Measurement Science
and Technology

Bearing Prognostics and Health Management Based on Hybrid Physical Mechanism and Data Models：A Systematic Review

To cite this article before publication: Shuo Wang *et al* 2025 *Meas. Sci. Technol.* in press https://doi.org/10.1088/1361-6501/adcce4

Bearing Prognostics and Health Management Based on Hybrid Physical Mechanism and Data Models: A Systematic Review

Shuo Wang [a, b], Liang Yan [a], Shichang Du [a, *], Shanshan Li [b], and Xianmin Chen [b]

[a] National Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

[b] National Key Laboratory of Strength and Structural Integrity, Xian, Shaanxi,710065,China

* Corresponding author (E-mail: lovbin@sjtu.edu.cn)

Abstract: Traditional physical models face significant challenges in parameter determination and adapting to complex systems, whereas data-driven models are constrained by data quality and quantity, making them susceptible to overfitting or underfitting problems. These limitations lead to deficiencies in robustness, physical interpretability, and generalization capabilities of existing models. In recent years, the fusion of physical mechanism-based and data-driven approaches has effectively addressed the shortcomings of both types of models, attracting widespread attention. However, there is no systematic review specifically on bearing prognosis and health management under hybrid physical mechanism and data-driven methods. To fill this gap, this paper comprehensively analyzes the research advancements in bearing prognosis and health management based on hybrid physical mechanism and data-driven methods. From the perspective of fusion strategy, the paper categorizes bearing prognosis and health management methods based on the fusion of physical mechanism and data-driven model into three levels: data level, network level, and model level, and further subdivides the research methods at each fusion level. In each subdivision field, this paper discusses the application of each research method in three main aspects: condition monitoring, fault diagnosis, and remaining useful life (RUL) prediction, summarizing the research methods employed by current scholars. Finally, this paper evaluates the advantages and disadvantages of each analytical method in practical applications, identifies current research challenges, and proposes future research directions. The aim is to provide guidance and in-depth insights for researchers and engineers in the field of bearing prognosis and health management.

Keywords: Bearing prognosis and health management; Hybrid physical mechanism and data-driven model; Data level; Network level; Model level

1. Introduction

Under the background of Germany's Industry 4.0, America's Industrial Internet, Made in China 2030, and Japan's Super Intelligent Society 5.0, manufacturing machinery and equipment are developing rapidly [1, 2]. Among these mechanical systems, rotating machinery is a significant category with extensive applications [3, 4] . Bearings, as vital components in rotating machinery [5-7], play a variety of roles, including supporting the rotating shaft, minimizing friction, guiding motion, bearing loads, maintaining the center position of the shaft, absorbing vibrations and shocks, simplifying design, enhancing efficiency, and extending mechanical life [8, 9]. Therefore, it is applied to aerospace, automobiles, machine tools and other fields. Typically, mechanical system failures are often caused by the damage of some components [10, 11], in which the failure rate of bearings is close to 30% [12]. Damage to bearings can severely threaten the normal operation and safety of the mechanical system, leading to mechanical failures and increasing maintenance costs. For example, in 1992, Kansai Electric Power Company of Japan caused a loss of nearly 5 billion due to the failure of bearings [13]. Therefore, active health management of bearings is essential to ensure the normal operation of mechanical systems and prevent accidents [14, 15].

Driven by the need to maintain the normal operation of critical systems, prognostic and health management (PHM) has emerged as a prognostic technology for proactively preventing unexpected failures [16]. It mainly assesses equipment faults by monitoring the operating status of the mechanical system and predicts the RUL of the mechanical system based on the current status. The application of PHM technology significantly enhances the reliability, operational safety, and equipment maintenance efficiency of mechanical systems [17, 18]. Due to its importance and rapid development, PHM technology has been widely used in bearing health management to improve the safety of the system [19].

Therefore, to illustrate the importance and popularity of bearing PHM research, this section first summarizes the trend of bearing PHM research and details the main research methods in bearing PHM. Subsequently, the differences between this review and the existing review are expounded. Finally, the motivation of writing this article is explained to highlight the main contribution of this paper.

1.1 The trend of bearing PHM research

In recent years, PHM is a key technology to ensure the reliable, efficient and safe operation of engineering equipment and systems [19], which has attracted extensive attention in academic field. Most of these studies focus on engineering machinery. Among them, key components such as bearings have been extensively researched [20]. In the academic field, the Web of Science database contains academic papers in various research directions, which can be used to analyze the development of scientific research in a certain field. To more intuitively portray the academic research on bearing PHM, this paper uses the data from the Web of Science database to represent the popularity of research on bearing PHM technology. Since the update of the database will cause changes in the data, all the papers collected in this paper were published before December 2024 to avoid errors caused by data updates. To ensure that the retrieved literature is the same as the research field of this paper, the search criteria are defined as follows: [Topic = ("PHM" OR "fault diagnosis" OR "RUL" OR "Condition Monitoring" OR "remaining useful life")] AND [Topic = (bearings)] AND [Publication type = (article) OR (review) OR (conference paper)] AND [Publishing year = ("1998 – 2024")]. According to the above standards, this paper counts the number of publications on bearing PHM technology over the past 26 years, as shown in Figure 1.

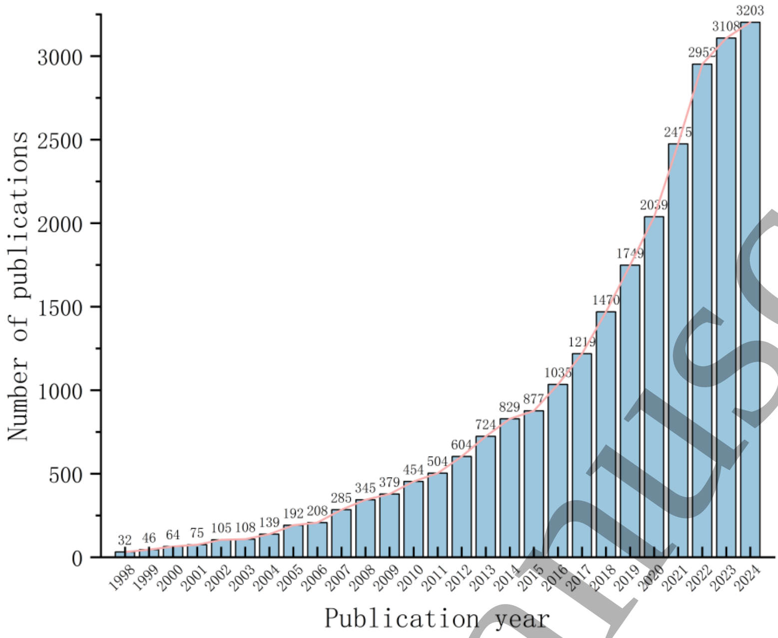

Figure 1 The number of publications on bearing PHM technology in Web of Science

Figure 1 illustrates that the research on bearing PHM has been on an upward trend since 1999. The growth trend has been even more obvious since 2015. At the same time, the research on bearing PHM has also been applied to many industrial fields, mainly including aerospace equipment [21, 22], machine tools [23], wind power generation systems [24, 25]. These data further illustrate that bearing PHM research has become a hot topic in current research.

1.2 Related works on bearings PHM technology

Generally speaking, PHM technology refers to real-time monitoring, analysis and diagnosis of the state of equipment or system to predict potential failures and adopt corresponding maintenance strategies [26, 27]. It mainly includes operating status monitoring, fault identification and diagnosis, health status assessment, and RUL prediction [28, 29]. Therefore, the main research directions of PHM technology are divided into mechanical system status monitoring, fault diagnosis, and RUL prediction. By conducting PHM research on mechanical systems, enterprises can prevent mechanical systems failure in advance. This approach ensures the reliability of equipment operation and facilitates timely maintenance. To achieve accurate PHM, numerous methods have been proposed by many scholars, roughly divided into three categories: physical mechanism methods, data-driven methods, and hybrid methods [30]. Among these methods, the method of physical mechanism is to model the failure form of equipment by using experimental results and prior knowledge. This model representing the relationship between equipment failure and external working conditions, so as to achieve the role of pre-prevention. The data-driven method is to substitute data into the constructed model, analyzing the relationship between data and equipment failure to characterize the state of the equipment, and thus achieving preemptive prevention. The core difference between the two methods lies in the degree of prior knowledge required during the modeling process. The physical mechanism method relies heavily on extensive knowledge related to failures to ensure the accuracy and reliability of the model, while the data-driven method focuses on extracting latent features from data without the need for extensive domain knowledge, although its physical interpretability is relatively weaker. However, both methods have their own strengths and limitations in practical applications. The hybrid approach aims to combine the advantages of both methods and reduce the modeling complexity associated with each individual method. Nevertheless, this also implies that the hybrid method will be influenced by both methods. The following subsections will provide detailed introductions to these three methods.

1.2.1 Physical mechanism methods

Physical mechanism methods mainly describe the degradation process of mechanical systems through natural physical laws, and construct relevant physical mechanism models to evaluate the current state of mechanical system. Through these physical mechanism models, it is possible to discover the law of how the health of mechanical systems changes with time under external loads, so as to carry out timely maintenance. These classic models mainly include crack growth models [31], bearing dynamic models [32], wear models [33], electrochemical models [34]. Generally, with sufficient experimental data and sufficient domain knowledge, physical mechanism methods can better complete PHM work. However, in most cases, the physical mechanism of mechanical systems is particularly complex, and their domain knowledge is incomplete or even difficult to obtain [35]. Therefore, most physical mechanism methods are carried out under certain assumptions [36, 37]. These physical mechanism models usually lead to low diagnostic efficiency or inaccurate life prediction. Moreover, even under the same external load, the failure process of different mechanical systems vary significantly, which limits the generalization ability of physical mechanism-based methods. All these aspects limit the widespread application of physical mechanism-based PHM technology.

1.2.2 Data-driven methods

With advancements in sensor technology, acquiring monitoring data related to system degradation has become both convenient and cost-effective. Such as, vibration signals, acoustic emission signals and temperature signals [38-42]. These data can distinguish the degradation information of the mechanical system in the whole life cycle [43], thereby determining the state and life stage of the mechanical system. Therefore, modeling the degradation data of the mechanical system enables relevant PHM research. This modeling method can not only get rid of the requirement of physical mechanism methods for the integrity of prior knowledge but also effectively alleviate the current situation of lack of physical experimental data. Driven by monitoring data, researchers have widely studied and developed various data-driven PHM research methods over the past few decades [44]. The general process of bearing PHM research using a data-driven approach is shown in Figure 2. First, the bearing is in different degrees of degradation due to various operating loads and environmental factors, and various sensors are used to obtain bearing degradation data. After that, some data processing algorithms are used to preprocess the data to obtain high-quality signals. Finally, a data-driven model is built based on the degradation data type to realize bearing status detection, fault diagnosis, RUL prediction, and other functions.

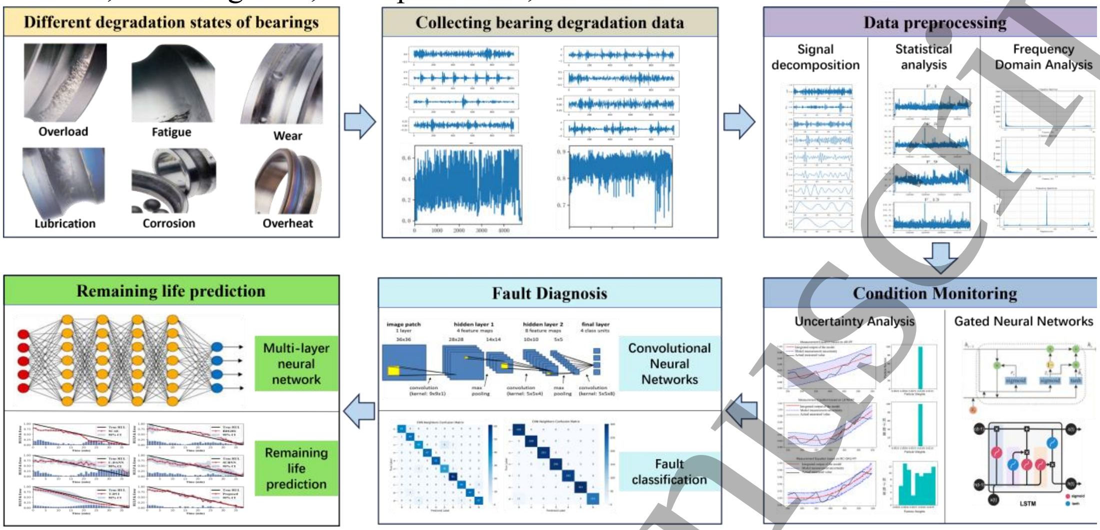

Figure 2 The general process of bearing PHM research under data-driven model

At present, data-driven PHM research methods mainly include statistical methods and machine learning (ML) methods [45]. The basic idea of the statistical method is to find potential faults by analyzing the operation data of equipment in normal and abnormal states. Among them, common analysis methods include regression analysis, time series analysis, hypothesis testing and probability model. Through these methods, the probability model of equipment performance degradation and fault occurrence is established. These models can be used to observe the health status of equipment and predict its future performance trends and possible failure times. The parameters of the probability model are estimated based on the monitoring data [46]. The main advantage of PHM research based on statistical method is that it can estimate model parameters based on data and give probabilistic prediction of the model. In this process, the confidence level of the prediction results can be given, thereby quantifying the uncertainty of the PHM technology itself. Another advantage is that statistical methods can optimize the real-time performance and responsiveness of PHM systems. For example, it can use probabilistic models and hypothesis testing to quickly extract key information from massive data sets, thereby reducing the time and resources required for data processing. However, this method is highly dependent on the quality of historical data. If the data is noisy, incomplete, or inaccurate, it will directly impact the reliability of the predictions. At the same time, statistical methods may need to make some assumptions about data distribution when modeling. These assumptions are not always valid and will affect the accuracy of the model.

With the development of the Internet of Things (IoT) [47] and cloud computing [48], human society has entered the big data era. This shift has resulted in a significant increase in both the volume and diversity of mechanical system monitoring data, creating a demand for more advanced technologies to address these challenges. Driven by this, artificial intelligence algorithms are gradually used by scholars to solve various problems. In this context, PHM technology based on ML methods have become a hot topic in PHM [49]. Generally, ML methods for achieving equipment PHM typically involves the following steps: data collection and acquisition, data preprocessing and feature extraction, feature evaluation and screening, and pattern recognition and model training. Common ML methods mainly include support vector machines(SVM) [50], K nearest neighbor algorithm [51], random forest [52]. Deep learning, a branch of ML methods, employs multi-layer Neural Networks to learn complex patterns, making it suitable for processing unstructured data and solving complex tasks. The specific classification of deep learning models is shown in Figure 3. Compared to ML methods, deep learning can discern and comprehend more complex patterns and relationships in data, rendering it suitable for processing large-scale, high-dimensional data. The advantage of this method is its strong generalization ability enabling diagnosis and prediction in many data sets. Moreover, this method does not require physical prior knowledge and can achieve PHM through a large amount of labeled data. Therefore, over the past five years, numerous scholars have conducted research in related directions and achieved a series of results. However, many ML models, such as Neural Networks, are black box models that lack interpretability [53]. At the same time, this model is highly dependent on the quality and quantity of collected data, which leads to low accuracy of ML models in the absence of data labels. Therefore, insufficient data can lead to decisions that deviate from reality. Due to the poor physical interpretability of ML methods, it is difficult for maintainers to identify wrong results. Even worse, subsequent predictions will inherit and amplify these wrong results, resulting in very serious consequences. Therefore, how to improve the physical interpretability of ML models and the universality of models in the case of small amounts of data remains a very challenging problem.

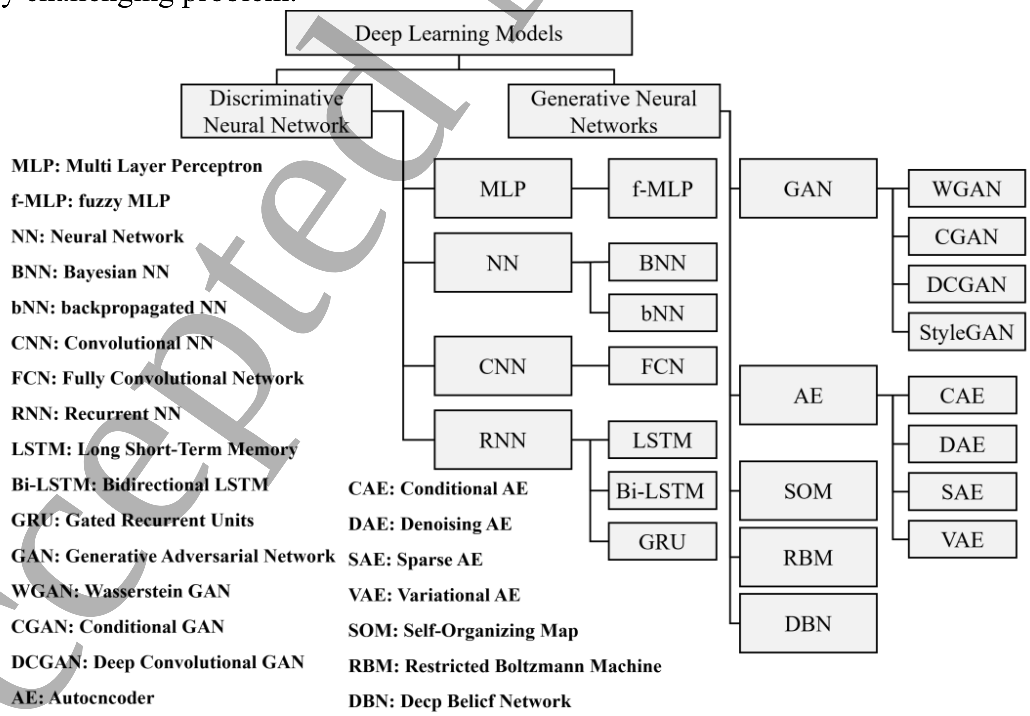

Figure 3 Classification of commonly used deep learning frameworks

1.2.3 Hybrid physical mechanism and data-driven methods

In general, physical mechanism methods and data-driven methods have their own characteristics in terms of generalization, accuracy, physical interpretability, and model building difficulty. However, each method may have certain limitations when applied individually to the PHM of complex mechanical systems. In this case, if the advantages of these two methods can be combined, the prediction accuracy and physical interpretability of PHM technology will be improved. In other words, the prior knowledge of physical mechanisms enhances the physical interpretability of data-driven models, while the data-driven models are used to improve the generalization ability and accuracy of physical mechanism models. Therefore, many scholars have put forward the method of physical mechanism guiding data-driven model [54, 55]. This method mainly combines physical knowledge with data-driven model in the process of training ML model. However, it should be noted that the implementation process of hybrid physical mechanisms and data-driven methods is often complicated. At the same time, there is no universal combination method for PHM research using hybrid methods, and scholars need to combine them according to actual engineering conditions.

This paper summarizes the advantages and disadvantages of physical mechanism-based methods, data-driven methods, and hybrid methods in application, as shown in Table 1. In addition, this paper also compares these three methods in calculation costs, usage scenarios and other dimensions, and the results are shown in Table 2.

Table 1 Comparison of advantages and disadvantages of existing methods

Methods	Advantage	Disadvantage
Physical mechanism-based Methods	• High accuracy • Explainable • Clear failure mechanism • Low model calculation	• Requires a lot of repeated experiments • Requires expert knowledge guidance • Necessary simplification is required for complex systems • Online diagnosis is not possible
Data-driven Methods	• No need for extensive experiments • No need for domain knowledge • Fast system response • Strong generalization ability	• Requires a large amount of historical data for training • No physical interpretability • Prone to overfitting • High computational complexity
Hybrid Methods	• Not prone to overfitting • Failure mechanism is clear • Model is interpretable • Reduces reliance on labeled data	• Balance the integration between physical knowledge and data • No general combination structure • Complex models require huge computational effort • Domain knowledge is required to build physical mechanism models

Table 2 Comparison of different dimensions of existing methods

Aspect	Physical Mechanism Methods	Data-Driven Methods	Hybrid Methods
Computation Cost	Low to moderate; complex models can be computationally expensive.	High; requires significant computational resources for training and inference.	High; combines the computational demands of both methods.
Data Requirements	Low; Requires minimal data; mainly relies on physical mechanism models and expert knowledge.	High; Requires large volumes of high-quality data for model training and validation.	Moderate; needs both data and physical mechanism model knowledge, but can work with smaller data sets due to model guidance.
Usage Scenarios	Best for systems with well-understood physical mechanism and relatively simple failure modes.	Suitable for systems with complex, nonlinear behavior where explicit physical mechanism models are difficult to construct.	Ideal for systems that benefit from both high data availability and the need for a robust, explainable model.
Physical Interpretability	High; models are based on physical laws, offering clear insight into system behavior and failure mechanisms.	Low; models are often "black boxes," making it hard to interpret the underlying processes.	High; combines physical mechanisms with data, leading to better interpretability than pure data-driven models.
Generalization Ability	Low; relies on assumptions and simplifications, often struggles with system variations or unseen conditions.	High; can generalize well across different operational conditions with sufficient data.	High; benefits from data-driven adaptability and physical mechanism model constraints for more robust predictions.
Model Building Difficulty	High; requires detailed domain knowledge and extensive model tuning.	Moderate; building models may be easier, but requires large amounts of high-quality data and careful model tuning.	High; requires expertise in both physical mechanism and data-driven modeling, and integration of both approaches can be challenging.

1.3 The difference from other PHM reviews

In recent years, numerous scholars have summarized PHM technology. Table 3 summarizes some influential PHM technology reviews, which have different focuses.

Specifically, Guo [56] systematically reviewed various methods for equipment RUL prediction with engineering systems as the research object. Zhang [57] comprehensively expounded the fatigue life prediction method of mechanical parts based on data-driven model, and introduced the application of various ML methods in component life prediction in detail. Wilhelm [58] comprehensively expounds the related progress of fault diagnosis research using hybrid physical mechanism and data-driven methods, and points out the application of these methods in practice. Wang [59] compared different adaptive mathematical models in deep learning, mainly through lithium-ion battery RUL prediction. Wang [60] reviewed current methods related to hybrid physical mechanism and data-driven model, and divided the hybrid models into (1) physical information ML methods, (2) ML methods assisted simulation, and (3) explainable artificial intelligence. Ferreira and Gonçalves [61] described the relevant structures and components of ML, proposed an analytical framework for the ML-based RUL prediction process, and discussed the advantages and disadvantages of related ML methods. Peng [62] reviewed the development of digital twin technology and studied related technologies for building bearing digital twins. Yin [63] summarized and analyzed the failure modes of wind turbine bearings, and explained the fault diagnosis methods of wind turbine bearings from qualitative and quantitative aspects. Based on the definition of digital twins, Liu [64] analyzed the research methods of the relevant components of digital twins and explained the current research results. Li [45] reviewed the physics-informed data-driven RUL prediction, centering on the Physics-Informed Machine Learning (PIML) method. Wang [65] comprehensively outlined how to build data-driven models and discussed how hybrid models are combined. Finally, the application of the hybrid method is discussed based on the fatigue life of metallic materials.

Table 3 Summary of the review article on PHM

References	Publication year	Mainly discussed methods	Whether the object is bearing
Guo et al	2020	Data-driven, physical mechanism-based, hybrid method	Partly
Zhang et al.	2021	ML based methods	Partly
Wilhelm	2021	Data-driven, physical mechanism-based, hybrid method	Partly
Wang et al.	2021	DL based methods	No
Wang et al.	2022	Hybrid method	Partly
Ferreira	2022	ML based methods	Partly
Peng	2022	Digital Twin Technology	Yes
Yin	2023	Data-driven, physical mechanism-based	Yes
Liu	2023	Digital Twin Technology	Partly
Li	2024	Hybrid methods	Partly
Wang et al.	2024	Data-driven, physical mechanism-based, hybrid method	No

Although the papers in Table 3 do not include all the relevant reviews that have appeared in recent years, they also reflect the importance of current PHM research to a certain extent. However, it is worth noting that few review papers address PHM technology specifically for bearings, with existing reviews primarily focusing on various components and complex mechanical systems. Yin [63] takes bearings as the research object and reviews their PHM methods, but the review methods focus on data-driven methods without emphasizing the fusion of physical knowledge. Although Peng [62] also takes bearings as the research object and integrates physical mechanism models, the relevant technologies are mainly concentrated on the digital twin method. Since the purpose of this paper is mainly to highlight the virtual-real combination of physical mechanism models and data-driven models, the methods presented in this review tend to emphasize the simulation of physical mechanism models. Wang and Li [45, 60] systematically discussed the hybrid method, but these papers all focused on one method in PHM research, without providing a comprehensive analysis of all PHM techniques from a hybrid physical mechanism and data-driven perspective. Therefore, the above two reviews are not universal at the methodological level. Last but not least, the existing reviews clearly show that there is insufficient discussion on the fusion of physical knowledge. At the same time, the hybrid methods stated in the existing reviews are basically classified according to the research direction, which leads to illogicality at the methodological level.

1.4 Motivation of this review paper

So far, PHM research has achieved great breakthroughs, but there are still many opportunities and challenges. In particular, the development of the fourth industrial revolution and the gradual enhancement of sensor technology have provided relevant data sources and intelligent detection and diagnosis technologies for PHM research [66]. In this context, the need to ensure the safe and stable operation of complex mechanical systems has also greatly accelerated the process of developing PHM technology. However, in the context of the new industrial revolution, the integration between systems is increasing, and mechanical systems are becoming more and more complex. When a mechanical system fails normally, the collected detection data often exhibit significant mutual interference [66], low value density [68], incompleteness [69], and low data quality. This poses a huge challenge to the application of PHM technology. In fact, mechanical system failures are mainly caused by damage to components, especially key components such as bearings [70-72]. The degradation process of components such as bearings during operation follows physical mechanisms. Therefore, using the existing physical mechanisms combined with the collected system degradation data can provide new modeling ideas for PHM technology. For example: using physical mechanisms to extract and analyze features of massive sensor data, thereby obtaining high-quality and stable data. Using these data to conduct PHM research to improve model accuracy. In addition, data can assist physical prior knowledge in modeling, which makes PHM model more consistent with physical reality during the diagnosis and prediction. Moreover, it can reduce the difficulty of physical modeling and enhance the generalization ability and prediction accuracy of the model, as shown in Figure 4.

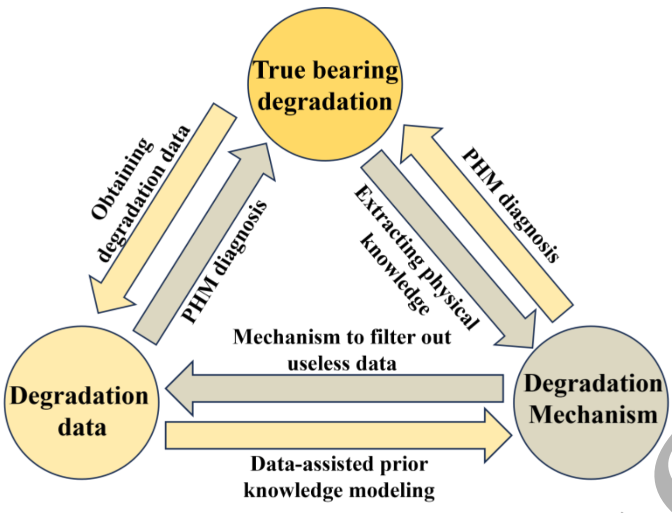

Figure 4 Relationship between physical mechanism, data-driven and bearing PHM

Through the above discussion and analysis, the hybrid physical mechanism and data-driven PHM technology will become a new hot spot in the future. At the same time, bearings are very important components of complex mechanical systems. Once damaged, they will directly affect the operation of the rotating system or even the entire mechanical equipment [73]. However, there is still no review of the research on PHM of bearings using fusion physical mechanisms and data-driven models, or at least no in-depth discussion. To fill this gap, this paper reviews the latest progress in hybrid physical mechanism and data-driven bearing PHM technology, aiming to inspire more scholars in this research area. At the same time, it promotes the development of fusion physical mechanism and data-driven PHM technology. Given the multitude of methods for classifying fusion physical mechanisms, each classification method has its own advantages and disadvantages. From the perspective of fusion strategy, this paper classifies the hybrid physical mechanism and data-driven bearing PHM methods according to the data level, network level, and model level. Under each level, the PHM technology related to bearings is reviewed. It aims to provide readers with a clear logical relationship so that relevant researchers can better understand the current research progress. Then, the advantages and disadvantages of each fusion strategy in practical applications are discussed, and the challenges and future development directions of current research are summarized.

To facilitate a clearer understanding of this paper's structure, it is summarized as follows. Chapter 2 outlines the bearing PHM methods under the three different fusion strategies discussed in this paper. Chapter 3 focuses on the research progress of bearing PHM at data level fusion. Chapter 4 focuses on the research progress of bearing PHM at the network level fusion. Chapter 5 focuses on the research progress of bearing PHM at the model level fusion. Chapter 6 discusses the advantages and disadvantages of the fusion model. Chapter 7 discusses the challenges and future directions of the current hybrid physical mechanism and data-driven model based on the existing progress, advantages and disadvantages. Section 8 is the conclusion of this paper.

2 Overview of bearing prognostics and health management based on fusion models

This section mainly provides an overview of the hybrid physical mechanism and data-driven bearing PHM fusion strategy discussed in this paper at the methodological level. In general, the PHM methods of fusion models can be broadly categorized into three types based on the distinct fusion strategies of physical prior knowledge and data during fusion. Firstly, from the perspective of the data level fusion, physical mechanisms are employed to process the collected data or expand the data set. Secondly, from the perspective of the network level fusion, physical knowledge can be directly integrated into the network structure or used as a constraint to guide model learning. Third, from the perspective of the model level fusion, a physical mechanism model is established based on the physical mechanism, and the parameters of the physical mechanism model are calibrated through monitoring data. Therefore, this paper divides the existing hybrid physical mechanism and data-driven bearing PHM methods into three categories, namely data level fusion, network level fusion, and model level fusion, and subdivides them under each fusion strategy. The detailed classification is shown in Figure 5. To enable readers to better understand each fusion strategy, in the following subsections, this paper mainly explains the fields that each fusion strategy is suitable for solving, and briefly discusses the overall idea of each method.

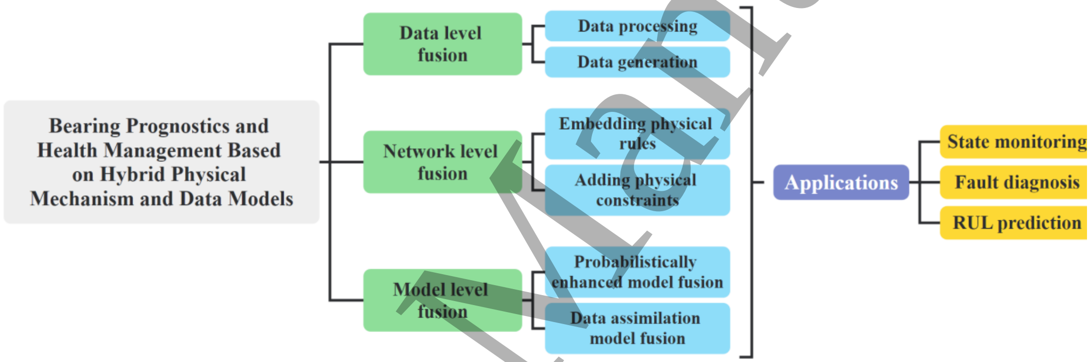

Figure 5 Classification of bearing PHM based on fusion models

2.1 Data level fusion

In actual industrial applications, bearings frequently operate under dynamic and complex operating conditions that change over time. During the actual data collection process, the coexistence of multiple devices may introduce additional variables. Such environments often result in significant variability in vibration signals, which challenges the reliability and diagnostic accuracy of pre-trained data-driven models. In addition, background noise and interference are inevitable due to the operation of other mechanical equipment. These factors may interfere with the data collection process, especially under test conditions with high noise levels, which may further reduce classification performance. At the same time, most rotating machinery operates under normal conditions, and it is difficult to collect fault data. In this case, there is not enough data to train and optimize the model [74]. These practical engineering challenges limit the ability to rely purely on data-driven approaches for effective diagnosis and prediction of bearings. In this case, using physical mechanisms to guide the collected data, that is, the fusion of physical mechanisms at the data level, can effectively deal with the problem of poor data quality in complex environments.

According to the different reasons for poor quality of collected data, data-level fusion models can be divided into two categories: data processing and data generation. At data level, common fusion methods include Generative Adversarial Network(GAN) [75], digital twin technology [76], wavelet transform [77], Domain Adaptation [78], spectrum negentropy [79], and fast Fourier Transform (FFT) [80]. GAN can generate high-quality synthetic data through the confrontation process between the generator and the discriminator. At data level fusion, it can be used to generate synthetic data of bearings under different working conditions. Digital twins create virtual copies of bearings, allowing different operating conditions and failure modes to be simulated without risk. At data level fusion, data generated by the digital twin can be combined with actual sensor data to provide a more comprehensive view of bearing performance. These two methods are mainly used to generate data and enhance the data set by obtaining more labeled data. Ultimately, the purpose of enhancing the generalization and accuracy of the model is achieved. The remaining methods primarily focus on data processing techniques designed to address noise interference in the collected data. Wavelet transform can decompose the signal into components of different frequencies, thereby extracting the time-frequency characteristics of the bearing vibration signal. The Domain Adaptation method is used to migrate the model from one data distribution to another at data level fusion. Spectral negentropy is a tool to measure the randomness of the signal. At data level fusion, it can be used to evaluate the information quality of different data sources. FFT is a Fourier transform algorithm with high computational efficiency, which can extract the frequency domain characteristics of the bearing vibration signal. The above methods can be used to combine the physical mechanism of the bearing during operation to improve the data quality.

From the general discussion of the above data level fusion methods, it can be seen that the common point of these methods is to extract physical knowledge from the actual situation, which is used to process and generate data. Therefore, when selecting methods at data level fusion, it is necessary to consider the actual degradation process and data type of the research object to find a suitable physical mechanism model and data processing method.

2.2 Network level fusion

In bearing PHM, Neural Networks are favored for their excellent data processing capabilities and fast response characteristics. These methods monitor and predict the health status of bearings by constructing a mapping from monitoring data to health status information. However, these data-driven models are often viewed as “black boxes” which are clearly inadequate in explaining bearing degradation mechanisms. Moreover, the internal mechanisms of black box models are extremely complex, making debugging and diagnosing errors quite challenging. If the output of the model is incorrect (for example, the prediction is inaccurate), it is difficult to trace the error back to the specific cause or source. To address this limitation, researchers have begun to integrate physical and domain knowledge into Neural Networks at the network level. This method can alleviate training bias and overfitting by introducing rule restrictions in the loss function of the Neural Network or by adding weight modules to guide the optimization direction of the model. In addition, interpretable feature learning can be achieved by introducing physical formulas to replace traditional neuron operators. By restricting the optimization direction of the loss function at the network level or adding interpretable neuron operators, the physical interpretability and transparency of the data-driven model can be effectively improved. This not only helps to alleviating the black box problem, but also reduces the occurrence of overfitting. The key to this approach is to combine the accuracy of the physical mechanism model with the generalization ability of the data-driven model to achieve more accurate bearing fault diagnosis and prediction. The network-level fusion method, with its advantages, has been widely used in many scientific fields, including fluid mechanics [81], biomedicine [82], fracture mechanics [83], power systems [84], geophysics [85], and other scientific fields. With the continuous advancement of data-driven technology and physical mechanism models, the trend of incorporating physical knowledge into Neural Networks has become increasingly obvious in PHM. At present, the network-level fusion method has become an increasingly popular paradigm Neural Networks due to its potential advantages, and the research trend of incorporating physical knowledge into Neural Networks to develop PHM based on hybrid physical mechanism and data-driven is unstoppable.

2.3 Model level fusion

Traditional physical mechanism models have significant advantages in characterizing bearing degradation processes and damage states. However, when faced with the modeling of complex systems, they often encounter the challenges of difficult parameter updates and high modeling complexity. To overcome these limitations, many scholars are committed to model-level research on fusing physical and domain knowledge into ML models. The data-driven model serves as the observation equation to construct the mapping relationship between the measurement signal and the state parameters, realizing the inverse process of the equivalent physical equation. Furthermore, a data-driven update approach enables rapid parameter calibration. Integrating data-driven observation calibration module into physical mechanism models not only improves the accuracy and real-time performance of physical modeling, but also effectively reduces the uncertainty and random interference of complex systems. Common physical mechanism models include the Paris model [86] and Forman model [87] of fatigue crack growth, Archard wear model [88], Herz contact model [89], etc. Common parameter update methods include Bayesian method [90], Kalman filter(KF) [90], particle filter (PF)[92], etc. Because the physical mechanism model has strong interpretability, the parameters in the physical mechanism model can be estimated and calibrated in real time by using the measured data. Therefore, this means that model level fusion can take into account both physical interpretability and modeling difficulty. Based on the above advantages, fusion at the model level has also become a hot spot in current research.

3 Bearing PHM research method based on data level fusion

This section first introduces the importance of bearing PHM technology at data level fusion, and explains the general principles of bearing PHM research at data level fusion. Based on existing research, the data-level fusion methods used in bearing PHM are categorized into two main directions: data processing and data generation. In the data processing direction, the discussion focuses on how physical mechanisms can be utilized to optimize the data preprocessing process, such as removing redundant information and reducing noise interference, thereby improving data quality. In the data generation direction, the emphasis is placed on generating high-quality virtual data through physical mechanism models and simulation technology to expand the training dataset of the model. Subsequently, the research progress of these two directions is introduced in detail with the relevant literature, aiming at providing scholars with a new perspective for their research.

3.1 Basic principles of PHM based on data level fusion

In recent years, data-driven models have made significant progress in PHM. However, the performance of such models is highly dependent on the quality and quantity of data, and in actual industrial environments, they often face problems such as data scarcity and severe noise interference. This not only limits the training effect of the model, but also brings huge challenges to its promotion in practical applications. To solve these problems, researchers introduce physical mechanisms at the data level. According to different implementation methods, data fusion strategies can be divided into two categories: data processing and data generation.

The core of data processing principles lies in how to combine the accuracy of physical mechanism models with the generalization capabilities of data-driven methods. This process usually starts with cleaning and preprocessing the raw data to ensure data quality and consistency. Subsequently, through feature extraction and selection, the data is converted into a format that reflects the essence of the physical phenomenon. For example, when bearing sensor data is accompanied by a lot of noise, and the noise masks key fault signatures, physical information can be used to identify and remove extraneous noise based on the bearing failure mechanism while retaining key health status signatures. Secondly, when the bearing inner ring, outer ring or rolling element fails, a vibration signal with a specific characteristic frequency will be generated. These characteristic frequencies can be calculated from bearing design parameters and operating conditions. Using physical knowledge to extract these frequency components and remove noise features in irrelevant frequency bands can significantly improve the pertinence of data processing and thereby improve model accuracy. Finally, in the signal analysis, the key frequencies of physical mechanism (such as fault characteristic frequency) are assigned higher weights, thus increasing the attention to these fault characteristics and improving the robustness of the model.

The principle of data generation focuses on using physical mechanism models and data-driven methods to jointly generate new data samples to expand and enrich the data set. The physical mechanism model provides a prior knowledge of system behavior in this process and defines the boundaries and rules for data generation. Data-driven methods generate new data samples based on these rules. The introduction of physical mechanism models ensures that the generated data samples are physically feasible, while the data-driven approach ensures that these samples are statistically consistent with real data. This fusion method is particularly effective in dealing with complex systems, such as bearing PHM, because it can simulate failure modes that are difficult to observe in actual operation, thus providing a more comprehensive data processing strategy for model training. The generated simulation data enables the PHM model to be effectively trained even when experimental data is limited, and by increasing the diversity of the data set, the fault prediction accuracy and stability of the model under different working conditions are improved. This method enhances the generalization ability and robustness of the model by synthesizing data, effectively alleviating the limitations caused by data scarcity.

Table 4 systematically compares various data-level fusion methods applied to bearing PHM. GANs and digital twin technology exhibit high computational costs and large parameter scales primarily due to their inherently complex model architectures: GANs rely on iterative adversarial training to accurately capture subtle and rare bearing fault characteristics, demanding extensive computation and large Neural Network architectures. Similarly, digital twins require detailed physical modeling and simulation of bearing dynamics, necessitating high computational resources and numerous parameters to precisely represent real-world conditions. Domain adaptation methods also entail significant computational complexity and large parameter scales, attributed to extensive optimization and fine-tuning efforts required to transfer learned diagnostic models robustly across diverse operational scenarios or different bearing setups. In contrast, wavelet transform achieves moderate computational complexity, as it efficiently decomposes bearing vibration signals into multi-resolution frequency components without requiring excessively large parameter scales, balancing analytical precision and computational efficiency. FFT further reduces computational burden due to its efficient algorithmic structure, offering rapid extraction of bearing fault frequency features suitable for real-time monitoring. Spectrum negentropy has the lowest computational cost and parameter scale since it leverages straightforward entropy-based calculations to rapidly assess signal complexity, making it particularly suitable for real-time anomaly detection. These approaches effectively address critical PHM challenges such as limited fault data availability, noise interference, and real-time processing demands, providing better data input for constructing bearing PHM system.

Table 4 Comparisons of data-level fusion methods

Method	Time Complexity	Parameter Scale	Technological Advantages	Usage Scenarios
GAN	High	High	Generates realistic synthetic data, useful for data augmentation and rare event simulation.	Useful when limited fault data are available, or when one needs to augment training data diversity. Especially helpful in scenarios with small real datasets to improve model robustness.

Continued table 4 Comparisons of data-level fusion methods

Method	Time Complexity	Parameter Scale	Technological Advantages	Usage Scenarios
Digital Twin Technology	High	High	Enables virtual simulations of real-world systems, improving predictive maintenance and system optimization.	Ideal for complex equipment or systems requiring continuous monitoring and prediction, e.g., engines in aerospace or wind turbines, allowing rapid virtual experimentation to reduce real-world risks and costs.
Wavelet Transform	Moderate	Moderate	Efficient for multi-resolution analysis, retaining both time and frequency information, especially useful for non-stationary signals.	Well-suited for diagnosing and extracting features from non-stationary signals, such as mechanical vibration or seismic data, particularly in rotating machinery or structural health monitoring.
Domain Adaptation	High	High	Improves generalization across domains, ensuring high accuracy on significantly different datasets.	Useful for multi-condition or cross-domain data analysis, such as diagnosing similar equipment in different environments, or transferring lab data to real production environments.
Spectrum Negentropy	Low	Low	Simple and cost-effective method for measuring complexity or randomness, useful for signal quality analysis.	Ideal when rapid signal quality monitoring is needed, such as checking noise levels or distortion in vibration tests.
Fast Fourier Transform	Low	Moderate	Widely used to extract frequency domain features, offering low computational cost for vibration and acoustic signals.	Commonly used in preliminary spectral analysis for rotating machinery fault diagnosis or acoustic signal processing. Also suitable when real-time frequency domain transformations are needed for large-scale data.

3.2 Data processing

This section first introduces some common data processing methods in bearing PHM. Broadly speaking, the data processing methods used for PHM diagnosis mainly include wavelet transformation, FFT, etc. The performance of several commonly used data processing methods on bearing fault vibration signals, along with comparisons to healthy bearing vibration signals, are presented in Figure 6. The following introduces these processing methods and gives the results achieved by current scholars at data level fusion.

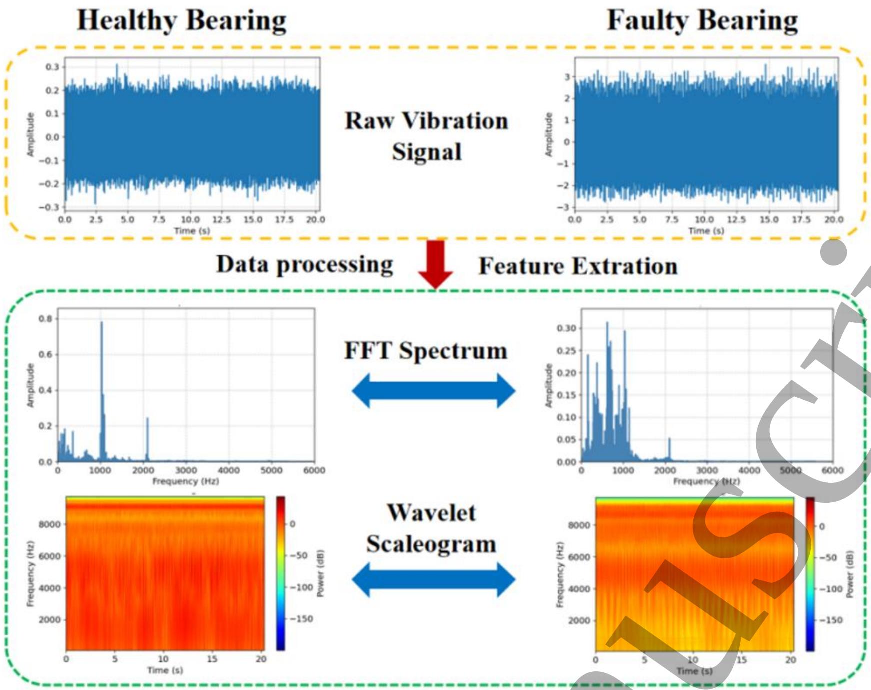

Figure 6 Performance of main methods for data processing

FFT is a characteristic frequency extraction method that analyzes and processes the frequency components of a signal by converting the signal from the time domain to the frequency domain. In the denoising process, a filter is used to remove or attenuate the frequency range where the noise is located, thereby retaining the main frequency components of the signal to achieve the purpose of denoising. Mathematically speaking, FFT can be implemented as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn} (k = 0,1,\ldots,N-1) \quad (1)$$

$$Y(k) = H(k) \cdot X(k) \quad (2)$$

$$y(n) = \frac{1}{N}\sum_{k=0}^{N-1} Y(k) \cdot e^{j\frac{2\pi}{N}kn} (n = 0,1,\ldots,N-1) \quad (3)$$

where $x(n)$ is the n-th sample of the time domain signal, $X(k)$ is the k-th sample of the frequency domain signal, j is an imaginary unit, $H(k)$ is the filter function, and $y(n)$ is the denoised time domain signal.

The wavelet transform is a powerful tool for analyzing signals in both time and frequency domains. Unlike the FFT, which provides a global frequency representation of the signal, the wavelet transform captures local features in both time and frequency, making it ideal for non-stationary signals. The wavelet transform is often used to analyze transient and non-stationary behavior of bearing vibrations. Mathematically, the discrete wavelet transform (DWT) is expressed as:

$$DWT(j,k) = \sum_{n=-\infty}^{\infty} x[n] \cdot \psi^*\left(\frac{n-k\cdot 2^j}{2^j}\right) \quad (4)$$

where $x[n]$ is the discrete-time signal, $\psi(t)$ is the wavelet function, j is the scale index controlling the wavelet's width, k is the translation index shifting the wavelet in time, and ψ^* is the complex conjugate of the wavelet function.

In the PHM field, many studies make full use of physical mechanisms to optimize the data processing process. Cheng [93] proposed an Adaptive Fault Attention Residual Network, which integrates physical mechanism-driven domain adaptation techniques for intelligent cross-domain bearing fault diagnosis of nuclear cooling pumps. By introducing physical mechanisms, the model can more effectively eliminate irrelevant noise, improving diagnostic feature alignment and model physical interpretability. Lu [94] proposed a physical mechanism-driven feature weighting method, which assigns higher weights to fault characteristic frequencies in the physical mechanism, significantly enhancing the accuracy and robustness of rolling bearing fault diagnosis under different operating conditions. Song [95] developed a hybrid kernel Support Vector Machine that combined a physical mechanism-driven Bayesian algorithm. By optimizing the kernel parameters of the vibration signal, the method reduced the impact of noise on fault diagnosis accuracy, further improving the model's performance. Sheng [96] combined data-driven models with degradation physical mechanism model, using physical information to track the bearing's degradation process, successfully improving the accuracy and reliability of RUL prediction. Zhu [97] proposed a method based on Multi-Scale Convolutional Neural Network(MSCNN) which combined the time-frequency representation of vibration signals with physical mechanisms to capture both global and local degradation features, significantly improving RUL prediction performance. Lu [98] used volatility and wavelet Neural Networks, combined with physical mechanisms, to optimize autoregressive filters for extracting bearing fault features. This approach showed superior performance in early-stage detection of rolling bearing degradation and failure cycle prediction. Pan [99] proposed a cloud-edge collaborative method for bearing RUL prediction. It denoises degradation data, extracts temporal and frequency features, and builds a feature space graph using the Pearson correlation coefficient for accurate and efficient predictions. Liu [100] proposed a cloud-edge collaborative method for bearing RUL prediction. The method processes vibration signals from both horizontal and vertical data sources through denoising and feature extraction, utilizing the parallel computing capabilities of the Transformer model for dual data source fusion prediction. Zhang [101] proposed an explicit dynamics-driven method for bearing fault diagnosis under variable conditions. Finite element method (FEM) simulations were used to generate samples across multiple conditions and fault types, and the Nuisance Attribute Projection method was employed to construct a projection matrix to remove redundant condition information from measured signals while preserving fault features. Walther [102] proposed a hybrid ML methods that combines FFT and frequency selection with deep learning models for rolling bearing fault classification. The method incorporates a physical feature extraction step during signal preprocessing to extract characteristic frequencies related to common bearing damages, enhancing the data features and enabling the model to perform well even with limited data. Cao [103] proposed a hybrid data- and model-driven framework for bearing RUL prediction. The method leverages an extended KF to estimate the parameters of a degradation physical mechanism model and extract degradation trend information. It integrates these estimates with sensor measurement data using a multi-head attention Transformer model. By fusing frequency-domain features and time-series data, the framework effectively combines physical mechanism model estimations and sensor features for accurate degradation modeling and prediction. Su [104] proposed a framework for bearing RUL prediction based on a feature pre-extraction mechanism and an adaptive Transformer model. The method designs a feature pre-extraction mechanism that utilizes Short-Time Fourier Transform to extract low-level features from vibration signals, retaining rich degradation information for deep learning models. Liu [105] proposed a bearing fault diagnosis method that uses fast Independent Component Analysis (ICA) and information fusion. Multi-sensor signal denoising and fusion, along with spectral kurtosis, are used to design adaptive filters that extract fault features even under low Signal-to-Noise Ratio (SNR) conditions. Deng [106] proposed the double-layer attention based adversarial network (DA-GAN) model with a double-layer attention mechanism to enhance partial transfer learning in fault diagnosis. Jia [107] proposed the Joint Distribution Adaptation with Diverse Feature Aggregation framework, using Diverse Feature Aggregation and Joint Maximum Mean Discrepancy to improve data alignment for bearing fault diagnosis across machines.

3.3 Data generation

In bearing PHM with the fusion of physical mechanism and data at the data level, commonly used data generation methods rely on physical simulations to generate large amounts of virtual data. These methods compensate for data scarcity and expand the model's training set. They share the common goal of generating high-quality virtual data to enrich model inputs and improve generalization ability, but differ in their implementation and focus. Common data generation methods include bearing dynamic model, GAN and so on. The details of the data generation method will be introduced one by one below.

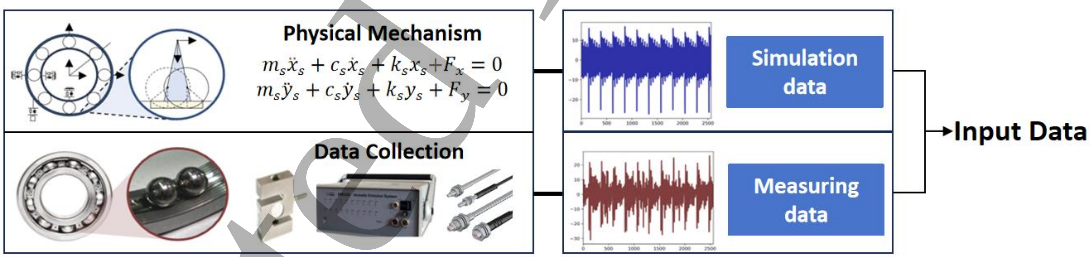

Figure 7 The principle of data generation

Bearing dynamics models use physical modeling methods based on key parameters such as the forces, temperature, and wear of the bearing to simulate its operation and generate highly realistic simulated data, as shown in Figure 7. For example, Fukata [108] first proposed a two-degree-of-freedom model to theoretically describe the vibration response of rolling bearings, with nonlinear behavior modeled through a mass-spring-damping system. The specific calculation method is as follows:

$$m\frac{d^2}{dt^2}\begin{bmatrix}x\\y\end{bmatrix} + c\frac{d}{dt}\begin{bmatrix}x\\y\end{bmatrix} + \begin{bmatrix}F_x\\F_y\end{bmatrix} = \begin{bmatrix}W_x\\W_y\end{bmatrix} \tag{5}$$

where m is the mass of the inner race and rotor, c is the equivalent viscous damping, W_x and W_y are the radial force components applied to the rotor shaft, and F_x and F_y are the nonlinear contact force components, which can be calculated based on Hertz contact deformation theory.

GANs are a deep learning technique that uses adversarial training between a generator and a discriminator to generate virtual data that closely resembles the distribution of real data. The generator is responsible for creating data from noise, while the discriminator tries to distinguish between generated data and real data. The objective function can be expressed as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{\text{data}}(x)}[log\ D(x)] + \mathbb{E}_{z\sim P_z(z)}[log\ (1-D(G(z)))] \tag{6}$$

where $D(x)$ represents the discriminator's output probability that the input x is real data, $G(z)$is the generator's output from noise z, and $D(G(z))$ is the discriminator's output probability that the generated data $G(z)$ is real. $x \sim P_{\text{data}}(x)$ represents the distribution of real data, and $z \sim P_z(z)$ represents the distribution of noise data.

When physical information is integrated with GANs, the generated data not only closely matches the real data in distribution but also aligns with actual fault patterns. This fusion significantly enhances the model's generalization ability, enabling it to handle various fault types and complex industrial scenarios.

In the bearing PHM field, researchers have effectively addressed data scarcity issues through physical mechanism-assisted data generation methods. By enhancing the diversity and realism of the data, these methods have significantly improved model performance. Multiple studies have leveraged physical mechanisms to assist in data generation for bearing fault diagnosis, enhancing accuracy and generalization.

(1) digital twin technology

The most common method is generating simulated data through digital twin technology. Gong [109] developed a Simulink-based digital twin to generate simulated rolling bearing data, thereby enriching training datasets. Yang [110] employed digital twin technology to create a virtual-real interaction model, integrating high-fidelity simulations with physical test data to improve diagnostic reliability. Zhang [111] used digital twin technology to create a virtual bearing model, generating simulation data for accurate diagnosis under low SNR conditions. Huang [112] fused low-fidelity FEM simulation data with high-fidelity experimental data using digital twin and multi-fidelity information fusion, enabling dynamic updates of diagnostic models. Li [113] introduced the Digital Twin-Assisted Dual Transfer method, merging simulation data with real data via Digital Twin-Driven Information Transfer and Digital-Analog-Driven Model Transfer, improving fault diagnosis accuracy and generalization in small samples. Zhang [114] used digital twin technology to generate simulation data, applied an improved Bidirectional Long Short-Term Memory (Bi-LSTM) with macro-micro attention mechanisms and CatBoost for RUL prediction, improving accuracy in small samples. Desai [115] developed a tribology-aware digital twin using four-ball test simulations and Convolutional Neural Network (CNN) to predict bearing RUL from sensor data, enhancing maintenance accuracy.

(2) GAN

Several studies utilized GANs and their variants. Xu [115] generated multi-modal fault data with a dynamic model and Multi-Agent Diverse GAN, aligning it with real data to enhance diagnosis in small samples. Cui [117] generated full-lifecycle degradation data with a dynamic model and built a Performance Degradation Dictionary by matching simulation and real data, enhancing RUL prediction. Peng [118] utilized Conditional Deep Convolutional GANs to generate multi-category simulation data with bearing fault labels and combined it with a one-dimensional convolutional Neural Network (1-D CNN) for accurate fault classification in small-sample condition.

(3) Finite Element Models

FEM also played a significant role in data generation. Dai [119] proposed the Sequential Hybrid of FEM and Neural Network method, generating high-quality simulation data with FEM and training Neural Networks for accurate wear prediction in small samples. Haj Mohamad [120] combined defect-free simulation data from a physical mechanism model with experimental fault data to extract physics-informed features, creating a hybrid feature set.

(4) Mixed methods

Some scholars have combined several of the aforementioned methods for data generation. Cheng [121] utilized dynamic modeling combined with GAN-based augmentation to produce high-quality simulation data, improving fault diagnosis in small-sample scenarios. Han [122] integrated a dynamic bearing model with Generative Adversarial Network based on Deep Autoencoder structure (GAN-DAE) to generate full-lifecycle degradation data and used Multiple source-and-target Domain Joint Adaption Network (MDJAN) to align simulation with real data, enhancing RUL prediction and cross-domain generalization. Zhao [123] combined digital twin simulation data with an improved CycleGAN for domain mapping, enhancing RUL prediction accuracy in small samples. Jantunen [124] combined physical mechanism-based wear models with data-driven approaches to predict rolling bearing RUL using a hybrid method. Yan [125] combined Support Vector Machine with a hybrid degradation model, using simulation data to extract dimensionless features for accurate RUL prediction across conditions. Song [126] integrated FEM, Wavelet Packet Transform, and Support Vector Machine to generate high-fidelity vibration signals, improving fault type identification. Gao [127] combined FEM simulations with GANs to generate synthetic samples for rotor-bearing fault classification, effectively addressing the small-sample issue. Lou [128] generated FEM-based simulation data and used GAN for domain adaptation to align with real signals, improving fault diagnosis accuracy in small samples. Piltan [129] combined digital twin Acoustic Emission signal simulations with SVM to classify bearing cracks, achieving high-accuracy diagnosis in small samples.

(5) Other methods

Other unique methods in which physical mechanisms assist data generation are as follows. Hou [130] combined fault simulation with normal data using transfer learning, enabling fault modeling without target fault data. Xiao [131] introduced the Joint Transfer Network, aligning simulated and experimental data with Joint Maximum Mean Discrepancy to enhance model adaptability. Dong [132] integrated a physical mechanism model, transfer learning and Joint Maximum Mean Discrepancy to align simulated and real fault data, enhancing diagnostic accuracy. Deng [133] used a five-degree-of-freedom dynamic model and PF calibration to predict bearing RUL across different machines, bridging simulation and real data. Dong [134] created a fault diagnosis framework by pre-training a CNN with large-scale simulation data and transferring knowledge to real scenarios, enhancing generalization in small samples. Liu [135] employed a knowledge-based Extreme Learning Machine to select samples from simulation data, augmenting small training sets and improving diagnosis accuracy. Tai [136] combined simulation data from physical modeling with experimental vibration data to train a Gate Recurrent Unit (GRU) network, achieving accurate fault type and location diagnosis in small samples. Wang [137] used a four-degree-of-freedom model and Coupled Hybrid Stochastic Resonance to generate and enhance weak signal features from simulation data, validating effectiveness under strong noise.

4 Bearing PHM research method based on network level fusion

In this section, Neural Networks' shortcomings are first introduced, explaining why it is necessary to integrate physical mechanisms with Neural Networks in PHM, and the feasibility of such integration. Then, based on existing research, network level fusion methods used in bearing PHM are divided into two main directions: embedding physical rules and adding physical constraints. Each direction is discussed in detail, highlighting the current mainstream methods and research progress, with references to relevant literature.

4.1 Overview of network level fusion for PHM

Currently, purely data-driven models are widely used in the PHM field. Among them, CNNs [138, 139] have been extensively utilized in bearing research. CNNs extract the time-frequency characteristics of signals by transforming one-dimensional signals into two-dimensional spectrograms, enabling fault classification and RUL prediction. The advantages of CNNs include the convolutional layers' ability to detect local features, the pooling layers' effectiveness in reducing dimensionality and enhancing robustness, and the fully connected layers' functionality for final classification or regression tasks. However, the black-box nature of CNN is a significant disadvantage. To overcome these challenges, researchers are exploring the fusion of physical knowledge and data-driven models and have made certain developments [140, 141]. Karpatne [142] conceptualized the Theory-guided data science (TGDS) paradigm for the first time and introduced a method for integrating knowledge from different fields into a data-driven model. After that, Karniadakis [143] successively proposed the physical information Neural Network method. By integrating the constraints of physical knowledge into the loss function of the Neural Network, the fusion at network level is achieved. This fusion method, which incorporates physical rules into the Neural Network's structure or loss function, allows the prediction results to capture the characteristics of the data while adhering to physical laws. Compared with traditional pure data-driven models, network level fusion greatly enhances the model's generalization ability, improves model physical interpretability and transparency, and improves data utilization efficiency. Due to the above-mentioned key advantages, fusing physical mechanisms and data at the network level has become a popular direction in bearing PHM.

In the following subsections, we divide network level fusion into two categories based on existing research methods: embedding physical rules and adding physical constraints. Among them, embedding physical rules is to directly integrate physical processes into the data-driven model structure, and enhance the physical interpretability of the model by taking physical knowledge as part of the model. Adding physical constraints is to introduce physical constraints to guide model learning, so that the output of the model can not only closely align with real-world operating conditions but also adhere to physical constraints.

Table 5 offers a comprehensive comparison of network-level fusion methods that incorporate physical knowledge into Neural Network architectures. The table categorizes key approaches of network-level fusions and evaluates them based on computational cost, time complexity, parameter scale, and other implementation factors.

Table 5 Comparisons of network-level fusion methods

Method	Time Complexity	Parameter Scale	Technological Advantages	Usage Scenario
Physical mechanism-guided Parameter Initialization	Low	Low to Medium	Accelerates convergence, reduces overfitting, and enhances physical consistency.	Ideal for early-stage training in PHM systems such as bearing fault diagnosis where physical properties (e.g., vibration characteristics) are well understood.
Physical mechanism-guided Hyperparameter Design	Medium	Medium	Improves hyperparameter selection accuracy, reduces tuning time, and enhances model stability and prediction performance.	Suitable for complex dynamic systems (e.g., aerospace or automotive applications) where model performance must align with known physical behaviors.
Physical mechanism-guided Feature Processing	Medium to High	Medium	Enhances the physical interpretability of features, improves data representativeness, and increases model accuracy.	Best applied in scenarios with noisy sensor data, such as fault detection in industrial equipment or structural health monitoring where key physical features are critical.
Adding Physical Constraints	Medium	Low to Medium	Strengthens the model's physical consistency, prevents overfitting, and improves reliability.	Essential in safety-critical applications like material fatigue or wear prediction, where model outputs must strictly follow physical laws to ensure reliability.

Physical mechanism-guided parameter initialization requires only a one-time computation of initial weights using predefined physical laws, which avoids iterative, computationally intensive processes during training and keeps the parameter scale low to medium. In contrast, physical mechanism-guided hyperparameter design involves an additional selection process driven by physical insights. Although this selection may require some iterative evaluations, it is typically performed offline, resulting in medium time complexity and a moderate parameter scale. Physical mechanism-guided feature processing integrates physical principles into the feature extraction stage, which increases the computational demand and adds extra parameters, yet it significantly enhances the interpretability and representativeness of the extracted data. Lastly, adding physical constraints augments the loss function with terms derived from physical laws, slightly increasing per-iteration computation but maintaining a low to medium parameter scale because it does not introduce new layers.

Collectively, these methods streamline the training process, reduce the risk of overfitting, and ensure that model outputs remain consistent with known physical laws. They are applicable across a range of usage scenarios—from early-stage fault diagnosis in systems with well-understood physical properties to safety-critical applications such as material fatigue and wear prediction.

4.2 Embedding physical rules

The purpose of embedding physical knowledge into data-driven models is to directly combine known physical mechanisms with Neural Network models so that the model can learn patterns from the data while also adhering to physical laws during model training. The principle is to integrate a computational module of physical mechanisms into the structure of Neural Network models, guiding the model's predictions with physical knowledge, as shown in Figure 8.

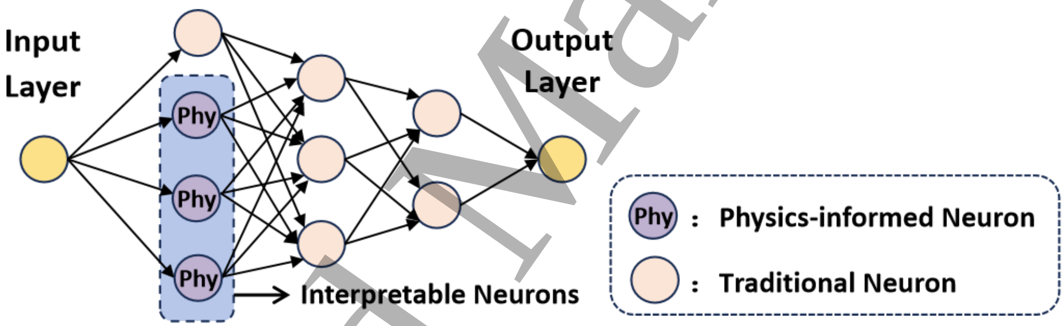

Figure 8 The principle of embedding physical rules in Neural Networks

Physical knowledge can be used to set the initial weights of Neural Networks, ensuring that the model has physical significance during the early stages of training. For example, by utilizing the fault characteristic frequencies of bearings, the convolutional weights in a CNN can be initialized to make the model more sensitive to vibrations at specific frequencies, accelerating convergence and improving training efficiency. He [144] designed an enhanced wavelet basis function, defined a scale smoothing factor to obtain more reasonable wavelet scales, and proposed a plug-and-play wavelet weight initialization method for deep Neural Networks based on physical wavelet priors. Also, a normalized activation map is designed to reveal the effectiveness of Z-scores from a visual perspective rather than experimental results.

Using the characteristics of physical signals, specific convolution kernel sizes, shapes, and numbers can be designed to better capture physical features. For example, based on the frequency distribution and time-frequency characteristics of bearing fault vibration signals, convolution kernels of corresponding scales can be designed to extract key fault features, improving the model's diagnostic capability and generalization performance. Ruan [145] developed a physical mechanism-driven CNN that utilizes the physical properties of bearing fault acceleration signals to guide the design of CNN parameters. This method dynamically determines input length and convolution kernel size based on fault cycles and signal decay, improving bearing fault diagnosis accuracy and reducing uncertainty. Sadoughi and Hu [80, 146] developed a physical mechanism-based CNN by embedding fault characteristic frequencies and rotational speeds into the convolutional layers for rolling bearing fault diagnosis. This approach improved fault detection accuracy and reduced the risk of model overfitting. Kim [147] incorporated physical knowledge into the design of model parameters to effectively enhance noise robustness. The use of attention mechanisms further improved the model's accuracy under noisy conditions.

By analyzing physical mechanisms, we can process features from physical formulas, variables simulated by physical models, or features extracted using signal processing methods (e.g., spectrum analysis, wavelet transform, short-time Fourier transform).Compared with feature extraction at the data level, feature processing is part of the network itself and requires embedding physical rules into the network for feature integration and analysis. Its processing depends on the network structure. In contrast, feature extraction only requires obtaining features at the data level, with steps independent of the network. The results of feature extraction can then be input into different networks for further processing. For example, in bearing PHM, in addition to extracting statistical features from vibration signals at the data level, one can directly embed physical rules, such as the relationship between rotational speed and fault frequencies, into the network level. By integrating these physical rules within the network architecture, Neural Network models can not only extract and analyze features tied to underlying physical mechanisms but also leverage fundamental physical principles when processing sensor data. Wen [148] proposed a hybrid model, GRU-AE-Wiener, which integrates the Wiener process model into a GRU and Auto-Encoder structure, assisted by GANs for synthetic data generation. This method incorporates physical constraints into the model architecture, using the Wiener process to describe non-monotonic degradation trends and leveraging GANs to enhance data quality, thereby improving prediction accuracy and robustness. He [149] integrated vibration analysis techniques such as time-synchronous resampling and discrete Fourier transform into a hybrid deep learning framework to improve the accuracy of fault diagnosis for rolling bearings based on vibration signals. Sun [150] used attention mechanisms to embed rotational machine fault characteristics from multimodal information, constructing a cross-modal correlation fusion module to adaptively build the correlations among multimodal information, highlighting the unique features of single-modal information and shared representations of multimodal data. Chen [151] used enhanced gradient-weighted class activation mapping to visualize regions of interest in convolutional layers. By combining this with prior knowledge of bearing fault mechanisms, the study improved the understanding of learned features and model decisions. Yang [152] proposed a bearing RUL prediction method based on physical information, called the Multi-State Time-Frequency Network. This method constructs a dynamic adaptive inverse discrete Fourier transform frequency domain module that maps known domain spectral information to unknown spectral intervals to extract periodic features from the sequences. They also developed a residual self-attention multi-state gating control unit, which integrates periodic and trend features to achieve compatibility between the inverse discrete Fourier transform (IDFT) physical mechanism model and the data-driven model for final RUL prediction. Deng [153] proposed the Theory-guided Progressive Transfer Learning Network (TPTLN) model, using a "distract-attract" strategy to address open-set diagnosis transfer in rotating machinery, improving fault detection and class alignment.

4.3 Adding physical constraints

Incorporating physical constraints into Neural Networks is a pivotal strategy to ensure their outputs adhere to the underlying physical laws of a system. Instead of modifying the network's structure, this method focuses on the training process, particularly the loss function. Typically, during training, the network updates its parameters to minimize discrepancies between predicted and observed data. By enriching the loss function with additional terms derived from physical equations, empirical formulas, or domain expertise, researchers can guide the network to satisfy certain physical requirements effectively.

A common technique is to create a composite loss function consisting of two primary components: a conventional data-driven term and a physical residual term. The physical residual term arises from known relationships, such as differential equations, conservation laws, or specific failure patterns in mechanical systems. In the bearing field, known fault frequencies or degradation models can be mathematically formulated and included in the loss function as penalties when the network's predictions deviate from these established benchmarks. By balancing the contributions of these two components, often through weighting factors, researchers ensure the network minimizes both its error with respect to training data and its deviation from key physical principles. The entire process is shown in Figure 9.

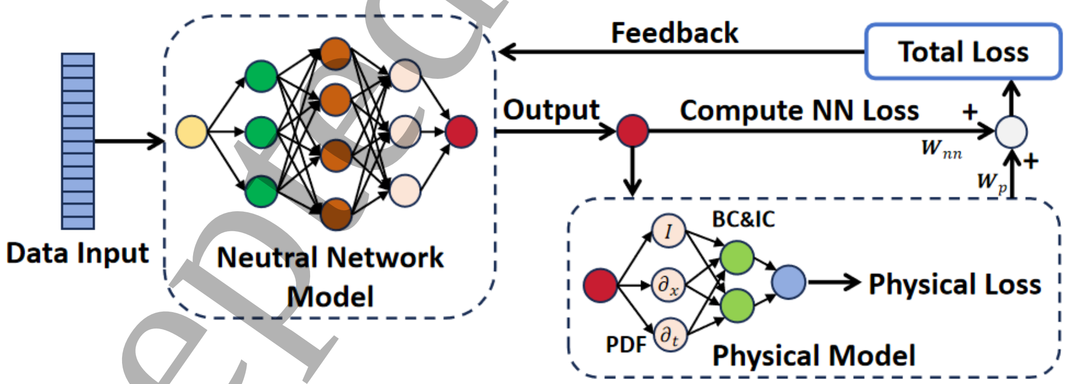

Figure 9 The principle of adding physical constraints in Neural Networks

By integrating physical mechanism-based penalty or regularization terms, researchers can retain existing Neural Network architectures and utilize standard backpropagation and gradient-based optimization techniques. These additional constraints merely adjust the scalar objective function that the training process seeks to minimize. This loss-based method enables the network to learn complex data patterns while consistently respecting physical insights specific to bearing systems. This approach has proven especially fruitful in bearing PHM. Yucesan [154] proposed a hybrid physics-informed Neural Network that combines a physical mechanism-based bearing fatigue model with a data-driven grease degradation model, embedding physical constraints into a recurrent Neural Network. The model incorporates a regularization term to ensure predictions adhere to the physical laws of grease degradation and uses a two-step probabilistic approach to quantify uncertainties in grease quality distribution. Shen [155] developed a physics-informed deep learning approach that combines a simple threshold model with a CNN for bearing fault detection. By designing input features based on fault characteristic frequencies and incorporating a customized loss function, this approach embeds physics constraints into model training, improving both classification consistency with physical laws and overall accuracy. Kim [156] developed a physics-informed Neural Network method that incorporates low-fidelity physical information, such as monotonicity and the sign of curvature, as constraints in the optimization process to reduce uncertainty and prevent unrealistic predictions. By incorporating physical constraints into the loss function, the authors quantified uncertainty and made more reliable predictions for cooling fan bearing degradation and crack growth without using run-to-failure data. Yang [157] proposed a feature-based transfer Neural Network model that integrates multi-layer domain adaptation and pseudo-label learning into a CNN to reduce the distribution discrepancy between laboratory and real-world bearing data. Physical constraints are incorporated as regularization terms during training to ensure consistency in cross-domain feature distributions and improve diagnostic accuracy. Chen [158] proposed a physics-informed deep Neural Network for predicting the RUL of bearings by embedding monotonic degradation characteristics into a recurrent Neural Network. The method uses temperature signals as degradation indicators, incorporates a positive increment recurrence relationship to ensure consistency with the degradation process, and optimizes model parameters through a physics-informed loss function to align predictions with the physical laws of bearing degradation. Lu [159] proposed a physical mechanism-guided Long Short-Term Memory(LSTM) network that incorporates the monotonic variation of feature indicators during bearing degradation as prior physical knowledge into the loss function, ensuring the network output aligns with physical laws. The method dynamically weights feature trend variations to optimize training, improving the accuracy and physical consistency of bearing RUL predictions.

5 Bearing PHM research method based on model level fusion

In hybrid physical mechanism and data-driven approaches, model level fusion is also one of the core methods in bearing PHM research. Fusion at the model level aims to solve problems such as limitations in data acquisition, lack of model generalization capabilities, and high requirements for real-time monitoring. It describes the physical behavior of the system by building a physical mechanism model based on physical laws and extracts potential patterns by building a data-driven model. Then, the physical mechanism model and data-driven model are fused together by using fusion strategies such as parameter calibration and state estimation. The accuracy and reliability of its prediction are ensured by strict calibration and verification procedures, and its principle can be shown in Figure 10. The fusion method based on the model level can integrate the interpretability of the physical mechanism model and the generalization ability of the data-driven model to achieve accurate prediction of bearing health status.

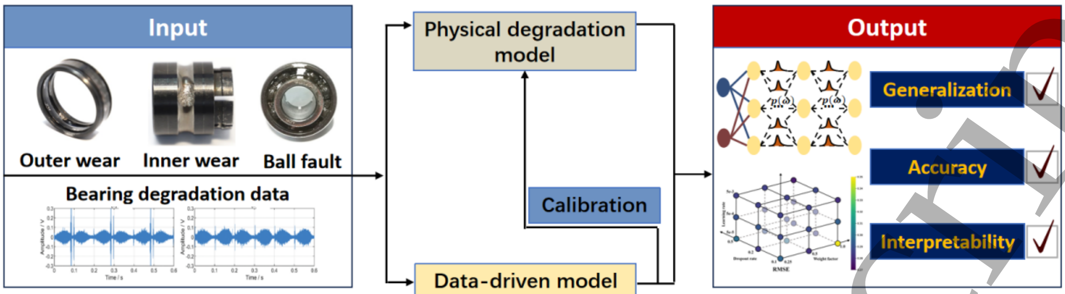

Figure 10 The principle of model level fusion

This section first introduces the general bearing PHM model at model level fusion. Then, based on existing research, the model-level fusion methods used in bearing PHM are divided into two main directions: probability-enhanced model fusion and data assimilation model fusion. In the direction of probability-enhanced model fusion, the core of the research is how to enhance the prediction ability of traditional physical mechanism models by introducing probabilistic models. This involves modeling the uncertainty of physical processes with probabilistic models and how to achieve real-time updating and optimization of model parameters through a probabilistic framework. Data assimilation model fusion focuses on combining observed data with physical mechanism models to improve the accuracy of state estimation. This section not only deeply analyzes the current methods and progress of these research directions, but also provides corresponding literature citations to support the discussion.

5.1 General bearing PHM model at model level fusion

5.1.1 General physical mechanism model

In this subsection, some physical mechanism models used to describe the bearing degradation mechanism are introduced. In a broad sense, the forms of bearing failure include fracture, wear, etc. The models used to describe these physical mechanisms mainly include the Paris model for crack propagation [86], the Archard model for wear degradation [88], and the Herz model for contact mechanics analysis [89]. The details of each model will be introduced one by one below.

Fatigue crack growth model

$$\frac{da}{dN} = c(\Delta K)^m \tag{7}$$

where c and m are material constants, a is the crack length, and ΔK is the stress intensity factor magnitude.

Archard wear model

$$\gamma = \frac{dh}{dt} = KP^m v^n \tag{8}$$

where γ is the wear rate, h is the wear thickness, t is the time, m is the influence index of P on γ, n is the influence index of v on γ, and K is the wear coefficient.

Herz contact model

$$\sigma_H = \sqrt{\frac{F_n\left(\frac{1}{\rho_1} \pm \frac{1}{\rho_2}\right)}{\pi L\left(\frac{1-\mu_1{}^2}{E_1} + \frac{1-\mu_2{}^2}{E_2}\right)}} \tag{9}$$

where σ_H is the contact stress, F_n is the normal force, L is the length of the contact line, ρ is the radius of curvature, the positive sign is used for external contact, and the negative sign is used for internal contact.

In fact, wear degradation and crack growth will affect the contact load of the bearing system. For example, the contact load generated by the bearing in the wear area is greater. At the same time, rolling bearings have multiple rolling elements and degrees of freedom, so the contact load of one rolling element will also affect other rolling elements. In addition, when wear or cracks occur, surface defects will disrupt the movement of the bearing and affect the dynamic response of the bearing system. In response to the above problems, many scholars have studied the physical mechanism model. Among them, Chen [160] put forward a friction dynamic model of coupling dynamic misalignment and wear, and evaluated the friction dynamic performance of bearings through translation and angular motion equations. Feng [161] considered the influence of bearing roughness and mixed lubrication on bearing wear, and then established the prediction framework of bearing wear profile. Belaid [162] studied the influence of vibration on crack propagation. By setting crack defects, the dynamic model of vibration on bearing natural frequency is constructed. However, these models currently face limitations such as difficulty in real-time parameter updates and high modeling complexity, which seriously hinder their large-scale application in actual factory mechanical systems. Therefore, the research on bearing PHM based on model level fusion is necessary.

5.1.2 General parameter calibration method

In this subsection, some methods for parameter updating are introduced. These methods mainly include Bayesian inference [90], KF [91], PF [92] and so on. The details of each model will be introduced one by one below.

Bayesian Inference

Bayesian inference is a statistical method, which calculates the posterior probability of parameters through Bayesian theorem and some prior knowledge, so as to achieve the goal of updating the parameters inside the model. According to prior knowledge, the parameter distribution of the physical mechanism model is $p(\theta|\alpha)$.On the basis of this prior knowledge, the data distribution of the model is $p(x|\theta)$. Its data distribution can also be expressed as a likelihood function, and the relationship can be written as follows

$$L(\theta|x) = p(x|\theta) \tag{10}$$

where $L(\theta|x)$ is likelihood function, θ is model parameter and x is observation data.

Therefore, the edge distribution of prior knowledge on parameters can be written as follows

$$p(x|\alpha) = \int_\theta p(x|\theta)\, p(\theta|\alpha) d\theta \tag{11}$$

Finally, according to Bayesian theorem, the posterior distribution of parameters can be written as follows

$$p(\theta|x,\alpha) = \frac{p(x|\theta)p(\theta|\alpha)}{p(x|\alpha)} \tag{12}$$

Kalman Filter

KF method estimates the state of the current moment (*k* moment) according to the data of the last moment (*k*-1 moment), so as to obtain the estimated value of the parameters at *k* moment. This method assumes that the system obeys linear Gaussian and consists of a state equation and an observation equation. Its state equation and observation equation can be written as follows

$$x_k = \boldsymbol{A}x_{k-1} + \boldsymbol{B}u_{k-1} + \eta_{k-1} \tag{13}$$

$$z_k = \boldsymbol{H}x_k + v_k \tag{14}$$

where x_k is the state variable at time *k*, $\boldsymbol{A}$ is the state transition matrix, x_{k-1} is the state variable at *k*-1, $\boldsymbol{B}$ is a matrix for converting the input into a state, u_{k-1} is the input signal at *k*-1, η_{k-1} is the noise in the process of *k*-1 time prediction, z_k is the observed variable at time *k*, $\boldsymbol{H}$ is the transformation matrix from state variables to measurement, v_k is the noise observed at *k* time.

After that, according to the state equation and observation equation, the process of KF prediction can be written as follows

$$\hat{x}_k^- = \boldsymbol{A}\hat{x}_{k-1} + \boldsymbol{B}u_{k-1} \tag{15}$$

$$\boldsymbol{P}_k^- = \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{A}^T + \boldsymbol{Q} \tag{16}$$

where $\hat{x}_k^-$ is the prior state estimation value at time *k*, $\hat{x}_{k-1}$ is the prior state estimation value at moment *k*-1, $\boldsymbol{P}_k^-$ is the covariance of prior estimation at time *k*, $\boldsymbol{P}_{k-1}$ is the covariance of posterior estimation at time *k*-1, $\boldsymbol{Q}$ is covariance of process excitation noise.

The state updating equation of KF can be written as follows

$$K_k = \frac{\boldsymbol{P}_k^-\boldsymbol{H}^T}{\boldsymbol{H}\boldsymbol{P}_k^-\boldsymbol{H}^T + \boldsymbol{R}} \tag{17}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \boldsymbol{H}\hat{x}_k^-) \tag{18}$$

$$\boldsymbol{P}_k = (\boldsymbol{I} - K_k\boldsymbol{H})\boldsymbol{P}_k^- \tag{19}$$

where K_k is the filter gain matrix, $\boldsymbol{R}$ is the measurement noise covariance, $\hat{x}_k$ is the posterior state estimation value at time *k*, $\boldsymbol{P}_k$ is the covariance of posterior estimation at time *k*, *I* is identity matrix.

Particle Filter

PF is a method to use Monte Carlo to sample a group of particles, and then calculate the approximate posterior probability distribution to update the model parameters. Usually, the equation of state and the observation equation can be written as follows

$$\theta_k = f(\theta_{k-1}) + \eta_{k-1},\ \ \eta_k \sim p(\eta_{k-1}) \tag{20}$$
$$y_k = h(\theta_k) + v_k,\ \ v_k \sim p(v_k) \tag{21}$$

where θ_k is the state variable at time k, $f(\cdot)$ is a state transition function; θ_{k-1} is the state variable at k-1, $\eta_{\eta-1}$ is the noise in the process of k-1 time prediction, y_k is the observed variable at time k, $h(\cdot)$ is the observation function, v_k is the noise observed at k time.

To update parameters by PF algorithm, it is necessary to sample particles through prior knowledge. Generally speaking, the distribution of particles adopts the prior of parameter evolution, and its expression can be written as following formula.

$$q(\theta_k|\theta_{k-1}^{(i)}, y_k) = p(\theta_k|\theta_{k-1}^{(i)}) \tag{22}$$

where $q(\cdot)$is the proposed distribution, $\theta_{k-1}^{(i)}$ is the parameter value of the i-th particle at time k-1.

After that, the approximate posterior distribution of particles can be written as follows

$$p(\theta_{k+1}|y_k) = \int p(\theta_{k+1}|\theta_k)p(\theta_k|y_k)d\theta_k \approx \sum_{i=1}^{N} \widetilde{w}_k^{(i)}\delta\left(\theta_k - \theta_k^{(i)}\right) \tag{23}$$

where $\widetilde{w}_k^{(i)}$ is the normalized weight; $\delta(\cdot)$ is Dirac function, which represents the discrete probability mass of each particle; N is the number of particles.

The normalized weight can be calculated by the following formula

$$w_k^{(i)} = w_{k-1}^{(i)} \cdot \frac{p(y_k|\theta_k^{(i)})p(\theta_k^{(i)}|\theta_{k-1}^{(i)})}{q(\theta_k^{(i)}|\theta_{k-1}^{(i)}, y_k)} = w_{k-1}^{(i)} \cdot (y_k|\theta_k^{(i)}) \tag{24}$$

$$\widetilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}} \tag{25}$$

Finally, in order to avoid the degradation of particle weight, the physical mechanism model parameters are updated by resampling the particle set.

The above are the main methods of parameter updating, and each method has its own advantages and usage scenarios. Table 6 summarizes the advantages and usage scenarios of the above three methods. Scholars can choose according to the data characteristics and calculation requirements of their own data.

Table 6 Comparisons of model-level fusion methods

Method	Core Assumptions	Computational Cost	Applicable Scenarios	Advantages
Bayesian Inference	Conjugate priors	Moderate	Offline calibration	Quantifies uncertainty, maintains physical interpretability
Kalman Filter	Linear Gaussian system	Low	Real-time state estimation	Computationally efficient, optimal linear estimation
Particle Filter	Nonlinear/non-Gaussian, large sample size	High	Complex dynamic systems	Handles strong nonlinearities and non-Gaussian noise

Bayesian inference constructs posterior parameter distributions through conjugate prior assumptions, with its core strength lying in quantifying uncertainties and preserving physical interpretability. However, it is constrained by computational complexity in high-dimensional parameter spaces or scenarios involving non-conjugate priors. The KF, grounded in linear-Gaussian system assumptions, achieves real-time state estimation (e.g., navigation and positioning) via the minimum mean square error (MMSE) criterion, offering theoretically optimal computational efficiency among comparable methods. Yet, it fails to address nonlinearities or non-Gaussian noise disturbances effectively. In contrast, PF employs a Monte Carlo sampling strategy, approximating posterior distributions of nonlinear/non-Gaussian systems through particle ensembles. This approach is well-suited for modeling strong nonlinearities in complex dynamic systems (e.g., target tracking) but inherently suffers from particle degeneracy and computational resource intensity.

These methodologies collectively advance rapid parameter calibration and accuracy enhancement, yet exhibit complementary performance boundaries and application scopes. Bayesian inference is ideal for offline parameter calibration and uncertainty quantification, while the KF balances efficiency and accuracy under linear-Gaussian assumptions. PF provides a universal framework for addressing nonlinear/non-Gaussian challenges. Researchers must carefully weigh system dynamics, computational resource constraints, and interpretability requirements when selecting an appropriate method.

5.2 Probabilistically enhanced model fusion

In bearing PHM, probabilistic enhanced model fusion is a key technology that improves the predictive capabilities of traditional physical mechanism models. This fusion is particularly suitable for handling failure modes such as bearing crack growth and wear degradation. The core of this method is to introduce probabilistic models to quantify the uncertainty and randomness in physical processes, thereby improving the accuracy of predictions. Fig. 11 is a probability enhancement model constructed by Bayesian calibration. Through probability-enhanced model fusion, the physical mechanism model can update parameters in real time, simplify the modeling process, and significantly reduce the negative impact of uncertainty on prediction accuracy. Wang [163] proposed an integrated fault diagnosis and prediction method for estimating the RUL of wind turbine bearings with limited data. The method first identifies early defects of bearings through wavelet transform and then uses the health index algorithm for feature fusion. Finally, the physical experience and statistical model under the Bayesian framework are used to achieve recursive prediction and uncertainty quantification of bearing RUL through PF. Gebraeel [164] developed a RUL prediction method based on Bayesian updating, which uses real-time condition monitoring information to update the stochastic parameters of the exponential degradation model. Through this method, the researchers constructed a closed remaining life distribution for the monitored equipment and verified the effectiveness of the developed degradation and RUL models through degradation signals obtained from bearing acceleration tests. Deng [133] proposed a rolling bearing RUL prediction model based on Bayesian calibration, aiming to improve data fidelity and model versatility. The study first constructed a five-degree-of-freedom dynamic model of the rolling bearing, and then maintained the high fidelity of the physical simulation through a particle filter-based calibration method. The researchers designed a physically informed Bayesian deep dual network, which fuses physical calibration simulations into an augmented input space to learn representative predictive features and integrate physical model parameters through adversarial learning. Make the transfer learning process interpretable. Sun [165] studied the problems of large fluctuations in time domain feature prediction accuracy and low data utilization and proposed a rolling bearing RUL prediction method based on the fusion of spectral kurtosis features and degradation physical mechanism model. This method constructs health indicators through Principal Component Analysis (PCA) and uses Bayesian theory and maximum likelihood functions to estimate the parameters of the exponential degradation model to predict the RUL of the bearing at each moment. Yang [166] proposed a RUL prediction method for wind turbine bearings based on multi-modal data fusion and Bayesian updating. This method integrates multiple feature parameters through PCA and uses the Wiener process for modeling. To reduce dependence on a large amount of historical data, the researchers used the Bootstrap sampling method to estimate the prior distribution parameters and used the Bayesian method to update the model parameters online. König [167] studied a RUL method for predicting the wear life of sliding bearings. This method constructs a statistical linear degradation model that couples Elastohydrodynamic lubrication and wear simulations. Bayesian inference is then used to update the linear degradation model throughout the operation to account for transient, system-dependent wear processes in RUL predictions. In addition, Zhao [168] proposed a multi-state health model to predict bearing failures. The model adopts a regression-based approach to detect health state transition points and applies an exponential random coefficient model with a Bayesian update process to estimate the failure time distribution. At the same time, the researchers introduced a model training framework to ensure that the model applies to more bearing instances in the same system setting.

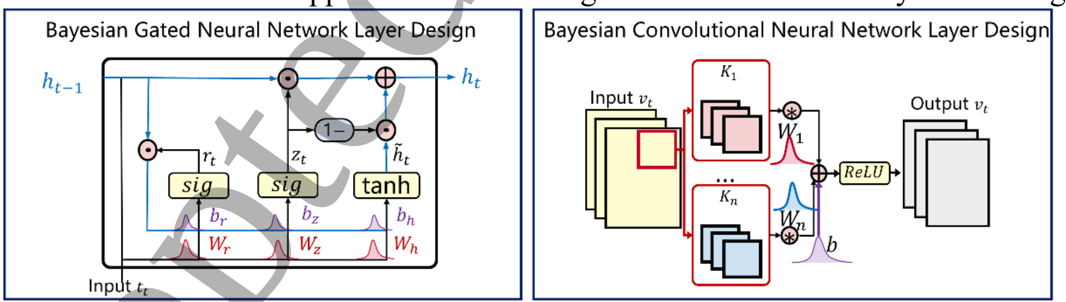

Figure 11 Probabilistically enhanced fusion model based on Bayesian calibration

In the above studies, the parameters of the physical mechanism model are usually not determined using physical mechanism knowledge, but monitoring data are used to calibrate the parameters of the physical mechanism model. At the same time, during the calibration process, the accuracy of the model is enhanced by introducing a probabilistic model. In addition, probabilistic enhanced model fusion will also quantify the uncertainty and randomness in the physical process and enhance the generalization of the bearing PHM model. However, the accuracy of prediction results is highly dependent on the correctness of the physical mechanism model. Therefore, future research needs to further explore how to reduce reliance on the accuracy of physical mechanism models.

5.3 Data assimilation model fusion

Data assimilation model fusion is equally crucial in bearing PHM. It improves the prediction accuracy of bearing operating status by combining physical mechanism models and real-time observation data. This method is particularly suitable for dynamically monitoring the health of bearings and predicting their performance changes. The data assimilation process starts with building a model based on physical laws, then collecting real-time data of bearing operation and preprocessing the data. Select a suitable data assimilation algorithm for model assimilation to obtain the optimal estimate of the bearing state, and continuously update the model based on new observation data. This process ensures that the model can reflect the actual state of the bearing in real time and improves the accuracy and reliability of the model prediction, ultimately achieving the application goal of real-time monitoring and health prediction of the bearing. Jiang [169] achieves accurate prediction of the RUL of bearings by constructing state space equations and iterating parameters using PF. Wang [170] constructed a new scalable two-stage linear/nonlinear composite model to represent various degradation behaviors and clarify the evolution law of individual degradation. At the same time, the degradation process is tracked by training the LSTM prediction network to learn the knowledge of multi-sample degradation behavior. Through the real-time matching of the hidden layer state, the interactive fusion of information and RUL prediction is realized. Zhan [171] first established a temperature prediction model based on the thermal network method. Then, SIAN and Sobol global sensitivity indices are used to extract parameters with global structural identifiability and estimability. These parameters are used as uncertainty parameters to construct the model's state space equation. In addition, a multi-layer PF method based on PF and sparse identification was proposed to improve the accuracy and robustness of thermal parameter estimation. Zhang [172] proposed a probabilistic fault detection method based on PF. By combining sensor data with a degradation model based on the Paris fatigue law, this method realizes early detection of faults and quantifies the probability of fault occurrence. It has a specific confidence level and improves the accuracy of different faults. Huang [173] proposed an enhanced deep learning-based fusion prediction method for RUL prediction by combining a Bi-LSTM network and a PF. This method fuses multi-sensor data to extract degradation features and perform RUL prediction, improving the prediction accuracy and uncertainty quantification ability of engineering systems. Deng [174] developed a hybrid prediction model based on a PF and a GRU network. This method combines physical wear models and data-driven techniques, and introduces Monte Carlo Dropout for uncertainty estimation, thereby improving the accuracy of bearing RUL prediction and degradation monitoring. Qian [175] and Liu [176] used optimization algorithms to update physical mechanism model parameters to achieve bearing RUL prediction. The above model-driven prediction methods require modeling of the structure, material, and working state of the bearing.

This method uses the dynamic characteristics of damage to extract relative damage indicators, which are then used by the dual exponential model in the PF to accurately estimate the RUL. Niu [13] proposed a life prediction method based on a physical-data hybrid model. First, a mathematical model of damage evolution was established. Then, the least squares method was used to estimate the parameters of the established damage evolution model. The model parameters were updated by the unscented KF method, and the future damage degree was predicted by continuous extrapolation.

Through the above discussion, data assimilation model fusion has many advantages. However, the process of data assimilation is very complicated and involves a lot of data processing and calculation, which leads to high computational costs. Secondly, the effect of data assimilation is limited by the quality and completeness of the observation data. Therefore, future research directions may include developing more efficient data assimilation algorithms to reduce computational costs and increase assimilation speed. At the same time, improving the robustness of data assimilation algorithms so that they can better handle incomplete or noisy data is also an important direction for future development.

6 Comprehensive analysis of different fusion methods

6.1 Advantages and limitations analysis

(1) Advantages

Holistically, multi-level physical mechanism-data fusion significantly enhances the accuracy, reliability, and transparency of bearing PHM [177-179]. This approach excels in detecting rare and highly variable faults. By incorporating physical mechanisms, the physical interpretability of the models is improved, and the predictions not only closely align with real-world operating conditions but also adhere to physical constraints. Integrating physical mechanisms allows deep learning models to more effectively discern the underlying patterns in the data, yielding more rational and reliable analytical results.

At the data level, leveraging domain-specific physical knowledge enables a more precise approach to data cleansing and feature extraction. By identifying the signals most directly tied to equipment health, large volumes of irrelevant or noisy data can be eliminated, easing computational burdens in subsequent steps and leading to a cleaner, more representative training set. Moreover, in situations where real-world data is limited or fault scenarios are insufficiently documented, physical mechanism-based simulations can be utilized to create virtual datasets. These datasets preserve high physical plausibility and better represent typical operating conditions. In order to quantify the advantages of data level fusion, this paper selects a digital twin-assisted dual transfer (DTa-DT) model based on data level fusion for data comparison with other 10 models. According to the results collected in relevant literature, the accuracy results of data-level fusion model and other 10 models are shown in Table 7 and Table 8 [113]. From these results, DTa-DT achieves average accuracies of 85.80% and 84.13% on the CWRU and XJTU-SQ datasets, respectively, significantly outperforming the other methods. This shows that the dual transfer framework can effectively reduce the distribution gap between the pure simulated dynamic response and the actual measurement by combining the simulated data generated by the physical mechanism model with the noise distribution transfer and using the deep branch network to adapt the model. This fusion model fully capitalizes on the respective strengths of physical and data models, thereby achieving higher diagnostic accuracy in scenarios with small sample sizes and missing fault types.

Table 7 Diagnostic accuracy of various models on the CWRU dataset[113]

Number	Method Name	1730rpm	1750rpm	1772rpm	1797rpm	Average Accuracy
1	No dual transfer (using dynamic model response directly)	33.33%	33.33%	33.33%	33.33%	33.33%
2	No model transfer (using only DTd-IT to generate twin data)	61.67%	53.33%	63.33%	50.00%	57.08%
3	CNN	79.83%	80.00%	76.33%	67.67%	75.96%
4	DBN	74.33%	76.67%	66.67%	61.67%	69.84%
5	WDCNN	82.71%	82.33%	76.67%	70.00%	77.93%
6	ResNet18	83.33%	82.17%	80.75%	76.25%	80.62%
7	TCA (Feature transfer)	56.67%	53.33%	54.71%	59.83%	56.45%
8	DANN (Feature transfer + adversarial learning)	69.70%	72.33%	66.70%	65.84%	67.65%
9	Improved ConvNext	85.53%	89.17%	81.33%	79.00%	83.76%
10	Improved CycleGAN (Data augmentation with subsequent CNN diagnosis)	80.33%	83.00%	79.33%	75.96%	79.66%
11	DTa-DT (Proposed dual transfer via information and model migration)	88.33%	89.83%	84.33%	80.71%	85.80%

Table 8 Diagnostic accuracy of various models on the XJTU-SQ dataset[113]

Number	Method Name	543rpm	1143rpm	1743rpm	2343rpm	Average Accuracy
1	No dual transfer (using dynamic model response directly)	33.33%	33.33%	33.33%	33.33%	33.33%
2	No model transfer (using only DTd-IT to generate twin data)	43.44%	51.20%	47.33%	56.33%	49.55%
3	CNN	74.83%	73.00%	70.33%	75.67%	73.46%
4	DBN	64.33%	66.67%	66.67%	61.67%	64.84%
5	WDCNN	76.71%	77.33%	70.67%	76.00%	75.18%
6	ResNet18	79.33%	80.17%	73.75%	76.25%	77.38%
7	TCA (Feature transfer)	54.83%	57.13%	52.25%	50.00%	54.38%
8	DANN (Feature transfer + adversarial learning)	69.70%	70.00%	67.67%	66.67%	67.54%
9	Improved ConvNext	82.33%	85.00%	79.00%	84.33%	82.67%

Continued table 8 Diagnostic accuracy of various models on the XJTU-SQ dataset[113]

Number	Method Name	543rpm	1143rpm	1743rpm	2343rpm	Average Accuracy
10	Improved CycleGAN (Data augmentation with subsequent CNN diagnosis)	79.96%	78.33%	75.67%	79.33%	78.32%
11	DTa-DT (Proposed dual transfer via information and model migration)	83.00%	85.17%	81.67%	86.67%	84.13%

At the network level, integrating physical constraints within the structure or loss function of deep learning models enforces stronger priors on the system's behavior. Purely data-driven approaches often falter under extreme or previously unseen conditions, whereas physics-informed networks retain fidelity to core principles, reducing the likelihood of nonsensical predictions. When faced with noisy or limited data, physical constraints can guide the model toward more plausible solutions, strengthening its robustness and diminishing the need for extensive datasets. In order to quantify the advantages of network-level fusion, this paper selects an innovative network-level physical mechanism fusion model for data comparison with the other five models. GRU encoder-decoder, Self-organizing map (SOM), LSTM encoder-decoder, CNN-Bi-LSTM and Deep convolutional Long short-term memory(CLSTM) are defined as M1, M2, M3, M4 and M5. Table 9[144] presents the predictive accuracy and error performance of the proposed network-level physical fusion method compared with five other methods (M1, M2, M3, M4, M5) across different bearing test sets. As can be seen from the table, the proposed method achieves mean squared errors of 582, 735, and 1086 on the three bearing test sets, which are far lower than those of the other traditional methods—whose mean square error (MSE) values are generally in the thousands or even tens of thousands. At the same time, the method also performs exceptionally well in terms of overall score, reaching 93.42 points, which is markedly higher than that of the control groups.

Table 9 Predictive accuracy and error performance across different test sets[144]

Method	Bearing A (MSE)	Bearing B (MSE)	Bearing C (MSE)	Average Score
M1	10726	8582	12033	78.54
M2	6428	7033	7635	81.07
M3	4230	5030	5255	83.50
M4	1845	2490	3150	87.12
M5	1320	2217	2912	88.00
Proposed	582	735	1086	93.42

At the model level, combining physical and data-driven modeling yields synergistic benefits that neither method alone can fully provide. Physical mechanism models offer a macro-level backbone grounded in fundamental laws, such as Archard wear model or Herz contact model, while data-driven models excel at capturing localized nonlinearities and transient phenomena. Through real-time calibration of the physical mechanism model's parameters using insights derived from data, the hybrid model can adapt to evolving operating conditions. Such adaptability proves particularly valuable in industrial contexts where system states shift frequently or unpredictably, and where decision-makers require transparent, science-based rationales to trust the model's outcomes. In order to quantify the advantages of model-level fusion, this paper selects a fusion model of physical simulation calibration and anti-migration mechanism for data comparison with the other five models. Table 10[133] presents a comparison of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coverage Probability of Prediction Interval (PICP) metrics for various models under the same machine (task A1) and cross-machine (task B1) conditions, clearly demonstrating the advantages of this hybrid method in reducing prediction errors and maintaining high uncertainty interval coverage.

Table 10 Comparisons for models under the same machine and cross-machine conditions [133]

Task	Indicator	SCAE	BDGRU	C-DANN	SCRNN	T-DVI	Proposed Method (HPDM+PI)
A1	RMSE	0.124	0.111	0.207	0.106	0.088	0.058
A1	MAE	0.104	0.096	0.176	0.091	0.080	0.043
A1	PICP	0.128	0.256	0.384	0.692	0.487	0.615
B1	RMSE	0.168	0.283	0.277	0.187	0.183	0.136
B1	MAE	0.145	0.243	0.239	0.165	0.141	0.119
B1	PICP	0.100	0.068	0.180	0.269	0.210	0.369

(2) Disadvantages

While multi-layer physical mechanism-data fusion can substantially elevate the accuracy and depth of insights for industrial fault diagnostics and health management, it also introduces a complex set of methodological and operational hurdles. These include high modeling costs, the need for specialized interdisciplinary collaboration, and significant computational expenditures. Achieving effective integration of physical mechanisms with data-driven approaches often requires cross-domain expert knowledge [180]. Additionally, it requires the support of advanced hardware to meet industrial demands for large-scale or near-real-time analytics [181]. Addressing these issues demands ongoing research to strike a suitable balance between theoretical rigor, practical feasibility, and the evolving needs of industry.

At the data level, for systems that exhibit multi-physics coupling or extreme nonlinearities, simplifying assumptions used in simulations may introduce substantial discrepancies between synthetic and real-world data. Consequently, virtual datasets might fail to provide meaningful augmentations to existing data, or worse, introduce systematic errors into the modeling process.

At the network level, incorporating detailed physical constraints can complicate both the model design and computational workflow. Balancing a neural architecture's learning capacity with the rigor of physical laws requires complex parameter tuning. When an industrial process undergoes unpredictable changes, previously embedded constraints may no longer apply. If the mechanistic assumptions are incomplete or partially incorrect, the model may experience training instability, slow convergence, even undermining the potential gains in robustness and physical interpretability.

At the model level, the coupling of physical mechanism and data-driven models heightens the complexity of system design, maintenance, and testing. If the operating conditions or failure modes of a machine evolve rapidly, frequent recalibration and parameter updates become indispensable yet challenging. When multiple models work in unison, the testing phase must ensure their collective performance remains reliable under a wide range of conditions, including rare or extreme scenarios.

6.2 Comparative analysis of fusion methods

In this section, the fusion methods of data layer, network layer and model layer are compared in detail. A comprehensive comparison of fusion methods at each level is presented in Table 11. By analyzing their respective application scenarios, computational complexity, data requirements, and physical interpretability, the advantages and challenges of each fusion method in bearing PHM are clarified. Data layer fusion focuses on improving data quality through processing and generation, network layer fusion integrates physical knowledge into the Neural Network structure to enhance model interpretability, while model layer fusion combines physical mechanism models with data-driven models to achieve more accurate health predictions. The comparison provides a clearer understanding of the applicability and potential of each method in practical applications.

Table 11 Comparison of different fusion methods

Aspect	Data Layer Fusion	Network Layer Fusion	Model Layer Fusion
Applicable Scenarios	Used when data quality is poor, such as noisy environments or insufficient labeled data. Physical mechanisms help process, clean, or augment data.	Used in deep learning models where physical knowledge is integrated to improve accuracy and interpretability.	Used when combining physical mechanism models (e.g., wear or crack propagation models) with data-driven models to improve prognostics.
Complexity	Moderate, involves preprocessing and data generation techniques.	High, as this involves modifying Neural Network architectures and embedding physical constraints into the learning process.	High, involves coupling physical mechanisms with machine learning models. May include parameter calibration and real-time updates.
Data Requirements	Relies on both raw and synthetic data (e.g., digital twin or simulation data).	Requires both high-quality data and robust physical mechanism models.	Requires both real-time data and reliable physical mechanism models. Calibration of physical mechanism models using sensor data is often necessary.
Physical Interpretability	High, as physical mechanisms guide data generation and processing.	Medium to high, as physical laws are embedded within the network's structure or loss function.	Very high, as physical mechanism models inherently provide interpretability.
Advantages	Enhances model generalization by increasing data quantity and quality, especially useful in small sample or noisy data scenarios.	Reduces overfitting, improves robustness, ensures predictions align with physical laws, enhancing interpretability.	Highly accurate predictions by merging the reliability of physical mechanism models with the adaptability of data-driven models. Reduces uncertainty.

Continued table 11 Comparison of different fusion methods

Aspect	Data Layer Fusion	Network Layer Fusion	Model Layer Fusion
Disadvantages	May introduce errors if simulations or data generation assumptions are not accurate.	Increased complexity in network design and training. Model instability if physical rules are not balanced properly.	High computational cost. The effectiveness depends on the accuracy of physical mechanism models, which can be hard to update or calibrate in real time.

In terms of application scenarios, data-layer fusion is best suited for noisy environments or insufficient labeled data because it focuses on leveraging physical mechanisms to clean, generate, and augment data. This approach not only increases data diversity but also improves model generalization at the source. Network-layer fusion, on the other hand, explicitly embeds physical rules into the Neural Network's structure or loss function. While it achieves higher interpretability and robustness, it also entails greater design complexity and requires higher-quality data. If real-time updates to the physical mechanism model are necessary—such as coupling wear or crack propagation models with data-driven systems—model-layer fusion becomes more appropriate. Although it provides high-fidelity degradation modeling under limited data conditions, it also demands accurate physical mechanism models and potentially substantial computational overhead for online updates.

From a complexity standpoint, data-layer fusion primarily operates before training, affecting the data rather than the core model structure; hence its complexity is moderate. In contrast, both network-layer and model-layer fusion demand significant architectural revisions or iterative procedures to incorporate physical constraints or co-estimate parameters, thereby raising the overall complexity.

Concerning data requirements, data-layer fusion relies on both real and synthetic datasets to mitigate issues like small samples or high noise; network-layer fusion also needs consistent, high-quality data plus reliable physical knowledge for custom network design; and model-layer fusion often uses real-time sensor data and high-fidelity physical simulations to continuously calibrate model parameters.

Regarding physical interpretability, data-layer fusion ensures a relatively high interpretability in data processing, network-layer fusion yields medium to high interpretability by embedding physical principles directly into the network, and model-layer fusion provides the highest level due to its inherent reliance on established physical mechanism models.

Finally, each level offers distinct advantages and disadvantages: data-layer fusion quickly enhances robustness to small sample or noisy conditions but risks introducing bias if the simulation assumptions are flawed; network-layer fusion effectively reduces overfitting and aligns predictions with physical laws yet is more complex to design and tune; and model-layer fusion yields highly accurate, low-uncertainty predictions but depends heavily on the veracity of the physical mechanism model and demands considerable computational resources for real-time deployment.

7 Challenges and future directions

Although this paper has outlined many advantages of fusion physical mechanisms and data-driven models for bearing health diagnosis, these models still face many challenges in practical applications. In this section, challenges of fusion models in terms of uncertain information, model attributes and industrial applications will be elaborated in detail. Subsequently, based on the current research status, the future development direction of fusion models will be further discussed, aiming to provide scholars with new insights for future research.

7.1 Challenges of fusion model

7.1.1 The challenge of uncertainty

In bearing PHM, uncertainties are ubiquitous, such as sensor noise and complexity of the working environment [182-184]. These uncertainties not only directly affect the quality and integrity of the data, but also increase the challenge for the model to accurately learn fault features. Due to the complexity and variability of bearing faults [185], the model needs to be able to handle and explain these uncertainties to achieve accurate fault diagnosis and prediction. Although the fusion physical mechanism and data-driven method improve the accuracy and generalization ability of pure data-driven model prediction to some extent. However, the pure data-driven model is a part of the fusion model, so the prediction of the fusion model will still be affected by the data-driven components. Therefore, these uncertainties also reduce the accuracy and generalization of the model, thereby limiting the development of the fusion model.

In addition, uncertainty factors pose a major challenge to the credibility of fusion physical mechanisms and data-driven models, greatly limiting their application in the industrial field. If predictive models are applied to industrial practice, the decision-making process of the model must be transparent and credible enough to ensure that engineers can understand the diagnostic results provided by the model. This credibility is a prerequisite for the model to be widely accepted and applied, as it is directly related to the quality and safety of maintenance decisions. However, when the model must deal with data with uncertainty, its decision-making process becomes less transparent and credible. Moreover, these uncertainties are also hidden in the complex calculations of the model, and even experts find it difficult to track and explain the specific decision-making basis of the model, further limiting the widespread application of fusion models.

7.1.2 The challenges of model attribute

The challenges of the fusion physical mechanism and data-driven bearing PHM method in model attributes mainly come from the complexity of data integration, the precise determination of model parameters and the high requirements of optimizer design. These challenges directly affect the performance, generalization ability, real-time and reliability of the model, thus limiting the effectiveness and efficiency of the fusion model in practical applications.

First, the bearing PHM system needs to process heterogeneous data from multiple sensors, and the differences in type, scale, sampling rate and time make data integration complicated [186, 187]. The challenge of data integration lies in how to fuse these data of different natures into a unified representation so that the model can understand and analyze it. For example, vibration data may need to be processed by Fourier transform and other methods before it can be effectively compared and integrated with temperature data. This need for data preprocessing increases the complexity of model training because complex data synchronization and feature extraction techniques need to be developed. In addition, the problem of spatiotemporal alignment of data is also a technical challenge because sensor data may be at different time and space scales, and precise alignment techniques are required to integrate this data. These factors limit the generalization ability of the fusion model and the feasibility of real-time applications because the model must be able to process and integrate data from different sources to provide accurate predictions.

The parameters in the fusion model include not only the parameters of the data-driven model, but also the physical parameters of the physical mechanism model. The determination of these parameters is a complex process, because the physical parameters of the physical mechanism model may be difficult to estimate directly from the data and need to rely on professional knowledge and assumptions. The difficulty of parameter estimation stems from the complexity of the physical process and the scarcity of data. For example, in the prediction of bearing fatigue life, the relationship between the fatigue limit of the material and the number of cycles may need to be determined jointly by experimental data and physical mechanism models. This difficulty in parameter estimation limits the accuracy and reliability of the model, because the performance of the model is highly dependent on the accuracy of the parameters. These factors limit the effectiveness and efficiency of the fusion model in practical application, because the model requires precise parameters to ensure the accuracy of its predictions.

The design of the model optimizer is crucial for the training of the fusion model. The optimizer needs to find a good enough solution in a limited time, which requires the optimizer to have high computational efficiency while maintaining the convergence and stability of the model. The non-convex optimization problem is a major challenge in the design of the optimizer, because the loss function constructed by the physical equations is generally highly non-convex, which makes it difficult to find the global optimal solution. At the same time, adding prior knowledge or constraints will bring additional complexity to the optimization and model training process. For example, the fusion model may need to incorporate physical laws or domain knowledge as constraints in the loss function, and these constraints need to be carefully designed to ensure the effectiveness of model training and the accuracy of the solution. In addition, the requirement of computational efficiency is also a challenge because in real-time or near-real-time PHM applications, the optimizer needs to be able to process new data quickly. These factors limit the application of fusion models in real-time monitoring and prediction, because they need to provide accurate predictions in a limited time, which is directly related to the real-time and reliability of the model.

7.1.3 The challenges in industrial applications

Hybrid methods, which integrate physical mechanisms with data-driven models, have shown significant potential in various industrial applications. In the aerospace industry, hybrid approaches combine FEM stress simulations with LSTM networks to predict fatigue cracks under variable loads. For instance, studies use physical guidance feature extraction from simulation data to improve the accuracy of fault detection, and reduce the occurrence of false alarms compared with pure data-driven methods. In the wind energy sector, gearbox bearing failures account for 23% of wind turbine downtime. Hybrid frameworks integrating Archard wear models with Transformer-based RUL prediction have reduced maintenance costs by 18% through early pitting detection. In rail transportation, wheelset bearing monitoring systems (e.g., employ Hertz contact mechanics coupled with 1D-CNNs to diagnose cage fractures from acoustic emissions. Despite the promising applications, hybrid methods face several challenges in industrial settings. The first is data heterogeneity, for example, asynchronous sampling between strain gauges (1 kHz) and infrared cameras (30 Hz) induces feature misalignment. Multimodal data fusion (vibration, thermal, oil debris) requires physics-informed normalization. The second is dynamic load adaptability. The calibration parameters of the physical mechanism model in cruise state can't cover the transient load changes in takeoff/landing phase, which leads to great fluctuation of prediction error. The third is real-time. The hybrid model of high-speed railway bearings requires reasoning time less than 50 milliseconds, but the FEM simulation based on physical mechanism usually exceeds 2 seconds per cycle. Finally, there are scalability restrictions. The model trained on laboratory-grade data (e.g., CWRU dataset) show 22-35% accuracy drops when deployed on industrial machinery with varying SNR levels. Addressing these challenges will be essential for the widespread adoption of hybrid methods in industrial bearing PHM.

7.2 Future directions of fusion model

In bearing PHM, methods that fusion physical mechanisms and data-driven approaches are becoming a hot topic. These methods aim to improve the accuracy and reliability of fault diagnosis by combining the deep understanding of physical mechanism models with the generalization capabilities of data-driven models. Through the discussion of the challenges of the above fusion model, this section mainly gives the future development direction of this field.

7.2.1 Multimodal data fusion

With the advent of the big data era, there are many types of data involved in bearing PHM research, including vibration data, acoustic emission data, torque data, temperature data [38-42], etc. Future research will focus on developing advanced fusion frameworks to integrate heterogeneous data such as vibration, acoustic emission, torque and temperature from different sensors, and use physical knowledge and data analysis technology to improve the model's comprehensive monitoring capabilities for bearing health status. The efficient fusion of physical mechanism methods and feature extraction methods will make extracting key features from raw data more efficient, which is also an important research direction for identifying equipment status changes and potential faults. At the same time, multi-source heterogeneous data fusion methods, including feature-level fusion and decision-level fusion, are also important directions for the development of bearing PHM. These research directions will jointly promote the advancement of bearing PHM technology, improve the accuracy of fault prediction, enhance the physical interpretability of the model, and provide more reliable diagnostic results and maintenance recommendations for engineering practitioners.

7.2.2 Data generation and information reconstruction methods

In bearing PHM, future research will focus on solving the challenges of small samples and missing modal data to improve the accuracy and reliability of the monitoring system. Researchers will explore small sample learning techniques, such as combining physical mechanism methods with classification algorithms, so that PHM research can effectively learn and classify under limited fault samples to improve the generalization ability of the model. At the same time, physical mechanism-driven generative models are also one of the main research directions in the future. By combining GANs[188, 189] and variational autoencoders (VAEs) [190, 191], new data samples will be generated by learning the probability distribution of data, which also provides some new solutions for small sample data. Meta-learning techniques will also give models the ability to quickly adapt to new tasks, which is particularly suitable for few-sample learning scenarios, thereby improving the model's adaptability to new environments and learning efficiency. For missing data reconstruction technology, the Gappy POD (proper orthogonal decomposition) [192, 193] algorithm and physical knowledge will be used to reconstruct the physical field from a small amount of local measurement data to enhance the robustness of the system. In addition, the joint feature extraction method based on 1-D CNN [187, 194] will be further studied to improve the accuracy of fault diagnosis and reduce diagnostic delay. In summary, the problem of PHM research needs to solve the current situation of small samples and lack of modalities in the industrial field. Therefore, improving the data generation ability and information reconstruction ability of the model is an trend for future development.

7.2.3 Innovative methods for parameter estimation

In future bearing PHM research, innovations in parameter estimation will promote technological progress through deep fusion of physical mechanisms and data-driven methods, especially physical information Neural Networks that combine physical laws with deep learning. In industrial applications, further enhancing the accuracy of parameter estimation will greatly enhance the accuracy of bearing PHM, based on the premise of using physical knowledge to improve the physical interpretability of the model. Among them, the application of Bayesian optimization will further optimize the fusion model parameter selection process, especially when dealing with high-dimensional non-convex problems. In addition, the combination of Nondominated Sorting Genetic Algorithm II (NSGA-II) dynamic update and digital twin will also be an important direction for bearing parameter estimation. By constructing physical parameters through digital twin technology and combining NSGA-II for real-time update, the dynamic response capability of the model is effectively solved, and the reliability and maintenance efficiency of industrial systems are improved. The above research directions will jointly shape the future of the bearing PHM field, ensuring that the technology continues to advance and is widely adopted in industrial applications.

7.2.4 Multi-scale physical modeling

In bearing PHM, future research directions will focus on enhancing the physical interpretability of fusing models through multi-scale physics modeling. The core goal of this direction is to provide in-depth theoretical support for bearing design optimization and performance improvement by accurately simulating the physical behavior of bearings under variable working environments, including but not limited to the interaction of multi-physics fields such as friction and wear, fracture mechanics and heat transfer. Then, by integrating data-driven models with physical mechanism models based on physical principles, the research will focus on extracting valuable features from data while ensuring the consistency of these features with physical laws to enhance the predictive accuracy and reliability of the model. In addition, in-depth exploration of multi-scale modeling will help to understand the failure mechanism of bearings and predict the durability and life of these materials. In this process, optimizing the allocation of computing resources, strengthening interdisciplinary cooperation, developing experimental techniques, and innovating intelligent fault diagnosis methods are important foundations for the rapid development of multi-scale physical modeling. Through these comprehensive research efforts, the transparency and credibility of bearing PHM models will be improved for widespread application in the engineering field. In summary, developing multi-physics coupling models is a challenging but very promising direction, and it has important application value in bearing PHM.

8 Conclusion

This paper comprehensively reviews the latest research progress of integration physical mechanisms and data-driven methods in bearing prognosis and health management. In order to deeply understand these fusion frameworks, this paper divides the bearing prediction and health management methods based on physical mechanisms and data-driven into three levels: data level, network level and model level from the perspective of fusion strategy, and thoroughly classifies and discusses the fusion methods at each level. In each sub-field, this paper explores the modeling principles, the latest research progress, and analyzes the application advantages and limitations of various fusion methods. By discussing the advantages and limitations of these individual methods, this paper summarizes the challenges and future development trends faced by current research, aiming to provide valuable guidance and deep insights for researchers and engineers in bearing prediction and health management.

Acknowledgment

This work was funded by National Natural Science Foundation of China (Grant No. 52275499, 92467101), and Open Fund of National Key Laboratory of Strength and Structural Integrity (No. LSSIKFJJ202402003).

References:

[1] Ho W. R., Tsolakis N., Dawes T., Dora M. and Kumar M. 2023 A Digital Strategy Development Framework for Supply Chains *IEEE Transactions On Engineering Management* **70** 2493-506

[2] Huang S., Wang G., Lei D. and Yan Y. 2022 Toward digital validation for rapid product development based on digital twin: a framework *The International Journal of Advanced Manufacturing Technology* **119** 2509-23

[3] Hu Q., Si X., Qin A., Lv Y. and Liu M. 2022 Balanced Adaptation Regularization Based Transfer

Learning for Unsupervised Cross-Domain Fault Diagnosis *IEEE Sensors Journal* **22** 12139-51

[4] Li W., Huang R., Li J., Liao Y., Chen Z., He G., Yan R. and Gryllias K. 2022 A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges *Mechanical Systems and Signal Processing* **167** 108487

[5] Miao J., Tian X. and Pu W. 2024 Dynamic characteristic analysis of lubricated bearing in flexible rotor system using real-time coupled finite element model *Mechanical Systems and Signal Processing* **218** 111550

[6] Niu J., Pan J., Qin Z., Huang F. and Qin H. 2024 Small-Sample Bearings Fault Diagnosis Based on ResNet18 with Pre-Trained and Fine-Tuned Method *Applied Sciences* **14** 5360

[7] Zou X., Zhang H., Jiang Z., Zhang K. and Xu Y. 2024 Toward accurate extraction of bearing fault modulation characteristics with novel time–frequency modulation bispectrum and modulation Gini index analysis *Mechanical Systems and Signal Processing* **219** 111629

[8] Gao Y. and Li X. 2024 Fault Diagnosis of Centrifugal fan Bearings Based on I-CNN and JMMD in the Context of Sample Imbalance *Eksploatacja I Niezawodność – Maintenance and Reliability* **26**

[9] Zhang M. 2024 MKurt-LIA: mechanical fault vibration signal measurement scheme with frequency tracking capability for bearing condition monitoring *Measurement Science and Technology* **35** 96124

[10] Wang X., Li Y., Noman K. and Nandi A. K. 2024 Multi-task learning mixture density network for interval estimation of the remaining useful life of rolling element bearings *Reliability Engineering & System Safety* **251** 110348

[11] Hu R. 2022 Early fault diagnosis of key parts of rolling mill based on cross-correlation energy ratio entropy and BIGRU-GRU *Computer Measurement &Control* **30** 95-102

[12] Liu J. and Shao Y. 2018 Overview of dynamic modelling and analysis of rolling element bearings with localized and distributed faults *Nonlinear Dynamics* **93** 1765-98

[13] Niu Q. 2018 Study on performance degeradation and life prediction of mechanical rotary components *PhD Thesis* Zhejiang University

[14] Gao Y. and Yu D. 2021 Intelligent fault diagnosis for rolling bearings based on graph shift regularization with directed graphs *Advanced Engineering Informatics* **47** 101253

[15] Liu Z., Wang H., Liu J., Qin Y. and Peng D. 2021 Multitask Learning Based on Lightweight 1DCNN for Fault Diagnosis of Wheelset Bearings *IEEE Transactions On Instrumentation and Measurement* **70** 1-11

[16] Huang C., Bu S., Lee H. H., Chan K. W. and Yung W. K. C. 2024 Prognostics and health management for induction machines: a comprehensive review *Journal of Intelligent Manufacturing* **35** 937-62

[17] Tahan M., Tsoutsanis E., Muhammad M. and Abdul Karim Z. A. 2017 Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review *Applied Energy* **198** 122-44

[18] Montero Jimenez J. J., Schwartz S., Vingerhoeds R., Grabot B. and Salaün M. 2020 Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics *Journal of Manufacturing Systems* **56** 539-57

[19] Hu Y., Miao X., Si Y., Pan E. and Zio E. 2022 Prognostics and health management: A review from the perspectives of design, development and decision *Reliability Engineering & System Safety* **217** 108063

[20] Lee J., Wu F., Zhao W., Ghaffari M., Liao L. and Siegel D. 2014 Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications *Mechanical Systems and*

Signal Processing **42** 314-34

[21] Zhang Z., Liu B., Liu Y. and Zhang H. 2022 Fault Feature-Extraction Method of Aviation Bearing Based on Maximum Correlation Re'nyi Entropy and Phase-Space Reconstruction Technology *Entropy* **24** 1459

[22] Ma J., Li Z., Zhan L., Li C. and Zhang G. 2022 Research on non-contact aerospace bearing cage-speed monitoring based on weak magnetic detection *Mechanical Systems and Signal Processing* **171** 108785

[23] Li H., Liu C., Yang F., Ma X., Guo N., Sui X. and Wang X. 2023 Dynamic Temperature Prediction on High-Speed Angular Contact Ball Bearings of Machine Tool Spindles Based on CNN and Informer *Lubricants* **11** 343

[24] Zhao Z., Wang Q., Shao C., Chen N., Liu X. and Wang G. 2023 A state detection method of offshore wind turbines' gearbox bearing based on the transformer and GRU *Measurement Science and Technology* **35** 25903

[25] Liu X., Zhang Z., Meng F. and Zhang Y. 2023 Fault Diagnosis of Wind Turbine Bearings Based on CNN and SSA–ELM *Journal of Vibration Engineering & Technologies* **11** 3929-45

[26] Wang Y., Li Y., Zhang Y., Lei J., Yu Y., Zhang T., Yang Y. and Zhao H. 2024 Incorporating prior knowledge into self-supervised representation learning for long PHM signal *Reliability Engineering & System Safety* **241** 109602

[27] Fink O., Wang Q., Svensén M., Dersin P., Lee W. and Ducoffe M. 2020 Potential, challenges and future directions for deep learning in prognostics and health management applications *Engineering Applications of Artificial Intelligence* **92** 103678

[28] Murgia A., Harsha C., Tsiporkova E., Nawghane C. and Vandevelde B. 2024 A Hybrid Model for Prognostic and Health Management of Electronic Devices *Electronics* **13** 642

[29] Zio E. 2022 Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice *Reliability Engineering & System Safety* **218** 108119

[30] Meng H. and Li Y. 2019 A review on prognostics and health management (PHM) methods of lithium-ion batteries *Renewable and Sustainable Energy Reviews* **116** 109405

[31] Chen J. and Liu Y. 2022 Fatigue modeling using neural networks: A comprehensive review *Fatigue & Fracture of Engineering Materials & Structures* **45** 945-79

[32] Zhang Y., Feng K., Ji J. C., Yu K., Ren Z. and Liu Z. 2023 Dynamic Model-Assisted Bearing Remaining Useful Life Prediction Using the Cross-Domain Transformer Network *IEEE/ASME Transactions On Mechatronics* **28** 1070-80

[33] Zhuang X., Yu T., Saraygord Afshari S., Sun Z., Song K. and Liang X. 2021 Remaining useful life prediction of a mechanism considering wear correlation of multiple joints *Mechanical Systems and Signal Processing* **149** 107328

[34] Khodadadi Sadabadi K., Jin X. and Rizzoni G. 2021 Prediction of remaining useful life for a composite electrode lithium ion battery cell using an electrochemical model to estimate the state of health *Journal of Power Sources* **481** 228861

[35] Wang X., Ye P., Liu S., Zhu Y., Deng Y., Yuan Y. and Ni H. 2023 Research Progress of Battery Life Prediction Methods Based on Physical Model *Energies* **16** 3858

[36] Yin K., Xu Y., Li J. and Zhou X. 2024 Gaussian process regression driven rapid life-cycle based seismic fragility and risk assessment of laminated rubber bearings supported highway bridges subjected to multiple uncertainty sources *Engineering Structures* **316** 118615

[37] Li Y., Cui Y. and Deng S. 2024 Dynamic characteristics of rigid-elastic-liquid-coupled ball bearings

considering elastohydrodynamic lubrication *Mechanism and Machine Theory* **201** 105727
[38] Maruthi G. S. and Hegde V. 2016 Application of MEMS Accelerometer for Detection and Diagnosis of Multiple Faults in the Roller Element Bearings of Three Phase Induction Motor *IEEE Sensors Journal* **16** 145-52
[39] Li K., Chen P. and Wang H. 2012 Intelligent Diagnosis Method for Rotating Machinery Using Wavelet Transform and Ant Colony Optimization *IEEE Sensors Journal* **12** 2474-84
[40] Kang M., Kim J. and Kim J. 2015 An FPGA-Based Multicore System for Real-Time Bearing Fault Diagnosis Using Ultrasampling Rate AE Signals *IEEE Transactions On Industrial Electronics* **62** 2319-29
[41] Ibrahim A., El Badaoui M., Guillet F. and Bonnardot F. 2008 A New Bearing Fault Detection Method in Induction Machines Based on Instantaneous Power Factor *IEEE Transactions On Industrial Electronics* **55** 4252-9
[42] Xiao X., Liu J., Liu D., Tang Y., Dai J. and Zhang F. SSAE-MLP: Stacked sparse autoencoders-based multi-layer perceptron for main bearing temperature prediction of large-scale wind turbines *Concurrency and Computation: Practice and Experience* DOI:10.1002/cpe.6315.
[43] Xu Z. and Saleh J. H. 2021 Machine learning for reliability engineering and safety applications: Review of current status and future opportunities *Reliability Engineering & System Safety* **211** 107530
[44] Li X., Yu D., Søren Byg V. and Daniel Ioan S. 2023 The development of machine learning-based remaining useful life prediction for lithium-ion batteries *Journal of Energy Chemistry* **82** 103-21
[45] Li H., Zhang Z., Li T. and Si X. 2024 A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities *Mechanical Systems and Signal Processing* **209** 111120
[46] Guan Q. and Wei X. 2023 The Statistical Data-driven Remaining Useful Life Prediction-A Review on the Wiener Process-based Method *2023 Prognostics and Health Management Conference (PHM) Paris France* pp 64-8
[47] Li G., Wu Y., Wang C., Peng S., Niu J. and Yu S. 2023 The SRVM: A Similarity-Based Relevance Vector Machine for Remaining Useful Lifetime Prediction in the Industrial Internet of Things *IEEE Intelligent Systems* **38** 45-55
[48] Ren L., Liu Y., Wang X., Lu J. and Deen M. J. 2021 Cloud–Edge-Based Lightweight Temporal Convolutional Networks for Remaining Useful Life Prediction in IIoT *IEEE Internet of Things Journal* **8** 12578-87
[49] Sawant V., Deshmukh R. and Awati C. 2023 Machine learning techniques for prediction of capacitance and remaining useful life of supercapacitors: A comprehensive review *Journal of Energy Chemistry* **77** 438-51
[50] Wang X., Qiu J. and Liu G. 2005 Research on SVM-Based Machine Fault Feature Selection *Mechanical Science and Technology* **24** 1122-5
[51] Zhou Z. 2016 Fault diagnosis of complex industrial processes based on k-nearest neighbors *PhD Thesis* Zhejiang University
[52] Shevchik S. A., Saeidi F., Meylan B. and Wasmer K. 2017 Prediction of Failure in Lubricated Surfaces Using Acoustic Time–Frequency Features and Random Forest Algorithm *IEEE Transactions On Industrial Informatics* **13** 1541-53
[53] Zhao J., Feng X., Pang Q., Wang J., Lian Y., Ouyang M. and Burke A. F. 2023 Battery prognostics and health management from a machine learning perspective *Journal of Power Sources* **581** 233474
[54] Liu Y., Liao Y., Zhang J., Wei J., Chen Z. and Zhang Y. 2024 Adaptive Gearshift Control for Dual Clutch Transmissions Based on Hybrid Physical and Data Driven Modeling *Automotive Innovation* **7**

529-43

[55] Zhang Y., Gu P., Duan B. and Zhang C. 2024 A hybrid data-driven method optimized by physical rules for online state collaborative estimation of lithium-ion batteries *Energy* **301** 131710

[56] Guo J., Li Z. and Li M. 2020 A Review on Prognostics Methods for Engineering Systems *IEEE Transactions On Reliability* **69** 1110-29

[57] Zhang X., Liu H., Zhu C., Wei P., Wu S. 2021 Current Situation and Developing Trend of Fatigue Life Prediction of Components based on Data-driven *Journal of Mechanical Transmission* **45** 1-14

[58] Wilhelm Y., Reimann P., Gauchel W. and Mitschang B. 2021 Overview on hybrid approaches to fault detection and diagnosis: Combining data-driven, physics-based and knowledge-based models *Procedia CIRP* **99** 278-83

[59] Wang S., Jin S., Bai D., Fan Y., Shi H. and Fernandez C. 2021 A critical review of improved deep learning methods for the remaining useful life prediction of lithium-ion batteries *Energy Reports* **7** 5562-74

[60] Wang J., Li Y., Gao R. X. and Zhang F. 2022 Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability *Journal of Manufacturing Systems* **63** 381-91

[61] Ferreira C. and Gonçalves G. 2022 Remaining Useful Life prediction and challenges: A literature review on the use of Machine Learning Methods *Journal of Manufacturing Systems* **63** 550-62

[62] Peng F., Zheng L., Peng Y., Fang C. and Meng X. 2022 Digital Twin for rolling bearings: A review of current simulation and PHM techniques *Measurement* **201** 111728

[63] Yin X., Wang L., Jiang X. 2023 Research Progress on Fault Modes and Diagnosis Methods for Wind Turbine Bearings *Bearing* 131-9

[64] Liu X., Jiang D., Tao B., Xiang F., Jiang G., Sun Y., Kong J. and Li G. 2023 A systematic review of digital twin about physical entities, virtual models, twin data, and applications *Advanced Engineering Informatics* **55** 101876

[65] Wang H., Li B., Gong J. and Xuan F. 2023 Machine learning-based fatigue life prediction of metal materials: Perspectives of physics-informed and data-driven hybrid methods *Engineering Fracture Mechanics* **284** 109242

[66] Yang T., Yi X., Lu S., Johansson K. H. and Chai T. 2021 Intelligent Manufacturing for the Process Industry Driven by Industrial Artificial Intelligence *Engineering (Beijing, China)* **7** 1224-30

[67] Arunan A., Qin Y., Li X. and Yuen C. 2024 A Federated Learning-Based Industrial Health Prognostics for Heterogeneous Edge Devices Using Matched Feature Extraction *IEEE Transactions On Automation Science and Engineering* **21** 3065-79

[68] Zhong K., Han M. and Han B. 2020 Data-driven based fault prognosis for industrial systems: a concise overview *IEEE/CAA Journal of Automatica Sinica* **7** 330-45

[69] Yang J., Xie G. and Yang Y. 2020 An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data *Control Engineering Practice* **98** 104358

[70] Tian S., Zhen D., Sun G., Liu X., Feng G. and Gu F. 2024 Few-shot condition diagnosis of rolling bearing using adversarial transfer network with class aggregation-guided *Measurement Science and Technology* **35** 66120

[71] Tang J., Wu J., Hu B. and Liu J. 2022 An intelligent diagnosis method using fault feature regions for untrained compound faults of rolling bearings *Measurement* **204** 112100

[72] Zhao H., Liu H., Xu J. and Deng W. 2020 Performance Prediction Using High-Order Differential Mathematical Morphology Gradient Spectrum Entropy and Extreme Learning Machine *IEEE*

Transactions On Instrumentation and Measurement **69** 4165-72

[73] Suh S., Lukowicz P. and Lee Y. O. 2022 Generalized multiscale feature extraction for remaining useful life prediction of bearings with generative adversarial networks *Knowledge-Based Systems* **237** 107866

[74] Zhang R., Tao H., Wu L. and Guan Y. 2017 Transfer Learning With Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions *IEEE Access* **5** 14347-57

[75] Xu K., Kong X., Wang Q., Yang S., Huang N. and Wang J. 2022 A bearing fault diagnosis method without fault data in new working condition combined dynamic model with deep learning *Advanced Engineering Informatics* **54** 101795

[76] Hu H., Feng Y., Hu Q. and Zhang Y. 2023 A Masked One-Dimensional Convolutional Autoencoder for Bearing Fault Diagnosis Based on Digital Twin Enabled Industrial Internet of Things *IEEE Journal On Selected Areas in Communications* **41** 3242-53

[77] Huang F., Li X., Zhang K., Zheng Q., Ma J. and Ding G. 2024 A novel simulation-assisted transfer method for bearing unknown fault diagnosis *Measurement Science and Technology* **35** 106127

[78] Ben-David S., Blitzer J., Crammer K., Kulesza A., Pereira F. and Vaughan J. W. 2010 A theory of learning from different domains *Machine Learning* **79** 151-75

[79] Feng Z., Ma H. and Zuo M. J. 2017 Spectral negentropy based sidebands and demodulation analysis for planet bearing fault diagnosis *Journal of Sound and Vibration* **410** 124-50

[80] Sadoughi M. and Hu C. 2019 Physics-Based Convolutional Neural Network for Fault Diagnosis of Rolling Element Bearings *IEEE Sensors Journal* **19** 4181-92

[81] Raissi M., Yazdani A. and Karniadakis G. E. 2020 Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations *Science* **367** 1026-30

[82] Sahli Costabal F., Yang Y., Perdikaris P., Hurtado D. E. and Kuhl E. 2020 Physics-Informed Neural Networks for Cardiac Activation Mapping *Frontiers in Physics* **8** DOI:10.3389/fphy.2020.00042.

[83] Goswami S., Anitescu C., Chakraborty S. and Rabczuk T. 2020 Transfer learning enhanced physics informed neural network for phase-field modeling of fracture *Theoretical and Applied Fracture Mechanics* **106** 102447

[84] Hu X., Hu H., Verma S. and Zhang Z. 2021 Physics-Guided Deep Neural Networks for Power Flow Analysis *IEEE Transactions On Power Systems* **36** 2082-92

[85] Reichstein M., Camps-Valls G., Stevens B., Jung M., Denzler J., Carvalhais N. and Prabhat 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195-204

[86] Paris P. Erdogan F. 1963 A critical analysis of crack propagation laws *J. Basic Eng.* **85** 528–533

[87] Forman R. G. 1972 Study of fatigue crack initiation from flaws using fracture mechanics theory *Engineering Fracture Mechanics* **4** 333-45

[88] Archard J. F. and HIRST W. 1956 The Wear of Metals under Unlubricated Conditions *Proceedings of the Royal Society of London Series a, Mathematical and Physical Sciences* **236** 397-410

[89] Hertz. On the Contact of Elastic Solids *Crelle's Journal* **92** 156-71

[90] Yan B., Ma X., Huang G. and Zhao Y. 2021 Two-stage physics-based Wiener process models for online RUL prediction in field vibration data *Mechanical Systems and Signal Processing* **152** 107378

[91] Pang Z., Si X., Hu C. and Zhang Z. 2022 An Age-Dependent and State-Dependent Adaptive Prognostic Approach for Hidden Nonlinear Degrading System *IEEE/CAA Journal of Automatica Sinica* **9** 907-21

[92] Yang J., Fang W., Chen J. and Yao B. 2022 A lithium-ion battery remaining useful life prediction method based on unscented particle filter and optimal combination strategy *Journal of Energy Storage*

55 105648

[93] Cheng W., Liu X., Xing J., Chen X., Ding B., Zhang R., Zhou K. and Huang Q. 2023 AFARN: Domain Adaptation for Intelligent Cross-Domain Bearing Fault Diagnosis in Nuclear Circulating Water Pump *IEEE Transactions On Industrial Informatics* **19** 3229-39

[94] Lu H., Nemani V. P., Barzegar V., Allen C., Hu C., Laflamme S., Sarkar S. and Zimmerman A. T. 2023 A physics-informed feature weighting method for bearing fault diagnostics *Mechanical Systems and Signal Processing* **191** 110171

[95] Song X., Wei W., Zhou J., Ji G., Hussain G., Xiao M. and Geng G. 2023 Bayesian-Optimized Hybrid Kernel SVM for Rolling Bearing Fault Diagnosis *Sensors* **23** 5137

[96] Sheng Y., Liu H. and Li J. 2023 Bearing performance degradation assessment and remaining useful life prediction based on data-driven and physical model *Measurement Science and Technology* **34** 55002

[97] Zhu J., Chen N. and Peng W. 2019 Estimation of Bearing Remaining Useful Life Based on Multiscale Convolutional Neural Network *IEEE Transactions On Industrial Electronics* **66** 3208-16

[98] Lu Y., Li Q. and Liang S. Y. 2018 Physics-based intelligent prognosis for rolling bearing with fault feature extraction *The International Journal of Advanced Manufacturing Technology* **97** 611-20

[99] Pan L., Tang X., Lu J., Liu F., Liu N. 2024 Spatio-Temporal Feature Fusion for Residual Life Prediction of Bearings Under Cloud-Edge Collaboration *Modular Machine Tool & Automatic Manufacturing Technique* 116-21,125

[100] Liu N., Tang X., Lu J., Liu F., Liu P. 2023 Remaining Life Prediction Method of Bearing Based on Dual Data Sources Fusion Under Cloud-Edge Collaboration *Modular Machine Tool & Automatic Manufacturing Technique* 89-94

[101] Zhang L., Zhang H., Wang C., Wang X., Wen P. and Zhao L. 2023 Explicit Dynamics Driving Fault Diagnosis Method for Bearing Variable Conditions *China Mechanical Engineering* **34** 982-92

[102] Walther S. and Fuerst A. 2022 Reduced Data Volumes through Hybrid Machine Learning Compared to Conventional Machine Learning Demonstrated on Bearing Fault Classification *Applied Sciences* **12** 2287

[103] Cao H., Xiao W., Sun J., Gan M. and Wang G. 2024 A hybrid data- and model-driven learning framework for remaining useful life prognostics *Engineering Applications of Artificial Intelligence* **135** 108557

[104] Su X., Liu H., Tao L., Lu C. and Suo M. 2021 An end-to-end framework for remaining useful life prediction of rolling bearing based on feature pre-extraction mechanism and deep adaptive transformer model *Computers & Industrial Engineering* **161** 107531

[105] Liu P., Liu T., Wang S. and Wu X. 2020 Bearing fault diagnosis method based on information fusion and fast ICA *Journal of Vibration and Shock* **39** 250-9

[106] Deng Y., Huang D., Du S., Li G., Zhao C. and Lv J. 2021 A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis *Computers in Industry* **127** 103399

[107] Jia S., Deng Y., Lv J., Du S. and Xie Z. 2022 Joint distribution adaptation with diverse feature aggregation: A new transfer learning framework for bearing diagnosis across different machines *Measurement* **187** 110332

[108] Fukata S., Gad E. H., Kondou T., Ayabe T. and Tamura H. 1985 On the Radial Vibration of Ball Bearings: Computer Simulation *Bulletin of JSME* **28** 899-904

[109] Gong S., Li S., Zhang Y., Zhou L. and Xia M. 2024 Digital twin-assisted intelligent fault diagnosis for bearings *Measurement Science and Technology* **35** 106128

[110] Yang N. and Wang Z. 2023 Research on Virtual-Real Interaction of Digital Twin Driven Rolling

Bearing *Modular Machine Tool & Automatic Manufacturing Technique* 160-3, 168

[111] Zhang X., Liao Y., Li Q. and Chen Y. 2023 Fault diagnosis of vehicle motor-bearings under safe running by digital-twin technology *Journal of Automotive Safety and Energy* **14** 232-8

[112] Huang X., Xie T., Luo S., Wu J., Luo R. and Zhou Q. 2024 Incremental learning with multi-fidelity information fusion for digital twin-driven bearing fault diagnosis *Engineering Applications of Artificial Intelligence* **133** 108212

[113] Li Z., Ding X., Song Z., Wang L., Qin B. and Huang W. 2024 Digital twin-assisted dual transfer: A novel information-model adaptation method for rolling bearing fault diagnosis *Information Fusion* **106** 102271

[114] Zhang R., Zeng Z., Li Y., Liu J. and Wang Z. 2022 Research on Remaining Useful Life Prediction Method of Rolling Bearing Based on Digital Twin *Entropy* **24** 1578

[115] Desai P. S., Granja V. and Higgs C. F. 2021 Lifetime Prediction Using a Tribology-Aware, Deep Learning-Based Digital Twin of Ball Bearing-Like Tribosystems in Oil and Gas *Processes* **9** 922

[116] Xu K., Kong X., Wang Q., Han B. and Sun L. 2023 Intelligent fault diagnosis of bearings under small samples: A mechanism-data fusion approach *Engineering Applications of Artificial Intelligence* **126** 107063

[117] Cui L., Wang X., Wang H. and Jiang H. 2020 Remaining useful life prediction of rolling element bearings based on simulated performance degradation dictionary *Mechanism and Machine Theory* **153** 103967

[118] Peng C., Zhang S. and Li C. 2022 A Rolling Bearing Fault Diagnosis Based on Conditional Depth Convolution Countermeasure Generation Networks under Small Samples *Sensors* **22** 5658

[119] Dai J. and Tian L. 2023 A Novel Prognostic Method for Wear of Sliding Bearing Based on SFENN. *ICIRA 2023* Berlin Heidelberg pp 212-25

[120] Mohamad T. H., Abbasi A., Kappaganthu K. and Nataraj C. 2023 On extraction, ranking and selection of data-driven and physics-informed features for bearing fault diagnostics *Knowledge-Based Systems* **276** 110744

[121] Cheng G. 2021 Research on rolling bearing fault diagnosis method based on mechanism and data fusion *MSc Thesis* Xidian University

[122] Cheng H., Kong X., Wang Q., Ma H., Yang S. and Xu K. 2023 Remaining useful life prediction combined dynamic model with transfer learning under insufficient degradation data *Reliability Engineering & System Safety* **236** 109292

[123] Zhao W., Zhang C., Fan B., Wang J., Gu F., García Peyrano O., Wang S. and Lv D. 2023 Research on rolling bearing virtual-real fusion life prediction with digital twin *Mechanical Systems and Signal Processing* **198** 110434

[124] Jantunen E., Hooghoudt J. O., Yang Y. and McKay M. 2018 Predicting the remaining useful life of rolling element bearings. In: *2018 IEEE International Conference on Industrial Technology (ICIT),* pp 2035-40

[125] Yan M., Wang X., Wang B., Chang M. and Muhammad I. 2020 Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model *Isa Transactions* **98** 471-82

[126] Song W., Xiang J. and Zhong Y. 2018 A simulation model based fault diagnosis method for bearings *Journal of Intelligent & Fuzzy Systems* **34** 3857-67

[127] Gao Y., Liu X., Huang H. and Xiang J. 2021 A hybrid of FEM simulations and generative adversarial networks to classify faults in rotor-bearing systems *Isa Transactions* **108** 356-66

[128] Lou Y., Kumar A. and Xiang J. 2022 Machinery Fault Diagnosis Based on Domain Adaptation to Bridge the Gap between Simulation and Measured Signals *IEEE Transactions On Instrumentation and Measurement* **71** 1-9

[129] Piltan F., Toma R. N., Shon D., Im K., Choi H., Yoo D. and Kim J. 2022 Strict-Feedback Backstepping Digital Twin and Machine Learning Solution in AE Signals for Bearing Crack Identification *Sensors* **22** 539

[130] Hou W., Zhang C., Jiang Y., Cai K., Wang Y. and Li N. 2023 A new bearing fault diagnosis method via simulation data driving transfer learning without target fault data *Measurement* **215** 112879

[131] Xiao Y., Shao H., Han S., Huo Z. and Wan J. 2022 Novel Joint Transfer Network for Unsupervised Bearing Fault Diagnosis From Simulation Domain to Experimental Domain *IEEE/ASME Transactions On Mechatronics* **27** 5254-63

[132] Dong S., Zhu P., Zhu S., Liu L., Xing B. and Hu X. 2023 Fault Diagnosis Method of Rolling Bearings Based on Simulation Data Drive and Domain Adaptation *China Mechanical Engineering* **34** 694-702

[133] Deng Y., Du S., Wang D., Shao Y. and Huang D. 2023 A Calibration-Based Hybrid Transfer Learning Framework for RUL Prediction of Rolling Bearing Across Different Machines *IEEE Transactions On Instrumentation and Measurement* **72** 1-15

[134] Dong Y., Li Y., Zheng H., Wang R. and Xu M. 2022 A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample problem *Isa Transactions* **121** 327-48

[135] Liu T., Kou L., Le Y., Fan W. and Wu C. 2020 A physical knowledge-based extreme learning machine approach to fault diagnosis of rolling element bearing from small datasets *Adjunct Proceedings of the 2020 ACM International Joint Conference On Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium On Wearable Computers* 553-559

[136] Tai C. and Altintas Y. 2023 A hybrid physics and data-driven model for spindle fault detection *Cirp Annals* **72** 297-300

[137] Wang S., Qiao Z. and Niu P. 2023 Coupled Hybrid Stochastic Resonance With Multiobjective Optimization for Machinery Dynamic Signature and Fault Diagnosis *IEEE Sensors Journal* **23** 11825-37

[138] Zhang K., Wang J., Shi H., Zhang X. and Tang Y. 2021 A fault diagnosis method based on improved convolutional neural network for bearings under variable working conditions *Measurement* **182** 109749

[139] Zhao W., Wang Z., Cai W., Zhang Q., Wang J., Du W., Yang N. and He X. 2022 Multiscale inverted residual convolutional neural network for intelligent diagnosis of bearings under variable load condition *Measurement* **188** 110511

[140] Wang J., Wu J. and Xiao H. 2017 Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data *Physical Review Fluids* **2** 34603

[141] Singh A. P., Medida S. and Duraisamy K. 2016 Machine Learning-augmented Predictive Modeling of Turbulent Separated Flows over Airfoils *Aiaa Journal* DOI:10.2514/1.J055595

[142] Karpatne A., Atluri G., Faghmous J. H., Steinbach M., Banerjee A., Ganguly A., Shekhar S., Samatova N. and Kumar V. 2017 Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data *IEEE Transactions On Knowledge and Data Engineering* **29** 2318-31

[143] Raissi M., Perdikaris P. and Karniadakis G. E. 2019 Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential

equations *Journal of Computational Physics* **378** 686-707

[144] He C., Shi H., Si J. and Li J. 2023 Physics-informed interpretable wavelet weight initialization and balanced dynamic adaptive threshold for intelligent fault diagnosis of rolling bearings *Journal of Manufacturing Systems* **70** 579-92

[145] Ruan D., Wang J., Yan J. and Gühmann C. 2023 CNN parameter design based on fault signal analysis and its application in bearing fault diagnosis *Advanced Engineering Informatics* **55** 101877

[146] Sadoughi M. and Hu C. 2018 A Physics-Based Deep Learning Approach for Fault Diagnosis of Rotating Machinery *2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA* pp 5919-23

[147] Kim Y. and Kim Y. 2024 Physics-Informed Time-Frequency Fusion Network With Attention for Noise-Robust Bearing Fault Diagnosis *IEEE Access* **12** 12517-32

[148] Wen L., Su S., Li X., Ding W. and Feng K. 2024 GRU-AE-wiener: A generative adversarial network assisted hybrid gated recurrent unit with Wiener model for bearing remaining useful life estimation *Mechanical Systems and Signal Processing* **220** 111663

[149] He M. and He D. 2020 A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals *Neurocomputing* **396** 542-55

[150] Sun D., Li Y., Liu Z., Jia S. and Noman K. 2024 Physics-inspired multimodal machine learning for adaptive correlation fusion based rotating machinery fault diagnosis *Information Fusion* **108** 102394

[151] Chen Z., Qin W., He G., Li J., Huang R., Jin G. and Li W. 2023 Explainable Deep Ensemble Model for Bearing Fault Diagnosis Under Variable Conditions *IEEE Sensors Journal* **23** 17737-50

[152] Yang S., Tang B., Wang W., Yang Q. and Hu C. 2024 Physics-informed multi-state temporal frequency network for RUL prediction of rolling bearings *Reliability Engineering & System Safety* **242** 109716

[153] Deng Y., Lv J., Huang D. and Du S. 2023 Combining the theoretical bound and deep adversarial network for machinery open-set diagnosis transfer *Neurocomputing* **548** 126391

[154] Yucesan Y. A. and Viana F. A. C. 2022 A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation *Mechanical Systems and Signal Processing* **171** 108875

[155] Shen S., Lu H., Sadoughi M., Hu C., Nemani V., Thelen A., Webster K., Darr M., Sidon J. and Kenny S. 2021 A physics-informed deep learning approach for bearing fault detection *Engineering Applications of Artificial Intelligence* **103** 104295

[156] Kim S., Choi J. and Kim N. 2022 Data-driven prognostics with low-fidelity physical information for digital twin: physics-informed neural network *Structural and Multidisciplinary Optimization* **65** 255

[157] Yang B., Lei Y., Jia F. and Xing S. 2019 An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings *Mechanical Systems and Signal Processing* **122** 692-706

[158] Chen X., Ma M., Zhao Z., Zhai Z. and Mao Z. 2022 Physics-informed Deep Neural Network for Bearing Prognosis with Multi-sensory Signals *Journal of Dynamics, Monitoring and Diagnostics* **1** 200-207

[159] Lu W., Wang Y., Zhang M. and Gu J. 2024 Physics guided neural network: Remaining useful life prediction of rolling bearings using long short-term memory network through dynamic weighting of degradation process *Engineering Applications of Artificial Intelligence* **127** 107350

[160] Chen Y., Zhang H., Li X., Shi Z. and Gu F. 2025 Tribo-dynamic performance analysis of engine main bearings with provision for wear progression and dynamic misalignment *Engineering Failure*

Analysis **169** 109196

[161] Feng Y., Shi X. J., Lu X. Q., Sun W., Liu K. P. and Fei Y. F. 2025 Predictions of friction and wear in ball bearings based on a 3D point contact mixed EHL model *Surface and Coatings Technology* **502** 131939

[162] Belaid S., Lecheb S., Chellil A., Mechakra H., Safi B. and Kebir H. 2022 Crack Growth Diagnostic of Ball Bearing Using Vibration Analysis *International Journal of Applied Mechanics and Engineering* **27** 35-45

[163] Wang J., Liang Y., Zheng Y., Gao R. X. and Zhang F. 2020 An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples *Renewable Energy* **145** 642-50

[164] Gebraeel N. Z., Lawley M. A., Li R. and Ryan J. K. 2005 Residual-life distributions from component degradation signals: A Bayesian approach *Iie Transactions* **37** 543-57

[165] Sun L., Zhao J., Yuan C., Peng Z., Zhou H., Ren X., Li L. 2023 RUL Prediction of Rolling Bearing Based on Fusion Feature of Time Domain and Spectral Kurtosis and Exponential Model *Machine Tool & Hydraulics* **51** 203-9

[166] Yang Z. and Liu J. 2021 Remaining life prediction of wind turbine bearing based on Wiener process *Acta Energiae Solaris Sinica* **42** 189-94

[167] König F., Wirsing F., Jacobs G., He R., Tian Z. and Zuo M. J. 2024 Bayesian inference-based wear prediction method for plain bearings under stationary mixed-friction conditions *Friction* **12** 1272-82

[168] Zhao Y., Toothman M., Moyne J. and Barton K. 2022 An Adaptive Modeling Framework for Bearing Failure Prediction *Electronics* **11** 257

[169] Jiang L., Sheng H., Yang T., Tang H., Li X. and Gao L. 2023 A New Strategy for Bearing Health Assessment with a Dynamic Interval Prediction Model *Sensors* **23** 7696

[170] Wang X., Cui L. and Wang H. 2022 Remaining Useful Life Prediction of Rolling Element Bearings Based on Hybrid Drive of Data and Model *IEEE Sensors Journal* **22** 16985-93

[171] Zhan Z., Fang B., Wan S., Bai Y., Hong J. and Li X. 2024 Thermal characterization of the spindle-bearing system under different working conditions based on a hybrid-driven framework combining data-driven and model-based methods *Journal of Manufacturing Processes* **118** 1-14

[172] Zhang B., Sconyers C., Byington C., Patrick R., Orchard M. E. and Vachtsevanos G. 2011 A Probabilistic Fault Detection Approach: Application to Bearing Fault Detection *IEEE Transactions On Industrial Electronics* **58** 2011-8

[173] Huang C., Yin X., Huang H. and Li Y. 2020 An Enhanced Deep Learning-Based Fusion Prognostic Method for RUL Prediction *IEEE Transactions On Reliability* **69** 1097-109

[174] Deng Y., Du S., Jia S., Zhao C. and Xie Z. 2020 Prognostic study of ball screws by ensemble data-driven particle filters *Journal of Manufacturing Systems* **56** 359-72

[175] Qian Y. and Yan R. 2015 Remaining Useful Life Prediction of Rolling Bearings Using an Enhanced Particle Filter *IEEE Transactions On Instrumentation and Measurement* **64** 2696-707

[176] Liu H., Yuan R., Lv Y., Li H., Gedikli E. D. and Song G. 2022 Remaining Useful Life Prediction of Rolling Bearings Based on Segmented Relative Phase Space Warping and Particle Filter *IEEE Transactions On Instrumentation and Measurement* **71** 1-15

[177] Karniadakis G. E., Kevrekidis I. G., Lu L., Perdikaris P., Wang S. and Yang L. 2021 Physics-informed machine learning *Nature Reviews Physics* **3** 422-40

[178] Deng W., Khanh N., Gogu C., Morio J. and Medjaher K. 2022 Physics-informed Lightweight Temporal Convolution Networks for Fault Prognostics Associated to Bearing Stiffness Degradation *Phm*

Society European Conference **7** 118-25

[179] Gitzel R., Kotriwala A., Fechner T. and Schiefer M. 2021 Using Synthetic Data to Train a Visual Condition Monitoring System for Leak Detection *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (Big Data Service), Oxford, United Kingdom* pp 196-200

[180] Sivarajah U., Kamal M. M., Irani Z. and Weerakkody V. 2017 Critical analysis of Big Data challenges and analytical methods *Journal of Business Research* **70** 263-86

[181] Yu Q., Li Q., Li L. and Wang Y. 2024 A survey of data-driven fault diagnosis methods for large-scale industrial production processes *Chinese Journal of Engineering* DOI: 10.13374/j.issn2095-9389.2024.05.24.002

[182] Shi P., Yu Y., Gao H. and Hua C. 2022 A novel multi-source sensing data fusion driven method for detecting rolling mill health states under imbalanced and limited datasets *Mechanical Systems and Signal Processing*

[183] Pan H., Li B., Zheng J., Tong J., Liu Q. and Deng S. 2024 Research on roller bearing fault diagnosis based on robust smooth constrained matrix machine under imbalanced data *Advanced Engineering Informatics* **62** 102667

[184] Wang D., Dong Y., Wang H. and Tang G. 2023 Limited Fault Data Augmentation With Compressed Sensing for Bearing Fault Diagnosis *IEEE Sensors Journal* **23** 14499-511

[185] Yin L. and Wang Z. 2024 Parallel quantized dual-level fully connected classifier for bearing fault diagnosis *Engineering Applications of Artificial Intelligence* **136** 109052

[186] Sun B., Li Y., Zhang Y. and Guo T. 2024 Multi-source heterogeneous data fusion prediction technique for the utility tunnel fire detection *Reliability Engineering & System Safety* **248** 110154

[187] Han D., Zhang Y., Yu Y., Tian J. and Shi P. 2024 Multi-source heterogeneous information fusion fault diagnosis method based on deep neural networks under limited datasets *Applied Soft Computing* **154** 111371

[188] Sun X., Zhou K., Shi S., Song K. and Chen X. 2022 A new cyclical generative adversarial network based data augmentation method for multiaxial fatigue life prediction *International Journal of Fatigue* **162** 106996

[189] Zhou K., Diehl E. J. and Tang J. 2023 Deep convolutional generative adversarial network with semi-supervised learning enabled physics elucidation for extended gear fault diagnosis under data limitations *Mechanical Systems and Signal Processing*

[190] Qin Y., Zhou J. and Chen D. 2022 Unsupervised Health Indicator Construction by a Novel Degradation-Trend-Constrained Variational Autoencoder and Its Applications *IEEE/ASME Transactions On Mechatronics* **27** 1447-56

[191] Yan X., She D., Xu Y. and Jia M. 2021 Deep regularized variational autoencoder for intelligent fault diagnosis of rotor–bearing system within entire life-cycle process *Knowledge-Based Systems* **226** 107142

[192] Kim Y., Choi Y., Yoo B. 2024 Gappy AE: A nonlinear approach for Gappy data reconstruction using auto-encoder *Computer Methods in Applied Mechanics and Engineering* **426** 116978

[193] Wang X., Li D., Zhao J., Cao Z. and Weng W. 2023 Indoor environment reconstruction algorithm based on gappy POD and finite sensors *Energy and Buildings* **297** 113463

[194] Wang Q., Yang C., Wan H., Deng D. and Nandi A. K. 2021 Bearing fault diagnosis based on optimized variational mode decomposition and 1D convolutional neural networks *Measurement Science and Technology* **32** 104007