Research Status, Challenges and Future Prospects of Bearing Fault Diagnosis

Song Qi*, Guoqing Zhu

China Academy of Machinery Seience &Technology Qingdao Branch Co., Ltd., Qingdao 266300, China

*15063052877@163.com

Abstract. As the core component of rotating machinery, bearing condition monitoring and fault diagnosis are very important to ensure the safety of equipment. This paper systematically reviews the development of bearing fault diagnosis technology, covering three major methods based on signal processing, machine learning and deep learning. Firstly, the typical fault mechanism of bearing is expounded. Secondly, the principles, representative technologies, advantages and disadvantages of the three mainstream diagnostic methods are analyzed in detail. It is pointed out that the signal processing method relies on expert experience, the machine learning method is limited by artificial features, and the deep learning method has data dependence. On this basis, the limitations of current research in variable condition generalization, small sample adaptability and multi-physics fusion are discussed in depth. Finally, the future development trend is prospected, and it is pointed out that the research will focus on multi-source heterogeneous data deep fusion, interpretable and small sample intelligent algorithm innovation, cross-domain diagnosis and edge real-time computing, and finally evolve to the closed-loop intelligent operation and maintenance mode of predictive maintenance and equipment life cycle management to cope with increasingly complex engineering challenges.

Keywords: Bearing fault diagnosis; Deep learning; Multi-source information fusion; Predictive maintenance; Intelligent diagnosis

1 Introduction

With the continuous improvement of industrial automation, the operation condition monitoring and fault diagnosis technology of mechanical equipment has been paid more and more attention. As a key component in rotating machinery, the running state of the bearing directly affects the overall performance and life of the equipment. Therefore, the early identification and diagnosis of bearing faults has become an important means to ensure the safe operation of equipment and improve production efficiency.

Y. Xu et al. (eds.), *Proceedings of the 2026 5th International Conference on Engineering Management and Information Science (EMIS 2026)*, Advances in Computer Science Research 130,
https://doi.org/10.2991/978-94-6239-652-4_11

2 Analysis of Bearing Failure Mechanism

As a key transmission component in mechanical equipment, the running state of rolling bearing directly affects the performance and life of the equipment. In practical applications, bearings often fail for a variety of reasons. The common fault types mainly include wear, fatigue, corrosion, fracture, plastic deformation and gluing.

Wear failure usually occurs on the surface of the inner ring, outer ring or rolling element of the bearing, mainly due to the friction between the materials, especially in the case of poor lubrication or large load.

Fatigue failure is due to the fact that under the action of long-term alternating stress, small cracks are generated on the surface of the rolling element or raceway. These cracks gradually expand, eventually leading to the peeling of the material and the formation of pits. In severe cases, the bearing will lose its normal function.

Corrosion failure is mostly caused by external environmental factors. For example, the lubricating oil contains water or impurities, or the bearing is exposed to a humid, acid-base environment for a long time, which causes chemical or electrochemical reactions, corrodes the bearing surface, and weakens its bearing capacity.

Fracture failure is mostly caused by overload, high speed or improper installation. When the stress of the bearing exceeds the strength limit of the material, it may cause the fracture of the component and cause serious accidents.

Plastic deformation generally occurs in the case of improper assembly or strong impact of the bearing, resulting in indentation or depression on the surface of the rolling element or raceway, affecting the normal operation of the bearing.

Adhesion failure is due to insufficient lubrication or high speed, so that a high temperature and high pressure environment is formed between the contact surfaces, resulting in adhesion of the two contact surfaces, which in turn leads to material tearing or surface damage.

These fault types are often interrelated. For example, poor lubrication may lead to wear and scuffing at the same time, while overload may cause fatigue and fracture. Therefore, in practical application, the bearing type and lubrication mode should be reasonably selected according to the operating environment, load conditions and maintenance requirements of the equipment, and the condition monitoring and maintenance should be carried out regularly to prolong the service life of the bearing and ensure the stable operation of the equipment.

3 Bearing Fault Diagnosis Method Classification

3.1 Diagnosis Method Based on Signal Processing

The diagnosis method based on signal processing is the most classical and basic technical route in the field of bearing fault diagnosis. The core idea is to extract the feature information that can reflect the fault state of the bearing from the collected original signal (mainly the vibration signal) through signal processing technology, and then judge the type and degree of the fault by analyzing these features.

The diagnosis method based on signal processing is the most classical and basic technical route in the field of bearing fault diagnosis. The core idea is to extract the feature information that can reflect the fault state of the bearing from the collected original signal (mainly the vibration signal) through signal processing technology, and then judge the type and degree of the fault by analyzing these features.

Among them, variational mode decomposition (VMD) is widely used to suppress modal aliasing due to its adaptive decomposition ability [1]. The improved empirical wavelet transform (EWT) filters sensitive components through kurtosis and correlation coefficient, and its envelope spectrum can more clearly identify the outer ring fault frequency and its frequency doubling [2]. Multivariate variational mode decomposition (MVMD) combined with 1.5-dimensional envelope spectrum can simultaneously diagnose the inner ring 132.67 Hz and outer ring 87.83 Hz composite faults, and its EECI parameter optimization method significantly improves the feature selection accuracy [3]. In the bearing monitoring of nuclear power plants, the EWT-GG clustering method uses the experimental data of Case Western Reserve University to verify the detectability of EDM defects [6].After the full-vector EEMD fusion of dual-channel signals, the hidden Markov model (HMM) is used to make the inner ring fault recognition rate reach 100 % [4].

The diagnosis method of signal processing has the advantages of clear physical meaning, strong interpretability, relatively small amount of calculation, and wide application in engineering practice. However, it also has limitations, relying heavily on expert experience for feature extraction and selection ; sensitive to noise and working condition changes ; the degree of automation is low, and it is difficult to deal with complex and changeable fault modes.

The diagnosis method of signal processing has the advantages of clear physical meaning, strong interpretability, relatively small amount of calculation, and wide application in engineering practice. However, it also has limitations, relying heavily on expert experience for feature extraction and selection ; sensitive to noise and working condition changes ; the degree of automation is low, and it is difficult to deal with complex and changeable fault modes.

3.2 Diagnosis Method Based on Machine Learning

The method based on machine learning is an extension and intelligent upgrade of traditional signal processing methods. It regards the diagnosis problem as a pattern recognition problem, and uses machine learning algorithms to automatically learn the mapping relationship from signal features to fault states.

BP neural network is the basic framework because of its simple structure and convenient training. The standard BP network is faced with the problem of gradient disappearance and local optimum in the diagnosis of aviation bearings. The accuracy rate is improved to 98.7 % by using the large mutation genetic algorithm (GMGA) [5]. Support vector machine (SVM) combines energy entropy to achieve 97.3 % classification accuracy in nuclear power plant bearing diagnosis by virtue of the advantage of kernel function in dealing with high-dimensional nonlinear features, but its performance is significantly affected by random assignment of kernel parameters [6]. Extreme learning machine (ELM) greatly shortens the training time by randomly initializing the weight of hidden layer. When combined with variational mode decomposition-multiscale permutation entropy (VMD-MPE), the fault recognition speed is 1.865 seconds higher than that of support vector machine, and the accuracy rate is 99.5 % [1].

The diagnosis method of machine learning has the advantages of automation and intelligence of the diagnosis process. Compared with the traditional method, it can deal with more complex feature relationships and has higher classification accuracy. However, it still relies on artificial feature extraction. The upper limit of model performance is subject to feature quality. It requires a large amount of labeled data for training and has limited adaptability to changes in working conditions.

3.3 Diagnostic Methods Based on Deep Learning

The method based on deep learning represents the most cutting-edge technology of bearing fault diagnosis. It realizes the ' end-to-end ' diagnosis, that is, automatically learning and extracting the optimal features from the original signal directly, and completing the fault classification, eliminating the tedious manual feature screening process, and significantly improving the automatic diagnosis level of bearing faults.

Among them, convolutional neural network (CNN) combines bidirectional gated recurrent unit (BiGRU) and attention mechanism in the diagnosis of SKF6205 bearing, so that the average accuracy under variable load conditions reaches 85 %, which is 20 % higher than that of traditional SVM [7]. Transformer architecture captures the spatio-temporal dependence of vibration signals by means of self-attention mechanism. The DPformer model of dual parallel design realizes the accurate detection of multi-scale fault modes by simultaneously processing time series and frequency domain features [8].

The diagnosis method of deep learning does not require manual feature screening, and has strong automatic learning ability. It can process massive data and learn deeper and more abstract fault features. Under complex working conditions and large data sets, it can usually achieve the highest diagnosis accuracy. At the same time, a large amount of labeled data is needed for training. The model structure is complex, the calculation cost is high, and GPU acceleration is usually required. The model has poor interpretability, high requirements for data quality and quantity, and is easy to over-fit in data scarcity scenarios.

4 Analysis of Applicability and Limitations of the Method

The current research on bearing fault diagnosis has made significant progress in both method innovation and engineering application. The signal processing method effectively suppresses modal aliasing through adaptive decomposition techniques such as VMD and EWT, and combines high-order transient extraction and synchronous compression transformation to achieve accurate positioning of microsecond-level impact components. Machine learning methods continue to make breakthroughs in feature selection and model optimization. G-DPSO algorithm and integrated learning strategy make the accuracy of compound fault diagnosis exceed 97 % [9]. The deep learning method achieves a recognition rate of nearly 100 % in cross-domain diagnosis by virtue of Transformer architecture and diffusion model. However, there are still three significant limitations in the existing research. First, the generalization ability of most methods is insufficient under variable working conditions. For example, the accuracy of traditional SVM is reduced by 30 % when the load changes [7].Although sometimes frequency domain resampling and other compensation strategies are used, the feature drift mechanism under the coupling of speed and load has not been fully elucidated. Second, the diagnostic reliability under the condition of small samples lacks systematic verification. The existing data augmentation methods rely on the prior distribution assumption of the target domain, and the problem of sample scarcity and labeling cost in the real industrial environment has not been fundamentally solved [10]. Thirdly, there is a theoretical gap in multi-physical field coupling diagnosis. Although the current 3D sound field analysis improves the recognition rate by 10 % [11], the quantitative evaluation criteria and weight distribution mechanism of multi-source information fusion such as vibration-temperature-acoustic emission have not been established.

5 Engineering Application Challenges and Future Development Trends

At present, bearing fault diagnosis technology still faces multiple challenges in engineering application, and its core contradiction is reflected in the gap between theoretical methods and actual needs. Specifically, it is mainly reflected in the following aspects :

(1) Signal feature extraction problem and low signal-to-noise ratio dilemma

It is difficult to extract bearing fault signals, and there is a common problem of low signal-to-noise ratio. During the operation of the bearing, the signal generated by the early fault is extremely weak. At the same time, the background noise inside the mechanical system is strong, which makes the fault signal easily submerged. This situation leads to huge obstacles for traditional methods to extract effective fault features, and it is difficult to accurately capture key information that can reflect the fault from complex signals.

(2) Interference caused by complex transfer path

Vibration signals need to go through complex mechanical structures in the process of transmission. This complex transmission path will not only lead to signal attenuation, but also cause multi-source coupling interference. The interaction of these interference factors makes the characteristic signal that can reflect the fault distorted, which increases the risk of misdiagnosis and missed diagnosis, and brings great uncertainty to the accurate diagnosis of bearing faults.

(3) The dual constraints of changing working conditions and sample scarcity

The actual equipment operating conditions are complex and changeable, such as speed and load and other factors are often in dynamic change. The fluctuation of this working condition will lead to the drift of the fault characteristic frequency, which makes it difficult for the diagnosis model based on the fixed characteristic frequency to adapt to the change of the actual working condition, and the generalization ability of the model is greatly reduced. In addition, it is extremely difficult to obtain a sufficient number of fault data with accurate labels in the industrial field, and the learning and training of supervised models are highly dependent on these labeled data. The scarcity of samples and the high cost of labeling make it difficult for supervised models to be fully trained. In the case of small samples, the diagnostic accuracy is severely limited.

(4) Shortcomings in model generalization and robustness

The existing diagnostic models have obvious deficiencies in generalization ability and robustness. The same diagnostic model often shows unstable performance under different equipment or new working conditions, and it is difficult to transplant and apply directly. Moreover, in the actual noise environment, the anti-interference ability of the traditional diagnostic method is weak, and the diagnostic accuracy will be significantly reduced, which seriously affects the reliability of the diagnostic results and greatly reduces the effect of the model in practical engineering applications.

(5) Field equipment dependence and localization technology promotion dilemma

Some enterprises rely too much on foreign products in the selection of bearing fault diagnosis equipment, which hinders the promotion of domestic diagnosis technology. This not only makes enterprises face higher cost pressure in equipment procurement and maintenance, but also is not conducive to the independent development and innovation of bearing fault diagnosis technology in China, which further limits the wide application of domestic technology in engineering practice.

Although the current research on bearing fault diagnosis has established three methods of signal processing, machine learning and deep learning, and has made significant breakthroughs in feature extraction, model optimization and cross-domain diagnosis, in the face of increasingly complex industrial scenarios and stringent equipment reliability requirements, this field still needs to be explored in depth. In the future, technology will evolve in the direction of intelligence, precision and generalization. Its core development trends are reflected in the following aspects :

(1) From single utilization to multi-source heterogeneous deep fusion

The key breakthrough in the future lies in the effective mining and integration of multi-source heterogeneous data. Bearing operation data have a wide range of sources and various types. The research will focus on deepening the fusion technology, such as constructing a multi-modal fusion model based on deep learning to reveal the intrinsic correlation between different data, so as to fundamentally improve the diagnostic accuracy. On this basis, a structured fault knowledge base will be constructed by using association rule mining and other technologies, combined with knowledge graph, to provide deep knowledge support for intelligent diagnosis and maintenance decision-making.

(2) From model application to innovation and credibility.

The innovation and optimization of intelligent algorithms is the core driving force. Although the current deep learning model has achieved remarkable results, the bottleneck of its poor interpretability and insufficient adaptability of small samples has become increasingly prominent. Future research will focus on developing interpretable models and exploring efficient learning methods for small sample scenarios. In addition, the collaborative integration of multiple intelligent algorithms will become an important trend, such as combining signal processing with deep learning, machine learning with optimization algorithms to form a hybrid intelligent model with complementary advantages, so as to comprehensively improve the comprehensive performance.

(3) From specific scenarios to cross-domain universal and real-time response

Cross-domain and cross-device diagnosis is the key to improve the versatility of technology. The complexity of the actual working conditions and environment leads to the drift of fault features. In the future, the adaptability of the model under different working conditions will be significantly improved through strategies such as transfer learning. In order to achieve cross-device diagnosis, we will focus on building a large-scale, standardized data set to train a universal diagnostic model. At the same time, real-time online monitoring and edge computing technology are indispensable. By combining the Internet of Things and cloud computing, a real-time monitoring system is constructed, and edge computing is used for efficient processing at the data source, which can not only reduce the transmission delay, but also ensure the immediacy of diagnosis.

(4) From fault diagnosis to predictive maintenance and life cycle management

Finally, the realization of technical value lies in the deep integration with the equipment maintenance strategy. Bearing fault diagnosis will no longer be an isolated link, but will extend upstream to predictive maintenance. Through accurate remaining useful life prediction, the optimal maintenance strategy will be formulated. Downstream through the whole life cycle management of equipment. This closed-loop intelligent operation and maintenance mode will provide solid technical support for ensuring the high reliability operation of industrial equipment and promoting the intelligent transformation of manufacturing industry.

6 Conclusion

This paper systematically reviews the research status, core challenges and future trends in the field of bearing fault diagnosis. Research shows that the technology has developed from a signal processing method that relies on expert knowledge to a machine learning that realizes automatic classification, and then to the current dominant end-to-end deep learning method. The technical path is clear and the diagnostic accuracy continues to improve. However, through in-depth analysis of existing research, there is still a gap between theory and engineering practice. The core challenges are mainly reflected in the difficulty of weak fault feature extraction under strong noise background, the lack of model generalization ability under variable working conditions, the reliability bottleneck of small sample learning, and the lack of theory of multi-source information fusion such as vibration-temperature-acoustic emission.

In the future, the research on bearing fault diagnosis is undergoing profound changes from single technology to systematization, from passive diagnosis to active prediction, and from isolated analysis to closed-loop management. Multi-source heterogeneous data fusion will build a more comprehensive device health portrait ; the interpretable and small sample intelligent algorithm will enhance the reliability and credibility of the model. Cross-domain diagnosis and edge computing will promote the generalization and real-time of technology. The deep integration with predictive maintenance strategy will eventually realize intelligent operation and maintenance throughout the whole life cycle of equipment, providing key support for industrial intelligent transformation. Overcoming these challenges will be the key to promote the bearing fault diagnosis technology from the laboratory to a wider and more reliable engineering application.

References

1. Zeng X, Sun W, Wang H G, et al. Fault Diagnosis Analysis of Rolling Bearing Based on VMD-MPE Combined with PNN and LEM[J]. Hot Working Technology, 2022,51(10):157-163.
2. Zhao G X, Li F, Wei Z H, et al. Bearing Fault Diagnosis Method Based on Improved EWT[J]. Coal Mine Machinery, 2020,41(07):157-159.
3. Tang G J, Xue G, Wang X L, et al. Fault Diagnosis for Rolling Bearing Based on Multivariate Variational Modal Decomposition and 1.5-Dimensional Spectrum[J]. Bearing, 2022(12): 74-82.
4. Gao S, Zhou Y P, Chen H, et al. Application of Full Vector-Spectrum EEMD in Bearing Fault Diagnosis[J]. Machinery Design & Manufacture, 2021(03):118-121.
5. Pi J, Liu P, Ma S, et al. Frequency Prediction Based on SVR-ARMA Combination Model Improved by Particle Swarm Optimization[J]. Journal of Vibration, Measurement & Diagnosis, 2020, 40(02): 381-388+423.
6. Wang Z C, Xia H, Zhu S M, et al. Bearing fault diagnosis method in nuclear power plants based on EWT-GG clustering[J]. Journal of Harbin Engineering University, 2020, 41(06): 899-906.
7. Wu D M, Wang F Q, Li X G, et al. Bearing intelligent fault diagnosis[J]. Journal of Mine Automation, 2022, 48(09): 49-55.
8. Tai Z Y, Hou T Y, Liu M, et al. Application of dual-path Transformer in bearing fault diagnosis[J]. Journal of Chongqing University of Technology (Natural Science), 2025, 39(05): 179-184.
9. Zhang Y L, Yan J Y. Fault Diagnosis of Bearing Based on G-DPSO and Decision Tree[J]. Industrial Engineering Journal, 2021, 24(06): 41-47.
10. Lai Z H, Luo L K, Xu D S, et al. Fault diagnosis with a cross-domain small sample based on the diffusion model[J]. Journal of Vibration and Shock, 2025, 44(18): 273-283.
11. Hou J J, Zhang Y Q, Fang Z P. Diagnosing Bearing Faults Using Spatial Features of Three-Dimensional Sound Field[J]. Journal of Vibration, Measurement & Diagnosis, 2021, 41(05): 933-938+1034.