Vol. 23 (2026): Proceedings of 2026 International Symposium on Engineering Technology Innovation and Engineering Management
(ISETIM 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/ISETIM-2026.2026.1418

Review of Data-Driven Fault Diagnosis for Rotating Machinery

Yan Wu*

School of Electronic Engineering, Heilongjiang University, Major in Electronic Science and Technology, Harbin, 150080, China

**Corresponding author: Yan Wu.*

Abstract

With the acceleration of industrialization in modern society, rotating machinery has become an indispensable and important part of various industries. Under long-term high-speed, heavy-load and variable working conditions, key components such as rotors, bearings and gearboxes are prone to failure. Traditional diagnostic methods based on models and experience are difficult to adapt to complex working conditions and multi-source data processing needs. With the advantages of no need for precise physical models, strong adaptability and high diagnostic accuracy, the data-driven method has become the mainstream technology of intelligent fault diagnosis of rotating machinery. This paper reviews the common fault types of rotating machinery, summarizes the research progress of the core technology of data-driven fault diagnosis, discusses its application status in typical industrial equipment, and discusses future research directions, so as to provide references for the research and engineering implementation of intelligent operation and maintenance technology of rotating machinery.

Keywords

fault diagnosis, data-driven, rotating machinery, deep learning

1. Introduction

With the development of modern industry, rotating machinery is essential equipment in industrial production in various industries and plays a crucial role in the whole industrial production. However, with the continuous expansion of the application scope of rotating machinery, the working environment of rotating machinery has become increasingly harsh, faults in rotating machinery are difficult to detect, and the problem of increasing wear of rotating machinery is also becoming more and more prominent. In traditional fault diagnosis for rotating machinery, it mainly relies on human experience and intuition to make judgments. This method not only consumes a lot of labor and resources, but also has low accuracy and reliability, which cannot meet the requirements of accuracy, reliability and safety of the current social production.

To meet the requirements for real-time monitoring, accurate early warning, and intelligent maintenance of modern equipment, fault diagnosis plays an essential role in ensuring safe and reliable operation. At present, there are three main methods for fault diagnosis of rotating machinery: model-driven method, knowledge-driven method and data-driven method. And among them, the data-driven method is the dominant one. Therefore, applying the fault diagnosis method to rotating machinery plays an important role in improving the

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

accuracy and effectiveness of rotating machinery maintenance. This paper will review the fault diagnosis of various parts of rotating machinery.

2. Common Faults in Rotating Machinery

The structure of rotating machinery is complex and the parts are closely related to each other. Most of the faults occur in three main areas such as rotors, bearings, gearboxes and supports. Different faults correspond to different forms of vibration, temperature, noise and other manifestations, which is the basis for fault diagnosis.

Rotor system failure is the most common type of failure in rotating machinery, mainly including rotor imbalance, shaft system misalignment, rotor cracks, rub-impact and other forms. The rotor imbalance is caused by processing error, material adhesion or parts falling off. It is characterized by a large amplitude of power frequency vibration, and it increases with the increase of rotational speed; The axial misalignment includes parallel misalignment, angular misalignment and a combination of the two, which will cause double-frequency vibration and axial vibration, which is also the reason for the large vibration of the equipment to a certain extent; Rotor cracks are generally caused by fatigue damage. Its occurrence and development are related to the operating speed and load, and nonlinear vibration is easy to occur; Rub-impact is caused by contact and friction between the rotor and stator and may lead to permanent rotor bending in severe cases.

Rolling bearings and gearboxes are important components of power transmission, and their failure rate is high. Rolling bearing faults include inner ring pitting, outer ring peeling, rolling element breakage and cage wear, etc. The fault signal is in the form of periodic impact and is easily disturbed by noise; The failure of the gearbox is mainly due to tooth breakage, tooth surface wear, tooth surface pitting, scuffing, etc. There will be gear meshing frequency and its double frequency, sidebands component, which shows strong modulation properties in vibration.

Support and structural failures mainly include mechanical loosening, oil film vortex and oil film oscillation. Loose base and loose bearing seat will produce low-frequency vibration, and its vibration signal is nonlinear and non-stationary; Oil film vortex and oil film oscillation are unique faults of sliding bearings. Due to the instability of the lubricating oil film, the oscillation frequency is close to the natural frequency of the rotor, which is easy to cause instability and damage to the shaft system; In addition, motor electrical faults, fluid-induced vibration, surge and other reasons will also affect the rotating machine, which needs to be judged by a variety of physical quantities.

3. Research on Data-Driven Fault Diagnosis for Rotating Machinery

Data-driven fault diagnosis is based on monitoring data. After preprocessing, feature extraction and classification and identification of the monitoring data, the process of precise location of the fault is completed. There is no need to establish a complex physical model. It has good robustness and generalization ability that is much higher than the traditional fault diagnosis method.

3.1 Data Preprocessing Techniques

Data preprocessing is crucial to ensure the accuracy of diagnosis. Rotary mechanical monitoring data includes vibration, temperature, current, sound and other types. Affected by external interference and sensor's own problems during collection, it is prone to missing values, noise, redundant information, etc. Therefore, the original data needs to be preprocessed, that is, data preprocessing includes data cleaning, signal noise reduction, normalization and data augmentation. Data cleaning is mainly used to fill in missing values with interpolation or mean values, and to remove abnormal values by threshold or clustering methods; signal denoising is carried out using wavelet transformation, empirical modal decomposition (EMD), variational mode decomposition (VMD) and other methods to suppress external noise while retaining fault impulse information. For example, Chen Yu, Cheng Long and Yang Bo apply simulation-based preprocessing methods in unmanned aerial vehicle fault diagnosis, which solves the impact of noise on data in small samples [1]; Data standardization and normalization eliminate the impact of different units and improve the speed of model training. For the problem of insufficient fault data, the method of oversampling or generative adversarial network (GAN) can be used to increase the amount of data to make the data more balanced. A method of

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

combining simulation experiments and data-driven proposed by Rongguang, Zhang Yexin, Tang Chao and others can provide good support for small sample failures of rolling bearings [2].

3.2 Fault FEATURE Extraction Methods

Feature extraction plays an important role in obtaining fault information and improving diagnostic accuracy. The traditional feature extraction method has three types: time domain, frequency domain and time-frequency domain: time domain feature extraction mean, variance, kurtosis, crest factor, etc., which can reflect the size of the signal amplitude; The frequency domain uses FFT and envelope spectrum to obtain the fault characteristic frequency and determine the position of the bearing or gear; the use of STFT and wavelet packet transformation in the time frequency domain is more suitable for non-stationary and nonlinear fault signals [3].

With the development of deep learning, automatic feature extraction is gradually being paid attention to. For example, CNN, LSTM, autoencoder (AE), etc. can automatically learn deep fault features from the original signal. The subjectivity and limitations brought about by the need for human feature selection greatly improve the representation ability of complex faults. When analyzing the early faults of the wind power transmission chain, Li Jun, Liu Yuanyuan and Wang Heng used the deep learning automatic feature extraction method to solve the disadvantage that traditional feature extraction is not easy to find weak fault signals [4]; In addition, some technical means of feature selection and dimension reduction, such as principal component analysis (PCA), linear discriminant analysis (LDA), etc., can also be used to reduce the feature space dimension to improve the training speed and accuracy.

3.3 Fault Classification and Identification Models

Classification identification is an important step in fault judgment, including traditional machine learning and deep learning. Traditional machine learning mainly includes SVM, random forest, K-nearest neighbors (KNN), XGBoost, etc., which has the characteristics of small calculation volume and fast in a small amount of data and simple fault detection. In Sun Dengyun's research on machine learning-based coal mine ventilation fault detection, he compared the effects of different algorithms such as SVM, random forest, neural network, etc., which shows that traditional machine learning can be applied to fault classification and identification in complex environments [5]; Due to its superior feature learning and pattern recognition ability, deep learning has a good diagnostic effect under variable working conditions, compound faults and large data. On this basis, the proposed hybrid models such as CNN-LSTM, Transformer, graph neural network, etc. make the model have better cross-device and cross-condition generalization ability.

For small samples and unbalanced problems, transfer learning, meta-learning, weakly supervised learning and other methods can be used to solve the problems caused by the small number of fault samples in the industrial site. Han Daiyu and others proposed the application of a method based on data-driven and multi-source domain transfer learning for the fault diagnosis of hydraulic pumps in the case of small samples [6]; Model evaluation generally uses accuracy, precision, recall, F1-score, etc. for comparison, and tests on CWRU and XJTU public data to ensure the effectiveness of the proposed method.

4. Applications of Data-Driven Fault Diagnosis in Rotating Machinery

Fault diagnosis technology based on data analysis has developed from the laboratory stage to the industrial site. It has been widely used in energy and power, petrochemical, rail transit, aerospace and other industries and promoted the transformation of equipment maintenance from corrective maintenance and scheduled maintenance to preventive maintenance.

In the energy and power industry, wind turbine generator sets, water pumps, electric motors, etc. all apply data-driven diagnostic strategies. Zhao Mengnan studied the fault characteristics of the wind power transmission system and carried out corresponding work [7]. Gao Zhikang proposed an early fault diagnosis technology for wind turbine transmission chain, which helps to reduce the number of unexpected downtimes of wind power equipment [8]; Li Hao and others established a wind power gearbox fault prediction model based on the operation mechanism and a hybrid mechanism- and data-driven approach, and provided reasonable suggestions for the maintenance of wind power equipment [9]. By monitoring the vibration, rotational speed, temperature, current, etc. of the unit, it can timely find problems such as blade cracking,

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

bearing damage, rotor imbalance, etc., so as to improve the power generation capacity of wind power and thermal power.

In the petrochemical industry, compressors, centrifuges, pump groups and other important rotary equipment have been running continuously. The data-driven diagnostic system can analyze a large amount of sensor information in time, effectively monitor problems in gearboxes, bearings, rotors and other parts, prevent accidental downtime, and ensure the smooth progress and safety and reliability of chemical production. Luo Hongjiao's research on data-driven fault diagnosis of hydroelectric power units has also enabled hydropower equipment to better play its role [10].

In the fields of rail transit and aerospace, train bearings, gearboxes and aircraft engine rotors have very high requirements for reliability. Data-driven methods integrate edge computing to carry out online fault detection onboard railway vehicles and aircraft to improve the safety of transportation. Li Xiangyi and Zhang Shijie's research on the hybrid eccentric fault of high-speed train axle box and permanent magnet synchronous motor provides a feasible method for rail transit equipment fault diagnosis [11][12]; In intelligent manufacturing, common industrial equipment such as CNC machine tool spindles, industrial fans, conveying pumps, etc. can be managed throughout the life cycle by using data-driven diagnosis to reduce the maintenance costs of enterprises.

5. Conclusions and Future Perspectives

The data-driven method solves the problem that traditional rotary mechanical fault diagnosis requires the establishment of physical models, relying on human experience and poor adaptation of working conditions. It has achieved good results in data preparation, feature extraction and classification. It has been widely used in many industries such as unmanned aerial vehicles, wind turbines and rail transit. It plays an important role in the intelligent maintenance of industrial equipment, proving that it has a wide application prospect.

At present, data-driven fault diagnosis still has problems such as weak robustness to complex working conditions, difficult to effectively identify composite faults, a small number of samples cannot be well promoted to different devices, model lacks interpretability, low data quality and difficult engineering technology to achieve. The future research direction should focus on the combination of knowledge and data, multiple types of information acquisition, rapid learning ability under a small number of samples, digital twin–based management and standardized engineering implementation throughout the product life cycle, so as to further improve the diagnosis level and achieve large-scale promotion and application, and help the manufacturing industry move towards high-end.

References

[1] Chen, Y., Cheng, L., & Yang, B. (2026). Research on a data-driven gas-path fault diagnosis method for turboshaft engines. *Journal of Chongqing University of Technology (Natural Science)*, *40*(1), 185–192.

[2] Rong, G., Zhang, Y., Tang, C., et al. (2025). Research on simulation data-driven fault diagnosis technology for unmanned aerial vehicles. *Aerospace Defense*, *8*(6), 73–84.

[3] Zhang, Y. (2025). Research on an intelligent fault diagnosis mechanism for mine ventilation based on machine learning. *Journal of Safety Science and Technology*, *21*(11), 159–168.

[4] Li, J., Liu, Y., Wang, H., et al. (2025). A data-driven method for early fault diagnosis of rolling bearings. *Cyber Security and Data Governance*, *44*(Suppl. 1), 341–346.

[5] Sun, D. (2025). *Research on data-driven fault feature analysis and diagnosis methods for wind turbine transmission systems* (Master's thesis, Yanshan University).

[6] Han, D. (2025). *Research on a generative data-driven fault diagnosis method for hydraulic pumps based on multi-source domain transfer learning* (Master's thesis, Yanshan University).

[7] Zhao, M. (2025). *Research on data-driven fault detection and diagnosis and symptom chain inference for heating, ventilation, and air conditioning systems* (Master's thesis, Beijing University of Civil Engineering and Architecture).

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group

[8] Gao, Z. (2025). *Research on rolling bearing fault diagnosis driven by data augmentation and multi-source information fusion* (Master's thesis, Beijing University of Civil Engineering and Architecture).

[9] Li, H. (2025). *Research on axle box bearing fault diagnosis for high-speed trains based on the integration of fault mechanism simulation and data-driven methods* (Master's thesis, Shijiazhuang Tiedao University).

[10] Luo, H. (2025). *Research on data-driven fault diagnosis and remaining useful life prediction methods for rolling bearings* (Master's thesis, Qingdao University of Technology).

[11] Li, X. (2025). *Research on data-driven diagnosis of hybrid eccentricity faults in permanent magnet synchronous motors* (Master's thesis, Harbin Institute of Technology).

[12] Zhang, S. (2025). *Research on data-driven fault diagnosis of civil aviation engine gas-path systems* (Master's thesis, University of Electronic Science and Technology of China).

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/).

Copyright © 2026 The Authors. All individual papers remain the copyright of their respective author(s) · Published by Zeus Academic Research Publishing Group