approximately independent or nonlinearly separable.

In terms of engineering usability, random forests and XGBoost are the most convenient: they require neither kernel selection nor feature scaling, their rules are readable, and their tuning is intuitive. The comparison by Zhang et al. [22] demonstrated their advantages in overfitting resistance and computation time. Interpretability is one of the few advantages of conventional methods over deep learning, because decision-tree rules can be translated directly into field criteria. Noise robustness, however, depends strongly on front-end feature engineering; if the features are noise-sensitive, even a powerful classifier cannot compensate. Overall, conventional machine learning remains valuable when labeled data are limited, features can be constructed from physical knowledge, and interpretability and rapid deployment are required. However, its shallow representational capacity and dependence on human expertise become increasingly apparent with massive datasets, cross-condition domain shifts, and incipient weak faults, shifting the research focus toward end-to-end deep learning.

## 4. Fault-Diagnosis Methods Based on Deep Learning

### 4.1 Convolutional Neural Network Methods

Convolutional neural networks (CNNs) integrate feature extraction and classification in a single jointly trained network, avoiding the expert-dependent manual feature engineering discussed in Section 3. Early one-dimensional CNNs used raw vibration signals directly as input. The wide first-layer kernel deep convolutional neural network (WDCNN) proposed by Zhang et al. [25] uses a large convolution kernel in its first layer to capture longer segments and suppress high-frequency noise, followed by small kernels for multilayer nonlinear mapping. Adaptive batch normalization (AdaBN) reduces differences in feature distributions across loads, allowing the model to outperform shallow networks based on frequency-domain statistics under variable loads and strong noise. In subsequent work [26], denoising preprocessing was eliminated, and end-to-end diagnosis was performed directly on raw time-domain signals. High accuracy was retained under varying conditions



© 2026 The Author(s). Authors retain copyright in their articles.

without a dedicated domain-adaptation algorithm, and feature visualization explained how the network distinguished impact components buried in noise.

Another approach transforms signals into time-frequency grayscale images by wavelet analysis and feeds them into a two-dimensional CNN [27]. Residual networks alleviate degradation caused by stacking deep layers. Zhang et al. [28] first decomposed raw signals by wavelet packets, used frequency-band attention to emphasize important wavelet-coefficient bands, and used channel attention to strengthen convolution-channel responses. Their time-frequency feature extraction was validated on simulated and measured wind-turbine gearbox data. To reduce parameter count and improve inference efficiency for deployment, Wang et al. [29] introduced $1 \times 1$ convolutions into a one-dimensional network to enhance nonlinear representation and replaced fully connected layers with global average pooling, reducing both model size and overfitting risk while identifying multiple bearing-fault states.

## 4.2 Recurrent Neural Network Methods

Unlike CNNs, which emphasize local spatial patterns, recurrent neural networks (RNNs) recursively update hidden states along the time axis to model dependencies between successive signal instants. Conventional RNNs suffer from vanishing gradients on long sequences. Long short-term memory (LSTM) networks and gated recurrent units (GRUs) use input, forget, and related gates to retain historical information selectively and capture long-term temporal patterns. Lei et al. [30] applied end-to-end LSTM to multivariate time-series fault diagnosis in wind turbines, learning features directly from multichannel monitoring variables. The method outperformed existing approaches on two wind-turbine datasets and remained robust with small samples.

A more typical use of this approach is degradation assessment and remaining-useful-life prediction. Zhang et al. [31] fed trend indicators such as waveform entropy into an LSTM recurrent network, optimized its parameters using particle swarm optimization, and integrated cumulative information on fault propagation over time, enabling relatively accurate identification of bearing-degradation stages and prediction of remaining useful life. Because RNNs unfold serially over time steps, training and inference are less efficient than in parallelizable convolutional networks, and performance is more sensitive to sample length and annotation volume. When the objective is to distinguish instantaneous impacts rather than track state evolution, their advantage over CNNs is limited.

## 4.3 Autoencoder and Generative-Model Methods

Autoencoders (AEs) learn latent representations without supervision through compression and reconstruction, matching the diagnostic reality that unlabeled data are readily available while fault labels are scarce. Jia et al. [32] used a stacked autoencoder to mine fault features layer by layer from raw vibration signals. By addressing the dependence of handcrafted features on experience and the limited expressiveness of shallow networks, the method outperformed existing approaches on large datasets from rolling bearings and planetary gearboxes and established the AE paradigm. Denoising, sparse, and related variants further constrain the encoding layer so that the network learns robust, compact, intrinsic features rather than memorizing individual samples. For diesel-engine valve-fault identification, Chen et al. [33] introduced dropout and batch normalization to suppress overfitting and used a harmony-search algorithm to optimize the hyperparameters of a stacked autoencoder adaptively. On data covering seven valve-health states, the method was more accurate than a fixed-parameter stacked autoencoder and conventional approaches.

Because an AE can measure reconstruction error, it is closely related to anomaly detection: normal samples have small reconstruction errors, whereas unknown faults produce large errors, enabling alarms without fault labels. On the generative side, generative adversarial networks (GANs) implicitly model the real-data distribution through a minimax game between a generator and a discriminator [34]. When fault samples are scarce, GANs can synthesize rare samples and improve classification generalization under class imbalance [35]. AEs and GANs thus complement each other in completing scarce samples; data augmentation is discussed further in a later section.

## 4.4 Attention-Mechanism and Transformer Methods

Attention mechanisms allow models to assign weights automatically to important locations, while self-attention enables direct interactions between any two-time steps, overcoming the local receptive field of CNNs



© 2026 The Author(s). Authors retain copyright in their articles.

and the serial long-range modeling of RNNs. The Vision Transformer (ViT) divides a time-frequency signal image into patches and models global relationships through self-attention. However, the original ViT extracts features equally from all patches and incurs quadratic self-attention complexity. Ji et al. [36] therefore proposed a deformable-convolution-based Vision Transformer (DEViT), adding a deformable-convolution feature-extraction module at the embedding stage and using probabilistic attention to reduce computation. Top-1 accuracies of 99.81% and 97.32% were achieved on the CWRU and Southeast University (SEU) gearbox benchmarks, respectively.

Pure Transformers are computationally expensive, making lightweight designs and convolution-Transformer hybrids mainstream. Deng et al. [37] fused vibration signals into multidomain two-dimensional time-frequency images using the Gramian angular field, bispectrum, and Chirplet transform. They designed a lightweight scale-aware modulated Swin Transformer that combines convolution with windowed self-attention and incorporates a feature pyramid. Accuracies of 99.88%, 99.92%, and 99.96% were achieved under three variable-speed conditions, outperforming ResNet, ViT, and the original Swin Transformer. Attention is also frequently used as a plug-in to strengthen CNNs; the frequency-band and channel attention in Ref. [28], for example, guides convolutional networks toward important frequency bands and channels.

### 4.5 Performance Comparison of Deep-Learning Methods

The four method families differ in signal format, data requirements, robustness to noise and variable conditions, computational cost, and interpretability. Selection should therefore be scenario-driven rather than novelty-driven. In terms of input, one-dimensional CNNs can read raw vibration signals directly without time-frequency transformation, and WDCNN and related models have demonstrated robustness under strong noise and varying loads [25, 26]. Two-dimensional CNNs, ViT, and Swin Transformer usually use time-frequency images and can exploit mature vision-modeling techniques, but they add a signal-transformation stage [37, 36]. RNNs and LSTMs are less suited to discriminating instantaneous impacts but are indispensable for degradation assessment and remaining-useful-life tasks that track state evolution over time [30, 31].

Regarding data requirements, the AE family relies on unlabeled pretraining and is suitable when annotations are scarce, whereas CNNs and Transformers are data-driven and generalize poorly with insufficient samples. In a review of CNN-based diagnosis, Jiao et al. [27] identified class imbalance, weak cross-condition generalization, and poor interpretability as common challenges; Refs. [6] and [8] likewise regard them as major barriers to deployment. In terms of computational cost, a one-dimensional CNN lightweighted with $1 \times 1$ convolutions and global average pooling is best suited to real-time edge deployment [29]. RNN inference is slower because of serial unfolding, while Transformers require the most computation and must be compressed through lightweight variants. All four method families are black boxes and require activation visualization or post hoc explanations [27]. Overall, CNN backbones are preferred for routine classification, LSTMs for temporal degradation and life prediction, AEs for feature pretraining or anomaly detection when unlabeled data are abundant, and lightweight Transformers when high accuracy is required and sufficient computing resources are available [5]. Figure 2 compares the fault-diagnosis accuracies of representative models from these four families on public benchmarks.

## 5. Emerging Methods for Complex Industrial Environments

### 5.1 Few-Shot and Zero-Shot Fault Diagnosis

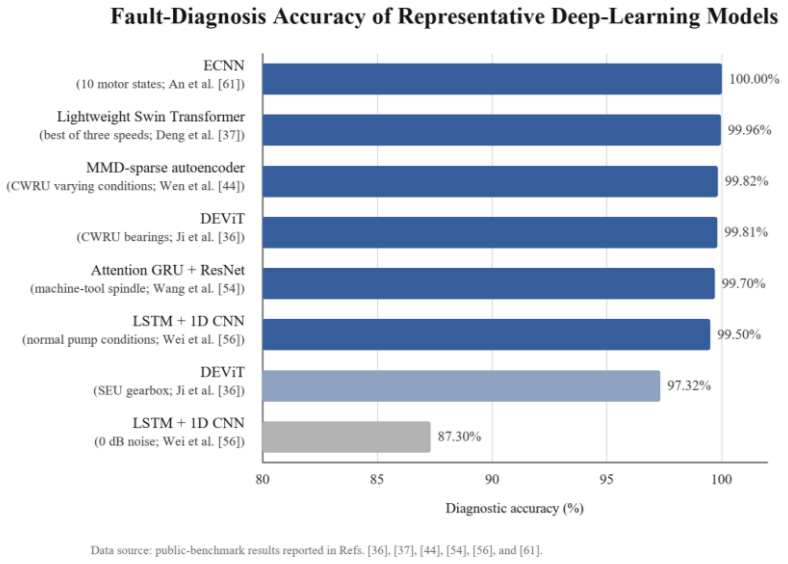
Deep models depend on large labeled datasets, whereas industrial sites present the opposite conditions: units operate normally for long periods, rare defects are difficult to reproduce, and only tens of labeled fault samples may be available. Wu et al. [38] grouped few-shot strategies for rotating machinery into meta-learning, transfer learning, domain generalization, data augmentation, and self-supervised learning, with the central bottleneck being the combination of scarce labels and operating-condition drift. Meta-learning is represented by the model-agnostic meta-learning (MAML) method of Finn et al. [39], which learns an initialization that is sensitive to fine-tuning so that a new fault class can be adapted using only a few gradient steps. Prototypical networks use the mean of the support samples in each class as a class prototype in embedding space and perform N-way K-shot classification according to the Euclidean distance from a query sample to each prototype [40]. Jiang et al. [41] proposed a two-branch prototypical network (TBPN) in which time-domain



© 2026 The Author(s). Authors retain copyright in their articles.

and frequency-domain signals learn fault prototypes separately before the two distance measures are fused, demonstrating complementary benefits on enterprise data. Limitations remain: meta-learning still requires many source tasks and considerable computing resources, while prototypes become distorted under substantial interdomain shifts [38]. Zero-shot diagnosis uses semantic attributes to identify unseen fault classes, but stable mappings between mechanical faults and semantic labels are lacking, leaving the approach far from engineering deployment.

*Figure 2: Comparison of fault-diagnosis accuracy among representative deep-learning models (Data source: references cited in Sections 4 and 6)*




## 5.2 Transfer-Learning and Domain-Adaptation Methods

Data distributions shift when the same equipment operates at different speeds or loads, and the difference becomes greater when a model is transferred to a similar unit. Laboratory models therefore often suffer a sharp loss of accuracy in the field, a cross-condition and cross-equipment domain-shift problem. Yan et al. [42] categorized knowledge transfer by its object into parameter, instance, feature, and relational transfer: parameter transfer reuses weights, instance transfer reweights source-domain samples, feature transfer seeks shared representations across domains, and relational transfer mines latent mappings among operating conditions. For feature alignment, the domain-adversarial neural network (DANN) of Ganin et al. [43] places a gradient-reversal layer and domain discriminator after the feature extractor, making features domain-invariant while preserving their discriminative power. Training requires only labeled source-domain data and unlabeled target-domain data. Maximum mean discrepancy (MMD) provides another explicit alignment route. Wen et al. [44] introduced MMD into a three-layer sparse autoencoder to minimize the distribution difference between the training and test domains, achieving 99.82% on CWRU bearings under varying conditions. When real fault samples are scarce, Liu and Gryllias [45] used a phenomenological bearing model to generate simulated source-domain signals in batches and applied DANN to align coarsely labeled simulated data with finely labeled real data, approaching fully supervised accuracy with only a small amount of real data. Domain generalization trains transferable models without access to target-domain data; it is more difficult than domain adaptation but better reflects field deployment [38].

## 5.3 Multisensor Fusion and Graph Neural Network Methods

A single vibration sensor captures only a local dynamic response. In coupled systems such as bearings, gearboxes, and motors, fault signals are spatially dispersed and overlap in frequency. Although installing multiple sensors increases wiring costs, complementary views generally provide stronger discrimination than a single channel. Traditional fusion often concatenates channels and cannot model spatial dependencies among sensors. Graph neural networks treat each sensor or feature channel as a node and relationships among nodes as edges, providing a structured paradigm for multisource-signal modeling. Li et al. [46] used the Grow-Shrink



© 2026 The Author(s). Authors retain copyright in their articles.

algorithm to extract dependencies among multisensor signals and construct a graph topology, and then used an improved graph attention network, GATv2, to assign adaptive weights to and aggregate neighboring nodes, improving diagnostic accuracy over single-sensor methods. Sun and Yin [47] treated multisensor data points as nodes, connected adjacent points from the same sensor to carry spatial and temporal information, and replaced conventional graph convolution with empirical-mode-decomposition graph convolution to extract multiple intrinsic-mode features adaptively, achieving more than 99% accuracy with limited samples. Graph construction still depends on prior knowledge or data-driven exploration, and the interpretability of attention weights remains unclear. As the number of sensors increases, graph sparsity and computational cost must be balanced.

### 5.4 Federated Learning and Distributed Diagnosis

In large factories and cross-enterprise alliances, workshop data cannot be centralized because of process confidentiality and regulatory requirements, creating data silos. Training locally in isolation, however, prevents experience from being shared across equipment. Federated learning enables collaborative modeling without moving raw data outside its local domain. FedAvg established the paradigm of multiple local gradient-descent rounds followed by equal-weight server aggregation. Yet participants often differ greatly in operating conditions and have non-independent and identically distributed (Non-IID) labels; equal averaging incorporates domain differences into the global model and degrades accuracy. Chen et al. [48] proposed discrepancy-based weighted federated averaging (D-WFA), in which each participant trains a generalized model using labeled source-domain and unlabeled target-domain data, while the server assigns dynamic weights according to interdomain distances measured by MMD. The method outperformed equal-weight FedAvg. In BCE-FL, Xiao et al. [49] constructed a feature-contrastive loss to mitigate Non-IID bias and added incentives and reputation-based Byzantine-fault-tolerance scoring, demonstrating security on chiller and wind-turbine gearbox datasets. Kang et al. [50] improved parameter transfer across rolling bearings of different specifications by introducing differential updates and parameter sparsification. After FedAvg aggregation, the server progressively unfroze layers to retain shared parameters, followed by local fine-tuning for each user, thereby enabling transfer across bearing specifications. Federated diagnosis still incurs many communication rounds and slow convergence, and the balance between personalization and global consistency remains unresolved.

### 5.5 Generative Adversarial Networks and Data Augmentation

Mechanical-fault data follow a long-tail distribution: normal samples are abundant, whereas incipient or compound faults may have only a few examples. Directly training a deep network therefore biases it toward majority classes and causes rare faults to be missed. Oversampling is a common remedy, but simply duplicating minority-class samples readily induces overfitting. The generative adversarial network (GAN) introduced by Goodfellow et al. [34] uses a minimax game between a generator and a discriminator so that the generator implicitly approximates the real distribution, providing a framework for synthesizing rare fault samples. Wang et al. [35] used spectra as input and generated rare-fault samples with an improved Wasserstein GAN (WGAN). They combined the synthetic and real samples to train a stacked autoencoder for state identification and visualized the model layer by layer to analyze the generator's mechanism, markedly improving classification generalization under class imbalance. Compared with generation from raw vibration signals, using spectra or time-frequency images as an intermediate representation converges more readily and integrates more easily with existing feature pipelines. Synthetic data nevertheless carry risks: the generator may suffer mode collapse and cover only a few modes of the real distribution, and subtle distribution shifts between generated and real samples may inflate test-set performance while causing field failure. Data augmentation with generated samples must therefore be accompanied by authenticity assessment and physical-consistency checks.

## 6. Representative Application Scenarios

### 6.1 Fault Diagnosis of Rotating Machinery

Rotating machinery is the most thoroughly developed application of intelligent diagnosis. Bearings, gearboxes, and rotors serve as validation platforms for nearly all public algorithms. Rolling bearings are the single-point components with the richest fault samples and most extensive research (Fig. 3), while planetary



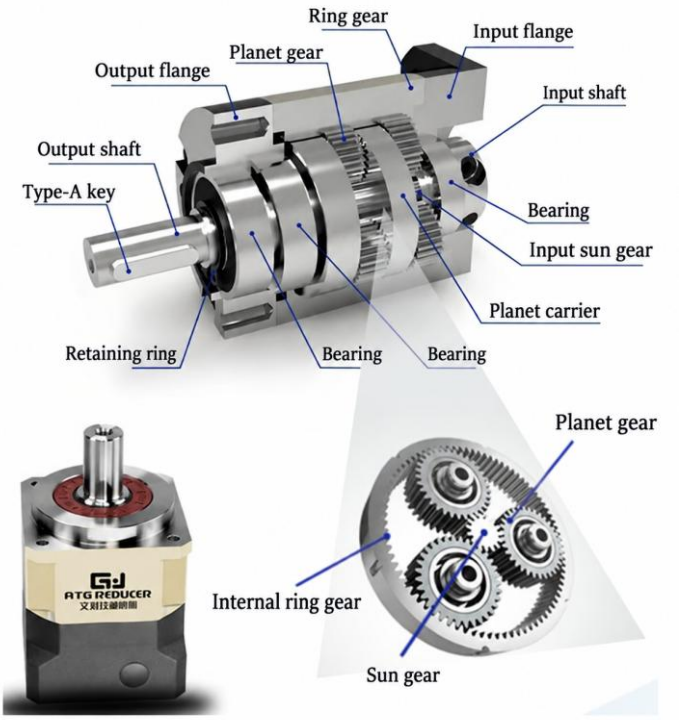
© 2026 The Author(s). Authors retain copyright in their articles.

gear reducers are representative coupled transmissions with multiple meshing gears (Fig. 4); together, they represent two core objects of intelligent rotating-machinery diagnosis. For bearings, Hoang and Kang [9] reviewed applications of autoencoders, restricted Boltzmann machines, and convolutional neural networks (CNNs). Jiao et al. [27] further grouped CNN studies into state classification, degradation prediction, and cross-domain transfer, enabling horizontal comparison of results on CWRU and related datasets. Gearbox environments are much more complex than those of single-point bearings because planetary-gear modulation spectra are coupled with variable loads. Zhang et al. [28] first applied wavelet-packet transformation to raw signals and then used frequency-band attention and channel attention to emphasize key frequency bands and convolutional-channel features, validating the resulting accuracy improvement on drivetrain-dynamics-simulator data and field measurements from a wind farm. Rotor faults such as misalignment, imbalance, and oil whirl are closely related to spectral structure. After comparing the suitability of deep learning for these three component classes, Shen et al. [51] noted that rotor diagnosis remains more dependent on prior physical knowledge and that the advantages of purely data-driven methods are less evident than for bearings. The field has progressed from manual features plus shallow classifiers to end-to-end deep models, but high laboratory accuracy does not guarantee field readiness; robustness under strong noise and variable loads remains critical.

*Figure 3: Rolling bearing (Source: publicly available product image)*



*Figure 4: Structure and transmission schematic of a planetary gear reducer (Source: publicly available material)*






© 2026 The Author(s). Authors retain copyright in their articles.

## 6.2 Fault Diagnosis of Motors and Power Equipment

Motor diagnosis can reuse the current-measurement circuit of a variable-frequency drive without installing additional vibration sensors, giving motor current signature analysis (MCSA) a cost advantage. The three-phase induction motor shown in Fig. 5 is the most widely installed and used induction-type rotating machine in industry and the principal target of MCSA. In 2000, Benbouzid [10] established the theoretical framework for detecting induction-motor faults through stator-current spectral analysis and defined sideband-based criteria for identifying broken rotor bars and bearing faults. For induction motors and permanent-magnet synchronous motors, Niu et al. [52] grouped methods for detecting bearing faults, stator interturn short circuits, broken rotor bars, and related defects into five categories: MCSA spectral analysis, demodulation, time-frequency analysis, parameter estimation, and artificial intelligence. AI is converting expert threshold rules into data-driven classification models. Kumar et al. [53] noted that induction-motor monitoring is moving from isolated signal processing toward the integration of models, signals, and data. Pure MCSA has limited sensitivity to incipient mechanical faults, a shortcoming alleviated by end-to-end deep learning. Interturn short circuits and demagnetization in permanent-magnet synchronous motors have weaker signatures and fewer samples, creating a more urgent need for few-shot and interpretable methods.

*Figure 5: Three-phase induction motor (Source: publicly available product image)*



## 6.3 Fault Diagnosis of CNC Machine Tools and Industrial Robots

The spindles and feed systems of computer numerical control (CNC) machine tools operate under varying cutting loads. The vertical CNC machining center shown in Fig. 6 is representative, and its faults include both slow global degradation and fine-grained defects such as local resonance and misalignment. Wang et al. [54] designed a division-of-labor attention framework: a gated recurrent unit (GRU) identifies global degradation, while ResNet identifies local fine-grained faults after sym8 wavelet soft-threshold denoising. The framework achieved approximately 99.7% accuracy for no-load misalignment, local resonance, and loaded machining faults, showing that denoising plus specialized identification better matches the hierarchical characteristics of machine-tool faults than a single end-to-end network. The serial multijoint structure of the six-axis industrial robot shown in Fig. 7 creates another difficulty: installing a vibration sensor on every joint significantly increases deployment cost and wiring complexity. Long et al. [55] proposed an economical strategy that installs only one attitude sensor on the end joint and uses its output as the input to a hybrid stacked-autoencoder and support-vector-machine (SAE-SVM) model for reducer-fault identification. This demonstrates that under a constrained sensor budget, a single-source attitude signal and hybrid architecture can still capture joint-level faults, providing a feasible paradigm for low-cost monitoring of mass-produced robots.



© 2026 The Author(s). Authors retain copyright in their articles.

*Figure 6: Vertical CNC machining center (Source: publicly available product image)*



*Figure 7: Six-axis industrial robot (Source: publicly available product image)*




## 6.4 Fault Diagnosis of Pumps and Reciprocating Machinery

Cavitation and wear in hydraulic and plunger pumps are often accompanied by fluctuations in inlet pressure and flow, producing nonstationary signals. The high-speed axial piston pump shown in Fig. 8 is representative. On a high-speed axial piston-pump test rig, Wei et al. [56] combined an LSTM with a one-dimensional CNN to identify four cavitation levels under different inlet pressures. Accuracy was approximately 99.5% under normal conditions and remained approximately 87.3% at a 0 dB signal-to-noise ratio, demonstrating the adaptability of combined temporal modeling and convolutional features to strongly noisy pump-source signals. High-reliability equipment such as nuclear circulating-water pumps poses a more difficult problem: fault samples are extremely scarce, and load variation makes the training and operating distributions inconsistent. Cheng et al. [57] proposed a Q-network calibration ensemble (QCE) to address class imbalance and distribution shift and used Grad-CAM to verify that the network focused on fault features rather than noise. This case shows that diagnosis of pumps and reciprocating machinery is not merely an accuracy problem but also requires robustness and credibility under scarce data and stringent safety requirements. Table 2 summarizes representative fault types and monitoring-signal sources for the equipment discussed in Section 6.



© 2026 The Author(s). Authors retain copyright in their articles.

*Figure 8: Axial piston pump (Source: publicly available product image)*



*Table 2: Summary of Intelligent-Diagnosis Applications for Representative Equipment*

| Equipment | Typical Faults | Monitoring Signals | Representative Methods | Main Findings |
|---|---|---|---|---|
| Rolling bearings | Spalling and damage | Vibration | CNN, AE, and attention networks | High benchmark accuracy and mature deployment |
| Gearboxes | Modulation-spectrum coupling and variable loads | Vibration | Attention ResNet and XGBoost | Validated with both simulated and measured data |
| Motors | Broken rotor bars, interturn short circuits, and bearing faults | Current (MCSA) | Spectral analysis + AI classification | Clear cost advantage |
| CNC machine tools | Misalignment, local resonance, and global degradation | Vibration | Specialized GRU and ResNet identification | Approximately 99.7% for no-load and loaded faults |
| Industrial robots | Reducer faults | End-effector attitude | Hybrid SAE-SVM model | Low-cost single-sensor monitoring |
| Plunger pumps | Cavitation and wear | Pressure, flow, and vibration | LSTM + one-dimensional CNN | Approximately 99.5% under normal conditions and 87.3% at 0 dB |
| Nuclear circulating-water pumps | Class imbalance and distribution shift | Vibration | QCE + explainability verification | Balances robustness and credibility |

*Note: Compiled from Refs. [9], [27], [28], and [51]-[57] cited in Section 6.*

## 7. Current Challenges and Development Trends

### 7.1 Data-Quality Issues

The fundamental difference between industrial field data and laboratory benchmarks is their unfavorable distribution. Abundant normal samples and scarce fault samples create severe class imbalance, while historical labels often contain manual-annotation errors. Jia et al. [58] proposed a deep normalized convolutional neural network (DNCNN) to address sample imbalance and used neuron-activation maximization to reveal the behavior of deep networks under imbalance. From a weakly supervised perspective, Liu et al. [59] used bagging to sample the majority class and train multiple SVMs, predicted unlabeled data, and then replenished the minority class. A bidirectional GRU and weighted cross-entropy were used to suppress noisy labels, and the approach was validated on the PHM 2008 and 2010 datasets. Although rare-sample synthesis with GANs, loss reweighting, and semisupervised replenishment can alleviate imbalance, the distribution shift encountered for nuclear circulating-water pumps in Ref. [57] shows that resampling alone cannot eliminate systematic bias caused by operating-condition drift.

### 7.2 Model-Generalization Issues

The sharp performance decline of high-accuracy laboratory models across operating conditions and equipment is the greatest barrier to field deployment. Its root cause is domain shift between the training and



© 2026 The Author(s). Authors retain copyright in their articles.

target domains: changes in load, rotational speed, or transmission path alter signal distributions and deprive learned fault features of discriminative power. Wen et al. [44] used a three-layer sparse autoencoder for feature extraction and minimized maximum mean discrepancy (MMD), increasing accuracy to 99.82% under varying CWRU conditions and demonstrating that explicit distribution alignment can narrow interdomain gaps. Yan et al. [42] reviewed transfer diagnosis across conditions, locations, equipment, and fault types in terms of parameter, instance, feature, and relational knowledge transfer. Zhang et al. [26] showed that with improved training strategies and feature visualization, a deep CNN can retain high accuracy under variable loads without a dedicated domain-adaptation algorithm. Overall, transfer learning and domain generalization are effective, but generalization across equipment and specifications remains constrained by interdomain similarity.

## 7.3 Model-Interpretability Issues

The lack of trust in deep networks at industrial sites largely stems from their black-box nature: maintenance personnel cannot trace which waveform segment led the model to a decision. In safety-critical settings such as nuclear power and aviation, this directly affects compliance and accountability. Explainable artificial intelligence (XAI) has therefore been introduced. Brusa et al. [60] applied SHAP to SVM- and k-nearest-neighbor-based bearing-fault detection and identified the most influential features, such as vibration-signal skewness and waveform indicators. After SHAP-based feature selection, both models achieved accuracies above 98.5%, showing that interpretability and accuracy can coexist. For convolutional networks, Jia et al. [58] used neuron-activation maximization to analyze features learned by convolution kernels, while Cheng et al. [57] used Grad-CAM to confirm that the network focused on fault-impact segments rather than environmental noise. Most of these methods provide post hoc explanations; they can expose incorrect evidence but cannot structurally guarantee the physical plausibility of a decision, which explains the growing interest in physics-data integration.

## 7.4 Edge-Deployment and Real-Time Issues

Field monitoring requires diagnosis to be completed locally at edge nodes under constraints on computing power, bandwidth, and latency. Returning high-frequency vibration signals to the cloud is both uneconomical and vulnerable to network interruption. The efficient convolutional neural network (ECNN) designed by An et al. [61] can be deployed directly on an edge node and achieved 100% accuracy for ten motor states, with an inference time below 14 ms and memory consumption below 44 KiB, substantially less than classical deep models. EdgeCog, developed by Fu et al. [62], goes further by compressing a CNN through energy-entropy-based distillation and quantization so that it fits a Cortex-M4 microcontroller with only 192 kB of RAM and 512 kB of ROM; attention combined with envelope-spectrum analysis verified the validity of the extracted features. Wang et al. [29] reduced the model structurally, using $1 \times 1$ convolutions to enhance nonlinear representation and replacing fully connected layers with global average pooling, thereby greatly reducing parameter count while maintaining accuracy. Lightweight design, distillation, and quantization have become the three principal tools for deploying diagnostic models on embedded terminals.

## 7.5 Future Development Trends

Three pathways are becoming clear for addressing data scarcity, inadequate generalization, and weak interpretability. The first is physics-informed machine learning (PIML), which integrates physical mechanisms with data. Deng et al. [63] classified PIML methods in PHM and noted that the small-data-plus-sparse-physics-knowledge paradigm requires fault mechanisms, conservation equations, or expert rules to be embedded explicitly as network constraints rather than relying solely on data fitting. The second is digital-twin-driven predictive maintenance. In a systematic review of 42 studies, Van Dinter et al. [64] summarized objectives, platforms, and design patterns while warning that computational burden, data diversity, and model complexity are major barriers to scaling. The third is the introduction of foundation models and large language models. Wang et al. [65] proposed DiagLLM, which injects envelope-spectrum images and expert knowledge into a multimodal large model through visual instruction tuning and parameter-efficient fine-tuning. It outperformed mainstream baselines in few-shot and cross-dataset settings and provided diagnostic rationales in natural language.



© 2026 The Author(s). Authors retain copyright in their articles.

## 8. Conclusions

Following a data-feature-model-application framework, this paper reviewed research progress in machine-learning-based fault diagnosis of mechanical equipment. A three-pronged system of conventional machine learning, deep learning, and emerging methods has taken shape, with accuracies above 90% and in some cases approaching 100% for bearings, gearboxes, motors, machine tools, robots, and pumps. The diagnostic paradigm has progressed from manual features plus shallow classifiers to end-to-end intelligent diagnosis.

Nevertheless, a gap remains between high laboratory accuracy and industrial deployment. The three principal engineering bottlenecks are data quality, generalization, and interpretability.

Future research is expected to advance along four paths: reducing dependence on real fault data through PIML and digital twins; improving cross-equipment and cross-condition generalization through transfer learning, domain generalization, and foundation models; increasing decision transparency through SHAP, Grad-CAM, and large-model reasoning; and enabling real-time edge deployment through lightweight design, distillation, and quantization.

In summary, the field is evolving from discriminative diagnosis dependent on large amounts of labeled data toward few-shot, cross-condition, interpretable, and deployable solutions. Data governance and engineering implementation, rather than further increases in benchmark accuracy, are the central requirements for intelligent diagnosis to create value.

## References

[1] ZHAO Z B, WU J Y, LI T F, et al. Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: a review[J]. Chinese Journal of Mechanical Engineering, 2021, 34: 66. DOI: 10.1186/s10033-021-00570-7.

[2] JARDINE A K S, LIN D, BANJEVIC D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J]. Mechanical Systems and Signal Processing, 2006, 20(7): 1483-1510. DOI: 10.1016/j.ymssp.2005.09.012.

[3] CHEN X F, WANG S B, QIAO B J, et al. Basic research on machinery fault diagnostics: past, present, and future trends[J]. Frontiers of Mechanical Engineering, 2018, 13(2): 264-291. DOI: 10.1007/s11465-018-0472-3.

[4] SHEN B M, CHEN B J, ZHAO C H, et al. A review of deep learning in fault prognostics and health management of mechanical equipment[J]. Machine Tool & Hydraulics, 2021, 49(19): 162-171. (in Chinese)

[5] ZHAO R, YAN R, CHEN Z, et al. Deep learning and its applications to machine health monitoring[J]. Mechanical Systems and Signal Processing, 2019, 115: 213-237. DOI: 10.1016/j.ymssp.2018.05.050.

[6] KHAN S, YAIRI T. A review on the application of deep learning in system health management[J]. Mechanical Systems and Signal Processing, 2018, 107: 241-265. DOI: 10.1016/j.ymssp.2017.11.024.

[7] LEI Y, YANG B, JIANG X, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587. DOI: 10.1016/j.ymssp.2019.106587.

[8] ZHANG S, ZHANG S B, WANG B N, et al. Deep learning algorithms for bearing fault diagnostics—A comprehensive review[J]. IEEE Access, 2020, 8: 29857-29881. DOI: 10.1109/ACCESS.2020.2972859.

[9] HOANG D T, KANG H J. A survey on deep learning based bearing fault diagnosis[J]. Neurocomputing, 2019, 335: 327-335. DOI: 10.1016/j.neucom.2018.06.078.

[10] BENBOUZID M E H. A review of induction motors signature analysis as a medium for faults detection[J]. IEEE Transactions on Industrial Electronics, 2000, 47(5): 984-993. DOI: 10.1109/41.873206.

[11] LESSMEIER C, KIMOTHO J K, ZIMMER D, et al. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set



© 2026 The Author(s). Authors retain copyright in their articles.

for data-driven classification[C]//Proceedings of the 3rd European Conference of the PHM Society. Bilbao, Spain: PHM Society, 2016: 1-10.

[12] SMITH W A, RANDALL R B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study[J]. Mechanical Systems and Signal Processing, 2015, 64-65: 100-131. DOI: 10.1016/j.ymssp.2015.04.021.

[13] LEI Y G, HAN T Y, WANG B, et al. XJTU-SY rolling element bearing accelerated life test datasets: a tutorial[J]. Journal of Mechanical Engineering, 2019, 55(16): 1-9. DOI: 10.3901/JME.2019.16.001.

[14] WU Z, HUANG N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41. DOI: 10.1142/S1793536909000047.

[15] YANG Y, YU D, CHENG J. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM[J]. Measurement, 2007, 40(9-10): 943-950. DOI: 10.1016/j.measurement.2006.10.010.

[16] SAMANTA B, AL-BALUSHI K R. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features[J]. Mechanical Systems and Signal Processing, 2003, 17(2): 317-328. DOI: 10.1006/mssp.2001.1462.

[17] SAIDI L, BEN ALI J, FNAIECH F. Application of higher order spectral features and support vector machines for bearing faults classification[J]. ISA Transactions, 2015, 54: 193-206. DOI: 10.1016/j.isatra.2014.08.007.

[18] GUO W C, ZHAO H S, LI C, et al. Enhanced diagnosis of bearing fault features based on wavelet-packet energy spectrum and principal component analysis[J]. Acta Armamentarii, 2019, 40(11): 2370-2377. DOI: 10.3969/j.issn.1000-1093.2019.11.022. (in Chinese)

[19] WANG S X, HUANG Q, WANG X S, et al. Diagnosis of rolling-bearing fault severity based on WPT, PCA, and SVM[J]. Machinery Design & Manufacture, 2022(4): 5-9. DOI: 10.3969/j.issn.1001-3997.2022.04.002. (in Chinese)

[20] WIDODO A, YANG B S. Support vector machine in machine condition monitoring and fault diagnosis[J]. Mechanical Systems and Signal Processing, 2007, 21(6): 2560-2574. DOI: 10.1016/j.ymssp.2006.12.007.

[21] XU G, LIU M, JIANG Z, et al. Bearing Fault Diagnosis Method Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning[J]. Sensors, 2019, 19(5): 1088. DOI: 10.3390/s19051088.

[22] ZHANG R T, CHEN Z G, LI B B, et al. Gearbox fault diagnosis based on a deep convolutional neural network model and XGBoost[J]. Journal of Mechanical Strength, 2020, 42(5): 1059-1066. DOI: 10.16579/j.issn.1001.9669.2020.05.007. (in Chinese)

[23] WANG D. K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited[J]. Mechanical Systems and Signal Processing, 2016, 70-71: 201-208. DOI: 10.1016/j.ymssp.2015.10.007.

[24] LIU R, YANG B, ZIO E, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review[J]. Mechanical Systems and Signal Processing, 2018, 108: 33-47. DOI: 10.1016/j.ymssp.2018.02.016.

[25] ZHANG W, PENG G, LI C, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals[J]. Sensors, 2017, 17(2): 425. DOI: 10.3390/s17020425.

[26] ZHANG W, LI C, PENG G, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load[J]. Mechanical Systems and Signal Processing, 2018, 100: 439-453. DOI: 10.1016/j.ymssp.2017.06.022.

[27] JIAO J, ZHAO M, LIN J, et al. A comprehensive review on convolutional neural network in machine fault diagnosis[J]. Neurocomputing, 2020, 417: 36-63. DOI: 10.1016/j.neucom.2020.07.088.



© 2026 The Author(s). Authors retain copyright in their articles.

[28] ZHANG K, TANG B, DENG L, et al. A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox[J]. Measurement, 2021, 179: 109491. DOI: 10.1016/j.measurement.2021.109491.

[29] WANG Q, DENG L F, ZHAO R Z. Rolling-bearing fault identification based on an improved one-dimensional convolutional neural network[J]. Journal of Vibration and Shock, 2022, 41(3): 216-223. (in Chinese)

[30] LEI J, LIU C, JIANG D. Fault diagnosis of wind turbine based on Long Short-term memory networks[J]. Renewable Energy, 2019, 133: 422-432. DOI: 10.1016/j.renene.2018.10.031.

[31] ZHANG B, ZHANG S, LI W. Bearing performance degradation assessment using long short-term memory recurrent network[J]. Computers in Industry, 2019, 106: 14-29. DOI: 10.1016/j.compind.2018.12.016.

[32] JIA F, LEI Y, LIN J, et al. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data[J]. Mechanical Systems and Signal Processing, 2016, 72-73: 303-315. DOI: 10.1016/j.ymssp.2015.10.025.

[33] CHEN K, MAO Z W, ZHANG J J, et al. Diesel-engine fault diagnosis using a stacked autoencoder optimized by harmony search[J]. Journal of Mechanical Engineering, 2020, 56(11): 132-140. DOI: 10.3901/JME.2020.11.132. (in Chinese)

[34] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems (NeurIPS). Montreal: Curran Associates, 2014: 2672-2680.

[35] WANG J, LI S, HAN B, et al. Generalization of deep neural networks for imbalanced fault classification of machinery using generative adversarial networks[J]. IEEE Access, 2019, 7: 111168-111180. DOI: 10.1109/ACCESS.2019.2924003.

[36] JI M, ZHAO G. DEViT: Deformable convolution-based vision transformer for bearing fault diagnosis[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-13. DOI: 10.1109/TIM.2024.3440383.

[37] DENG F Y, ZHENG S X, HAO R J. Axle-box bearing fault diagnosis using a lightweight scale-aware modulated Swin Transformer[J]. Journal of Xi'an Jiaotong University, 2024, 58(9): 83-93. (in Chinese)

[38] WU K, WU J, SHU Q M, et al. Review of few-shot fault-diagnosis methods for rotating machinery based on deep learning[J]. Chinese Journal of Ship Research, 2025, 20(2): 3-19. DOI: 10.19693/j.issn.1673-3185.04175. (in Chinese)

[39] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//Proceedings of the 34th International Conference on Machine Learning (ICML). Sydney: PMLR, 2017: 1126-1135.

[40] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[C]//Advances in Neural Information Processing Systems (NeurIPS). Long Beach: Curran Associates, 2017: 4077-4087.

[41] JIANG C, CHEN H, XU Q, et al. Few-shot fault diagnosis of rotating machinery with two-branch prototypical networks[J]. Journal of Intelligent Manufacturing, 2023, 34(4): 1667-1681. DOI: 10.1007/s10845-021-01904-x.

[42] YAN R, SHEN F, SUN C, et al. Knowledge transfer for rotary machine fault diagnosis[J]. IEEE Sensors Journal, 2020, 20(15): 8374-8393. DOI: 10.1109/JSEN.2019.2949057.

[43] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.

[44] WEN L, GAO L, LI X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 49(1): 136-144. DOI: 10.1109/TSMC.2017.2754287.

[45] LIU C, GRYLLIAS K. Simulation-driven domain adaptation for rolling element bearing fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2022, 18(9): 5760-5770. DOI: 10.1109/TII.2021.3103412.



© 2026 The Author(s). Authors retain copyright in their articles.

[46] LI C, KWOH C K, LI X, et al. Rotating machinery fault diagnosis based on multi-sensor information fusion using graph attention network[C]//2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV). Singapore: IEEE, 2022: 678-683. DOI: 10.1109/ICARCV57592.2022.10004378.

[47] SUN K, YIN A. Multi-sensor temporal-spatial graph network fusion empirical mode decomposition convolution for machine fault diagnosis[J]. Information Fusion, 2025, 114: 102708. DOI: 10.1016/j.inffus.2024.102708.

[48] CHEN J, LI J, HUANG R, et al. Federated transfer learning for bearing fault diagnosis with discrepancy-based weighted federated averaging[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3518211. DOI: 10.1109/TIM.2022.3180417.

[49] XIAO Y, SHAO H, LIN J, et al. BCE-FL: A secure and privacy-preserving federated learning system for device fault diagnosis under Non-IID condition in IIoT[J]. IEEE Internet of Things Journal, 2024, 11(8): 14241-14252. DOI: 10.1109/JIOT.2023.3340745.

[50] KANG S Q, XIAO Y, WANG Y J, et al. Fault diagnosis of rolling bearings of different specifications based on federated model transfer[J]. Journal of Vibration and Shock, 2023, 42(22): 184-192. DOI: 10.13465/j.cnki.jvs.2023.22.021. (in Chinese)

[51] SHEN T, LI S M, XIN Y. Review of deep-learning-based fault diagnosis for rotating machinery[J]. Computer Measurement & Control, 2020, 28(9): 1-8. (in Chinese)

[52] NIU G, DONG X, CHEN Y. Motor fault diagnostics based on current signatures: a review[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 3506419. DOI: 10.1109/TIM.2023.3285999.

[53] KUMAR R R, ANDRIOLLO M, CIRRINCIONE G, et al. A comprehensive review of conventional and intelligence-based approaches for the fault diagnosis and condition monitoring of induction motors[J]. Energies, 2022, 15(23): 8938. DOI: 10.3390/en15238938.

[54] WANG W P, WANG Q, YU Y. Fault identification of machine-tool spindle systems based on attention mechanisms and deep-learning algorithms[J]. Acta Armamentarii, 2022, 43(4): 861-875. (in Chinese)

[55] LONG J, MOU J, ZHANG L, et al. Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots[J]. Journal of Manufacturing Systems, 2021, 61: 736-745. DOI: 10.1016/j.jmsy.2020.08.010.

[56] WEI X L, CHAO Q, TAO J F, et al. Fault diagnosis of high-speed plunger pumps based on LSTM and CNN[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(3): 429-439. (in Chinese)

[57] CHENG W, WANG S, LIU Y, et al. A novel planetary gearbox fault diagnosis method for nuclear circulating water pump with class imbalance and data distribution shift[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-13. DOI: 10.1109/TIM.2023.3238752.

[58] JIA F, LEI Y, LU N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization[J]. Mechanical Systems and Signal Processing, 2018, 110: 349-367. DOI: 10.1016/j.ymssp.2018.03.025.

[59] LIU H, LIU Z, JIA W, et al. A novel imbalanced data classification method based on weakly supervised learning for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2022, 18(3): 1583-1593. DOI: 10.1109/TII.2021.3084132.

[60] BRUSA E, CIBRARIO L, DELPRETE C, et al. Explainable AI for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring[J]. Applied Sciences, 2023, 13(4): 2038. DOI: 10.3390/app13042038.

[61] AN K, LU J, ZHU Q, et al. Edge solution for real-time motor fault diagnosis based on efficient convolutional neural network[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 3508912(1-12). DOI: 10.1109/TIM.2023.3276513.



© 2026 The Author(s). Authors retain copyright in their articles.

[62] FU L, YAN K, ZHANG Y, et al. EdgeCog: a real-time bearing fault diagnosis system based on lightweight edge computing[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 3515011(1-11). DOI: 10.1109/TIM.2023.3298403.

[63] DENG W, NGUYEN K T P, MEDJAHER K, et al. Physics-informed machine learning in prognostics and health management: state of the art and challenges[J]. Applied Mathematical Modelling, 2023, 124: 325-352. DOI: 10.1016/j.apm.2023.07.011.

[64] VAN DINTER R, TEKINERDOGAN B, CATAL C. Predictive maintenance using digital twins: a systematic literature review[J]. Information and Software Technology, 2022, 151: 107008. DOI: 10.1016/j.infsof.2022.107008.

[65] WANG J, LI T, YANG Y, et al. DiagLLM: multimodal reasoning with large language model for explainable bearing fault diagnosis[J]. Science China Information Sciences, 2025, 68(6): 161103. DOI: 10.1007/s11432-024-4333-7.

## 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.

## Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).



© 2026 The Author(s). Authors retain copyright in their articles.