


REVIEW

Open Access



Challenges and Opportunities of AI-Enabled Monitoring, Diagnosis & Prognosis: A Review

Zhibin Zhao, Jingyao Wu, Tianfu Li, Chuang Sun, Ruqiang Yan*  and Xuefeng Chen

Abstract

Prognostics and Health Management (PHM), including monitoring, diagnosis, prognosis, and health management, occupies an increasingly important position in reducing costly breakdowns and avoiding catastrophic accidents in modern industry. With the development of artificial intelligence (AI), especially deep learning (DL) approaches, the application of AI-enabled methods to monitor, diagnose and predict potential equipment malfunctions has gone through tremendous progress with verified success in both academia and industry. However, there is still a gap to cover monitoring, diagnosis, and prognosis based on AI-enabled methods, simultaneously, and the importance of an open source community, including open source datasets and codes, has not been fully emphasized. To fill this gap, this paper provides a systematic overview of the current development, common technologies, open source datasets, codes, and challenges of AI-enabled PHM methods from three aspects of monitoring, diagnosis, and prognosis.

Keywords: Monitoring, Diagnosis, Prognosis, PHM, Artificial intelligence, Deep learning

1 Introduction

As the key ingredient in the modern industry, mechanical equipment, such as helicopters, high-speed rail, aero engines, etc., is chronically operating in an increasingly harsh environment and its structure is becoming increasingly complex as well, which may result in sudden equipment failure, long maintenance cycles, high maintenance costs, and large downtime losses. Different from traditional maintenance methods (corrective maintenance and periodical maintenance), Prognostic and Health Management (PHM) uses the integration of advanced sensors as well as various intelligent approaches to monitor the status of the mechanical system, which realizes timely and optimal maintenance via reducing manual labor, spares, and maintenance cost.

PHM mainly consists of monitoring, diagnosis, prognosis, and health management [1, 2], whose relationships are summarized in Figure 1. Monitoring refers to fault detection, and the purpose is to determine whether the

system is in a normal operating state, in which anomaly detection is one of the most important tools to trace the corresponding health state. Diagnosis refers to the identification of the fault type and its corresponding degree. Prognosis makes use of appropriate models to assess the degree of performance degradation and further predicts the remaining useful life (RUL). Health management integrates outputs from monitoring, diagnosis, and prognosis and makes optimal maintenance and logistic decisions via considering economic costs and other available resources. In general, PHM will greatly improve the operational safety, system reliability, and maintainability of equipment, and reduce the cost of equipment throughout its life cycle at the same time.

Traditional maintenance methods generally rely on experts to observe and diagnose equipment artificially and determine the fault type and its location by reasonably mounting the sensors and analyzing the result using appropriate algorithms. This type of method increases manual labor, and the efficiency of maintenance largely depends on expert experience. With the development of the sensor technology, a large number of sensors are installed on mechanical equipment to

*Correspondence: yanruqiang@xjtu.edu.cn

The State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

collect multi-source data, including vibration, temperature, images, etc., which provides a base preparation for potential implementation of PHM. However, due to the fact that mechanical equipment is chronically operating in an extremely complex environment, the measured signal often contains heavy background noise, that is, the related fault features are often submerged in the interference. Traditional signal processing methods, such as FFT (fast Fourier transform) and simple metric construction, cannot extract and analyze the feature information with high efficiency and precision. Advanced signal processing methods, such as sparse representation [3, 4] and time-frequency analysis [5], often have some parameters that need to be adjusted carefully, resulting in huge workload. With the development of big data techniques and artificial intelligence (AI) algorithms, AI-enabled PHM is becoming increasingly popular and has already achieved wide success in both academia and industry. The main superiority of AI-enabled PHM is that it can perform monitoring, diagnosis, and prognosis at a high level of automation, and requires little intervention and expert knowledge.

AI-enabled PHM is mainly about using traditional machine learning (ML) or deep learning (DL) methods to perform the final health management. Traditional ML algorithms such as K-nearest neighbor (KNN), artificial neural network (ANN), support vector machine (SVM), etc., have been successfully applied to PHM and also have achieved considerable progress. However, their applications, to a large extent, still depend on hand-crafted feature extraction. As long as the extracted features can represent fault features effectively, traditional ML models can also establish the mapping between features and the mechanical health status successfully. However, hand-crafted feature extraction still relies on expert knowledge, which also differs considerably between different signals or equipment. Moreover, when handling massive heterogeneous data, these methods based on hand-crafted features are obviously time-consuming, and

such experience-oriented methods are easy to drop their accuracy in the context of big data. Therefore, it remains a challenging problem about how to establish AI-enabled PHM with high efficiency and precision. Since Hinton et al. [6] first proposed and realized a DL model in 2006, DL has become a subversive technology in AI. DL has achieved a significant breakthrough and extensive applications in a wide range of fields, especially computer vision and natural language processing. In 2015, Nature organized a special issue to deeply summarize the development process of AI and took DL as one of the six breakthrough technologies in this field [7]. Because of its strong representation learning ability, DL is very suitable for automatic data analysis, which can establish the mapping from the data side to the task side via learning the representation features automatically from a large number of data. Consequently, the application of DL in PHM is becoming increasingly popular because of providing a technology with the potential to process a large number of data, extract features from high-dimension data, and form an “end-to-end” monitoring, diagnosis, and prognosis system automatically.

To explain the popularity of AI-enabled PHM, we conducted a literature search using Web of Science with a database called the web of science core collection in the past five years. It is worth mentioning that it is impossible to cover all the related papers because the names of AI-enabled algorithms are often different. As shown in Figure 2, we can observe that research about AI-enabled PHM has increased rapidly and it is of great importance to embed AI into PHM. To summarize the research of AI-enabled PHM, many scholars published their review papers from a different angle. Hamadache et al. [8]

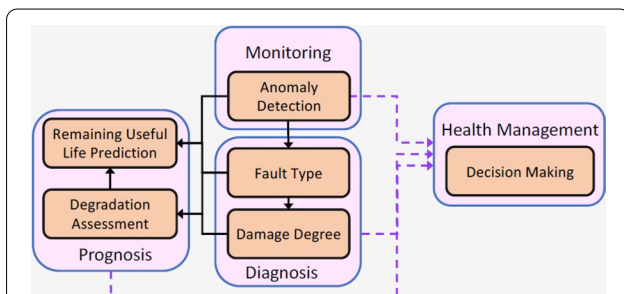


Figure 1 Relationship between monitoring, diagnosis, prognosis, and health management

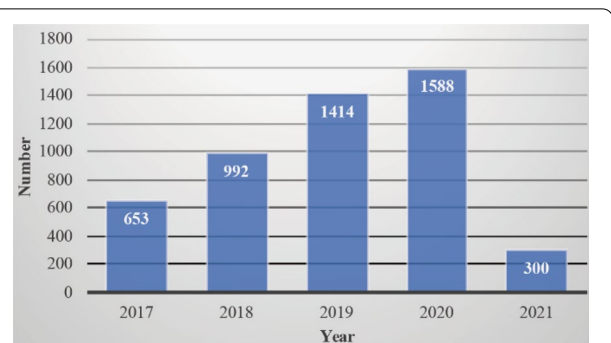


Figure 2 Relationship between the number of published papers and publication years covering the last five years (as of March 2021). The basic descriptor is \TI= ((AI OR artificial intelligence OR machine learning OR support vector machine OR SVM OR data-driven OR deep OR autoencoder OR convolutional network* OR neural network*) AND (fault detection OR fault isolation OR fault diagnosis OR intelligent diagnosis OR prognosis OR residual useful life prediction OR condition monitoring OR health management))

introduced the fundamentals of PHM techniques for rolling element bearings (REBs) and reviewed contemporary techniques including modern AI techniques and DL approaches for fault detection, diagnosis, and prognosis of REBs. Lee et al. [9] detailed previous and on-going efforts in PHM for rotary machinery systems, and introduced a systematic PHM design methodology and its applications. Ellefsen et al. [10] reviewed four DL-based techniques applied to PHM for autonomous and semi-autonomous ships. However, these reviews mainly discussed the applications of PHM for a specific object. Liu et al. [11] and Lei et al. [12] reviewed applications of AI techniques to machine fault diagnosis. Lei et al. [13] and Zhang et al. [14] reviewed recent advances on machinery prognostics systematically. These papers mainly focused on one aspect of monitoring, diagnosis, and prognosis. Khan et al. [15] and Fink et al. [1] provided a systematic review of the DL and its applications in PHM. However, there is neither emphasis on the importance of anomaly detection in monitoring nor the summary of open source datasets and codes in these papers. Therefore, it needs a review to cover monitoring, diagnosis, and prognosis based on AI techniques with the emphasis on DL and the requirement of an open source community.

To fill the aforementioned gap, this paper systematically reviews the current development, common technologies, open source datasets, codes, and challenges of AI-enabled PHM from three aspects of monitoring, diagnosis, and prognosis. We focus on the applications of AI-enabled algorithms, especially DL in monitoring, diagnosis, and prognosis. More importantly, we emphasize the importance of open source datasets and codes for the benign development of the research community of AI-enabled PHM. Last but not least, this paper provides some promising future directions in the field of AI-enabled PHM. It is worth mentioning that this review paper does not cover another important part of PHM, e.g., health management.

2 AI-Enabled Monitoring

2.1 Introduction to AI-Enabled Monitoring

As the key and basic task of PHM, monitoring of machinery has not attracted enough attention, as shown in Figure 3. What is more, many existing studies about monitoring are based on the supervised methods [16–19]. It means that both normal data and anomaly data are required for model training, which is usually inconsistent with the scenario of monitoring, since anomaly data with faults is not always available, and the form and location of the failure are even unknown. Thus, methods relying on the existing faults would fail when confronting the new fault, which would result in the catastrophic missing

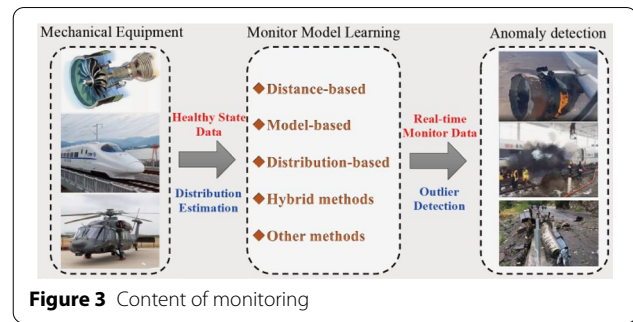


Figure 3 Content of monitoring

alarm. In this section, we specially define the monitoring task as anomaly detection and review the related papers.

2.2 Anomaly Detection

The generalized concept of anomaly detection can be divided into three categories, including supervised learning, semi-supervised learning, and unsupervised learning. As described in the previous sub-section, supervised learning methods are not suitable for the monitoring task, and data from the healthy state is often available. So in this paper, anomaly detection refers to semi-supervised anomaly detection in the following discussion, specifically.

Semi-supervised anomaly detection can be regarded as a one-class classification problem, which means that only the data at the health state is available. The goal of anomaly detection is to detect the fault that may occur in the future based on the existing data. The failure would occur on any component of the machinery with any external manifestation, so it is a typical open-set task.

According to the different strategies of anomaly determination, anomaly detection methods for monitoring are divided into three categories in this paper, including distance-based methods, model-based methods, distribution-based methods (also called density-based methods), hybrid methods, and others. We will review these methods in the following subsections.

2.2.1 Distance-Based Methods

Distance-based methods pay attention to the distance between data collected on the anomaly state and health state. The distance to be measured is calculated in the signal space or in the latent feature space after feature extraction. It is based on the assumption that collected data at the health state would be close to each other in the signal space or in latent feature space, and the collected data at the anomaly state would be naturally far away from the former data. Various metrics can be applied for distance calculation, including Euclidean distance, Manhattan distance, cosine distance, Chebyshev distance, etc.

Meanwhile, to consider the contributions of different features to the distance calculation and the compactness of the feature space, plentiful pre-processing and representation learning strategies can also be applied before the distance calculation.

In Ref. [20], a comb filtering was applied to smooth the original signal, Gini-guided residual singular value decomposition and Principle Component Analysis (PCA) were used for feature extraction, and then iterative Mahalanobis distance was calculated to obtain an anomaly score. In Ref. [21], Short-Time Fourier Transform (STFT), Hidden Markov Model (HMM), and dimension reduction were applied to the original signal for feature extraction and a distance-based strategy was used for anomaly score calculation. In Ref. [22], a classical one-class classification method, Support Vector Data Description (SVDD) was applied together with a Genetic Algorithm (GA) for parameter optimization. An improved SVDD with artificially generated outliers was proposed for rolling element bearings detection [23]. Multi-sensor data was utilized in Ref. [24] with a correlation-based anomaly detection method for predictive maintenance. In Ref. [25], self-organizing map and KNN were used for cooling fan bearing monitoring. In Ref. [26], the K-means cluster method was utilized to obtain cluster center points, and an anomaly score was calculated based on the distance from center points. In Ref. [27], one class SVM was utilized for kinematic chain monitoring using data processed by Laplacian score ranking guided features selection.

2.2.2 Model-Based Methods

Model-based methods try to establish a prediction model to reflect the intrinsic regularity between parameters or on the timeline based on the health state data. It is assumed that, when an internal or external failure occurs on equipment, the intrinsic regularity of the data would deviate from the original model. So the occurrence of anomaly can be represented by the degree of deviation between the model prediction and actual data.

In Ref. [28], a Long Short-Term Memory (LSTM) model was trained based on features extracted by Stacked AutoEncoders (SAE) to predict the vibration signal in next N time steps. The residual between predicted and actual signals was utilized to indicate the occurrence of anomaly for rotary machinery. In Ref. [29], Generative Adversarial Networks (GAN) was trained to discriminate the fake data from real data, and the output of the discriminator was further regarded as the anomaly indicator. Similarly, the reconstruction model based on LSTM and GAN was trained with normal data in Refs. [30, 31], and the anomaly score was influenced by the output of the discriminator and the reconstruction error

simultaneously. In Ref. [32], AutoEncoders (AE) based GAN was trained to generate artificial normal data. Then the test data was fed into AE to get reconstructed latent features and reconstructed signals. Finally, an anomaly score was calculated based on the reconstruction error. In Refs. [33, 34], the Yet Another Segmentation Algorithm (YASA) was utilized for data segmentation, and the segmentation results were fed into one-class SVM for offshore oil extraction turbo machine anomaly detection. In Refs. [35, 36], a LSTM prediction model was trained, and the residual between predictions and actual signals was utilized as an anomaly score for water treatment system and aircraft anomaly detection, respectively. In Refs. [37–39], an AE model for data reconstruction was built and the reconstruction error was used as the anomaly indicator. In Ref. [40], a rotary speed to vibration regression was trained to predict the vibration signal, and the residual was further used for anomaly detection. A comparison between autoregressive-based models and network-based models was implemented in Ref. [41] for wind turbine fault detection. In Ref. [42], the HMM model was trained for screw compressors anomaly detection after dimension reduction by PCA. In Refs. [43, 44], the autoregressive integrated moving average (ARIMA) process was proposed for data prediction and anomaly detection with multi-sensors. In Ref. [45], the AE model was trained with data processed by the series-to-image transform for anomaly detection.

2.2.3 Distribution-Based Methods

Distribution-based methods (density-based methods) try to estimate the distribution of normal data. It is assumed that anomaly data would be subject to a different distribution from normal data. So if the anomaly data is input into the distribution model, a low probability will be obtained. This type of method can also be understood as density-based methods. Anomaly events can be regarded as low-probability events and they exhibit low density characteristics in the sample space. Therefore, without distribution model estimation, the sample density of the test data can also represent the probability of an anomaly.

In Ref. [46], a multivariate Gaussian distribution model was built after a set of feature extraction processes, including Hodrick-Prescott (HP) filtering and Gradient of Change (GoC), for anomaly detection. The Gaussian distribution was also used in Ref. [47] after feature extraction with SAE for gas turbine engine gas path anomaly detection. Generalized Extreme Value (GEV) distribution was applied in Ref. [48] for power generation monitoring. The martingale-test was performed in Ref. [49] to detect the change point of gearboxes based on the graph model. In Ref. [50], correlation coefficients of segmented signals were calculated and the derivation of the anomaly score

was based on the Probability Density Function (PDF) of correlation coefficients.

2.2.4 Hybrid Methods

In order to break the limitation of a single algorithm, some other research took the advantages of the above methods and constructed hybrid methods for anomaly detection.

In Refs. [51, 52], model-based and distance-based methods were combined for anomaly detection. In Ref. [51], a PCA matrix was obtained for dimension reduction based on the normal data, and the residual of PCA was calculated as the input of the SVDD method for fault detection. Similarly, the model based on AE and LSTM was combined with SVDD in Ref. [52] for bearings initial fault detection. In Refs. [53, 54], model-based and distribution-based methods were combined for anomaly detection. In Ref. [53], the probability of the anomaly state was defined as the combination of the reconstruction error and the latent feature with an AE model, and a fault-attention factor was implied to re-weight the anomaly score. In Ref. [55], a set of anomaly detection methods were compared, including Gaussian Mixture Model (GMM), Parzen window density estimation, Local Outlier Factor (LOF), k-means clustering, PCA-based methods, and SVDD-based methods. In Ref. [56], a Hierarchical Temporal Memory (HTM) model was built and the distribution of the model prediction error was estimated for real-time, continuous, online detection of streaming data.

2.2.5 Others Methods

Different from methods we described above, some researchers paid more attention to representation learning or the relationship between equipment groups and provided novel perspectives for condition monitoring.

In Ref. [57], advantages of equipment groups were taken by utilizing a clustering algorithm on electrical machine fleet after domain specific pre-processing. It is similar to the density-based methods, the clustering objects changed from the samples of time scale to the different equipment of space scale. But it also put forward a high requirement for the consistency of operation states of different equipment in the group. In Ref. [58], a non-parametric k^2 decomposition method was used to isolate the fault from multi-variate processes by measuring the relative contribution of an individual variable. In Ref. [59], an AR-based model was proposed to detect the condition change between adjacent periods, but this method was not applicable to non-stationary working conditions. In Ref. [60], speed-energy spectra of the jet engine were calculated to reflect the operation state, and the difference between spectra was regarded as the

anomaly indicator. In Ref. [61], a hybrid feature selection method based on ReliefF and an adaptive GA was proposed, and recursive one-class SVM was trained after pre-processing with Extended Kalman Filter (EKF) to realize an online undated detection for chillers. In Ref. [62], dictionary learning was used, and the change of the dictionary was monitored for condition monitoring of rotating machinery.

2.3 Open Source Datasets and Codes

2.3.1 Date Type Summary

Among different monitoring scenarios, data used for anomaly detection is diverse. Although the photograph of equipment can directly reflect the healthy condition of equipment, it is hard to obtain surface photographs of all the components of equipment and most of the key components are invisible due to the complexity of the structure. A better alternative is the vibration signal of the rotating machinery, in which the global health information for entire equipment can be obtained simultaneously. Meanwhile, vibration signals are also the most commonly used data in recent research work. Besides, temperature signals [24, 51, 61], electric current signals [25, 57], sound signals [49, 59], pressure signals [24], speed signals [40, 41, 63], voltage signals [64], and acoustic emission signals [62] are also used as the source data for monitoring. What is more, parameters related to equipment operation conditions are also used in many large scale equipment [29, 31, 36, 47].

2.3.2 Open Source Datasets

Most of the methods proposed above are evaluated on real industrial data, but there is no way for performance comparison due to the privacy of datasets. So open source datasets are necessary for performance comparison between different methods. Here, we summarize a list of open source datasets utilized in existing papers.

Many researchers used fault classification datasets, such as CWRU datasets [65], Tennessee Eastman Process datasets [66], and SEU datasets [67], for anomaly detection. The categories with the healthy condition are regarded as normal data and the categories with fault occurring are regarded as anomaly data. The run-to-failure datasets, such as IMS datasets [68] and PHM2012 datasets [69] were also used in condition monitoring task with artificial division of normal and anomaly states based on the degradation state of components.

There are also some datasets specially made for monitoring. For example, Airbus Helicopter Accelerometer dataset [70] used in Ref. [45] was collected by Airbus SAS with vibration measurements of helicopter in different directions (longitudinal, vertical, and lateral). Numanta Anomaly Benchmark [71] with over 50 labeled real-world

and artificial time series data files was used in Ref. [56]. Secure Water Treatment (SWaT) dataset [72] was established for the research of the protection of Cyber Physical Systems (CPS) such as those for water treatment, power generation and distribution, and oil and natural gas refinement, and was widely used for the performance evaluation of anomaly detection method. The details of these datasets are listed in Table 1.

2.3.3 Open Source Codes

Although a large number of methods have been proposed for monitoring, a few source codes of these methods are publicly available, which is not conducive to the sustainability of research. In this subsection, we summarize the online available codes of related papers to provide a convenient way for researchers to get started in this field and more open source studies are also required in the future.

The whole code of Fault-attention Generative Probabilistic Adversarial Autoencoder based on Pytorch framework proposed by Ref. [53] for anomaly detection with SEU datasets was released online in Ref. [73]. Numenta Platform for Intelligent Computing (NuPIC), proposed by Ref. [56], based on HTM learning algorithms for anomaly detection and prediction of streaming data sources was online available in Ref. [74]. Multivariate Anomaly Detection with GAN (MAD-GAN) framework proposed by Ref. [31] was also online available in Ref. [75].

2.4 Challenges

All of the above methods are more or less based on certain assumptions, but in real-world applications, these assumptions would fail due to the inherent characteristics of equipment or the complexity of the external environment. To realize reliable and accurate monitoring and abnormal early warning in industrial applications, more realistic challenges need to be considered. The main challenges hindering the implementation of anomaly detection in reality can be summarized as follows.

2.4.1 Balance between Recall and Precision

During the construction of anomaly detection models, the choice of the decision threshold or the decision boundary is an inevitable process. It is essentially a trade-off between recall and precision. Low recall would lead to the omission of fault and cause catastrophic damages. Low precision would cause excessive and unnecessary maintenance and inspection, and the associated cost of condition monitoring will increase. Therefore, the optimal threshold can be chosen via comparing the cost of omission and the cost of over-maintenance and minimizing the expected cost.

2.4.2 Unified Benchmark Datasets

Although there are many datasets which can be used for anomaly detection research, benchmark datasets specially designed for AI-enabled monitoring are still required. Unified benchmark datasets can provide a normalized data processing flow and make performance comparisons between different models more convenient. Meanwhile, realistic anomaly detection datasets for monitoring can make related research based on these datasets practically significant.

2.4.3 Quick Alarm for Early Failure

In the stage of early failure for machinery, only weak fault features can be shown in collected data and there is little difference between the signal characteristics under the health state and the fault state. To quickly detect early failure and avoid the fault extension, the sensibility of methods for early failure is of the utmost importance. The existing methods do not take this perspective into consideration.

2.4.4 Adaptability under Variable Working Conditions

In reality, the operating condition of equipment, such as rotary speeds, loads, and temperatures, is variable and the corresponding data characteristics would also change accordingly. An ideal monitoring algorithm should be robust for different operating conditions. However, due

Table 1 Open source datasets for condition monitoring

Dataset Name	Research object	Description
Airbus helicopter accelerometer [70]	Helicopter	Data were acquired from different accelerometers placed at different positions of the helicopter, including 1677 training sequences from normal flights and 594 sequences from normal and anomalous flights
Numenta anomaly benchmark [71]	Various objects	It is composed of over 50 labeled real-world and artificial time series data sets plus a novel scoring mechanism designed for real-time applications
SWaT [72]	CPS	Data were acquired from a water treatment testbed for research in the area of cyber security, and the data systematically generated from the testbed over 7 days under normal operation and 4 days with attack scenarios is provided

to the complexity of equipment, it is impossible to train the model with data under all possible operating conditions. Thus, models are required to recognize unseen operating conditions out of the training sets and do not make a false alarm under unseen operating conditions.

3 AI-Enabled Diagnosis

3.1 Introduction to AI-Enabled Diagnosis

Fault diagnosis plays an important role in exploring the relationship between measured data and machine health states [76, 77], which has been the research hot-spot of PHM. Traditionally, the relationship is found via expert knowledge and engineering experience. However, in engineering scenarios, people would like to shorten the maintenance cycle and improve the diagnostic accuracy through an automatic method. Especially with the help of AI, fault diagnosis is expected to become smart enough to automatically detect and identify health states.

The AI-enabled diagnosis aims to diagnose health states with applications of ML theories, such as SVM [78, 79], ANN [80, 81], and deep neural networks (DNN) [82, 83]. These methods utilize ML theories to capture features hidden in measured data with less expert knowledge. It attempts to build a bridge that automatically detects health states from the collected data. In recent years, AI-enabled methods have swept the field of mechanical fault diagnosis [84, 85]. It has been widely used to solve problems in fault diagnosis, such as class imbalance, variable working conditions, and fault diagnosis under strong background noise. Therefore, in this section, we mainly review how the existing AI-enabled diagnostic methods solve these problems.

3.2 Diagnosis

3.2.1 Vanilla Fault Diagnosis

Generally, when we explore intelligent diagnosis algorithms, the used datasets will not encounter the issues of class imbalance, low signal-to-noise ratio, and variable operating conditions. We call the diagnosis under this situation vanilla fault diagnosis. AI-enabled diagnosis algorithms are mainly divided into two categories: traditional ML-based methods and DL-based methods. DL-based methods are closer to the expectations of automatic fault classification as shown in Figure 4. It can extract features automatically without human intervention, and can establish the relationship between the learned features and the fault patterns. Therefore, in the following section, we will mainly review DL-based algorithms that are widely used in fault diagnosis.

(1) AE-based methods

In the past five years, AE has made tremendous development in the field of PHM. AE has a strong ability to learn feature representation, and by inputting the extracted features into the classifier, fault diagnosis can be realized. For example, Lu et al. [86] introduced stacked denoising autoencoder (SDAE) to fault diagnosis, and Liu et al. [87] proposed a rolling bearing fault diagnosis method by using SAE to extract features and adopted a fully connected layer to classify the fault modes. Ma et al. [88] proposed a deep coupling AE to achieve multi-modal data fusion and fault diagnosis. Shi et al. [89] proposed a fault diagnosis method based on SAE, which integrated compression sensing and wavelet

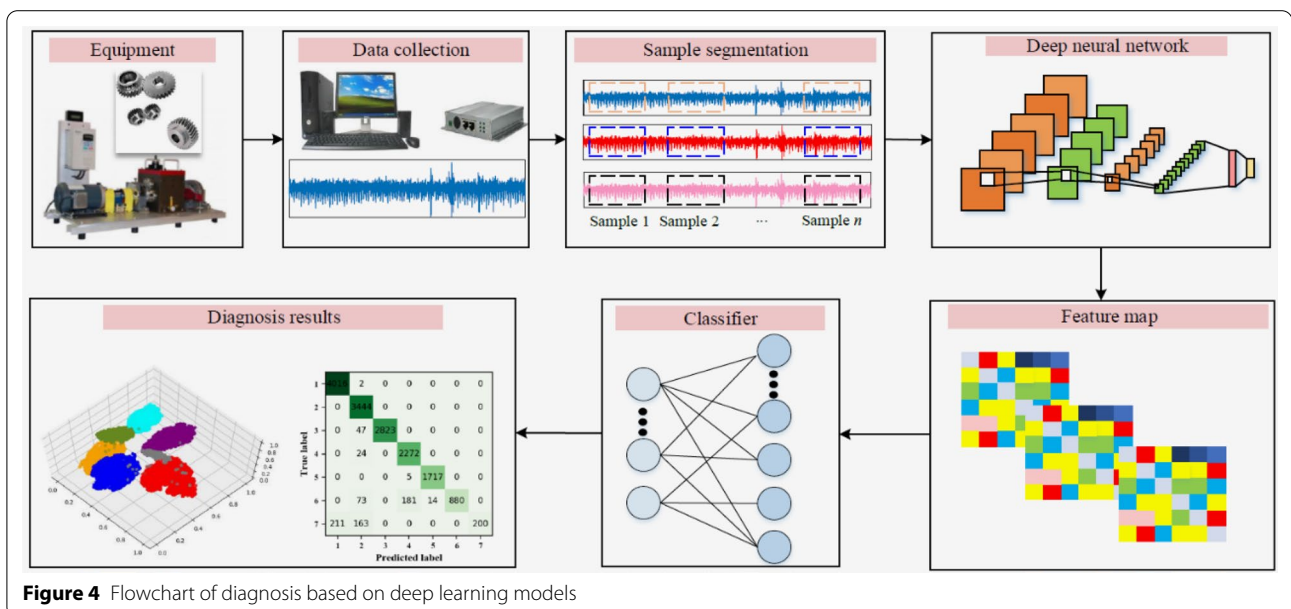


Figure 4 Flowchart of diagnosis based on deep learning models

packet energy entropy for feature dimension reduction.

(2) DBN-based methods

Deep belief network (DBN) has achieved great success in mechanical fault diagnosis in past five years. For example, Han et al. [90] used DBN and Teager-Kaiser energy operator for feature extraction, and then used the particle swarm optimization based SVM for bearing fault classification. Jiang et al. [91] proposed a feature fusion DBN for the intelligent fault diagnosis of rolling bearing. Shao et al. [92] proposed a continuous DBN with locally linear embedding for rolling bearing fault diagnosis. Wang et al. [93] proposed a hybrid method, including DBN and feature uniformization, and impulsive signals were used as the input to achieve real-time and accurate fault diagnosis. Gao et al. [94] proposed a self-adaptive optimized DBN, which was pre-trained by a mini-batch stochastic gradient descent. Zhang et al. [95] combined the advantages of DBN and variational mode decomposition (VMD) for rolling bearing fault diagnosis.

(3) CNN-based methods

Convolutional neural network (CNN) can directly learn features from the original monitoring data through sparse connection and weight sharing, which reduces the number of training parameters to accelerate convergence and suppress over-fitting. Many end-to-end frameworks based on CNN have been successfully constructed for fault diagnosis by using the vibration signal as the input. From the perspective of network inputs, CNN used in fault diagnosis can be classified into one-dimensional CNN (1D-CNN) and two-dimensional CNN (2D-CNN). Their applications in mechanical fault diagnosis are listed in the following Table 2, where the

category, motor, includes induction motors, permanent magnet synchronous motors, etc.

(4) RNN-based methods

Due to the one-way non-feedback connection of DNN, they cannot learn the temporal dependencies containing in the signal. Recurrent neural network (RNN) can store the data information (short-term memory) of the most recent periods in the form of excitation, which makes them suitable for processing time series.

For example, Liu et al. [138] proposed a low-speed lightweight RNN, which has a small storage space occupancy rate and low calculation delay. Miki et al. [139] proposed a LSTM-based method for time-series analysis and a training method for weakly supervised training. Rao et al. [140] proposed a many-to-many-to-one bi-directional LSTM to automatically extract the rotating speed from vibration signals. Shao et al. [141] proposed a method based on the enhanced deep gated recurrent unit and the complex wavelet packet energy moment entropy for early fault diagnosis of bearings. Shi et al. [52] presented a fault diagnosis framework based on SDAE and LSTM, which can effectively detect initial anomalies of rolling bearing and accurately describe the deterioration trend. To improve the diagnostic accuracy, Zhang et al. [142] presented an attention-based equitable segmentation gated recurrent unit network, which consists of an equitable segmentation approach and an improved deep model.

3.2.2 Fault Diagnosis under Imbalanced Dataset

During the operation of the machine, the collected datasets are often highly imbalanced, which contain many samples in the normal state but a paucity of samples from the fault state. Facing the imbalanced datasets, intelligent fault diagnosis approaches are biased towards the major

Table 2 Publications about the application of CNN in fault diagnosis

Object	Architectures	References
Bearing	1D-CNN	Chuya et al. [96], Eren et al. [97], Gonzalez-Muniz et al. [98], Li et al. [99], Wang et al. [100], Yin et al. [101], Zhang et al. [102]
	2D-CNN	Gong et al. [103], Guo et al. [104], Huang et al. [105], Khodja et al. [106], Li et al. [107], Wang et al. [108], Wang et al. [109], Wang et al. [110], Xu et al. [111], Zhang et al. [112], Islam et al. [113], Kumar et al. [114]
Gear	1D-CNN	Liang et al. [115], Wu et al. [116], Zhao et al. [117], Li et al. [118]
	2D-CNN	Chen et al. [119], Chen et al. [120], Han et al. [121], Li et al. [122], Qiu et al. [123], Xin et al. [124], You et al. [125], Tang et al. [126]
Motor	1D-CNN	Ince et al. [127], Kao et al. [128]
	2D-CNN	Hsueh et al. [129], Lee et al. [130], Yang et al. [131], Liu et al. [132], Wen et al. [133], Shao et al. [134]
Turbine	1D-CNN	Zhong et al. [135], Chang et al. [136]
	2D-CNN	Fu et al. [137], Jiang et al. [85]

classes and hence show very poor classification accuracy on the minor classes.

From current research, there are three ways to solve this problem, including the data synthesis methods, designing a powerful feature extractor, and designing the corresponding loss function, which are summarized as follows.

(1) Data synthesis based methods

Data synthesis based methods are the most direct way for solving the class-imbalanced problem. Traditional data synthesis methods are SMOTE [143] and cost sensitivity based methods [144]. For example, Razavi-Far et al. [145] used an imputation-based oversampling technique for class-imbalanced learning and the proposed scheme was evaluated on three experimental scenarios with different imbalance ratios. Zhang et al. [146] adopted a weighted minority oversampling strategy to balance the data distribution, and used a data synthesis strategy to avoid generating incorrect or unnecessary samples.

Recently, GANs [147] have been applied to generate artificial data for the minor classes or for data augmentation [148, 149]. For example, Mao et al. [150] used GAN to generate synthetic samples for minority fault classes and improved the generalization ability of the fault diagnosis model. Luo et al. [151] proposed a conditional deep convolutional GAN. By using the conditional auxiliary generative samples as the input, fault diagnosis under the imbalanced dataset was achieved. Zhang et al. [152] utilized GAN to learn the mapping between the distributions of noise and real machinery temporal vibration data, and then used the generated samples to balance the minor classes. Wang et al. [153] used the encoder network of conditional variational AE to learn the distribution of fault samples, and then generated a large number of fault samples of minor classes through the decoder network. Zheng et al. [154] proposed a dual discriminator conditional GAN to learn data distributions from signals on multi-modal fault samples, and automatically synthesized realistic 1D signals of each fault. Besides, there are also other data synthesis models. For example, Zhou et al. [155] proposed a nonlinear auto-regressive neural network to synthesize the small number of samples.

(2) Powerful feature extractor based methods

Designing a powerful feature extractor via DNN models to extract the discriminant features from signals can also achieve fault diagnosis under the imbalanced dataset. For example, Zhao et al. [156]

proposed a normalized CNN to estimate the feature distribution difference and to diagnose the fault severity under the data imbalanced situation. Jia et al. [157] also used a normalized CNN model for the imbalanced fault classification of machinery and used a neuron activation maximization algorithm to explain what DNN learned. Zhao et al. [158] proposed a deep Laplacian AE to extract deep sensitive features and trained the model with a Laplacian regularization for rotating machinery fault imbalanced diagnosis. Lan et al. [159] used the weighted extreme learning machine as the feature extractor, and used the gravitational search algorithm to optimize the extractor to further extract the significant features.

(3) Corresponding loss function based method

Designing a proper loss function for DNN models is also a promising methodology for imbalanced fault diagnosis of machinery. For example, Li et al. [160] proposed an adaptive channel weighted CNN (ACW-CNN) and used Focal loss for condition monitoring of the helicopter transmission system function. With the help of Focal loss, the ACW-CNN could reduce the weight of easily classified categories and increase the weight of categories that were not easy to classify, so that the model could pay more attention to the minor classes. Xun et al. [161] proposed a deep cost adaptive CNN based intelligent classification method for imbalanced data, which used the cost adaptive loss function to adaptively assign different misclassification costs for all categories.

3.2.3 Fault Diagnosis under Variable Working Condition

Fault diagnosis under variable working conditions is still a challenge due to the domain discrepancy problem. To achieve fault diagnosis under variable working conditions, there are currently two widely adopted methods, that is, discriminative feature extraction based methods and transfer learning based methods.

(1) Discriminative feature extraction based methods

Designing a DNN model that can extract discriminative features is a common way for intelligent diagnosis under variable working conditions. For example, Peng et al. [162] proposed a multi-branch and multi-scale CNN to learn discriminative features from multiple signals and time scales of vibration signals. Qiao et al. [163] also proposed an adaptive weighted multi-scale CNN to adaptively extract robust and discriminative multi-scale features from raw vibration signals. With the help of these extracted features, the model could achieve supe-

rior performance under variable working conditions. Cheng et al. [164] proposed a hybrid method to extract discriminative features for bearing fault diagnosis, including the recurrence plot transform, speed up robust feature extraction and isometric mapping. Guo et al. [165] proposed a CNN model with Pythagorean spatial pyramid pooling to extract features from the input signals after continuous wavelet transform and used the extracted features to achieve fault diagnosis under variable working conditions. Different from the aforementioned approaches, Xiang et al. [166] used a Teager energy operator demodulation to process the raw signal, and then input the obtained Teager computed order spectra to a stacking AE for bearing fault diagnosis under variable working conditions.

(2) Transfer learning based method

Transfer learning based methods have been widely used in mechanical fault diagnosis under variable working conditions. The existing transfer learning methods for transferring the learned knowledge between multiple working conditions can mainly be divided into four categories [167], that is, the model-based transfer learning methods, instanced-based transfer learning methods, mapping-based transfer learning methods, and adversarial-based transfer learning methods.

Model-based transfer learning methods mean that the model first uses the data in the source domain for pre-training, and then fine-tunes the partial network parameters using the data in target domain. Hasan et al. [168] used a discrete orthonormal Stockwell transform to process the raw signal, and trained a CNN model with the obtained vibration images under different working conditions. Then partial parameters of pre-trained CNN were frozen and transferred to the target network for fault diagnosis. Du et al. [169] employed STFT to transform bearing vibration signals to time-frequency images, and used the processed data as the input of a deep residual network. Then, the model-based transfer learning strategy was used to achieve the high performance in another working condition. He et al. [170] trained the model using sufficient auxiliary data in the source domain and used multi-wavelet as an activation function for discriminative feature extraction, and then the model parameters were transferred to the target domain. Wu et al. [171] proposed a model based few-shot transfer learning method by considering the variability of working conditions and the scarcity of fault samples in the real working condition. Shao et al. [172] developed a novel DL framework using

transfer learning and the pre-trained network was finetuned by time-frequency images of vibration signals.

Instance-based transfer learning methods explore the way to reweight instances in the source domain to improve the diagnostic accuracy or align the distribution between the target domain and source domain. For example, Zhang et al. [173] used wide kernels in the first layer to extract more informative features and used small convolutional kernels in the latter layers for the multi-layer nonlinear mapping. Xiao et al. [174] trained a CNN with data from the target domain and source domain, and used a modified TrAdaBoost algorithm to update the weight of each training sample to form a stronger diagnostic model.

Mapping-based transfer learning methods refer to mapping the data from the source and target domains into the same feature space. For example, Azamfar et al. [175] and Singh et al. [176] used a DL-based domain adaption method for intelligent fault diagnosis by minimizing the cross-entropy loss in the source domain and maximum mean discrepancies between the source and target domains, simultaneously. Che et al. [177] and An et al. [178] used multi-kernel maximum mean discrepancies to match features between the source and target domain, and optimized with a combined transfer learning method. Qian et al. [179] reduced the input dimension by sparse filtering, and proposed a joint distribution adaptation to align the data distribution of the source and target domain, which helps capture discriminative features. Li et al. [180] proposed a representation clustering algorithm to minimize the distance between intra-class and maximize the distance between the inter-class simultaneously, and domain adaptation was used to adapt the maximum mean discrepancies between source and target domains. Li et al. [181] used knowledge mapping to explore domain-invariant knowledge between the source domain and the target domain, which helps to obtain a powerful feature extractor.

Adversarial-based transfer learning methods refer to using adversarial training to enable the domain discriminator to reduce the feature distribution of the source and the target domain, which makes the feature extractor can extract more robust features [182, 183]. For example, Lu et al. [184] and Han et al. [185] used adversarial domain adaptation to train the proposed DNN to extract representative information. Xu et al. [186] used adversarial domain adaptation to train a two-branch network to extract domain-invariant features, and used a scaled exponential linear unit activation function for the nonlinear activation.

3.2.4 Fault Diagnosis for Low Signal-to-Noise Ratio Signals

In real industrial scenario, the fault patterns are often overwhelmed by heavy background noise. As a result, algorithms with excellent performance under ideal conditions are often severely degraded in practical applications, showing weak generalization ability. Therefore, it is necessary to develop some advanced methods to enhance generalization ability of current algorithms. According to the current publications, there are two mainstreams to address this issue, that is, robust feature extraction based methods and building robust models. In real industrial scenario, the fault patterns are often overwhelmed by heavy background noise. As a result, algorithms with excellent performance under ideal conditions are often severely degraded in practical applications, showing weak generalization ability. Therefore, it is necessary to develop some advanced methods to enhance generalization ability of current algorithms. According to the current publications, there are two mainstreams to address this issue, that is, robust feature extraction based methods and building robust models.

(1) Robust feature extraction based methods

AE has strong ability for feature extraction, recently, researchers have developed many AE variants, such as deep auto-encoder (DAE), SDAE, and contractive auto-encoder (CAE), to automatically extract high-level representative features from data collected under the noisy environment.

For example, Chen et al. [187] used a deep SAE trained with Gaussian noise to avoid over-fitting and learned more robust features from a noisy working environment. Guo et al. [188] employed the SDAE to denoise random noise and to extract fault features from the vibration signals. Jiang et al. [189] proposed a feature learning approach named stacked multilevel-denoising AE, which is able to learn more robust and discriminative fault features to improve diagnosis accuracy on vibration signals with abundant noise. Shen et al. [190] constructed a stacked CAE model to extract more robust features than a standard stacked AE. Wang et al. [191] proposed a hybrid method by combining GAN and SDAE, where SDAE was used as the discriminator of GAN to automatically extract effective fault features from input samples and to discriminate their authenticity. Liu et al. [192] trained an 1D denoising convolutional AE model with noisy signals to perform fault classification. Qi et al. [193] combined SAE and CAE to obtain sparser and robust features under noise interference. Zhang et al. [194]

designed a deep CAE to automatically learn invariant feature representation from raw signals.

(2) Constructing robust model

Extracting highly robust features from low signal-to-noise ratio signals are time-consuming and labor-intensive. Therefore, it is necessary to establish an end-to-end fault diagnosis model with high robustness. For example, Gan et al. [195] proposed a hierarchical diagnostic network, which stacked multiple DBN layers to overcome the overlapping problem caused by noise or other interference. Shao et al. [196] proposed an improved convolutional deep placement network with compressed sensing to improve the generalization performance of the constructed deep model. You et al. [197] proposed a hybrid technique, which used CNN as feature extraction under noise environment and SVM as the classifier. Zhang et al. [173] proposed a deep CNN with wide first-layer kernels, which used the wide kernels to extract features and to suppress high-frequency noise. Zhang et al. [198] designed a deep CNN with new training methods to achieve pretty high accuracy in a noisy environment. Peng(a) et al. [199] constructed a deep residual learning network, which can adaptively learn the deep fault features from the original vibration signals to achieve high diagnostic accuracy under a strong noise environment. Peng(b) et al. [200] proposed a deep CNN to identify the failure modes of rotating vector reducer under strong background noise. Zan et al. [201] presented a fault diagnosis model based on a multi-dimension input CNN, which used multiple input layers to fuse the original signal and to learn the signal characteristics automatically for improving recognition accuracy and anti-jamming ability. Jin et al. [202] designed an adaptive anti-noise DNN framework to deal with the diagnosis problem under heavy noise without manual feature selection or denoising procedures. Peng(a) et al. [162] proposed a multi-branch and multi-scale CNN that could automatically learn and fuse abundant and complementary fault information from high complexity, strong coupling, and low signal-to-noise ratio vibration signals.

3.3 Open Source Datasets and Codes

3.3.1 Open Source Datasets

In the field of AI-enabled fault diagnosis, it is quite difficult to obtain high-quality datasets from real industrial scenarios and it also lacks open source codes. Fortunately, some institutions have released the datasets and codes for research and applications. Therefore, we collect

these commonly used datasets and the description of these datasets are listed in Table 3.

3.3.2 Open Source Codes

There are relatively a few open source codes for intelligent diagnosis. In this subsection, we summarize some online available codes of related papers as follows. A CNN-based method for bearing fault diagnosis was provided by Ref. [207], and in Ref. [208], the author released a code for rolling bearing faults. In Ref. [209], the author released an interpretable DNN for industrial intelligent diagnosis. In Ref. [210], the author released a multi-receptive field graph convolutional network for machine fault diagnosis. In Ref. [172], the author released a code for few-shot transfer learning for intelligent fault diagnosis of machine. In Ref. [168], the author released a unified intelligent fault diagnosis library based on unsupervised deep transfer learning and provided the corresponding comparative study. Besides, in Ref. [211], the author provided the baseline (lower bound) accuracy and released a unified intelligent fault diagnosis library based on various DL-based models. In Ref. [212], a CNN based on LeNet-5 was proposed for fault diagnosis.

3.4 Challenges

AI-enabled diagnosis has achieved great development, it releases the dependence of manpower and can automatically identify the health states from the past to the present. However, there are still some issues that need to be further discussed. In this section, we attempt to discuss the challenges and give some feasible solutions.

3.4.1 Interpretability

Interpretability helps users understand the results generated by the model. A main limitation of AI-enabled methods in mechanical fault diagnosis is that they operate as a “black box” and are not interpretable, which does not offer insight into how and why they can make the final decision. To bridge the gap, there are two research interests worthy of further study:

- (1) Most of the current AI-enabled diagnosis algorithms are migrated from the field of image processing and lack expert knowledge in the field of fault diagnosis. Therefore, we can combine prior knowledge commonly used in fault diagnosis to design our network. For example, we can design a convolution kernel that can extract useful features in vibration signals [209], or design a network structure that can be interpreted.
- (2) We can combine signal processing methods or traditional ML algorithms with DL algorithms to obtain a deep model with interpretable output. Sparse coding [213] may be a good choice to achieve this goal.

3.4.2 Transfer Learning

Transfer learning based methods have achieved a breakthrough in fault diagnosis under variable working conditions. However, there are still some challenges that need to be further discussed:

Table 3 Publications about the application of CNN in fault diagnosis

Dataset Name	Research Object	Description
CWRU [65]	Bearing	This dataset consists of four sub-datasets, and the operating conditions of each sub-dataset are 0 hp - 1797 r/min, 1 hp - 1772 r/min, 2 hp - 1750 r/min, and 3 hp - 1730 r/min, respectively. Single-point motor bearing faults include ball fault, inner ring fault, and outer ring fault
MFPT [203]	Bearing	This dataset consists of four sets of bearing vibration data. In the first sub-dataset, it contains three baseline conditions. In the second sub-dataset, it contains three outer race fault conditions. In the third sub-dataset, it contains seven outer race fault conditions for seven different loads. In the fourth sub-dataset, it contains seven inner race fault conditions for seven different loads
PU [204]	Bearing	It contains 32 sub-les, that is, 26 faulty bearings and 6 healthy bearings. The faulty bearings consist of 12 artificial damages generated by electrical discharge machining and 14 real damages caused by accelerated lifetime test
JNU [205]	Bearing	This dataset contains four kinds of bearing fault type, including normal state, ball fault, inner ring fault, and outer ring fault. Three different working conditions are investigated with the motor rotate speed set to 600, 800, and 1000, respectively
SEU [67]	Gearbox	This dataset contains two subsets, that is, a gear dataset, and a bearing dataset. The gear dataset contains four kinds of fault types, i.e., chipped tooth, missing tooth, root crack, and tooth surface wear. The bearing dataset consist of four kinds of fault types, including ball fault, inner ring fault, outer ring fault, and a mixed fault
UoC [206]	Gearbox	In this data set, nine different gear conditions are introduced into the pinion on the input shaft, including root cracks, health conditions, spalling, missing teeth, and sharpening with five different severity levels

- (1) The backbones of transfer learning based algorithms are often different, which makes it difficult to directly compare the results, and the impact of different backbones has not been thoroughly studied.
- (2) If the assumptions related to the source and target domains are invalid, transfer learning based algorithms might use diagnosis knowledge from the source domain to carry out a negative transfer, thereby reducing the transfer performance of the model.

3.4.3 Class Imbalance and Few-Shot Learning

In real engineering scenarios, the collected data, especially for the key components, is far from the big data and the amount of data is highly imbalanced, which makes it difficult to train AI-enabled models. Although, there are many algorithms to solve a class-imbalanced problem, it is still difficult to synthesize data with only a few samples. Therefore, how to use few-shot learning to solve the imbalanced problem still needs to be further discussed.

4 AI-Enabled Prognosis

4.1 Introduction to AI-Enabled Prognosis

Prognosis aims to evaluate the current health state of the equipment, which is known as degradation assessment (DA) and predicts its future failure time, which is known as remaining useful life (RUL) estimation, so as to provide the basis for subsequent predictive maintenance. In the industry, the operating condition of the critical equipment is highly concerned, as its sudden shutdown or failure would bring huge economic losses, and even endanger the life safety of operators. Compared with the traditional scheduled maintenance strategy, the prognostic based maintenance strategy provides proactive decision making capability that can effectively avoid downtime and costs, improve manufacturing productivity, and more importantly, provide early warning for catastrophic system failure.

According to the literature statistics, the prognosis methods generally fall into four groups, i.e. physics-based, statistics-based, data-driven, and hybrid methods. Physics-based methods usually rely on dynamic modeling, such as the finite element model [214] and simulation [215], etc., to calculate the dynamic response and degradation process of the system with a given input. However, physics-based methods require accurate mathematical models and expert knowledge about the specific system, which is difficult to implement on complex mechanical equipment. Statistics-based methods commonly assume that the RUL of equipment obeys an empirical distribution, such as a Weibull distribution

[216]. It is worth noting that statistics-based methods need data to update the parameters of the empirical distribution to fit the degradation process of the device, which is in fact data-dependent. Data-driven methods mine the characteristics of the device degradation process from the historical run-to-failure data to identify the degradation pattern of current equipment. The hybrid methods are formed by the combination of the above three methods, thus obtaining the corresponding merits.

The focus of this section is to review the data-driven DA and RUL estimation methods based on AI, especially those based on DL. Since the research areas of DA and RUL estimation partially overlap, this section will summarize these problems from different horizons, to provide more diverse information and discussions. For the former aspect, a hierarchical overview is given by categories of DA methods. For the latter aspect, the motivations of the RUL approaches are discussed. Additionally, a brief introduction to the open source datasets and codes will be given since we believe the open source behavior will drive the prognostic community to grow rapidly. Last but not least, to provide more accurate information for the predictive maintenance, many pain points deserve attention, so the challenges of prognosis will be given at the end of this section.

4.2 Degradation Assessment

4.2.1 Overview

Mechanical equipment usually has four states: normal state, performance degradation state, maintenance state, and decommissioning state. From the deterioration of equipment performance to the complete failure of equipment, it usually goes through a series of different performance degradation stages. DA of mechanical equipment is to synthesize the state indexes of mechanical equipment, evaluate the degree of performance degradation, formulate maintenance plan, and make targeted treatment. Scholars in the field of mechanical equipment health management have done a lot of research on DA. This subsection will review the methods of DA for mechanical equipment, which can be divided into two categories: traditional ML-based methods and DL-based methods, and summarize the merits and shortcomings of these methods simultaneously.

4.2.2 Traditional ML-Based Methods

Since ML algorithms play an important role in most respects of DA, scholars engaged in prognosis have carried out a lot of research in this area. From the analysis of the experimental results, the ML techniques, such as data dimension reduction, feature fusion and pattern recognition, etc., are very effective for DA problems.

(1) Fuzzy C-means clustering

Tong et al. [217] proposed a bearing DA model based on information theory metric learning and fuzzy C-means (FCM) clustering. The constructed degradation index showed superior performance. To solve the instability problem of bearing in the initial stage of operation, Liu et al. [218] proposed a method based on wavelet packet decomposition and autoregressive (AR) model to calculate the entropy of the health factor index, through FCM of bearing performance degradation process. Zhou et al. [219] proposed a rolling bearing DA method based on auto-associative neural network (AANN) and FCM. The features were extracted by wavelet packet decomposition and AR, and the features after dimension reduction were input into AANN. Then, the difference between the output vector and the input vector of AANN was input into FCM as the feature vector. In order to find the bearing fault in real-time, Zhou et al. [220] proposed a method based on wavelet packet Tsallis entropy and FCM to evaluate the performance degradation state of bearings.

(2) HMM

Jiang et al. [221] proposed a bearing DA method based on HMM and nuisance attribute projection (NAP). Aiming at the problem of poor robustness in DA, Jiang et al. [222] proposed a method of NAP based on student t-hidden Markov model, and removed interference components from performance degradation features by NAP. Hu et al. [223] proposed a first-order Markov state space model. For better expression in the state space model, the degraded state was transformed into PDF which formed HMM and Bayesian recursive estimation mechanism. Wang et al. [224] proposed a method of bearing DA based on hierarchical Dirichlet process (HDP)-HMM, in which HDP was used to obtain the state number of equipment in operation and HMM was used to evaluate performance degradation. To establish the index with an obvious trend, Li et al. [225] proposed the negative log likelihood probability based on the two-dimensional HMM as the bearing performance degradation index, showing the sensitivity to weak defects. Liu et al. [226] proposed a bearing DA method based on orthogonal local preserving projection (OLPP) and continuous HMM. The continuous HMM was used to train the data after dimension reduction by OLPP, and then the performance could be evaluated quantitatively by calculating the logarithmic likelihood of the data.

(3) PCA

To adapt the application of signal decomposition and feature extraction to wind turbine under high background noise, Pan et al. [227] proposed a DA method of vibration signal denoising fusion performance based on complete ensemble EMD with adaptive noise and kernel PCA. Ma et al. [228] proposed a DA method based on multi-sensor information fusion, which was extracted by the proposed method, to extract features and to establish the relevant DA model. Feng et al. [229] proposed a bearing DA method based on integrated EMD and PCA, which showed good effect in denoising and degradation evaluation.

(4) SVDD

Wang et al. [230] proposed a DA method of the rolling bearing based on VMD and SVDD. The characteristic vectors combined with VMD singular values, root mean square values, and sample entropy values, were selected as the evaluation indexes of the degradation degree, and then the performance degradation index of the test samples could be obtained by SVDD. Zhou et al. [231] proposed a bearing DA method based on lifting wavelet packet symbolic entropy (LWPSE) and SVDD. The SVDD was trained by fitting the hyper-sphere around the normal samples, and then the relative distance between LWPSE and the hyper-sphere boundary of the test signal was calculated as the bearing DA indicator.

(5) Clustering

Ding et al. [232] used manifold learning to extract features, achieved the comparison between abnormal data and health data, and calculated the feature clustering index to evaluate the degree of performance degradation. Tiwari et al. [233] proposed a DA method based on local mean decomposition (LMD) and spectral clustering to solve the problem of the high-dimensional feature space in rolling bearing DA. LMD was used to decompose signals, and spectral clustering was used to classify features. Lu et al. [234] proposed a compact Gaussian mixture clustering algorithm based on complementary ensemble EMD, which could distinguish the scattered features and obtain better DA results. Zhang et al. [235] proposed a bearing DA model based on the multi-scale entropy and K-medians clustering. The multi-scale entropy of bearing vibration signals was extracted from the original data, and the test data was input into the established K-medians clustering model, then the bearing failure degree could be quantitatively evaluated by the membership degree of the model output. Wang et al. [236]

proposed a bearing DA method based on the basic scale entropy and Gath-Geva fuzzy clustering. Gath-Geva fuzzy clustering was used to divide the degradation stage and further to evaluate the degradation degree of bearing performance. Akhand et al. [237] proposed an evaluation method of bearing performance degradation based on EMD and K-medians clustering. The K-medians clustering was applied to features extracted from bearing signals by EMD, and then the dissimilarity between the test data and normal state was taken as the bearing performance degradation index.

(6) Hybrid methods

Zhou et al. [238] proposed a performance degradation evaluation method of the wind turbine bearing based on HMM and FCM. The FCM and HMM models were constructed via using the features extracted by wavelet packet and AR, and could better describe the decline trend of the bearing.

(7) Other methods

Prashant et al. [239] proposed a method for evaluating the performance degradation of ball bearings based on curve component analysis and self-organizing mapping (SOM) network, which was more sensitive to weak degradation. Akhand et al. [240] proposed a kind of bearing DA index based on SOM. The time-domain and frequency-domain features were extracted from the original bearing vibration signals and input into the SOM classifier to achieve the degradation metric by minimizing the quantization error of SOM. Because the global trend of the signal could not accurately reflect the running state of the rolling bearing, Zhu et al. [241] proposed a bearing DA method based on the improved fuzzy entropy. The baseline part was not removed when calculating the fuzzy entropy, but used as the index of DA of rolling bearing. Qin et al. [242] proposed a method based on segmentation vote and SVM. LMD and PCA were used to obtain effective indexes of bearing performance degradation.

4.2.3 DL-Based Methods

Many scholars verified the effectiveness of DL algorithms, such as CNN and RNN, in DA. Here, we review DL-based methods from four perspectives, including CNN, RNN, hybrid methods, and other methods.

(1) CNN

For better DA, Zhang et al. [243] proposed a method for health index (HI) construction based

on deep multi-layer perceptron CNN. In order to improve the DA of rolling bearings, Dong et al. [244] proposed a method based on the DAE, t-distribution stochastic neighborhood embedding (t-SNE) and improved CNN. The features were constructed by the DAE and t-SNE, and the degree of bearing performance degradation was characterized by Mahalanobis distance. In order to solve the problem of outliers in HI, Zhang et al. [245] proposed a method combining deep convolution inner ensemble learning with outlier removal to evaluate the degradation degree of bearing performance. The deep convolution internal integration learning was used to extract features from the original vibration signals, and then outlier removal based on the sliding threshold was used to remove outliers in HI. Guo et al. [246] proposed a method of bearing HI construction based on deep convolution feature learning, which used convolution kernels to extract features from the original vibration signals, and mapped the features into HI through the nonlinear transformation.

(2) RNN

Akpudo et al. [247] proposed a LSTM model, and the root mean square (RMS) statistical features in time domain were used as the key features to evaluate the degree of bearing degradation. Zhang et al. [248] proposed a bearing DA method based on RNN, which evaluated the bearing performance degradation degree through the waveform entropy index, and identified the bearing running state via inputting the waveform entropy index into RNN. Cheng et al. [249] proposed a DA method based on adaptive kernel spectral clustering (AKSC) and RNN. The DA method constructed a DA feature based on Euclidean distance, and used AKSC and RNN to identify machine faults. Shi et al. [52] proposed a bearing failure DA method based on SDAE. SDAE was used to reconstruct the rolling bearing signal processed by the sliding window. LSTM was used to predict the vibration value of rolling bearing in the next cycle based on the reconstructed signals. Meanwhile, the performance degradation degree of a bearing was evaluated by the reconstructed error.

(3) Hybrid methods

Wang et al. [250] proposed a structure based on CNN and LSTM. CNN was used to extract local features of the original sensor, and LSTM was used to extract the sequence features of the original signals. H-statistics calculated by d-statistics and q-statistics were used to evaluate the performance degradation of rolling bearings.

(4) Other methods

Xu et al. [251] proposed an improved unsupervised deep trust network model named median filtering DBN. The absolute amplitude of the original vibration signals was used as the direct input for less dependence on the artificial experience. Pan et al. [252] proposed a DA method based on DBN and SOM, which defined the minimum quantization error as the HI of early fault detection of the wind power transmission. Tu et al. [253] proposed a method combining ANN and AR to evaluate the degradation degree of rolling bearing. ANN was used to evaluate the performance degradation degree of bearing, and AR was used to evaluate the bearing performance according to the bearing DA results. Gai et al. [254] proposed a bearing DA method combining EMD-SVD (singular value decomposition) and the fuzzy neural network. Li et al. [255] proposed a DA method based on DNN and wavelet packet decomposition. After extracting wavelet coefficients and energy features from vibration signals, DNN was used to predict the performance degradation degree of rotating machinery.

4.3 RUL Estimation

Several definitions of RUL estimation have been introduced in Refs. [256–258]. To avoid confusion, this paper followed the definition from the International Standard Organization, that the RUL estimation is defined as the estimation of the time to failure. Data-driven RUL estimation methods can be divided into two strategies: matching-based and regression-based. For the former one, the library of HI (also called degradation index) needs to be constructed offline firstly, and then the HI is calculated and matched from the library [259–264]. The key point of the matching-based RUL estimation is to construct a monotone, smooth and obviously trending HI library. Since the construction of HI has been discussed in the previous subsection in detail, matching-based RUL estimation methods will not be included in this subsection.

The regression-based data-driven RUL estimation methods mainly leverage the historical run-to-failure data to model the degradation process of equipment, and then evaluate the health status of the current operating equipment.

As shown in Figure 5, the regression-based data-driven RUL estimation framework generally includes the following aspects:

1) Data acquisition: This procedure is to collect and save data measured by appropriate sensors.

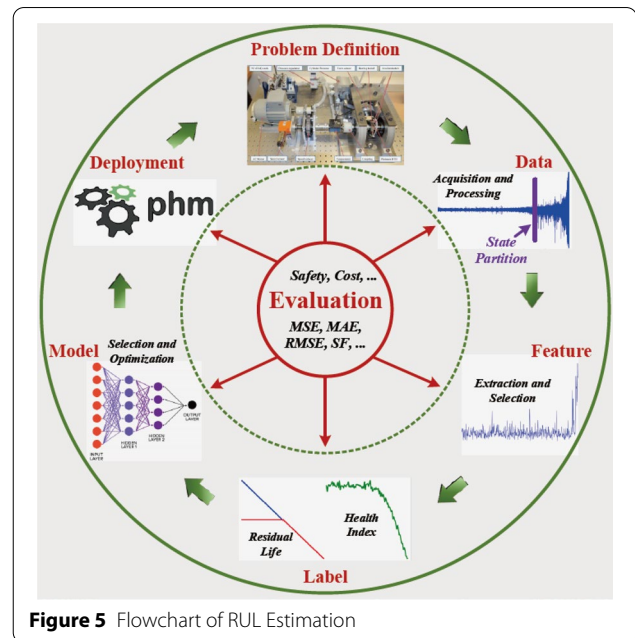


Figure 5 Flowchart of RUL Estimation

- 2) Data processing: Data needs to be cleaned at this stage for a higher data quality, and common methods include denoising and interpolation.
- 3) Feature extraction and selection: Features are extracted and selected to reflect the degradation trend. Traditional methods often use statistical features, such as Kurtosis, RMS, while DL-based methods can use unsupervised learning to extract deep features automatically.
- 4) States partition: This step is to divide the run-to-failure data into the health state and degradation state, which is also called the determination of first predicting time [265] or fault occurrence time [266].
- 5) HI construction: This procedure is to fuse the previously extracted features to construct a monotone, smooth and obviously trending HI. Some methods take HI as the label of the regression model, while others directly take the RUL as the label. Therefore, HI construction is not necessary for the RUL estimation. It is worth noting that there may be a confusing concept that the HI is not the same as the RUL, and there is still a mapping relationship between them. The HI is usually represented as a curve with jitter, while the RUL is represented as a linear function or piecewise linear function.
- 6) Model building and optimization: This process is to build the regression model offline based on historical data, and then the trained model will be deployed online.
- 7) Performance evaluation: This procedure is to evaluate the performance of the model. For the RUL

estimation, common evaluation metrics include mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean relative error (MRE), scoring function (SF) and some other variants. Evaluation metrics guide the training of the model. Elsheikh et al. [267] proposed a safety-oriented metric which was biased towards the earlier estimation. Therefore, an appropriate evaluation metric should be designed according to various application scenarios.

4.3.1 Traditional ML-Based Methods

The RUL estimation based on traditional ML methods has been developed for more than 20 years and a variety of traditional ML methods, like SVM [268–270], HMM [271, 272], and ANN [273–276], have been widely studied to solve this task. Since the theories and applications of these technologies are relatively mature, there have been many excellent reviews for traditional ML-based RUL estimation [13, 14]. Moreover, this section focuses on DL-based RUL estimation, so only the recent ML-based RUL prediction methods are briefly introduced.

Traditional ML-based RUL estimation generally consists of two steps. The primary step is to extract high-quality features from the original signals. The secondary step is to train the regression model based on the extracted features. Since the model theory is mature to a large extent, the difficulty of applying the model lies in how to match the specific task, to achieve better accuracy and efficiency, and to interpret the results.

In addition to the aforementioned classic methods, some scholars have done valuable works based on the pain points of the RUL estimation task. Considering the small sample size of run-to-failure data, transfer learning was used to solve the problem of inconsistent data distributions [266, 277]. Zhang et al. [278] proposed a prognostic method based on the dynamic Bayesian network with mixture of Gaussian output to deal with missing data in real scenarios. Compared with DL-based methods, traditional ML-based methods are more interpretable. However, due to the limitation of the model capacity, it is difficult for traditional ML-based methods to fit the massive and high-dimensional data.

4.3.2 DL-Based Methods

There have been several insightful reviews of the RUL estimation based on DL, such as Ref. [279]. Most of literatures were classified and reviewed according to the types of DL models, which made the paper hierarchical, but also easy to create an illusion. Although different DL-based models can accomplish specific tasks, being addicted to various DL models tends to ignore the RUL

estimation task itself. Therefore, this paper will not make a classification review according to the categories of DL models, but according to the motivations.

(1) Spatial-temporal feature extraction

DL-based methods can realize the automatic feature extraction and easily combine the extracted features with the subsequent regression model to construct an end-to-end optimization pipeline. As a result, most of DL-based RUL estimation methods leverage its powerful spatial-temporal feature extraction ability. These methods extract spatial-temporal features from the time domain, the frequency domain, and the time-frequency domain, in which spatial-temporal features benefit from CNN and RNN, respectively.

Original signal: It is feasible and convenient to directly input the original signal into DL models for the RUL estimation. Li et al. [280] directly input multi-source sensor signals into a LSTM. The CNN also showed the potential of processing raw signals for the RUL task [281]. With the exploration of the ability of 1D CNN to extract local features of time series, some scholars found that the spatial-feature extraction capability of CNN and the temporal-feature extraction capability of RNN can be obtained simultaneously via connecting the structures of CNN and RNN sequentially [282, 283]. For better feature representation, more complex parallel networks have been designed. Li et al. [284] built two feature extraction branches with LSTM and CNN respectively, and then modeled the fusion features with another LSTM. Thanks to CNN and LSTM, these methods combined spatial and temporal features by designing a network structure to automatically extract features from the original signals. Benefiting from the end-to-end training mode of DL methods, the feature extraction networks and prediction networks are optimized synchronously, which improves the efficiency and accuracy of fault prognosis.

Signal processing: Although the sliding window sampling strategy preserves partial non-stationary relations between different windows, the calculation of statistical features will lose non-stationary information in a single window. Zhao et al. [285] extracted manual features to form the feature sequences, and then modeled the sequences with a Gated Recurrent Unit network. Although more non-stationary relations can be retained by reducing the window length, the feature quality and

the window length are against each other, and the dense sliding window will bring a huge computation. Therefore, the introduction of time-frequency analysis is a natural idea, as this technology could relatively preserve the non-stationary relationship within a single sliding window, such as envelope spectrum [286], discrete wavelet transform [287], continuous wavelet transform [288], and STFT [289, 290]. In addition to these classical time-frequency analysis techniques, there were also some methods to extract frequency domain features directly [291–293]. These methods combined the advantages of both signal processing and DL, including more domain priors, and performed well in the RUL estimation task.

Unsupervised learning: Unsupervised learning can extract a good data representation without label information, and with the efforts of a large number of papers, the performance of unsupervised learning on some tasks has approached or surpassed that of supervised learning.

Thus, in addition to the above supervised learning methods to specifically extract features, unsupervised learning can also obtain excellent feature representation, mainly including DBN [294, 295], AE, and its variants [244, 296–299].

(2) Multi-modal and multi-task

Different types of data (temperature, pressure, vibration, etc.) from the same equipment are collected by several sensors simultaneously, and different sensors reflect various condition information, which requires the model to be able to process multi-modal data. As the task division of PHM becomes more and more elaborate, modeling each task separately is time-consuming, and there may be a problem that the decision results by two models for highly similar tasks are inconsistent, so using a unified model to accomplish multiple tasks is a promising approach. Multi-modal learning and multi-task learning both share a common trunk model respectively, despite their different motivations. For the former one, the trunk model requires multiple sources of the input to extract redundant features from a broad perspective, and this strategy ensures the model can work effectively in the absence of some modal data. For the latter one, the trunk model has multiple output branches, each of which corresponds to a specific task, while the trunk model provides a shared feature subspace.

Multi-modal learning: In most methods with algorithm validation on C-MAPSS dataset, multi-

sensor data was basically used as the input of the trunk model. However, the concept of multi-modal was vague in these papers, and this concept was only highlighted in a few papers [300, 301]. The C-MAPSS dataset had 21 sensors and 3 operational conditions (altitude, Mach number, and sea-level temperature), so it was natural to consider multi-modal on this dataset. For other specific tasks, the multi-modal data based approach was less popular because collecting additional data often meant higher costs. Herp et al. [302] proposed a prognostic model for the wind turbine main bearing based on multi-modal data (actual wind speed, temperature, active power, etc.). He et al. [303] considered 6 sensors data and 5 operational setting data in the RUL estimation task for the ion mill etching flow-cool system.

Multi-task learning: The purpose of multi-task learning is to explore the common feature subspace between tasks with a joint model. Miao et al. [304] completed the RUL estimation task together with the DA task, and trained a trunk LSTM network via weighting two loss items. Liu et al. [305] combined the RUL estimation task with the fault recognition task based on a trunk CNN model. Aggarwal et al. [256] discussed the relationship between short-term failure prediction and long-term RUL estimation, arguing that a joint model could prevent inconsistent results.

(3) States partition

The difficulty for the RUL estimation is to determine the label for the input at each time step. Some methods take the HI as the label for the regression model. However, there is still a mapping relationship between the HI and the actual RUL, and since most HIs are not completely monotonous, one HI point may correspond to several RUL points. Therefore, many methods directly use the actual RUL as the label, such as linear or piece-wise linear functions. If the degradation mechanism of a physical system is unknown, it is natural to use the linear function due to the fact that the actual RUL of the system decreases step by step. However, the piece-wise linear function will be more consistent with the data distribution if we know the degradation mechanism. The run-to-failure data typically contains the health state and degradation state of the equipment. Generally, the distribution of health state data is concentrated, while the distribution of degradation state data tends to be scattered [291]. To provide corresponding labels for each state, the piece-wise linear function assumes that the expected value of the RUL in the health state is con-

stant and that in the degradation state is linearly decreasing.

The challenge for the piece-wise linear function is to determine the degradation occurrence time, which means how to divide run-to-failure data into the health state and the degradation state. Most methods rely on empirical rules to select an appropriate threshold, such as observing the trend of HI. However, the selection of a threshold mainly depends on expert experience. Mao et al. [297] calculated the Pearson correlation coefficient between the health state and the degradation state based on features from SDAE. Although the step of feature extraction was skipped, the setting of a threshold was still inevitable. Therefore, some scholars considered how to reduce the impact of expert knowledge on states partition to improve the generalization ability of the RUL model. Li et al. [265] regarded the state partition as an anomaly detection problem and used GAN to learn the health state data distribution, and then determined the first predicting time. Xia et al. [306] divided the run-to-failure data into different degradation stages and then classified these states by DL models. Yang et al. [307] also regarded states partition as a classification problem, but added more rules of engineering experience, such as rapid or slow degradation patterns, and a “3/5” principle. States partition is still challenging for the RUL estimation task as it is impossible to construct a common model or rule for all devices and application scenarios.

(4) Transfer learning

In general, the RUL estimation task constructs an offline model on existing historical run-to-failure data and then makes an online prediction for a new object. However, collecting large amounts of run-to-failure data in the real-world scenario is expensive and time-consuming. For industry, major equipment would not be allowed to operate in a near-failure condition, which means it is also difficult to collect complete degradation process data. As a result, the RUL estimation task naturally faces a few-shot sample and generalization problem.

Therefore, some scholars have conducted research on this issue based on transfer learning. In the relatively early study, Zhang et al. [308] adopted the fine-tuned strategy, that the model was pre-trained with a related dataset with a large data size, and then the pre-trained model was fine-tuned in the target dataset with only few samples. Sun et al. [296] proposed a deep transfer learning method based on SAE using three transfer strategies (weight

transfer, feature transfer, and weight update) for the tool RUL estimation. Mao et al. [297] aligned the features of source and target domains by the transfer component analysis. Yu et al. [309] proposed a transfer learning method to reduce the distribution discrepancy between source and target domains based on maximum mean discrepancy for the RUL estimation. Meanwhile, the feature alignment strategy based on adversarial learning for RUL estimation has been applied in Refs. [265, 310].

(5) Uncertainty modeling

If the cost and risk of decision need to be considered, it is necessary to estimate the uncertainty of the RUL estimation. Sankararaman et al. [311] argued that the traditional source classification of uncertainty (physical variability and lack of knowledge) may not be applicable to the RUL estimation, so he proposed a different classification method, that the sources of uncertainty in the RUL estimation were as follows: the current unknown state of the system, future uncertainty (i.e., the loading, operating, environmental, and usage conditions), modeling uncertainty, and the actual data distribution. Purposeful modeling can be performed if the factors are clearly known. The wandering set-points, input current, and fault magnitude were used to model the uncertainty for the sensor prognosis based on Gaussian process in Ref. [312]. Wiener process is also a common uncertainty modeling method [313, 314]. If the uncertainty factors are too complex to be recognized, latent modeling can be carried out, and the mainstream method is based on the Bayesian theory. Peng et al. [315] and Wang et al. [316] proposed to quantify the uncertainty of the RUL estimation using the Bayesian DL network based on Monte Carlo (MC) dropout, which was proved to be an effective Bayesian approximation in Ref. [317]. In addition to the Bayesian method, the sampling-based ensemble learning can also realize the uncertainty modeling and quantification for the RUL estimation [318, 319]. By training the sub-models with various sub-datasets, an ensemble learning method can be formed, and the actual distribution of few-shot data can be estimated effectively. In fact, latent modeling avoids the challenge of identifying the source of uncertainty, but it becomes more difficult to interpret the uncertainty. Sankararaman et al. [311] discussed in detail the significance, interpretation, and quantification of uncertainty in the RUL estimation, and then compared several methods of uncertainty propagation.

4.4 Open Source Datasets and Codes

4.4.1 Open Source Datasets

The AI-enabled prognosis requires a large amount of high-quality run-to-failure data, which is difficult to satisfy in real-world scenarios. There are several factors as follows:

- 1) The device will not be allowed to operate near the failure time for security and economic reasons, which means that the full degradation process is rare;
- 2) Different conditions and fault types will produce various degradation processes, and the cost of traversal experiments is obviously unacceptable;
- 3) Individual differences lead to inconsistencies in the distribution of data from historically failed equipment and those from current operating equipment, and it is expensive to perform experiments on a large number of subjects.

Fortunately, as shown in Table 4, several mechanical prognosis datasets have been shared from a few institutions, and NASA collected some of the open source datasets to build the Prognostics Center of Excellence (PCoE) database [320]. Additionally, the open source datasets provide a baseline standard for validation of various algorithms.

4.4.2 Open Source Codes

The open source code behavior in the prognosis field is very necessary and important to promote the theoretical research and application, and will also have a positive effect on the upstream and downstream tasks of PHM. However, there are only a few open source projects for prognosis because of the difficulty of this research. Oyharcabal et al. [325] coupled the convolution kernel to the operation of RNN and verified it on the C-MAPSS dataset. Lahiru et al. [326] described the overall process of the RUL estimation for C-MAPSS dataset in detail,

including data structures, labels, data augmentation, etc. Libera et al. [327] applied Bayesian and Frequentist DL models to the RUL estimation. Chen et al. [328] used attention mechanism to model the importance of extracted features and also released the source codes.

4.5 Challenges

4.5.1 Generalization Ability

As previously mentioned, the few-shot data makes it difficult to accurately predict the RUL of a new object. Transfer learning has been used to enhance the generalization ability of the RUL estimation, but mainly from the feature or model perspective. A natural way to enhance this ability is to increase the data volume. So how to generate more high-quality samples based on the existing data is still a challenge. Although it is impossible to obtain the actual data distribution, we can estimate the data distribution, such as using the resampling strategy or adversarial learning strategy to generate high-quality virtual samples. Additionally, with the digital twin model of a specific mechanical system, a large number of degradation processes under different conditions can be generated by changing the working conditions, fault types, and other variables.

4.5.2 Prognosis in Real-World Scenarios

There are many limitations and uncertainties in the real open world, such as restricted computing resources, variable working conditions, unknown failure modes, etc. The lack of computational power means that AI-enabled prognosis methods cannot be directly applied to the real scenario. Meanwhile, to ensure real-time prognosis, it is necessary to design a lightweight model, and common methods include model compression and pruning. Additionally, the open scenario requires the model to be able to continuously update parameters online, since the data distribution of the open scenario and that of training data are often inconsistent.

Table 4 Publications about the application of CNN in fault diagnosis

Dataset Name	Research Object	Description
C-MAPSS [321]	Simulated Turbofan engine	Engine degradation simulation was carried out using C-MAPSS. Four different sets were simulated under different combinations of operational conditions and fault modes. The dataset was provided by the PCoE at NASA Ames
Milling Dataset [322]	Actual milling insert	Experiments on a milling machine for different speeds, feeds, and depth of cut. The dataset was provided by the BEST lab at UC Berkeley
IMS Dataset [68]	Actual bearings	The dataset with bearing run-to-failure experiments was provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati
FEMTO Dataset [323]	Actual bearings	Experiments on bearing accelerated life tests provided by FEMTO-ST Institute, Besancon, France
XJTU-SY [324]	Actual bearings	Experiments on bearing accelerated life tests containing complete run-to-failure data of 15 rolling element bearings. The dataset was provided by the Xi'an Jiaotong University and the Changxing Sumyoung Technology Company

4.5.3 Combination of Data-Driven and Model-Driven Methods

As equipment becomes increasingly complex, a single method is usually difficult to accurately evaluate and predict the RUL. Combining multiple models based on data-driven and model-driven methods to establish more effective health indicators, we can make full use of the powerful feature extraction ability based on data-driven methods and the advantages of interpretability of model-driven methods. For example, Yucesan et al. [329] designed a physics-informed layer based on damage increment within deep neural networks to predict wind turbine main bearing fatigue. In addition, Wang et al. [330] proposed a cross physics-data fusion scheme and a loss function which embeds physical discipline for machine tool wear prediction.

5 Conclusions

In this paper, we mainly review the current development of AI-enabled approaches, especially DL-based approaches in monitoring, diagnosis, and prognosis, which are three essential ingredients of PHM. Besides, we emphasize the importance of open source datasets and codes for the benign development of the research community of AI-enabled PHM. For monitoring, we summarize the main challenges, containing balance between recall and precision, unified benchmark datasets, quick alarm for early failure, and adaptability under variable working conditions. For diagnosis, we conclude that the main challenges are interpretability, transfer learning, class imbalance learning, and few-shot learning. For prognosis, we further summarize the challenges as generalization ability, prognosis in real-world scenarios, and combination of data-driven and model-driven methods. We hope this review paper could provide some valuable discussions of future research and attract enough attention from researchers to the construction of the open source community.

Acknowledgements

The authors sincerely thanks to Zheng Zhou, Zuogang Shang, Chenye Hu, Hongbing Shang for all their help on this work.

Authors' contributions

ZZ: Writing—review & editing; JW: Writing—review & editing; TL: Writing—review & editing; CS: Supervision; RY: Review & supervision; XC: Supervision. All authors read and approved the final manuscript.

Authors' Information

Zhibin Zhao received the B.S. degree in *Tsien Hsue-Shen Honor Class, Xi'an Jiaotong University, China*, in 2015, and the Ph.D. degree in mechanical engineering from *Xi'an Jiaotong University, China*, in 2020. He was also a visiting Ph.D. candidate in AI for Healthcare at the *University of Manchester, UK*, from 2019 to 2020. He is now a lecturer in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University, China*. His current research is focused on sparse signal processing and machine learning algorithms for machinery health monitoring and healthcare.

Jingyao Wu, born in 1996, is currently a PhD candidate in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University, China*. He received his bachelor degree from *Xi'an Jiaotong University, China*, in 2018. His research interests include anomaly detection, fault intelligent diagnosis.

Tianfu Li, born in 1995, is currently a PhD candidate in mechanical engineering at *School of Mechanical Engineering, Xi'an Jiaotong University, China*. He received his bachelor degree from *Chongqing University, China* in 2018. His research interests include deep learning, mechanical fault diagnosis and prognosis.

Chuang Sun received the Ph.D. degree in mechanical engineering from *Xi'an Jiaotong University, China*, in 2014. From 2015 to 2016, he held a Postdoctoral position at *Case Western Reserve University, USA*. He is currently an associate professor at *School of Mechanical Engineering, Xi'an Jiaotong University, China*. His current research interests include manifold learning, deep learning, sparse representation, mechanical fault diagnosis and prognosis, and remaining useful life prediction.

Ruqiang Yan (Senior Member, IEEE) received the Ph.D. degree in mechanical engineering from the *University of Massachusetts, USA*, in 2007. From 2009 to 2018, he was a Professor with the *School of Instrument Science and Engineering, Southeast University, China*. He joined the *School of Mechanical Engineering, Xi'an Jiaotong University, China*, in 2018. His research interests include data analytics, machine learning, and energy-efficient sensing, and sensor networks for the condition monitoring and health diagnosis of large-scale, complex, dynamical systems. Dr. Yan is a Fellow of *American Society of Mechanical Engineers* (2019). He was the recipient of several awards and honors, including the IEEE Instrumentation and Measurement Society Technical Award in 2019, New Century Excellent Talents in University Award from the Ministry of Education in China in 2009, and multiple best paper awards. He is an Associate Editor-in-Chief for the *IEEE Transactions on Instrumentation and Measurement* and an Associate Editor for the *IEEE Systems Journal*.

Xuefeng Chen (Senior Member, IEEE) received the Ph.D. degree in mechanical engineering from *Xi'an Jiaotong University, China*, in 2004. He is currently a Full Professor and the Dean of *School of Mechanical Engineering, Xi'an Jiaotong University, China*. He has authored more than 100 SCI publications in areas of composite structure, aeroengine, wind power equipment, etc. Dr. Chen was the recipient of the National Excellent Doctoral Thesis Award in 2007, First Technological Invention Award of Ministry of Education in 2008, Second National Technological Invention Award in 2009, First Provincial Teaching Achievement Award in 2013, First Technological Invention Award of Ministry of Education in 2015, and Science and Technology Award for Chinese Youth in 2013. He hosted a National Key 973 Research Program of China as a Principal Scientist in 2015. He is the Executive Director of the Fault Diagnosis Branch in *China Mechanical Engineering Society*.

Funding

Supported by National Key Research and Development Program of China (Grant No. 2018YFB1702400) and National Natural Science Foundation of China (Grant Nos. 51835009, 51705398).

Competing interests

The authors declare that they have no competing interests.

Received: 20 August 2020 Revised: 23 March 2021 Accepted: 11 May 2021

Published online: 09 June 2021

References

- [1] O Fink, Q Wang, M Svensen, et al. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, 2020, 92:103678.
- [2] R Zhao, R Yan, Z Chen, et al. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 2019, 115: 213-237.
- [3] Z Zhao, S Wu, B Qiao, et al. Enhanced sparse period-group lasso for bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2018, 66(3): 2143-2153.

- [4] Z Zhao, B Qiao, S Wang, et al. A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis. *Journal of Sound and Vibration*, 2019, 446: 429-452.
- [5] S Wang, X Chen, C Tong, et al. Matching synchrosqueezing wavelet transform and application to aeroengine vibration monitoring. *IEEE Transactions on Instrumentation and Measurement*, 2016, 66(2): 360-372.
- [6] G E Hinton, R R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 2006, 313(5786): 504-507.
- [7] Y LeCun, Y Bengio, G Hinton. Deep learning. *Nature*, 2015, 521(7553): 436-444.
- [8] M Hamadache, J H Jung, J Park, et al. A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning. *JMST Advances*, 2019, 1(1): 125-151.
- [9] J Lee, F Wu, W Zhao, et al. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 2014, 42(1-2): 314-334.
- [10] A L Ellefsen, V Æsøy, S Ushakov, et al. A comprehensive survey of prognostics and health management based on deep learning for autonomous ships. *IEEE Transactions on Reliability*, 2019, 68(2): 720-740.
- [11] R Liu, B Yang, E Zio, et al. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 2018, 108: 33-47.
- [12] Y Lei, B Yang, X Jiang, et al. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 2020, 138: 106587.
- [13] Y Lei, N Li, L Guo, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 2018, 104: 799-834.
- [14] Z Zhang, X Si, C Hu, et al. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. *European Journal of Operational Research*, 2018, 271(3): 775-796.
- [15] S Khan, T Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 2018, 107: 241-265.
- [16] M Strozzi, R Rubini, M Cocconcelli. Condition monitoring techniques of ball bearings in non-stationary conditions. *International Conference on Design, Simulation, Manufacturing: The Innovation Exchange*, Springer, Cham, 2019: 565-576.
- [17] A Mauricio, S Sheng, K Gryllias. Condition monitoring of wind turbine planetary gearboxes under different operating conditions. *Journal of Engineering for Gas Turbines and Power*, 2020, 142(3): GTP-19-1317.
- [18] S Schmidt, P S Heyns. Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring. *Measurement*, 2020, 149: 106964.
- [19] P S Ambika, P K Rajendrakumar, R Ramchand. Vibration signal based condition monitoring of mechanical equipment with scattering transform. *Journal of Mechanical Science and Technology*, 2019, 33(7): 3095-3103.
- [20] M Zhao, J Jiao, J Lin. A data-driven monitoring scheme for rotating machinery via self-comparison approach. *IEEE Transactions on Industrial Informatics*, 2018, 15(4): 2435-2445.
- [21] B W Harris, M W Milo, M J Roan. A general anomaly detection approach applied to rolling element bearings via reduced-dimensionality transition matrix analysis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2016, 230(13): 2169-2180.
- [22] X M Tao, W H Chen, B X Du, et al. A novel model of one-class bearing fault detection using SVDD and genetic algorithm. *2007 2nd IEEE Conference on Industrial Electronics and Applications*. IEEE, 2007: 802-807.
- [23] C Liu, K Gryllias. A semi-supervised Support Vector Data Description-based fault detection method for rolling element bearings based on cyclic spectral analysis. *Mechanical Systems and Signal Processing*, 2020, 140: 106682.
- [24] P Zhao, M Kurihara, J Tanaka, et al. Advanced correlation-based anomaly detection method for predictive maintenance. *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2017: 78-83.
- [25] J Tian, M H Azarian, M Pecht. Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. *Proceedings of the European Conference of the Prognostics and Health Management Society*, Citeseer, 2014: 1-9.
- [26] D A Clifton, P R Bannister, L Tarassenko. Application of an intuitive novelty metric for jet engine condition monitoring. *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, Berlin, Heidelberg, 2006: 1149-1158.
- [27] J A Cariño-Corralles, J J Saucedo-Dorantes, D Zurita-Millán, et al. Vibration-based adaptive novelty detection method for monitoring faults in a kinematic chain. *Shock and Vibration*, 2016, 2016: 2417856.
- [28] Z Li, J Li, Y Wang, et al. A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment. *The International Journal of Advanced Manufacturing Technology*, 2019, 103(1): 499-510.
- [29] S Plakias, Y S Boutalis. Exploiting the generative adversarial framework for one-class multi-dimensional fault detection. *Neurocomputing*, 2019, 332: 396-405.
- [30] D Li, D Chen, J Goh, et al. Anomaly detection with generative adversarial networks for multivariate time series. *arXiv preprint*, 2018, arXiv:1809.04758.
- [31] D Li, D Chen, B Jin, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. *International Conference on Artificial Neural Networks*, Springer, Cham, 2019: 703-716.
- [32] W Jiang, Y Hong, B Zhou, et al. A GAN-based anomaly detection approach for imbalanced industrial time series. *IEEE Access*, 2019, 7: 143608-143619.
- [33] L Martí, N Sanchez-Pi, J M Molina, et al. Anomaly detection based on sensor data in petroleum industry applications. *Sensors*, 2015, 15(2): 2774-2797.
- [34] L Martí, N Sanchez-Pi, J M López, et al. On the combination of support vector machines and segmentation algorithms for anomaly detection: A petroleum industry comparative study. *Journal of Applied Logic*, 2017, 24: 71-84.
- [35] J Inoue, Y Yamagata, Y Chen, et al. Anomaly detection for a water treatment system using unsupervised machine learning. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, 2017: 1058-1065.
- [36] A Nanduri, L Sherry. Anomaly detection in aircraft data using Recurrent Neural Networks (RNN). *2016 Integrated Communications Navigation and Surveillance (ICNS)*, IEEE, 2016: 5C2-1-5C2-8.
- [37] M Sakurada, T Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, 2014: 4-11.
- [38] T H Dwiputranto, N A Setiawan, T B Aji. Machinery equipment early fault detection using Artificial Neural Network based Autoencoder. *2017 3rd International Conference on Science and Technology-Computer (ICST)*, IEEE, 2017: 66-69.
- [39] C Y Lin, C P Weng, L C Wang, et al. Edge-based RNN anomaly detection platform in machine tools. *Smart Science*, 2019, 7(2): 139-146.
- [40] F Pittino, M Puggl, T Moldaschl, et al. Automatic anomaly detection on in-production manufacturing machines using statistical learning methods. *Sensors*, 2020, 20(8): 2344.
- [41] M Schlechtingen, I F Santos. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing*, 2011, 25(5): 1849-1875.
- [42] P Arpaia, U Cesaro, M Chadli, et al. Fault detection on fluid machinery using Hidden Markov Models. *Measurement*, 2020, 151: 107126.
- [43] X Wang, G Lu, P Yan. Multiple regression analysis based approach for condition monitoring of industrial rotating machinery using multi-sensors. *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, IEEE, 2019: 1-5.
- [44] T Wang, G Lu, P Yan. Multi-sensors based condition monitoring of rotary machines: An approach of multidimensional time-series analysis. *Measurement*, 2019, 134: 326-335.
- [45] G R Garcia, G Michau, M Ducoffe, et al. Time series to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. *arXiv preprint*, 2020, arXiv:2005.07031.
- [46] Y Zhang, P Hutchinson, N Lieven, et al. Adaptive event-triggered anomaly detection in compressed vibration data. *Mechanical Systems and Signal Processing*, 2019, 122: 480-501.
- [47] H Luo, S Zhong. Gas turbine engine gas path anomaly detection using deep learning with Gaussian distribution. *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, IEEE, 2017: 1-6.

- [48] D Toshkova, M Asher, P Hutchinson, et al. Automatic alarm setup using extreme value theory. *Mechanical Systems and Signal Processing*, 2020, 139: 106417.
- [49] G Lu, J Liu, P Yan. Graph-based structural change detection for rotating machinery monitoring. *Mechanical Systems and Signal Processing*, 2018, 99: 73-82.
- [50] T Krogerus, M Hyvönen, P Multanen, et al. Joint probability distributions of correlation coefficients in the diagnostics of mobile work machines. *Mechatronics*, 2016, 35: 82-90.
- [51] G Li, Y Hu, H Chen, et al. An improved fault detection method for incipient centrifugal chiller faults using the PCA-R-SVDD algorithm. *Energy and Buildings*, 2016, 116: 104-113.
- [52] H Shi, L Guo, S Tan, et al. Rolling bearing initial fault detection using long short-term memory recurrent network. *IEEE Access*, 2019, 7: 171559-171569.
- [53] J Wu, Z Zhao, C Sun, et al. Fault-attention generative probabilistic adversarial autoencoder for machine anomaly detection. *IEEE Transactions on Industrial Informatics*, 2020, 16(12): 7479-7488.
- [54] N N Thi, N A Le-Khac. One-class collective anomaly detection based on LSTM-RNNs. *Transactions on Large-Scale Data-and Knowledge-Centered Systems XXXVI*. Springer, Berlin, Heidelberg, 2017: 73-85.
- [55] T Ko, J H Lee, H Cho, et al. Machine learning-based anomaly detection via integration of manufacturing, inspection and after-sales service data. *Industrial Management & Data Systems*, 2017, 117(5): 927-945.
- [56] S Ahmad, A Lavin, S Purdy, et al. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 2017, 262: 134-147.
- [57] K Hendrickx, W Meert, Y Mollet, et al. A general anomaly detection framework for fleet-based condition monitoring of machines. *Mechanical Systems and Signal Processing*, 2020, 139: 106585.
- [58] S B Kim, T Sukchotrat, S K Park. A nonparametric fault isolation approach through one-class classification algorithms. *IIE Transactions*, 2011, 43(7): 505-517.
- [59] G Lu, Y Zhou, C Lu, et al. A novel framework of change-point detection for machine monitoring. *Mechanical Systems and Signal Processing*, 2017, 83: 533-548.
- [60] D A Clifton, L Tarassenko. Novelty detection in jet engine vibration spectra. *International Journal of Condition Monitoring*, 2015, 5(2): 2-7.
- [61] K Yan, Z Ji, W Shen. Online fault detection methods for chillers combining extended Kalman filter and recursive one-class SVM. *Neurocomputing*, 2017, 228: 205-212.
- [62] S Martin-del-Campo, F Sandin. Online feature learning for condition monitoring of rotating machinery. *Engineering Applications of Artificial Intelligence*, 2017, 64: 187-196.
- [63] W Yang, R Court, J Jiang. Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy*, 2013, 53: 365-376.
- [64] A Bzimek. Application of selected method of anomaly detection in signals acquired during welding process monitoring. *International Journal of Materials and Product Technology*, 2017, 54(4): 249-258.
- [65] Case Western Reserve University (CWRU) Bearing Data Center, [online]. (Accessed on August 1, 2020). <https://csegroups.case.edu/ZhibinZhao> et al. Page 23 of 29 bearingdatacenter/pages/download-data-file/.
- [66] Tennessee Eastman Process Datasets, [online]. (Accessed on August 1, 2020). <https://ieee-dataport.org/documents/tennessee-eastman-simulation-dataset>.
- [67] SEU Gearbox Datasets, [online]. (Accessed on August 1, 2020). <https://github.com/cathysiyu/Mechanical-datasets>.
- [68] IMS Bearing Datasets, [online]. (Accessed on August 1, 2020). <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository>.
- [69] PHM IEEE 2012 Data Challenge, [online]. (Accessed on August 1, 2020). <https://github.com/wkzs111/phm-ieee-2012-data-challenge-dataset>.
- [70] Airbus Helicopter Accelerometer Dataset, [online]. (Accessed on August 1, 2020). <https://www.research-collection.ethz.ch/handle/20.500.11850/415151>.
- [71] Numenta Anomaly Benchmark, [online]. (Accessed on August 1, 2020). <https://github.com/numenta/NAB>.
- [72] Secure Water Treatment Datasets, [online]. (Accessed on August 1, 2020). <https://itrust.sutd.edu.sg/testbeds/secure-water-treatment-swat/>.
- [73] Fault-Attention Generative Probabilistic Adversarial Autoencoder, [online]. (Accessed on August 1, 2020). <https://github.com/a1018680161/FGPAA>.
- [74] Numenta Platform for Intelligent Computing, [online]. (Accessed on August 1, 2020). <https://github.com/numenta/nupic>.
- [75] GAN-AD, [online]. (Accessed on August 1, 2020). <https://github.com/LiDan456/GAN-AD>.
- [76] J R Jiang, J E Lee, Y M Zeng. Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life. *Sensors*, 2020, 20(1): 166.
- [77] S B Jiang, P K Wong, Y C Liang. A fault diagnostic method for induction motors based on feature incremental broad learning and singular value decomposition. *IEEE Access*, 2019, 7: 157796-157806.
- [78] T Lobato, R da Silva, E da Costa, et al. An integrated approach to rotating machinery fault diagnosis using, EEMD, SVM, and augmented data. *Journal of Vibration Engineering & Technologies*, 2020, 8: 403-408.
- [79] A K Panda, J S Rapur, R Tiwari. Prediction of flow blockages and impending cavitation in centrifugal pumps using Support Vector Machine (SVM) algorithms based on vibration measurements. *Measurement*, 2018, 130: 44-56.
- [80] R Razavi-Far, E Hallaji, M Farajzadeh-Zanjani, et al. Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems. *IEEE Transactions on Industrial Electronics*, 2018, 66(8): 6331-6342.
- [81] Y Wang, Q Jin, G Sun, et al. Planetary gearbox fault feature learning using conditional variational neural networks under noise environment. *Knowledge-Based Systems*, 2019, 163: 438-449.
- [82] Z Chen, A Mauricio, W Li, et al. A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, 2020, 140: 106683.
- [83] L Kou, Y Qin, X Zhao. A multi-dimension end-to-end CNN model for rotating devices fault diagnosis on high-speed train bogie. *IEEE Transactions on Vehicular Technology*, 2019, 69(3): 2513-2524.
- [84] Z He, H Shao, P Wang, et al. Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples. *Knowledge-Based Systems*, 2020, 191: 105313.
- [85] G Jiang, H He, J Yan, et al. Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Transactions on Industrial Electronics*, 2018, 66(4): 3196-3207.
- [86] C Lu, Z Y Wang, W L Qin, et al. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 2017, 130: 377-388.
- [87] H Liu, L Li, J Ma. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration*, 2016, 2016: 6127479.
- [88] M Ma, C Sun, X Chen. Deep coupling autoencoder for fault diagnosis with multimodal sensory data. *IEEE Transactions on Industrial Informatics*, 2018, 14(3): 1137-1145.
- [89] P Shi, X Guo, D Han, et al. A sparse auto-encoder method based on compressed sensing and wavelet packet energy entropy for rolling bearing intelligent fault diagnosis. *Journal of Mechanical Science and Technology*, 2020, 34:1445-1458.
- [90] D Han, N Zhao, P Shi. A new fault diagnosis method based on deep belief network and support vector machine with Teager-Kaiser energy operator for bearings. *Advances in Mechanical Engineering*, 2017, 9(12): 1687814017743113.
- [91] H Jiang, H Shao, X Chen, et al. A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery. *Journal of Intelligent & Fuzzy Systems*, 2018, 34(6): 3513-3521.
- [92] H Shao, H Jiang, X Li, et al. Rolling bearing fault detection using continuous deep belief network with locally linear embedding. *Computers in Industry*, 2018, 96: 27-39.
- [93] X Wang, Y Qin, A Zhang. An intelligent fault diagnosis approach for planetary gearboxes based on deep belief networks and uniformed features. *Journal of Intelligent & Fuzzy Systems*, 2018, 34(6): 3619-3634.
- [94] S Gao, L Xu, Y Zhang, et al. Rolling bearing fault diagnosis based on intelligent optimized self-adaptive deep belief network. *Measurement Science and Technology*, 2020, 31(5): 055009.

- [95] C Zhang, Y Zhang, C Hu, et al. A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network. *IEEE Access*, 2020, 8: 36293-36312.
- [96] J C Sumba, I R Quinde, L E Ochoa, et al. Intelligent fault diagnosis for rotating machines using deep learning. *Smart and Sustainable Manufacturing Systems*, 2019, 3(2): 27-40.
- [97] L Eren, T Ince, S Kiranyaz. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems*, 2019, 91(2): 179-189.
- [98] A González-Muñiz, I Díaz, A A Cuadrado. DCNN for condition monitoring and fault detection in rotating machines and its contribution to the understanding of machine nature. *Heliyon*, 2020, 6(2): 03395.
- [99] G Li, C Deng, J Wu, et al. Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks and S-transform. *Sensors*, 2019, 19(12): 2750.
- [100] Y Wang, S Huang, J Dai, et al. A novel bearing fault diagnosis methodology based on SVD and one-dimensional convolutional neural network. *Shock and Vibration*, 2020, 2020: 17.
- [101] H Yin, Z Li, J Zuo, et al. Wasserstein generative adversarial network and convolutional neural network (WG-CNN) for bearing fault diagnosis. *Mathematical Problems in Engineering*, 2020:1850286.
- [102] C Zhang, J Feng, C Hu, et al. An intelligent fault diagnosis method of rolling bearing under variable working loads using 1-D stacked dilated convolutional neural network. *IEEE Access*, 2020, 8: 63027-63042.
- [103] W Gong, H Chen, Z Zhang, et al. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors*, 2019, 19(7): 1693.
- [104] S Guo, B Zhang, T Yang, et al. Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization. *IEEE Transactions on Industrial Electronics*, 2019, 67(9): 8005-8015.
- [105] T Huang, S Fu, H Feng, et al. Bearing fault diagnosis based on shallow multi-scale convolutional neural network with attention. *Energies*, 2019, 12(20): 3937.
- [106] A Y Khodja, N Guersi, M N Saadi, et al. Rolling element bearing fault diagnosis for rotating machinery using vibration spectrum imaging and convolutional neural networks. *The International Journal of Advanced Manufacturing Technology*, 2020, 106(5): 1737-1751.
- [107] G Li, C Deng, J Wu, et al. Rolling bearing fault diagnosis based on wavelet packet transform and convolutional neural network. *Applied Sciences*, 2020, 10(3): 770.
- [108] H Wang, J Xu, R Yan, et al. A new intelligent bearing fault diagnosis method using SDP representation and SE-CNN. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(5): 2377-2389.
- [109] J Wang, Z Mo, H Zhang, et al. A deep learning method for bearing fault diagnosis based on time-frequency image. *IEEE Access*, 2019, 7: 42373-42383.
- [110] Y Wang, J Yan, Q Sun, et al. Bearing intelligent fault diagnosis in the industrial Internet of Things context: A lightweight convolutional neural network. *IEEE Access*, 2020, 8: 87329-87340.
- [111] G Xu, M Liu, Z Jiang, et al. Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. *Sensors*, 2019, 19(5): 1088.
- [112] J Zhang, S Yi, G Liang, et al. A new bearing fault diagnosis method based on modified convolutional neural networks. *Chinese Journal of Aeronautics*, 2020, 33(2): 439-447.
- [113] M M Islam, J M Kim. Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Computers in Industry*, 2019, 106: 142-153.
- [114] A Kumar, Y Zhou, C Gandhi, et al. Bearing defect size assessment using wavelet transform based Deep Convolutional Neural Network (DCNN). *Alexandria Engineering Journal*, 2020, 59(2): 999-1012.
- [115] P Liang, C Deng, J Wu, et al. Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform. *Computers in Industry*, 2019, 113: 103132.
- [116] C Wu, P Jiang, C Ding, et al. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Computers in Industry*, 2019, 108: 53-61.
- [117] D Zhao, T Wang, F Chu. Deep convolutional neural network based planet bearing fault classification. *Computers in Industry*, 2019, 107: 59-66.
- [118] T Li, Z Zhao, C Sun, et al. Multi-scale CNN for multi-sensor feature fusion in helical gear fault detection. *Procedia Manufacturing*, 2020, 49: 89-93.
- [119] H Chen, N Hu, Z Cheng, et al. A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes. *Measurement*, 2019, 146: 268-278.
- [120] R Chen, X Huang, L Yang, et al. Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform. *Computers in Industry*, 2019, 106: 48-59.
- [121] Y Han, B Tang, L Deng. An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes. *Computers in Industry*, 2019, 107: 50-58.
- [122] X Li, J Li, Y Qu, et al. Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals. *Applied Sciences*, 2019, 9(4): 768.
- [123] G Qiu, Y Gu, Q Cai. A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions. *Measurement*, 2019, 145: 94-107.
- [124] Y Xin, S Li, J Wang, et al. Intelligent fault diagnosis method for rotating machinery based on vibration signal analysis and hybrid multi-object deep CNN. *IET Science, Measurement & Technology*, 14(4): 407-415.
- [125] W You, C Shen, D Wang, et al. An intelligent deep feature learning method with improved activation functions for machine fault diagnosis. *IEEE Access*, 2019, 8: 1975-1985.
- [126] S Tang, S Yuan, Y Zhu. Convolutional neural network in intelligent fault diagnosis toward rotatory machinery. *IEEE Access*, 2020, 8: 86510-86519.
- [127] T Ince. Real-time broken rotor bar fault detection and classification by shallow 1D convolutional neural networks. *Electrical Engineering*, 2019, 101(2): 599-608.
- [128] I H Kao, W J Wang, Y H Lai, et al. Analysis of permanent magnet synchronous motor fault diagnosis based on learning. *IEEE Transactions on Instrumentation and Measurement*, 2018, 68(2): 310-324.
- [129] Y M Hsueh, V R Ittagihai, W B Wu, et al. Fault diagnosis system for induction motors by CNN using empirical wavelet transform. *Symmetry*, 2019, 11(10): 1212.
- [130] J H Lee, J H Pack, I S Lee. Fault diagnosis of induction motor using convolutional neural network. *Applied Sciences*, 2019, 9(15): 2950.
- [131] Y Yang, H Zheng, Y Li, et al. A fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and convolutional neural network. *ISA transactions*, 2019, 91: 235-252.
- [132] R Liu, F Wang, B Yang, et al. Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions. *IEEE Transactions on Industrial Informatics*, 2019, 16(6): 3797-3806.
- [133] L Wen, X Li, L Gao. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Computing and Applications*, 2019, 32: 6111-6124.
- [134] S Shao, R Yan, Y Lu, et al. DCNN-based multi-signal induction motor fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(6): 2658-2669.
- [135] S S Zhong, S Fu, L Lin. A novel gas turbine fault diagnosis method based on transfer learning with CNN. *Measurement*, 2019, 137: 435-453.
- [136] Y Chang, J Chen, C Qu, et al. Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels. *Renewable Energy*, 2020, 153: 205-213.
- [137] J Fu, J Chu, P Guo, et al. Condition monitoring of wind turbine gearbox bearing based on deep learning model. *IEEE Access*, 2019, 7: 57078-57087.
- [138] W Liu, P Guo, L Ye. A low-delay lightweight recurrent neural network (LLRNN) for rotating machinery fault diagnosis. *Sensors*, 2019, 19(14): 3109.
- [139] W J Lee, K Xia, N L Denton, et al. Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery. *Journal of Intelligent Manufacturing*, 2020, 32: 393-406.
- [140] M Rao, Q Li, D Wei, et al. A deep bi-directional long short-term memory model for automatic rotating speed extraction from raw vibration signals. *Measurement*, 2020, 158: 107719.
- [141] S Haidong, C Junsheng, J Hongkai, et al. Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for

- early fault prognosis of bearing. *Knowledge-Based Systems*, 2020, 188: 105022.
- [142] W Zhang, D Yang, H Wang, et al. An attention-based temporal correlation approach for end-to-end machine health perception. *IEEE Access*, 2019, 7: 141487-141497.
- [143] N V Chawla, K W Bowyer, L O Hall, et al. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002, 16: 321-357.
- [144] C Drummond, R C Holte. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. *Workshop on Learning from Imbalanced Datasets II*, Aug. 21, 2003, Washington DC: Citeseer, 2003, 11: 1-8.
- [145] R Razavi-Far, M Farajzadeh-Zanjani, M Saif. An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors. *IEEE Transactions on Industrial Informatics*, 2017, 13(6): 2758-2769.
- [146] Y Zhang, X Li, L Gao, et al. Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning. *Journal of Manufacturing Systems*, 2018, 48: 34-50.
- [147] M Salvaris, D Dean, W H Tok. Generative adversarial networks. *Deep Learning with Azure*, Apress, Berkeley, CA, 2018: 187-208.
- [148] S Shao, P Wang, R Yan. Generative adversarial networks for data augmentation in machine fault diagnosis. *Computers in Industry*, 2019, 106: 85-93.
- [149] J Wu, Z Zhao, C Sun, et al. Ss-InfoGAN for class-imbalance classification of bearing faults. *Procedia Manufacturing*, 2020, 49: 99-104.
- [150] W Mao, Y Liu, L Ding, Li, et al. Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study. *IEEE Access*, 2019, 7: 9515-9530.
- [151] J. Luo, J Huang, H Li. A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. *Journal of Intelligent Manufacturing*, 2020, 32: 407-425.
- [152] W Zhang, X Li, X D Jia, et al. Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. *Measurement*, 2020, 152:107377
- [153] J Wang, W Zhang, J Zhou. Fault detection with data imbalance conditions based on the improved bilayer convolutional neural network. *Industrial & Engineering Chemistry Research*, 2020, 59(13): 5891-5904.
- [154] T Zheng, L Song, J Wang, et al. Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings. *Measurement*, 2020, 158:107741.
- [155] Q Zhou, Y Li, Y Tian, et al. A novel method based on nonlinear auto-regression neural network and convolutional neural network for imbalanced fault diagnosis of rotating machinery. *Measurement*, 2020, 161:107880.
- [156] B Zhao, X Zhang, H Li, et al. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. *Knowledge-Based Systems*, 2020, 199:105971.
- [157] F Jia, Y Lei, N Lu, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mechanical Systems and Signal Processing*, 2018, 110: 349-367.
- [158] X Zhao, M Jia, M Lin. Deep Laplacian auto-encoder and its application into imbalanced fault diagnosis of rotating machinery. *Measurement*, 2020, 152:107320.
- [159] Y Lan, X Han, W Zong, et al. Two-step fault diagnosis framework for rolling element bearings with imbalanced data based on GSA-WELM and GSA-ELM. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2018, 232(16): 2937-2947.
- [160] T Li, Z Zhao, C Sun, et al. Adaptive channel weighted CNN with multi-sensor fusion for condition monitoring of helicopter transmission system. *IEEE Sensors Journal*, 2020, 20(15): 8364-8373.
- [161] X Dong, H Gao, L Guo, et al. Deep cost adaptive convolutional network: A classification method for imbalanced mechanical data. *IEEE Access*, 2020, 8: 71486-71496.
- [162] D Peng, H Wang, Z Liu, et al. Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition. *IEEE Transactions on Industrial Informatics*, 2020, 16(7): 4949-4960.
- [163] H Qiao, T Wang, P Wang, et al. An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions. *IEEE Access*, 2019, 7: 118954-118964.
- [164] Y Cheng, B Zhou, C Lu, et al. Fault diagnosis for rolling bearings under variable conditions based on visual cognition. *Materials*, 2017, 10(6): 582.
- [165] S Guo, T Yang, W Gao, et al. An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN. *Sensors*, 2018, 18(11): 3857.
- [166] Z Xiang, X Zhang, W Zhang, et al. Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO Spectrum and Stacking Auto-encoder. *Measurement*, 2019, 138: 162-174.
- [167] Z Zhao, Q Zhang, X Yu, et al. Unsupervised deep transfer learning for intelligent fault diagnosis: An open source and comparative study. *arXiv preprint*, 2019, arXiv:1912.12528.
- [168] M J Hasan, M M M Islam, J Kim. Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions. *Measurement*, 2019, 138: 620-631.
- [169] Y Du, A Wang, S Wang, et al. Fault diagnosis under variable working conditions based on STFT and transfer deep residual network. *Shock and Vibration*. 2020, 2020: 1274380.
- [170] Z He, H Shao, X Zhang, et al. Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples. *IEEE Access*, 2019, 7: 115368-115377.
- [171] J Wu, Z Zhao, C Sun, et al. Few-shot transfer learning for intelligent fault diagnosis of machine. *Measurement*, 2020, 166: 108202.
- [172] S Shao, S McAleer, R Yan, et al. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Transactions on Industrial Informatics*, 2018, 15(4): 2446-2455.
- [173] W Zhang, G Peng, C Li, et al. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 2017, 17(2): 425.
- [174] D Xiao, Y Huang, C Qin, et al. Transfer learning with convolutional neural networks for small sample size problem in machinery fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2019, 233(14): 5131-5143.
- [175] M Azamfar, J Singh, X Li, et al. Cross-domain gearbox diagnostics under variable working conditions with deep convolutional transfer learning. *Journal of Vibration and Control*, 2020, 27(7-8): 854-864.
- [176] J Singh, M Azamfar, A Ainapure, et al. Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions. *Measurement Science and Technology*, 2020, 31(5): 055601.
- [177] C Che, H Wang, Q Fu, et al. Deep transfer learning for rolling bearing fault diagnosis under variable operating conditions. *Advances in Mechanical Engineering*, 2019, 11(12): 1-11.
- [178] Z An, S Li, J Wang, et al. Generalization of deep neural network for bearing fault diagnosis under different working conditions using multiple kernel method. *Neurocomputing*, 2019, 352: 42-53.
- [179] W Qian, S Li, P Yi, et al. A novel transfer learning method for robust fault diagnosis of rotating machines under variable working conditions. *Measurement*, 2019, 138: 514-525.
- [180] X Li, W Zhang, Q Ding. A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning. *Neurocomputing*, 2018, 310: 77-95.
- [181] Q Li, C Shen, L Chen, et al. Knowledge mapping-based adversarial domain adaptation: A novel fault diagnosis method with high generalizability under variable working conditions. *Mechanical Systems and Signal Processing*, 2021, 147: 107095.
- [182] Y Ganin, E Ustinova, H Ajakan, et al. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 2016, 17(1): 2096-2030.
- [183] T Li, Z Zhao, C Sun, et al. Domain adversarial graph convolutional network for fault diagnosis under variable working conditions. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 3515010.
- [184] W Lu, B Liang, Y Cheng, et al. Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2016, 64(3): 2296-2305.

- [185] T Han, C Liu, W Yang, et al. A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-Based Systems*, 2019, 165: 474-487.
- [186] K Xu, S Li, J Wang, et al. A novel convolutional transfer feature discrimination network for unbalanced fault diagnosis under variable rotational speeds. *Measurement Science and Technology*, 2019, 30(10): 105107.
- [187] R Chen, S Chen, M He, et al. Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2017, 231(6): 666-679.
- [188] X Guo, C Shen, L Chen. Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Applied Sciences*, 2017, 7(1): 41.
- [189] G Jiang, H He, P Xie, et al. Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 2017, 66(9): 2391-2402.
- [190] C Shen, Y Qi, J Wang, et al. An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder. *Engineering Applications of Artificial Intelligence*, 2018, 76: 170-184.
- [191] Z Wang, J Wang, Y Wang. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing*, 2018, 310: 213-222.
- [192] X Liu, Q Zhou, J Zhao, et al. Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network. *Sensors*, 2019, 19(4): 972.
- [193] Y Qi, C Shen, J Zhu, et al. A new deep fusion network for automatic mechanical fault feature learning. *IEEE Access*, 2019, 7: 152552-152563.
- [194] Y Zhang, X Li, L Gao, et al. Ensemble deep contractive auto-encoders for intelligent fault diagnosis of machines under noisy environment. *Knowledge-Based Systems*, 2020, 196: 105764.
- [195] M Gan, C Wang. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems and Signal Processing*, 2016, 72: 92-104.
- [196] H Shao, H Jiang, H Zhang, et al. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mechanical Systems and Signal Processing*, 2018, 100: 743-765.
- [197] W You, C Shen, X Guo, et al. A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery. *Advances in Mechanical Engineering*, 2017, 9(6): 1-17.
- [198] W Zhang, C Li, G Peng, et al. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, 2018, 100: 439-453.
- [199] D Peng, Z Liu, H Wang, et al. A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access*, 2018, 7: 10278-10293.
- [200] P Peng, J Wang. NOSCNN: A robust method for fault diagnosis of RV reducer. *Measurement*, 2019, 138: 652-658.
- [201] T Zan, H Wang, M Wang, et al. Application of multi-dimension input convolutional neural network in fault diagnosis of rolling bearings. *Applied Sciences*, 2019, 9(13): 2690.
- [202] G Jin, T Zhu, M W Akram, et al. An adaptive anti-noise neural network for bearing fault diagnosis under noise and varying load conditions. *IEEE Access*, 2020, 8: 74793-74807.
- [203] E Bechhofer. Society for Machinery Failure Prevention Technology, [Online]. Available: <https://mfpt.org/fault-data-sets/> (accessed on August 1, 2020).
- [204] C Lessmeier, J K Kimotho, D Zimmer, et al. KAT-DataCenter, Chair of Design and Drive Technology, Paderborn University. <https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/> (accessed on August 1, 2020).
- [205] K Li. School of Mechanical Engineering, Jiangnan University. <http://mad-net.org:8765/explore.html?t=0.5831516555847212>. (accessed on August 1, 2020).
- [206] P Cao, S Zhang, J Tang. Gear fault data, [Online]. Available: <https://doi.org/10.6084/m9.figshare.6127874.v1> (accessed on August 1, 2020).
- [207] W Zhang, G Peng, C Li. Rolling element bearings fault intelligent diagnosis based on convolutional neural networks using raw sensing signal. *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, Springer, Cham, 2017: 77-84.
- [208] M Zhang, Z Jiang, K Feng. Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump. *Mechanical Systems and Signal Processing*, 2017, 93: 460-493.
- [209] T Li, Z Zhao, C Sun, et al. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021. doi:<https://doi.org/10.1109/TSMC.2020.3048950>.
- [210] T Li, Z Zhao, C Sun, et al. Multi-receptive field graph convolutional networks for machine fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2020. doi:<https://doi.org/10.1109/TIE.2020.3040669>.
- [211] Z Zhao, T Li, J Wu, et al. Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study. *ISA Transactions*, 2020, 107: 224-255.
- [212] L Wen, X Li, L Gao, et al. A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, 2017, 65(7): 5990-5998.
- [213] K Gregor, Y LeCun. Learning fast approximations of sparse coding. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010: 399-406.
- [214] G W Bartram, S Mahadevan. Integrating heterogeneous information in diagnosis and prognosis. *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2013: 1941.
- [215] P C C Berri, M D L Dalla Vedova, L Mainini. Real-time fault detection and prognostics for aircraft actuation systems. *AIAA Scitech 2019 Forum*, 2019: 2210.
- [216] J B Ali, B Chebel-Morello, L Saidi, et al. Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network. *Mechanical Systems and Signal Processing*, 2015, 56: 150-172.
- [217] Q Tong, J Hu. Bearing performance degradation assessment based on information-theoretic metric learning and fuzzy C-means clustering. *Measurement Science and Technology*, 2020, 31(7): 075001.
- [218] G Liu, J Zhao, H Li, et al. Bearing degradation assessment based on entropy with time parameter and fuzzy c-means clustering. *Journal of Vibroengineering*, 2019, 21(5): 1322-1329.
- [219] J Zhou, C Zhang, L Zhang, et al. Rolling bearing performance degradation assessment based on AANN-FCM. *Machine Design and Research*, 2019, 35(1): 96-99. (in Chinese)
- [220] J Zhou, Q Xu, L Zhang, et al. Rolling bearing performance degradation assessment based on the wavelet packet Tsallis entropy and FCM. *Journal of Mechanical Transmission*, 2016, 40(5): 110-113. (in Chinese)
- [221] H Jiang, J Chen, G Dong, et al. An intelligent performance degradation assessment method for bearings. *Journal of Vibration and Control*, 2017, 23(18): 3023-3040.
- [222] H Jiang, J Yuan, Q Zhao, et al. A robust performance degradation modeling approach based on Student's T-HMM and nuisance attribute projection. *IEEE Access*, 2020, 8: 49629-49644.
- [223] Y Hu, S Liu, H Lu, et al. Remaining useful life model and assessment of mechanical products: a brief review and a note on the state space model method. *Chinese Journal of Mechanical Engineering*, 2019, 32(1): 1-20.
- [224] H Wang, Y Zhou, J Qu, et al. A prognostic method of mechanical equipment based on HDP-HMM. *Journal of Vibration and Shock*, 2019, 38(8): 173-179. (in Chinese)
- [225] L Li, T Ming, S Liu, et al. An effective health indicator based on two dimensional hidden Markov model. *Journal of Mechanical Science and Technology*, 2017, 31(4): 1543-1550.
- [226] T Liu, X Wu, Y Guo, et al. Bearing performance degradation assessment by orthogonal local preserving projection and continuous hidden Markov model. *Transactions of the Canadian Society for Mechanical Engineering*, 2016, 40(5): 1019-1030.
- [227] Y Pan, R Hong, J Chen, et al. Performance degradation assessment of a wind turbine gearbox based on multi-sensor data fusion. *Mechanism and Machine Theory*, 2019, 137: 509-526.
- [228] Y Ma, J Chen, R Hong, et al. Performance degradation assessment of wind turbine generator gearbox based on multi-sensor information fusion. *Computer Integrated Manufacturing Systems*, 2019, 25(2): 318-325. (in Chinese)

- [229] Y Feng, X Huang, R Hong, et al.: A multi-dimensional data-driven method for large-size slewing bearings performance degradation assessment. *Journal of Central South University (Science and Technology)*, 2017, 48(03): 684-693. (in Chinese)
- [230] F Wang, L Fang, Y Zhao, et al. Rolling bearing early weak fault detection and performance degradation assessment based on VMD and SVDD. *Journal of Vibration and Shock*, 2019, 38(22): 224-230+256. (in Chinese)
- [231] J Zhou, H Guo, L Zhang, et al. Bearing performance degradation assessment using lifting wavelet packet symbolic entropy and SVDD. *Shock and Vibration*, 2016, 2016: 3086454.
- [232] X Ding, L Wang, W Huang, et al. Feature clustering analysis using reference model towards rolling bearing performance degradation assessment. *Shock and Vibration*, 2020, 2020: 6306087.
- [233] W Mao, L He, Y Yan, et al. Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mechanical Systems and Signal Processing*, 2017, 83: 450-473.
- [234] Y Lu, R Xie, S Y Liang. CEEMD-assisted bearing degradation assessment using tight clustering. *The International Journal of Advanced Manufacturing Technology*, 2019, 104(1): 1259-1267.
- [235] L Zhang, C Song, C Wang, et al. Bearing performance degradation assessment based on a combination of multi-scale entropy and K-medoids clustering. *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, IEEE, 2019: 1-6.
- [236] B Wang, X Hu, D Sun, et al. Bearing condition degradation assessment based on basic scale entropy and Gath-Geva fuzzy clustering. *Advances in Mechanical Engineering*, 2018, 10(10): 1-11.
- [237] A Rai, S H Upadhyay. Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering. *Mechanical Systems and Signal Processing*, 2017, 93: 16-29.
- [238] J Zhou, C Zhang, F Wang. A method for performance degradation assessment of wind turbine bearings based on hidden Markov model and fuzzy C-means model. *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, IEEE, 2019: 1-5.
- [239] P Tiwari, S H Upadhyay. Degradation assessment of ball bearings utilizing curvilinear component analysis. *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-Body Dynamics*, 2019, 233(3): 714-730.
- [240] A Rai, S H Upadhyay. Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2018, 232(6): 1118-1132.
- [241] K Zhu, X Jiang, L Chen, et al. Performance degradation assessment of rolling element bearings using improved fuzzy entropy. *Measurement Science Review*, 2017, 17(5): 219.
- [242] Y Qin, D Wang, X Zhao, et al. Performance degradation assessment of train rolling bearings based on SVM and segmented vote method. *2016 Prognostics and System Health Management Conference (PHM-Chengdu)*, IEEE, 2016: 1-6.
- [243] D Zhang, E Stewart, J Ye, et al. Roller bearing degradation assessment based on a deep MLP convolution neural network considering outlier regions. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(6): 2996-3004.
- [244] S Dong, W Wu, K He, et al. Rolling bearing performance degradation assessment based on improved convolutional neural network with anti-interference. *Measurement*, 2020, 151: 107219.
- [245] D Zhang, E Stewart, M Entezami, et al. Degradation assessment of bearings using deep convolutional inner-ensemble learning with outlier removal. *2019 Prognostics and System Health Management Conference (PHM-Paris)*, IEEE, 2019: 315-319.
- [246] L Guo, Y Lei, N Li, et al. Deep convolution feature learning for health indicator construction of bearings. *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, IEEE, 2017: 1-6.
- [247] U Akpudo, J W Hur. A deep learning approach to prognostics of rolling element bearings. *International Journal of Integrated Engineering*, 2020, 12(3): 178-186.
- [248] B Zhang, S Zhang, W Li. Bearing performance degradation assessment using long short-term memory recurrent network. *Computers in Industry*, 2019, 106: 14-29.
- [249] Y Cheng, H Zhu, J Wu, et al. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks. *IEEE Transactions on Industrial Informatics*, 2018, 15(2): 987-997.
- [250] Z Wang, H Ma, H Chen, et al. Performance degradation assessment of rolling bearing based on convolutional neural network and deep long-short term memory network. *International Journal of Production Research*, 2020, 58(13): 3931-3943.
- [251] F Xu, Z Fang, R Tang, et al. An unsupervised and enhanced deep belief network for bearing performance degradation assessment. *Measurement*, 2020, 162: 107902.
- [252] Y Pan, R Hong, J Chen, et al. A hybrid dbn-som-pf-based prognostic approach of remaining useful life for wind turbine gearbox. *Renewable Energy*, 2020, 152: 138-154.
- [253] W Tu, T Liu, H Liu. Application of bp and ar model in performance degradation assessment and prediction of bearings. *Journal of Electronic Measurement and Instrument*, 2019, 33(11): 79-88.
- [254] J Gai, Y Hu, J Shen. A bearing performance degradation modeling method based on EMD-SVD and fuzzy neural network. *Shock and Vibration*, 2019, DOI: <https://doi.org/10.1155/2019/5738465>.
- [255] Z Li, Y Wang, K Wang. A deep learning driven method for fault classification and degradation assessment in mechanical equipment. *Computers in Industry*, 2019, 104: 1-10.
- [256] K Aggarwal, O Atan, A Farahat, et al. Two birds with one network: Unifying failure event prediction and time-to-failure modeling. *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, 2018: 1308-1317.
- [257] C Byington, M Roemer, T Galie. Prognostic enhancements to diagnostic systems for improved condition-based maintenance [military aircraft]. *Proceedings, IEEE Aerospace Conference*, IEEE, 2002, 6: 6-6.
- [258] A Muller, M Suhner, B lung. Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *Reliability Engineering & System Safety*, 2008, 93(2): 234-253.
- [259] W Yu, I Kim, C Mechefske. Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mechanical Systems and Signal Processing*, 2019, 129: 764-780.
- [260] C Che, H Wang, Q Fu, et al. Combining multiple deep learning algorithms for prognostic and health management of aircraft. *Aerospace Science and Technology*, 2019, 94: 105423.
- [261] D Belmiloud, T Benkedjough, M Lachi, et al. Deep convolutional neural networks for bearings failure prediction and temperature correlation. *Journal of Vibroengineering*, 2018, 20(8): 2878-2891.
- [262] Y Li, H Li, B Wang, et al. Research on the feature selection of rolling bearings' degradation features. *Shock and Vibration*, 2019, 22: 6450719.
- [263] D She, M Jia. Health indicator construction of rolling bearings based on deep convolutional neural network considering phase degradation. *2019 Prognostics and System Health Management Conference (PHM-Paris)*, IEEE, 2019: 373-378.
- [264] Y Yoo, J Baek. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Applied Sciences*, 2018, 8(7): 1102.
- [265] X Li, W Zhang, H Ma, et al. Data alignments in machinery remaining useful life prediction using deep adversarial neural networks. *Knowledge-Based Systems*, 2020, 197: 105843.
- [266] J Zhu, N Chen, C Shen. A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. *Mechanical Systems and Signal Processing*, 2020, 139: 106602.
- [267] A Elsheikh, S Yacout, M Ouali. Bidirectional handshaking LSTM for remaining useful life prediction. *Neurocomputing*, 2019, 323: 148-156.
- [268] M Souto, M Moura, I Lins. Particle swarm-optimized support vector machines and pre-processing techniques for remaining useful life estimation of bearings. *Eksplotacja i Niezawodność*, 2019, 21(4): 610-619.
- [269] C Ordóñez, F Lasheras, F Roca-Pardinas, et al. A hybrid ARIMA-SVM model for the study of the remaining useful life of aircraft engines. *Journal of Computational and Applied Mathematics*, 2019, 346: 184-191.
- [270] A Rai, S Upadhyay. Intelligent bearing performance degradation assessment and remaining useful life prediction based on self-organising map and support vector regression. *Proceedings of the Institution of Mechanical Engineers. Journal of Mechanical Engineering Science*, 2018, 232(6): 1118-1132.

- [271] D Tobon-Mejia, K Medjaher, N Zerhouni, et al. A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models. *IEEE Transactions on Reliability*, 2012, 61(2): 491-503.
- [272] Z Wu, H Luo, Y Yang, et al. K-pdm: Kpi-oriented machinery deterioration estimation framework for predictive maintenance using cluster-based hidden Markov model. *IEEE Access*, 2018, 6: 41676-41687.
- [273] S Laddada, T Benkedjouh, M Si-Chaib, et al. A data-driven prognostic approach based on wavelet transform and extreme learning machine. *2017 5th International Conference on Electrical Engineering-Bourmerdes (ICEE-B)*, IEEE, 2017: 1-4.
- [274] K Javed, R Gouriveau, N Zerhouni. Novel failure prognostics approach with dynamic thresholds for machine degradation. *IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2013: 4404-4409.
- [275] Z Li, Z Zheng, R Outbib. A prognostic methodology for power mosfets under thermal stress using echo state network and particle filter. *Microelectronics Reliability*, 2018, 88: 350-354.
- [276] P Lim, C Goh, K Tan. A time window neural network based framework for remaining useful life estimation. *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016: 1746-1753.
- [277] Y Fan, S Nowaczyk, T Rögnavaldsson. Transfer learning for remaining useful life prediction based on consensus self-organizing models. *Reliability Engineering & System Safety*, 2020, 203: 107098.
- [278] Z Zhang, F Dong, L Xie. Data-driven fault prognosis based on incomplete time slice dynamic Bayesian network. *IFAC-PapersOnLine*, 2018, 51(18): 239-244.
- [279] Y Wang, Y Zhao, S Addepalli. Remaining useful life prediction using deep learning approaches: A review. *Procedia Manufacturing*, 2020, 49: 81-88.
- [280] X Li, G Xie, H Liu, et al. Predicting remaining useful life of industrial equipment based on multivariable monitoring data analysis. *2018 Chinese Automation Congress (CAC)*, IEEE, 2018: 1861-1866.
- [281] X Li, Q Ding, J Sun. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, 2018, 172: 1-11.
- [282] A Ellefsen, S Ushakov, V Æsøy, et al.: Validation of data-driven labeling approaches using a novel deep network structure for remaining useful life predictions. *IEEE Access*, 2019, 7: 71563-71575.
- [283] R Zhao, R Yan, J Wang, et al. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors*, 2017, 17(2): 273.
- [284] J Li, X Li, D He. A directed acyclic graph network combined with CNN and LSTM for remaining useful life prediction. *IEEE Access*, 2019, 7: 75464-75475.
- [285] R Zhao, D Wang, R Yan, et al. Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Transactions on Industrial Electronics*, 2017, 65(2): 1539-1548.
- [286] P Shan, P Hou, H Ge, et al. Image feature-based for bearing health monitoring with deep-learning method. *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, IEEE, 2019: 1-6.
- [287] J Hong, Q Wang, X Qiu, et al. Remaining useful life prediction using time-frequency feature and multiple recurrent neural networks. *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, 2019: 916-923.
- [288] J Zhu, N Chen, W Peng. Estimation of bearing remaining useful life based on multiscale convolutional neural network. *IEEE Transactions on Industrial Electronics*, 2018: 66(4): 3208-3216.
- [289] X Li, W Zhang, Q Ding. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability Engineering & System Safety*, 2019, 182: 208-218.
- [290] M Ma, Z Mao. Deep recurrent convolutional neural network for remaining useful life prediction. *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2019: 1-4.
- [291] W Mao, J He, J Tang, et al. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Advances in Mechanical Engineering*, 2018, 10(12): 1687814018817184.
- [292] L Ren, J Cui, Y Sun, et al. Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *Journal of Manufacturing Systems*, 2017, 43: 248-256.
- [293] Q Wang, B Zhao, H Ma, et al. A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion. *Journal of Mechanical Science and Technology*, 2019, 33(6): 2561-2571.
- [294] J Deutsch, D He. Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 48(1): 11-20.
- [295] A Ellefsen, E Bjørlykhaug, V Æsøy, et al. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliability Engineering & System Safety*, 2019, 183, 240-251.
- [296] C Sun, M Ma, Z Zhao, et al. Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. *IEEE Transactions on Industrial Informatics*, 2018, 15(4): 2416-2425.
- [297] W Mao, J He, M Zuo. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. *IEEE Transactions on Instrumentation and Measurement*, 2019, 69(4): 1594-1608.
- [298] M Sadoughi, H Lu, C Hu. A deep learning approach for failure prognostics of rolling element bearings. *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2019: 1-7 (2019).
- [299] R He, Y Dai, J Lu, et al. Developing ladder network for intelligent evaluation system: Case of remaining useful life prediction for centrifugal pumps. *Reliability Engineering & System Safety*, 2018, 180: 385-393.
- [300] C Huang, H Huang, Y Li. A bidirectional lstm prognostics method under multiple operational conditions. *IEEE Transactions on Industrial Electronics*, 2019, 66(11): 8792-8802.
- [301] A Al-Dulaimi, S Zabihi, A Asif, et al. A multimodal and hybrid deep neural network model for remaining useful life estimation. *Computers in Industry*, 2019, 108: 186-196.
- [302] J Herp, N Pedersen, E Nadimi. A novel probabilistic long-term fault prediction framework beyond scada data-with applications in main bearing failure. *Journal of Physics: Conference Series*, 2019, 1222: 012043.
- [303] A He, X Jin. Failure detection and remaining life estimation for ion mill etching process through deep-learning based multimodal data fusion. *Journal of Manufacturing Science and Engineering*, 2019, 141(10): 101008.
- [304] H Miao, B Li, C Sun, et al. Joint learning of degradation assessment and RUL prediction for aeroengines via dual-task deep LSTM networks. *IEEE Transactions on Industrial Informatics*, 2019, 15(9): 5023-5032.
- [305] R Liu, B Yang, A Hauptmann. Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network. *IEEE Transactions on Industrial Informatics*, 2019, 16(1): 87-96.
- [306] M Xia, T Li, T Shu, et al. A two-stage approach for the remaining useful life prediction of bearings using deep neural networks. *IEEE Transactions on Industrial Informatics*, 2018, 15(6): 3703-3711.
- [307] B Yang, R Liu, E Zio. Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Transactions on Industrial Electronics*, 2019, 66(12): 9521-9530.
- [308] A Zhang, H Wang, S Li, et al. Transfer learning with deep recurrent neural networks for remaining useful life estimation. *Applied Sciences*, 2018, 8(12): 2416.
- [309] S Yu, Z Wu, X Zhu, et al. A domain adaptive convolutional LSTM model for prognostic remaining useful life estimation under variant conditions. *2019 Prognostics and System Health Management Conference (PHM-Paris)*, IEEE, 2019: 130-137.
- [310] P Costa, A Akcay, Y Zhang, et al. Remaining useful lifetime prediction via deep domain adaptation. *Reliability Engineering & System Safety*, 2020, 195: 106682.
- [311] S Sankararaman. Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, 2015, 25: 228-247.
- [312] S Sankararaman, C Teubert. Impact of uncertainty on the diagnostics and prognostics of a current-pressure transducer. *2015 IEEE Aerospace Conference*, IEEE, 2015: 1-10.
- [313] M Djezir, S Benmoussa, M Benbouzid. Data-driven approach augmented in simulation for robust fault prognosis. *Engineering Applications of Artificial Intelligence*, 2019, 86: 154-164.
- [314] Y Deng, A Bucchianico, M Pechenizkiy. Controlling the accuracy and uncertainty trade-off in RUL prediction with a surrogate Wiener propagation model. *Reliability Engineering & System Safety*, 2020, 196: 106727.
- [315] W Peng, Z Ye, N Chen. Bayesian deep-learning-based health prognostics toward prognostics uncertainty. *IEEE Transactions on Industrial Electronics*, 2019, 67(3): 2283-2293.

- [316] B Wang, Y Lei, T Yan, et al. Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery. *Neurocomputing*, 2020, 379: 117-129.
- [317] Y Gal, Z Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*, PMLR, 2016: 1050-1059.
- [318] Y Liao, L Zhang, C Liu. Uncertainty prediction of remaining useful life using long short-term memory network based on bootstrap method. *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*, IEEE, 2018: 1-8.
- [319] VTV, P Malhotra, L Vig, G Shroff, et al. Data-driven prognostics with predictive uncertainty estimation using ensemble of deep ordinal regression models. *arXiv preprint*, 2019, arXiv:1903.09795.
- [320] NASA. Prognostics Center of Excellence Database. Available: <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository> (accessed on August 1 2020).
- [321] A Saxena, K Goebel. Turbofan engine degradation simulation data set. Available: <https://ti.arc.nasa.gov/c/6/> (accessed on August 1 2020).
- [322] A Agogino, K Goebel. Milling data set. Available: <https://ti.arc.nasa.gov/c/4/> (accessed on August 1 2020).
- [323] P Nectoux, R Gouriveau, K Medjaher, et al. An experimental platform for bearings accelerated degradation tests. *Proceedings of the IEEE International Conference on Prognostics and Health Management*, IEEE, Beijing, China, 2012: 23-25. Available: <https://ti.arc.nasa.gov/c/18/> (accessed on August 1 2020).
- [324] Y Lei, B Wang. XJTU-SY bearing datasets. Available: https://mekhub.cn/machine-diagnosis/XJTU-SY_Bearing_Datasets (accessed on August 1 2020).
- [325] N Oyharcabal. Convolutional recurrent neural networks for remaining useful life prediction in mechanical systems. 2018, Available: https://github.com/nicolasoyharcabal/ConvRNN_for_RUL_estimation (accessed on August 1 2020).
- [326] L Jayasinghe, T Samarasinghe, C Yuen, et al. Temporal convolutional memory networks for remaining useful life estimation of industrial machinery. *2019 IEEE International Conference on Industrial Technology (ICIT)*, 2019: 915-920.
- [327] L D Libera. A comparative study between Bayesian and frequentist neural networks for remaining useful life estimation in condition-based maintenance. *arXiv preprint*, 2019, arXiv:1911.06256.
- [328] Z Chen, M Wu, R Zhao, et al. Machine remaining useful life prediction via an attention based deep learning approach. *IEEE Transactions on Industrial Electronics*, 2020, 68(3): 2521-2531.
- [329] Y Yucesan, F Viana. A physics-informed neural network for wind turbine main bearing fatigue. *International Journal of Prognostics and Health Management*, 2020, 125: 103386.
- [330] J Wang, Y Li, R Zhao, et al. Physics guided neural network for machining tool wear prediction. *Journal of Manufacturing Systems*, 2020, 57: 298-310.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
